Ethics issues across data life cycle

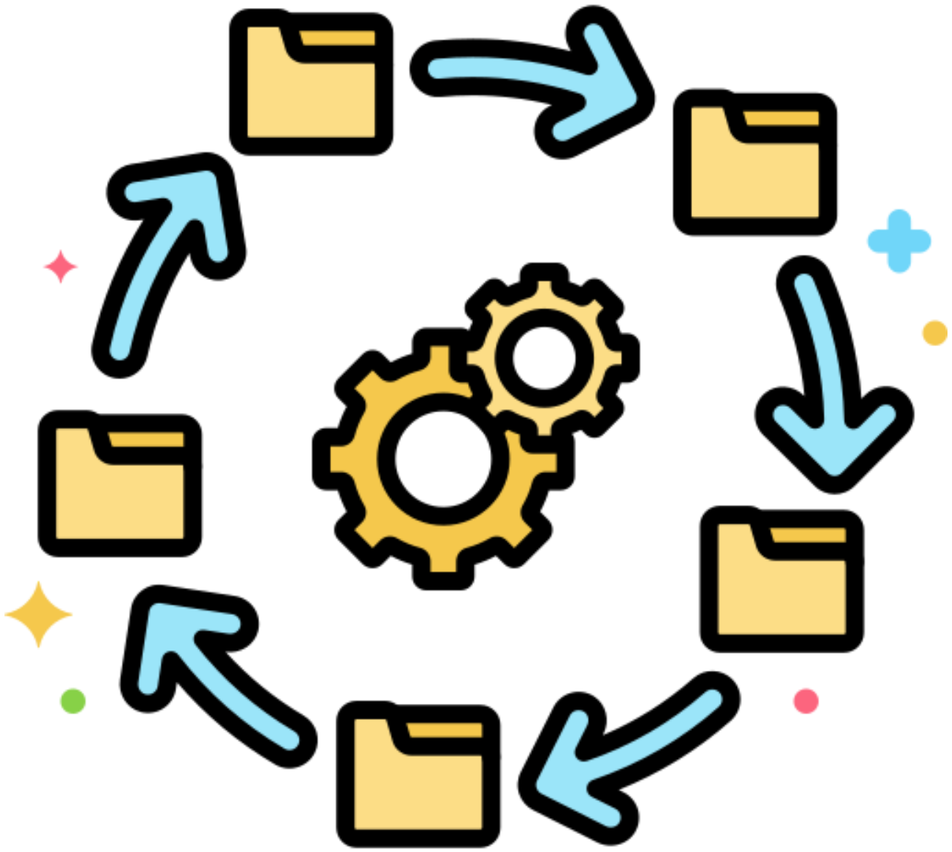
INTRODUCTION TO DATA ETHICS



Shalini Kurapati, PhD
Co-founder and CEO, Clearbox AI

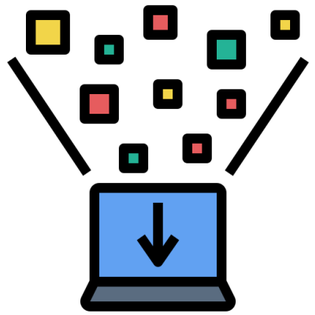
The data life cycle

- 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



[1] https://www.ibm.com/topics/data-life cycle-management#Data%20life cycle%20management%20and%20IBM [2] Icon made by Flat Icons on from www.flaticon.com

Data acquisition



- Many ways - surveys, mobile apps, sensors, wearables, web scraping, third parties
- 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

[1] Icon made by Parzival 1997 from www.flaticon.com

Data preparation



- 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

[1] https://time.com/6247678/openai-chatgpt-kenya-workers/

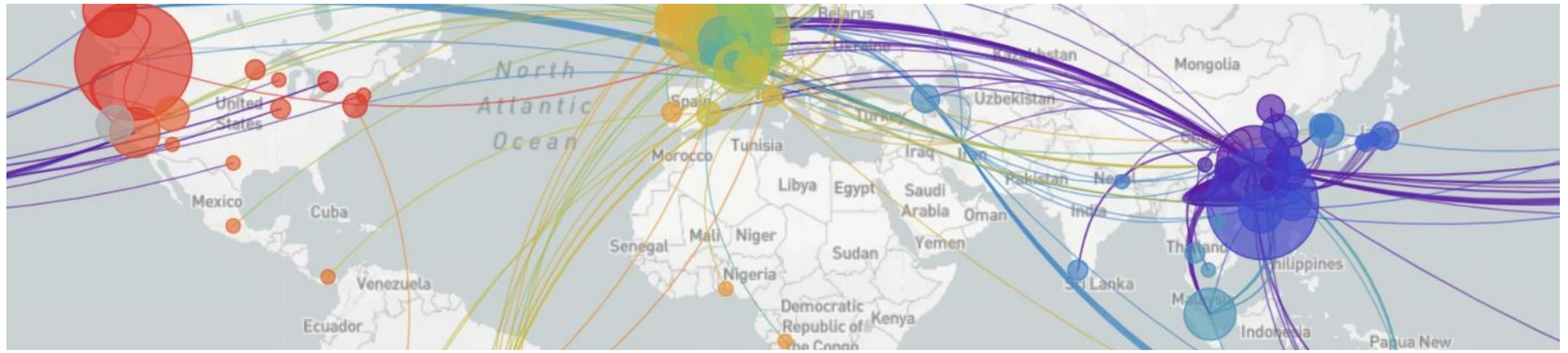
Data storage



- Confidentiality and integrity- prevent data breaches or accidental losses
- Data security- no unauthorized access
- Technical:
  - Infrastructure, methods, techniques, and devices for data storage
- Organizational:
  - Companies policies, training

[1] Icon made by HJ studio from www.flaticon.com

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

[1] https://www.ga4gh.org/news/regulatory-ethics-perspective-on-covid-19-data-sharing-an-interview-with-johan-ordish/

Let's practice!

INTRODUCTION TO DATA ETHICS

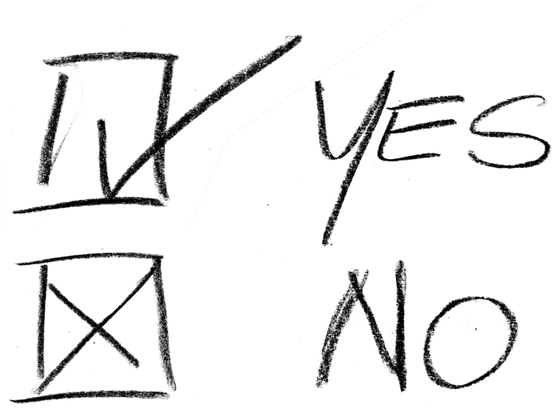
Valid consent

INTRODUCTION TO DATA ETHICS



Shalini Kurapati, PhD
Co-founder and CEO, Clearbox AI

Why consent in data ethics?



- Having control or say over your data
- Data ownership
- Personal/privacy sensitive data
- Intellectual property- copyright

Data ownership and IP

Getty Images sues AI art generator Stable Diffusion in the US for copyright infringement



/ Getty Images has filed a case against Stability AI, alleging that the company copied 12 million images to train its AI model 'without permission ... or compensation.'

- Important to understand who owns the data
- Copyright- photos, quotes. Someone else's data? you need permission
- Stability AI example- data used without permission, no royalties

[1] https://www.theverge.com/2023/1/17/23558516/ai-art-copyright-stable-diffusion-getty-images-lawsuit

Even for the non-tech stuff

	I wish to review the notes, transcripts, or other data collected during the research pertaining to my participation.
	I agree to be quoted directly.
	I agree to be quoted directly if my name is not published and a made-up name (pseudonym) is used.
	I agree that the researchers may publish documents that contain quotations by me.

- Clarification, granular consent
- Interviews- direct quotes
- Plagiarism- copying other's work
- Subjective data- chance to review the accuracy

[1] Interview consent form by School of GeoSciences, The University of Edinburgh

Personal data- focus on the valid



- Ethical and legal considerations for valid consent for data processing
- Informed- clear, specific information
- Freely given- real choice, no power imbalance
  - A photo editing app uses GPS location to activate the account, not freely given consent
  - Can an employee safely refuse employer data request?
- Special considerations for children

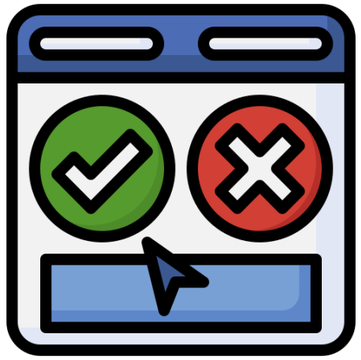
[1] Meme source: Andrea Minieri, Clearbox AI

What is in the consent request form?



- Name and contact details of your organization
- The purpose of data collection
- How will you use or reuse the data
- Third party data sharing
- Statement that individuals can withdraw consent at any time

Many ways to record consent



- A clear action like ticking a box
- Signing a declaration/form
- Through an email
- Selecting Yes/No or granular
- Oral statement (recording)
- Best at collection stage

[1] Icon made by surang from www.flaticon.com

Consent management and individual requests



- Systematic approach for record keeping
- Automate consent withdrawal process
- E.g. unsubscribe option for marketing emails
- Treat consent as personal data, period reviews

Let's practice!

INTRODUCTION TO DATA ETHICS

Data privacy and security

INTRODUCTION TO DATA ETHICS



Shalini Kurapati, PhD
Co-founder and CEO, Clearbox AI

datacamp

Aren't they the same?

Data Privacy

- Data protection
- Individual rights
- Ethical impact, compliance
- Data security- means to privacy
  - Technical and organizational measures

Data Security

- Not limited to personal data
- Information security
- Confidentiality
- Integrity
- Availability

Confidentiality



- Safe from unauthorized access
- Access controls, right person has access
- Encryption, secure VPN
- Awareness and training

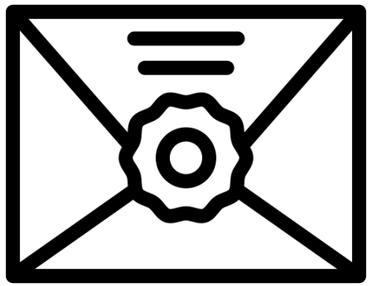
[1] Icon made by Freepik from www.flaticon.com

Data breaches

- Insufficient technology and unwanted user behavior
- Lost or stolen unprotected/unencrypted devices
- Naive security culture
- Measure to prevent and mitigate
  - Update tech
  - Cyber security
  - Incidence and response plan
  - Training



Integrity



- Trustworthy data
- Data quality
- Not accidentally modified
- Change management
- Data validation

[1] Icon made by Freepik from www.flaticon.com

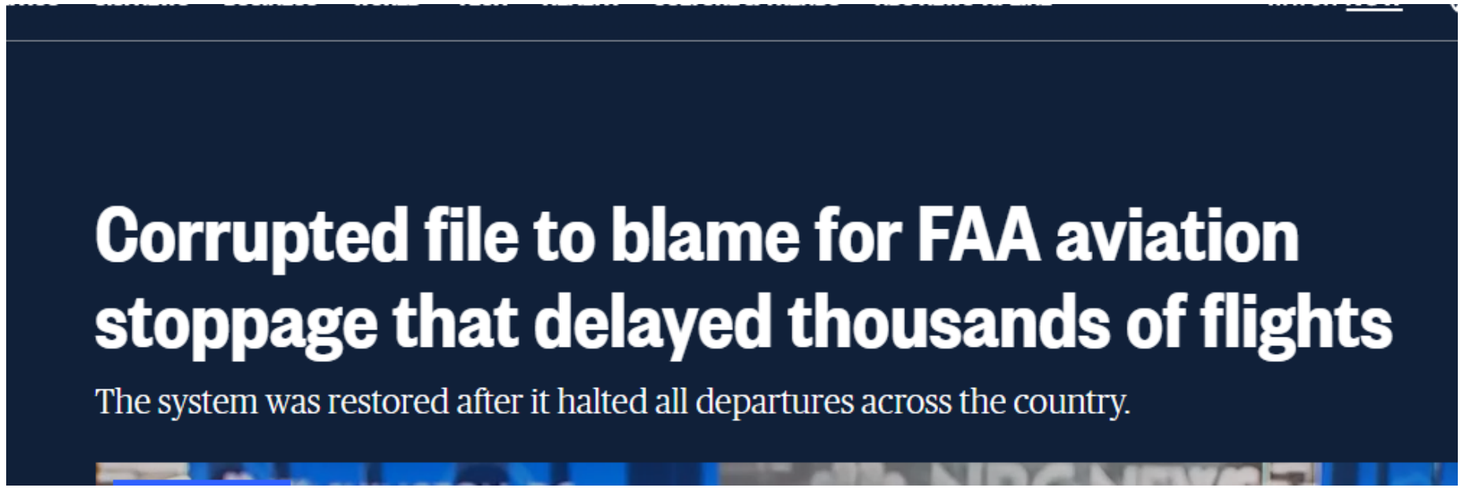
Availability



- You should access data when you need it
- System failures, power outages
- Backups and recovery

[1] Icon made by surang from www.flaticon.com

Backup anyone?



- Flights grounded due to computer outage
- Corrupted data file
- Unavailability of an efficient backup
- Impact on individuals and economy

[1] https://www.nbcnews.com/news/us-news/us-flights-grounded-faa-outage-rcna65243

Let's practice!

INTRODUCTION TO DATA ETHICS

Transparency and accountability

INTRODUCTION TO DATA ETHICS



Shalini Kurapati, PhD
Co-founder and CEO, Clearbox AI

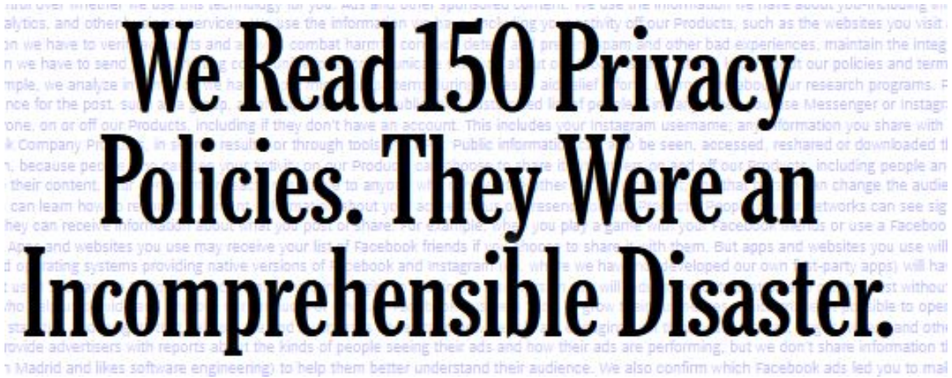
Transparency

- Clarity, disclosure of data use
- Increases trust, a proof of accountability
- Communication and engagement with stakeholders
- How?
  - Privacy policies/statements
  - Certifications and third party audits
  - Disclosure of mishaps



[1] Photo by Bud Helisson on Unsplash

Good example


We Read 150 Privacy Policies. They Were an Incomprehensible Disaster.

- Complicated statements are useless
- Concise, clear, plain language, high-school level


Using the BBC

Your Information and Privacy | Cookie and Browser Settings | Creating and Using Your Account | Terms of Use

The BBC Privacy and Cookies Policy (outside the UK)

Contents

1. What's in this policy?
2. What does this policy cover?

Page updated: September 2022
Version 1.0

1. What's in this policy?

- Reads well, short, clear, declarative
- Examples:
  - "We'll only send you marketing emails if you've agreed to this."
  - "You can delete your account. Your account information will be deleted immediately."

[1] https://www.nytimes.com/interactive/2019/06/12/opinion/facebook-google-privacy-policies.html [2] https://www.bbc.co.uk/usingthebbc/privacy-policy-outside-the-uk/

Don't hide or mislead

```
01110010111001111010 11
10001100101010010101 01
10101101101011011011
11101011HACKED11110110
00010101001000011111 11
10010101010101010100
11111001111111011001000
```

- Transparency is being honest with your users
- When a data breach happens, don't hide or mislead
- Have a robust communication- affected parties
- Advice, provide mitigation measures and channels to communicate

Reality check



- McKinsey report from 2021 on State of AI on data and AI practices
- 1843 professionals across the globe and across industries
- 30% aware of data ethics risks
- 17% have data governance teams
- 27% check for biased datasets

[1] https://www.mckinsey.com/capabilities/quantumblack/our-insights/global-survey-the-state-of-ai-in-2021

What does being accountable mean?

- Core principle necessary implementing all others
- Responsible for third party collaborators
- Data protection and individual rights
- Social implications at large

Who is accountable?



- Everyone :)
- Not limited to a single role
- Data scientists, legal and risk, IT, C-suite
- Corporate Digital Responsibility- the ESR for data ethics

[1] **CDR**: Lobschat, L., et al. (2021). Corporate digital responsibility. Journal of Business Research, 122, 875-888. [2] Icon made by Eucalyp from www.flaticon.com

Let's practice!

INTRODUCTION TO DATA ETHICS