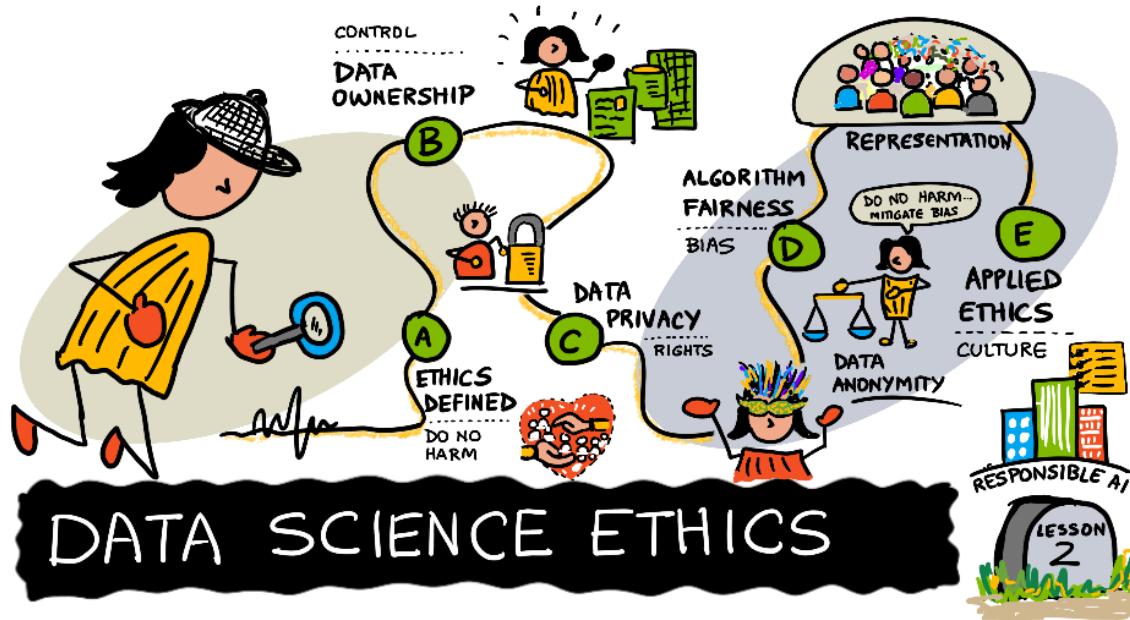


Introduction to Data Ethics



Data Science Ethics - Sketchnote by [@nitya](#)

We are all data citizens living in a datafied world.

Market trends tell us that by 2022, 1-in-3 large organizations will buy and sell their data through online [Marketplaces and Exchanges](#). As **App Developers**, we'll find it easier and cheaper to integrate data-driven insights and algorithm-driven automation into daily user experiences. But as AI becomes pervasive, we'll also need to understand the potential harms caused by the [weaponization](#) of such algorithms at scale.

Trends also indicate that we will create and consume over [180 zettabytes](#) of data by 2025. As **Data Scientists**, this gives us unprecedented levels of access to personal data. This means we can build behavioral profiles of users and influence decision-making in ways that create an [illusion of free choice](#) while potentially nudging users towards outcomes we prefer. It also raises broader questions on data privacy and user protections.

Data ethics are now *necessary guardrails* for data science and engineering, helping us minimize potential harms and unintended consequences from our data-driven actions. The [Gartner Hype Cycle for AI](#) identifies relevant trends in digital ethics, responsible AI, and AI governance as key drivers for larger megatrends around *democratization* and *industrialization* of AI.

Hype Cycle for Artificial Intelligence, 2020



gartner.com/SmarterWithGartner

Source: Gartner
© 2020 Gartner, Inc. and/or its affiliates. All rights reserved. Gartner and Hype Cycle are registered trademarks of Gartner, Inc. and its affiliates in the U.S.

Gartner®

In this lesson, we'll explore the fascinating area of data ethics - from core concepts and challenges, to case studies and applied AI concepts like governance - that help establish an ethics culture in teams and organizations that work with data and AI.

Pre-lecture quiz

Basic Definitions

Let's start by understanding the basic terminology.

The word "ethics" comes from the [Greek word "ethikos"](#) (and its root "ethos") meaning *character or moral nature*.

Ethics is about the shared values and moral principles that govern our behavior in society. Ethics is based not on laws but on widely accepted norms of what is "right vs. wrong". However, ethical considerations can influence corporate governance initiatives and government regulations that create more incentives for compliance.

Data Ethics is a [new branch of ethics](#) that "studies and evaluates moral problems related to *data, algorithms and corresponding practices*". Here, "**data**" focuses on actions related to generation, recording, curation, processing, dissemination, sharing, and usage, "**algorithms**" focuses on AI, agents, machine learning, and robots, and "**practices**" focuses on topics like responsible innovation, programming, hacking, and ethics codes.

Applied Ethics is the [practical application of moral considerations](#). It's the process of actively investigating ethical issues in the context of *real-world actions, products and processes*, and taking corrective measures to make that these remain aligned with our defined ethical values.

Ethics Culture is about [operationalizing applied ethics](#) to make sure that our ethical principles and practices are adopted in a consistent and scalable manner across the entire organization. Successful ethics cultures define organization-wide ethical principles, provide meaningful incentives for compliance, and reinforce ethics norms by encouraging and amplifying desired behaviors at every level of the organization.

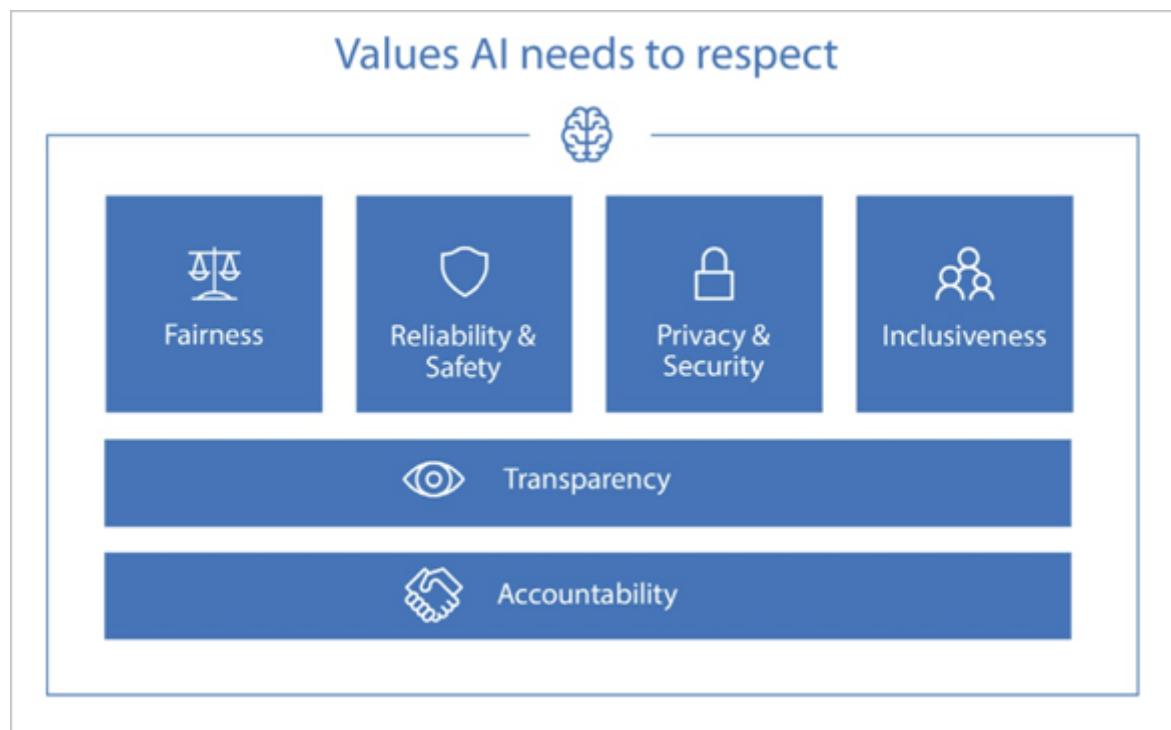
Ethics Concepts

In this section, we'll discuss concepts like **shared values** (principles) and **ethical challenges** (problems) for data ethics - and explore **case studies** that help you understand these concepts in real-world contexts.

1. Ethics Principles

Every data ethics strategy begins by defining *ethical principles* - the "shared values" that describe acceptable behaviors, and guide compliant actions, in our data & AI projects. You can define these at an individual or team level. However, most large organizations outline these in an *ethical AI* mission statement or framework that is defined at corporate levels and enforced consistently across all teams.

Example: Microsoft's [Responsible AI](#) mission statement reads: "*We are committed to the advancement of AI-driven by ethical principles that put people first*" - identifying 6 ethical principles in the framework below:



Let's briefly explore these principles. *Transparency* and *accountability* are foundational values that other principles built upon - so let's begin there:

- **Accountability** makes practitioners *responsible* for their data & AI operations, and compliance with these ethical principles.
- **Transparency** ensures that data and AI actions are *understandable* (interpretable) to users, explaining the what and why behind decisions.
- **Fairness** - focuses on ensuring AI treats *all people* fairly, addressing any systemic or implicit socio-technical biases in data and systems.
- **Reliability & Safety** - ensures that AI behaves *consistently* with defined values, minimizing potential harms or unintended consequences.
- **Privacy & Security** - is about understanding data lineage, and providing *data privacy and related protections* to users.
- **Inclusiveness** - is about designing AI solutions with intention, adapting them to meet a *broad range of human needs & capabilities*.

💡 Think about what your data ethics mission statement could be. Explore ethical AI frameworks from other organizations - here are examples from [IBM](#), [Google](#), and [Facebook](#). What shared values do they have in common? How do these principles relate to the AI product or industry they operate in?

2. Ethics Challenges

Once we have ethical principles defined, the next step is to evaluate our data and AI actions to see if they align with those shared values. Think about your actions in two categories: *data collection* and *algorithm design*.

With data collection, actions will likely involve **personal data** or personally identifiable information (PII) for identifiable living individuals. This includes [diverse items of non-personal data](#) that *collectively* identify an individual. Ethical challenges can relate to *data privacy*, *data ownership*, and related topics like *informed consent* and *intellectual property rights* for users.

With algorithm design, actions will involve collecting & curating **datasets**, then using them to train & deploy **data models** that predict outcomes or automate decisions in real-world contexts. Ethical challenges can arise from *dataset bias*, *data quality* issues, *unfairness*, and *misrepresentation* in algorithms - including some issues that are systemic in nature.

In both cases, ethics challenges highlight areas where our actions may encounter conflict with our shared values. To detect, mitigate, minimize, or eliminate, these concerns - we need to ask moral "yes/no" questions related to our actions, then take corrective actions as needed. Let's take a look at some ethical challenges and the moral questions they raise:

2.1 Data Ownership

Data collection often involves personal data that can identify the data subjects. [Data ownership](#) is about *control* and [user rights](#) related to the creation, processing ,and dissemination of data.

The moral questions we need to ask are:

- Who owns the data? (user or organization)
- What rights do data subjects have? (ex: access, erasure, portability)

- What rights do organizations have? (ex: rectify malicious user reviews)

2.2 Informed Consent

Informed consent defines the act of users agreeing to an action (like data collection) with a *full understanding* of relevant facts including the purpose, potential risks, and alternatives.

Questions to explore here are:

- Did the user (data subject) give permission for data capture and usage?
- Did the user understand the purpose for which that data was captured?
- Did the user understand the potential risks from their participation?

2.3 Intellectual Property

Intellectual property refers to intangible creations resulting from the human initiative, that may *have economic value* to individuals or businesses.

Questions to explore here are:

- Did the collected data have economic value to a user or business?
- Does the **user** have intellectual property here?
- Does the **organization** have intellectual property here?
- If these rights exist, how are we protecting them?

2.4 Data Privacy

Data privacy or information privacy refers to the preservation of user privacy and protection of user identity with respect to personally identifiable information.

Questions to explore here are:

- Is users' (personal) data secured against hacks and leaks?
- Is users' data accessible only to authorized users and contexts?
- Is users' anonymity preserved when data is shared or disseminated?
- Can a user be de-identified from anonymized datasets?

2.5 Right To Be Forgotten

The **Right To Be Forgotten** or **Right to Erasure** provides additional personal data protection to users.

Specifically, it gives users the right to request deletion or removal of personal data from Internet searches and other locations, *under specific circumstances* - allowing them a fresh start online without past actions being held against them.

Questions to explore here are:

- Does the system allow data subjects to request erasure?
- Should the withdrawal of user consent trigger automated erasure?
- Was data collected without consent or by unlawful means?
- Are we compliant with government regulations for data privacy?

2.6 Dataset Bias

Dataset or [Collection Bias](#) is about selecting a *non-representative* subset of data for algorithm development, creating potential unfairness in result outcomes for diverse groups. Types of bias include selection or sampling bias, volunteer bias, and instrument bias.

Questions to explore here are:

- Did we recruit a representative set of data subjects?
- Did we test our collected or curated dataset for various biases?
- Can we mitigate or remove any discovered biases?

2.7 Data Quality

[Data Quality](#) looks at the validity of the curated dataset used to develop our algorithms, checking to see if features and records meet requirements for the level of accuracy and consistency needed for our AI purpose.

Questions to explore here are:

- Did we capture valid *features* for our use case?
- Was data captured *consistently* across diverse data sources?
- Is the dataset *complete* for diverse conditions or scenarios?
- Is information captured *accurately* in reflecting reality?

2.8 Algorithm Fairness

[Algorithm Fairness](#) checks to see if the algorithm design systematically discriminates against specific subgroups of data subjects leading to [potential harms](#) in *allocation* (where resources are denied or withheld from that group) and *quality of service* (where AI is not as accurate for some subgroups as it is for others).

Questions to explore here are:

- Did we evaluate model accuracy for diverse subgroups and conditions?
- Did we scrutinize the system for potential harms (e.g., stereotyping)?
- Can we revise data or retrain models to mitigate identified harms?

Explore resources like [AI Fairness checklists](#) to learn more.

2.9 Misrepresentation

[Data Misrepresentation](#) is about asking whether we are communicating insights from honestly reported data in a deceptive manner to support a desired narrative.

Questions to explore here are:

- Are we reporting incomplete or inaccurate data?
- Are we visualizing data in a manner that drives misleading conclusions?
- Are we using selective statistical techniques to manipulate outcomes?
- Are there alternative explanations that may offer a different conclusion?

2.10 Free Choice

The [Illusion of Free Choice](#) occurs when system "choice architectures" use decision-making algorithms to nudge people towards taking a preferred outcome while seeming to give them options and control. These [dark patterns](#) can cause social and economic harm to users. Because user decisions impact behavior profiles, these actions potentially drive future choices that can amplify or extend the impact of these harms.

Questions to explore here are:

- Did the user understand the implications of making that choice?
- Was the user aware of (alternative) choices and the pros & cons of each?
- Can the user reverse an automated or influenced choice later?

3. Case Studies

To put these ethical challenges in real-world contexts, it helps to look at case studies that highlight the potential harms and consequences to individuals and society, when such ethics violations are overlooked.

Here are a few examples:

Ethics Challenge	Case Study
Informed Consent	1972 - Tuskegee Syphilis Study - African American men who participated in the study were promised free medical care <i>but deceived</i> by researchers who failed to inform subjects of their diagnosis or about availability of treatment. Many subjects died & partners or children were affected; the study lasted 40 years.
Data Privacy	2007 - The Netflix data prize provided researchers with <i>10M anonymized movie rankings from 50K customers</i> to help improve recommendation algorithms. However, researchers were able to correlate anonymized data with personally-identifiable data in <i>external datasets</i> (e.g., IMDb comments) - effectively "de-anonymizing" some Netflix subscribers.
Collection Bias	2013 - The City of Boston developed Street Bump , an app that let citizens report potholes, giving the city better roadway data to find and fix issues. However, people in lower income groups had less access to cars and phones , making their roadway issues invisible in this app. Developers worked with academics to <i>equitable access and digital divides</i> issues for fairness.
Algorithmic Fairness	2018 - The MIT Gender Shades Study evaluated the accuracy of gender classification AI products, exposing gaps in accuracy for women and persons of color. A 2019 Apple Card seemed to offer less credit to women than men. Both illustrated issues in algorithmic bias leading to socio-economic harms.
Data Misrepresentation	2020 - The Georgia Department of Public Health released COVID-19 charts that appeared to mislead citizens about trends in confirmed cases with non-chronological ordering on the x-axis. This illustrates misrepresentation through visualization tricks.
Illusion of free choice	2020 - Learning app ABCmouse paid \$10M to settle an FTC complaint where parents were trapped into paying for subscriptions they couldn't cancel. This illustrates dark patterns in choice architectures, where users were nudged towards potentially harmful choices.

Ethics Challenge	Case Study
Data Privacy & User Rights	2021 - Facebook Data Breach exposed data from 530M users, resulting in a \$5B settlement to the FTC. It however refused to notify users of the breach violating user rights around data transparency and access.

Want to explore more case studies? Check out these resources:

- [Ethics Unwrapped](#) - ethics dilemmas across diverse industries.
- [Data Science Ethics course](#) - landmark case studies explored.
- [Where things have gone wrong](#) - deon checklist with examples

💡 Think about the case studies you've seen - have you experienced, or been affected by, a similar ethical challenge in your life? Can you think of at least one other case study that illustrates one of the ethical challenges we've discussed in this section?

Applied Ethics

We've talked about ethics concepts, challenges ,and case studies in real-world contexts. But how do we get started *applying* ethical principles and practices in our projects? And how do we *operationalize* these practices for better governance? Let's explore some real-world solutions:

1. Professional Codes

Professional Codes offer one option for organizations to "incentivize" members to support their ethical principles and mission statement. Codes are *moral guidelines* for professional behavior, helping employees or members make decisions that align with their organization's principles. They are only as good as the voluntary compliance from members; however, many organizations offer additional rewards and penalties to motivate compliance from members.

Examples include:

- [Oxford Munich Code of Ethics](#)
- [Data Science Association Code of Conduct](#) (created 2013)
- [ACM Code of Ethics and Professional Conduct](#) (since 1993)

💡 Do you belong to a professional engineering or data science organization? Explore their site to see if they define a professional code of ethics. What does this say about their ethical principles? How are they "incentivizing" members to follow the code?

2. Ethics Checklists

While professional codes define required *ethical behavior* from practitioners, they [have known limitations](#) in enforcement, particularly in large-scale projects. Instead, many data Science experts [advocate for checklists](#), that can **connect principles to practices** in more deterministic and actionable ways.

Checklists convert questions into "yes/no" tasks that can be operationalized, allowing them to be tracked as part of standard product release workflows.

Examples include:

- [Deon](#) - a general-purpose data ethics checklist created from [industry recommendations](#) with a command-line tool for easy integration.
- [Privacy Audit Checklist](#) - provides general guidance for information handling practices from legal and social exposure perspectives.
- [AI Fairness Checklist](#) - created by AI practitioners to support the adoption and integration of fairness checks into AI development cycles.
- [22 questions for ethics in data and AI](#) - more open-ended framework, structured for initial exploration of ethical issues in design, implementation, and organizational, contexts.

3. Ethics Regulations

Ethics is about defining shared values and doing the right thing *voluntarily*. **Compliance** is about *following the law* if and where defined. **Governance** broadly covers all the ways in which organizations operate to enforce ethical principles and comply with established laws.

Today, governance takes two forms within organizations. First, it's about defining **ethical AI** principles and establishing practices to operationalize adoption across all AI-related projects in the organization. Second, it's about complying with all government-mandated **data protection regulations** for regions it operates in.

Examples of data protection and privacy regulations:

- [1974, US Privacy Act](#) - regulates *federal govt.* collection, use ,and disclosure of personal information.
- [1996, US Health Insurance Portability & Accountability Act \(HIPAA\)](#) - protects personal health data.
- [1998, US Children's Online Privacy Protection Act \(COPPA\)](#) - protects data privacy of children under 13.
- [2018, General Data Protection Regulation \(GDPR\)](#) - provides user rights, data protection ,and privacy.
- [2018, California Consumer Privacy Act \(CCPA\)](#) gives consumers more *rights* over their (personal) data.
- [2021, China's Personal Information Protection Law](#) just passed, creating one of the strongest online data privacy regulations worldwide.

 The European Union defined GDPR (General Data Protection Regulation) remains one of the most influential data privacy regulations today. Did you know it also defines [8 user rights](#) to protect citizens' digital privacy and personal data? Learn about what these are, and why they matter.

4. Ethics Culture

Note that there remains an intangible gap between *compliance* (doing enough to meet "the letter of the law") and addressing [systemic issues](#) (like ossification, information asymmetry, and distributional unfairness) that can speed up the weaponization of AI.

The latter requires [collaborative approaches to defining ethics cultures](#) that build emotional connections and consistent shared values *across organizations* in the industry. This calls for more [formalized data ethics cultures](#) in organizations - allowing *anyone* to [pull the Andon cord](#) (to raise ethics concerns early in the process) and making *ethical assessments* (e.g., in hiring) a core criteria team formation in AI projects.

Post-lecture quiz

Review & Self Study

Courses and books help with understanding core ethics concepts and challenges, while case studies and tools help with applied ethics practices in real-world contexts. Here are a few resources to start with.

- [Machine Learning For Beginners](#) - lesson on Fairness, from Microsoft.
- [Principles of Responsible AI](#) - free learning path from Microsoft Learn.
- [Ethics and Data Science](#) - O'Reilly EBook (M. Loukides, H. Mason et. al)
- [Data Science Ethics](#) - online course from the University of Michigan.
- [Ethics Unwrapped](#) - case studies from the University of Texas.

Assignment

[Write A Data Ethics Case Study](#)