**Topic 8. Big Data, Data Mining, and Machine Learning: Ethical and Privacy Issues**

Part 1

**8.1.1 Big Data for business**

1. **Discuss the potential and the concerns of big data for business**

Potential. ID.

* 1. Innovation
  2. Development Enabler

Concerns. PA.

1. Privacy
2. Anonymization
3. **Explain how the new term “data trust deficit” developed**

 Note: They earn it

Research shows that the public trust in companies which use data is lower than trust in general.

In this big data world, companies and government have to EARN citizens' trust in how they will manage and use data - and those that get it wrong will pay the price.

* + 1. **Ethical issues**

1. **List 5 methods of protecting Human Rights in the ‘Era of Big Data.’**



Pro-Law **🡪PEF** (to Protect & Enhance Fairness)

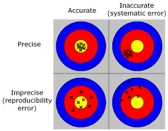
1. Stop high tech Profiling.



1. Respect the Law and its constitutional principles.



1. Protect people from inaccurate data.



1. Enhance individual control of personal information.



1. Ensure Fairness in automated decisions

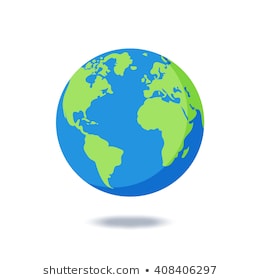


1. **Provide an example of a concern for each of 3 main areas of privacy issues:** 
   1. Customer Profiling

A picture containing clipart

Description automatically generatedFamous pregnancy story

* 1. Group Privacy

Global privacy laws.

* 1. Data Security

 Big breach

1. **Discuss what constitutes informed consent**

 Note: How news feed affects emotional state

Traditional methods of data collection require the explicit consent of respondents, stating clearly the purpose and objectives of the data collection.

The advent of IoT has challenged this approach, blurring the borders of what can be considered informed consent to the use of personal data.

Users are signing consent forms by checking the box and not really reading the consent statement.

Facebook Newsfeed example: Facebook did a social experiment that could alter the emotional state of users. Negative newsfeeds could cause the user to post negative comments. It was carried out without informing these users.

1. **Provide an example of how to improve the VERACITY of data.**



Note: Street Bump

In Boston, there was a case of implementing the **‘Street Bump’**. An app to report potholes in the streets. Where it was most needed was in lower income areas where people did not have smart phones to report potholes.

The 4 V’s of Big Data: volume, variety, velocity, and veracity. Think track race image. There is one “V” that we stress the importance of over all the others—veracity.

The trustworthiness of data. The digital divide is causing problems with this.

Data veracity is the one area that still has the potential for improvement and poses the biggest challenge when it comes to big data.

With so much data available, ensuring it’s relevant and of high quality is the difference between those successfully using big data and those who are struggling to understand it.

* + 1. **Ethics Test**
    2. **List 6 Questions which Ethics Professionals within an organization/company using big data can ask themselves.**

’s

HP’s Are Safe & Effective… Third parties say.

1. How does the company use Big Data, and to what extent is it integrated into strategic planning?
2. Does the organization send a Privacy Notice when personal data are collected?
3. Does my organization Assess the risks linked to the specific type of data my organization uses?
4. Does my organization have Safeguards in place to mitigate these risks?
5. Do we make sure that the tools to manage these risks are Effective and measure outcomes?
6. Do we conduct appropriate due diligence when sharing or acquiring data from third parties?

Part 2

**8.2.1 The nature of business risks of Artificial Intelligence (AI)**

1. **List three main features characterizing artificial intelligence (AI)**

LRI.

1. Learning – the ability to acquire information
2. Reasoning – the ability to apply information and reach conclusions
3. Iterative – the ability to change process when the new information is acquired
4. **List three immediate risks of artificial intelligence (AI)**

EWT. Ewait.

1. Ethics Risk – certain applications can lead to ethical lapses
2. Workforce Risk – automation can lead to deskilled workforce
3. Technology Risk – black box algorithms are difficult to ID in cyber-attacks.

**8.2.2** **Values that form the cornerstone of an ethical framework or artificial intelligence (AI) in business**

1. **Discuss each of the following as they impact the ethical nature of applications of artificial intelligence (AI) in business.**

**ARTIFICIAL**

1. Accurate results
2. Respect of privacy
3. Transparency and openness
4. Interpretability of algorithms
5. Fairness to stakeholders
6. Integrity and due diligence
7. Control of humans relative to machines
8. Impact of a new technology
9. Accountability assignment
10. Learning about how the AI technologies work

**8.2.3 The role of business decision makers**

1. **List 5 measures organizations can take to minimize the risk of ethical lapses due to improper use of AI technologies**



Note: skyRET. Research, Ethics, and Tools will Empower & Engage the machines.

1. Research. R
2. Ethics tests for AI machines. E
3. Design Tools for big decisions. T
4. Engage with 3rd parties for the design of AI algorithms only if they comply
5. Empower people through specific training courses
6. **List some questions addressing the use of AI that could be included in a code of ethics**

Code of ethics

A close up of a newspaper

Description automatically generated

Part 3

* + 1. **General Data Protection Regulation (GDPR)**

1. **Describe the primary purpose of the GDPR**

The most important change in data privacy in the last 20 years. It has updated legislation for the digital age. It sets articles for the legal capture, use and transition of personal data through organizations.

1. Give back control to the individual over how organizations use personal data
2. Harmonize privacy laws across Europe

This should make it easier for EU citizens to understand how their **data** is being used, and also raise any complaints, even if they are not in the country where its located.

1. **Describe the key changes in data protection regulation including the meaning of:**

A screenshot of a cell phone

Description automatically generated

* + 1. **Separating ethics and compliance**

1. **Distinguish between two types of threats of personal data breaches**
   * 1. External
     2. Internal
2. **Discuss people ‘risk’**

Securing physical systems fully still cannot mitigate a poorly trained individual with a desire or capacity to act against the values.



1. **List key questions around the role an ethical CULTURE plays in preventing data breaches**

KEY CULTURE QUESTIONS BOX 1 IN ARTICLE

* Have we decided at a Board level how to respond to threats and breaches?
* Do individuals understand the communication standards?
* Do we make decisions with openness, transparency, and honesty?
* Do we have a strategy to maintain trust with stakeholders?
* What monitoring is in place to identify threats to culture?
* Are all individuals trained and monitored appropriately?
* Are individuals rewarded for doing the right thing?
* Do we have culture aligned well with third parties?
* Which employees can be granted special access?
* If an employee speaks up, do we have procedures in place?
  + 1. **Maintaining privacy of personal data**

1. **Describe how an organization must build awareness regarding employee’s roles in protecting data**

* It is crucial that all individuals are aware of their role in protecting data.
* All employees should know how to handle the risk and how to respond.
* Establishing boundaries and clarity to all parties involved is important

Example of how the organization should react for request to be forgotten…

A screenshot of a cell phone

Description automatically generated

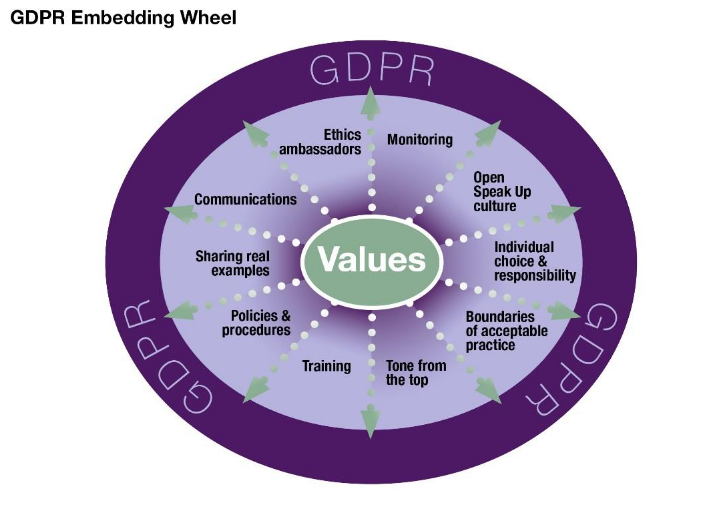
1. **Discuss liability if the 72-hour notification is missed**

One of the more notable provisions of the GDPR is Article 33 or the mandatory 72-hour breach reporting requirement.

Article 33 dictates that, in the event of a personal data breach, data controllers notify the appropriate supervisory authority “*without undue delay and, where, feasible,****not later than 72 hours****after having become aware of it.*”

A breach requires notification to the national regulator. (ICO for the UK).

**8.3.4 The GDPR Embedding Wheel (General Data Protection Regulation)**



1. **Describe how the tone from the top can help foster an ethical culture and compliance with the GDPR**

Leaders should convey, through their actions, that the organization’s reputation and long-term success are more important than short-term gains.

An organization’s board of directors and management must comply with the organization’s rules. Creating clear boundaries for risks and ethical business standards provides a framework for employees to make the right decisions.

1. **Describe how establishing the boundaries and standards can help foster an ethical culture and compliance with the GDPR**

What policies and procedures are in place to protect the ethical standards of The Speak Up process? The policies should provide clarity.

1. **Describe how communication and training can help foster an ethical culture and compliance with GDPR**

What information will be collected and from where.

1. **Describe how choice/hire of the individual can help foster or hinder an ethical culture and compliance with the GDPR**

Character evaluation should be a part of hiring, retention, and promotion practices. Employees need to understand what the organization values from the start, and this should be embedded into the performance evaluation and reward system.

1. **Describe how monitoring outcomes can help foster and ethical culture and compliance the GDPR**

Leaders should regularly examine complaints by employees and other stakeholders to determine whether the organization’s operations are in alignment with its values.

**8.4.1 Doing Good Data Science** New.

1. **Identify aspects of putting ethical principles into practice**

﻿﻿Over the past year, there has been a great discussion of data ethics:

1. Fake news
2. Targeted advertising
3. Algorithmic bias
4. The effect of data products on individuals and society
5. **Identify appropriate tools for implementing sound ethical practices.**
6. OATHS and codes of conduct are helpful but the process in developing the oath is most important.



1. CHECKLISTS are a simple and effective way of developing the process.

 Note: The checklist manifesto.

**8.4.3 The five C’s.**

1. **Recognize the main point behind each of the five ‘C’s.**  See below for main points.

Note: Consent clearly for consistent and controlled consequence.

* 1. Consent
  2. Clarity
  3. Consistency
  4. Controls and transparency
  5. Consequences and harm

1. **Recognize when an organization does and does not follow one of the “C’s” framing guidelines**

Consent

﻿﻿There should be an agreement at start. Hospital forms. Many collect without consent. Equifax and Experian do not get consent. Google maps, Comcast, Samsung voice commands do not ask for consent.



Clarity

Related to consent. You have to be clear about what you are consenting. The need isn’t just to get consent but to inform users what they’re consenting to. That’s clarity. Eventbrite screwed up. They would send teams of videographers to events.

A rainbow over a crowd of people

Description automatically generated Eventbrite event

Consistency

﻿You can’t trust someone if they are unpredictable. A major part of the shift in data privacy rights is moving to give users greater control of their data. Cambridge Analytica was an inconsistency for Facebook. Misuse of data after game “this is your digital life.”



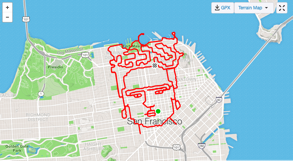
Controls and transparency

﻿﻿Individual should have control of their data. Understand what is happening to your data and how you can control it. Can you change your statuses? Remove your data? GDPR was created in 2018 to help with this. Remove data at your request.



Consequences and harm

**﻿**What are the consequences of sharing data?In 2018, STRAVA opened up their data to allow users to discover new places to run or bike. Strava didn’t realize that members of the US military were using GPS-enabled wearables, and their activity exposed the locations of bases and patrol routes in Iraq and Afghanistan.

 Note: A Frida jog

1. **Explain how to IMPLEMENT the five “C’s”.**

* **﻿**It’s the responsibility of the entire TEAM. Designers, Marketing, Sales, Product Managers, and Executives.
* **﻿**The five Cs need to be part of every organization’s culture. Product and design reviews should go over the five Cs regularly. They should consider developing a checklist before releasing a product to the public.
* In short, it’s about taking responsibility for the products that are built. The five Cs are a mechanism to foster dialogue to ensure the products “DO NO HARM.”

**8.4.4 Taking responsibility for our creation** (Ch. 4. Reckoning)

1. **Identify major issues in ethics and security training.**

* Ethics and security must be at the heart of the curriculum and not as an elective.
* They need to be coupled with ethical tools.
* It can’t stop at graduation. Employers must have regular forums concerning ethics.

1. **Argue for a method of developing guiding principles**

Have a checklist.

 Note: The checklist manifesto. Atul Gawande.

1. **Describe how to build ethics into a data-driven Culture. Louikies, Manson, Patil.** 
   1. Identify four methods for doing so … plus two more…
   2. An individual most be empowered to stop the process.

* Andon cord



* 1. Anyone should be able to escalate.
* Department of State has a Dissent Channel



* 1. An ethical challenge should be a part of the hiring process.
* Interview question: What proxy data can you use for race?



* 1. ﻿Product reviews must answer questions about the product’s impact.
* Teams need to think about impact and consequences of data



* 1. Describe the ideal role of teams and corporations

1. **﻿**Teams must reflect diversity of thought, experiences, race, and background.

* Prevent notorious case of Asian passports



1. ﻿Corporations must make their own principles clear.

* Google states “Don’t be evil.”
* Minimal viable projects (MVPs) should be tested with a diverse set of candidates.



1. **Discuss the regulatory environment for data and new technologies in terms of consent**

* Nuremberg Code
* 1974 National Research Act
* 1975 Declaration of Helsinki
* Federal Trade Commission. (FTC).
* Nuclear Regulatory Commission. (NRC).
* Federal Food and Drug administration. (FDA).
* Consumer Finance Protection Bureau. (CFPB).
* Institutional Review Board (IRB)
* Common Rule. Revision of Rule of Helsinki.
* International Review Board (IRB)

﻿It’s impossible for policy makers to iterate quickly enough to catch up with the newest technology.

﻿Another problem is that the committees that make policy often lack experts with the necessary technical background.