DATA301
29/07/2024

Assignment 1

Viy Moodley (Student ID: 300565283)

Question 1

a. The Reserve Bank of New Zealand (RBNZ) is introducing a new data collection called Loan-Level Data (LLD). The aim of this collection is to gather highly detailed, anonymised information about individual loans issued by banks in New Zealand. The LLD's primary purpose is to enhance RBNZ's analytical capabilities and data-driven decision-making, particularly for monitoring financial stability and understanding monetary policy impacts. The key items collected are:
    - ***Loan Dates:*** Maturity date (when the final loan payment is due) and next interest rate change date (when the interest rate will be reviewed)
    - ***Customer Information:*** Customer type (individual or business) and customer residency (whether the customer lives in New Zealand or overseas) ("Loan-level data collection - Reserve Bank of New Zealand - Te Pūtea Matua")
    - ***Loan Information:*** Interest rate type (floating or fixed) and loan balances (amount left to be repaid). (Reserve Bank of New Zealand, 2024)

b. The LLD collection will not include items that can directly identify individuals, such as names or account numbers. This is to maintain privacy and confidentiality by protecting personal information from exposure. Direct identifiers could lead to the identification of specific people, which would pose significant privacy concerns for individuals. The focus on anonymised data allows RBNZ to analyse trends and risks at a granular level while ensuring privacy since individual borrowers cannot be identified or traced through the collected data. (Reserve Bank of New Zealand, 2024)

c. Risks
    - **Privacy**:

- *Re-identification Risk*: Even though direct identifiers are removed, there is still a possibility that individuals could be re-identified by combining the anonymised data with other data sources. For example, unique combinations of loan dates, customer type, residency, and loan information could be used to trace back to specific individuals. Additionally, if anonymised LLD data is linked to other datasets containing identifiable information, there is a potential risk that individual privacy could be compromised.

- **Security**:
    - *Data Breaches:* Large volumes of loan-level data (both at rest and in transit) tend to be vulnerable to cyberattacks or data breaches. Unauthorised access to this data could result in sensitive information being exposed or misused.
    - *Insider Threats:* Employees or contractors with access to the data might intentionally or unintentionally misuse the data. This could involve leaking data to unauthorised parties, tampering with or using the data for personal gain.

- **Confidentiality:**
    - *Misuse of Data:* Anonymised data might still be sensitive, especially if it involves large or influential businesses. If accessed, such information could be used outside of its intended scope and could harm the business interests of the data subjects.
    - *Unregulated Compliance:* Ensuring that the LLD collection follows all relevant data protection and privacy regulations is critical. Making this data available to entities who do not comply could result in legal repercussions and loss of trust from the public and stakeholders.
    - *Third-Party Access:* If the data is shared with third parties (e.g., researchers, international banks or other government agencies), there is a risk that these parties may not adhere to the same standards of confidentiality and data protection, leading to potential data misuse or leaks.

Question 2

a. Ransomware is a type of malicious software that holds a user's data or device 'hostage', with access only being restored after a ransom is paid to the attacker.
b. Availability. I'd say that availability is the most likely classification of CIA to describe ransomware. When ransomware infects a system, it typically encrypts files or locks users out of the system entirely, making the data or system unavailable to the user. The attacker then demands a payment to restore access to the encrypted data.
c. In May 2021, Colonial Pipeline, a key fuel supplier for the East Coast of the United States, was hit by a major cyberattack. On May 7, 2021, the company detected a ransomware attack by the group DarkSide, believed to be based in Eastern Europe. DarkSide gained access to the Colonial Pipeline's network on May 6th and within two hours, they managed to steal 100 gigabytes of Data. The Next day the attackers used ransomware to encrypt Colonial Pipeline's data and demanded 75 Bitcoins (about $4.4 million) for the decryption key. In response, Colonial Pipeline shut down its entire pipeline operations to prevent further spread of the ransomware.
    - **Key dates:**
    - May 6th, 2021: DarkSide gains access to the Colonial Pipeline Network and steals 100 gigabytes of data
    - May 7th, 2021: Colonial Pipeline discovered the ransomware attack and shut down the entire pipeline to contain the threat.
    - May 9th, 2021: Joe Biden (US President in 2021) declares the attack a national emergency.
    - May 12th, 2021: Pipeline restarts and normal operations resume. (Kerner, 2022)

    The shutdown led to severe disruptions in fuel supply, causing fuel shortages, long gas station lines, soaring fuel prices, and panic buying across the East Coast. The transportation and aviation sectors were particularly affected due to their reliance on the pipeline's fuel. (Osborne, 2021)

    Colonial Pipeline decided to pay the ransom to regain access to their data. After receiving the decryption key, the company resumed operations on May 12, 2021. The U.S. Department of Justice later announced that it had recovered a significant portion of the ransom payment, reflecting ongoing efforts to combat cybercrime and recover stolen funds.
    (Sanger, Krauss, & Perlroth, 2021; U.S. Department of Justice, 2021)

Question 3

a. The purpose of the EU Artificial Intelligence (AI) Act is to regulate AI technologies by categorising them into three risk levels: unacceptable risk, high risk, and low risk. The Act aims to ensure that AI applications are safe and trustworthy by banning those with unacceptable risks, imposing strict requirements on high-risk applications, and leaving most low-risk applications unregulated.

The Act applies to any organisation or entity developing, selling, or using AI within the EU, as well as to non-EU organisations whose AI systems impact individuals within the EU. It has the potential to set global standards for AI regulation, like how the GDPR influenced global data protection practice

b. The EU AI Act is likely to have a significant impact on AI practices in New Zealand, especially for businesses and organisations that run or have markets in the EU. New Zealand companies developing or using AI technologies will need to ensure their products and services follow the requirements of the Act, particularly for high-risk applications. This could lead to increased pressure and adoption of more rigorous standards and practices in AI development and deployment, aligning with EU regulations to avoid legal and commercial barriers.

   Additionally, the principles of Māori Data Sovereignty, which emphasise the control and governance of data by Māori communities, could be influenced by the EU AI Act. The Act's focus on ethical AI and the protection of individual rights may resonate with the goals of Māori Data Sovereignty, with practices that respect and protect Māori data. In keeping with Māori Data Sovereignty, New Zealand will likely have to develop an AI Act of its own, to ensure that Māori data is considered, protected and fairly used. NZ will likely need to develop standards and processes that are compliant with international standards and culturally appropriate and respectful of indigenous data governance principles.

c. Social scoring involves evaluating individuals' behaviour and characteristics to assign them a score that can affect their access to services and opportunities. This score is often based on data collected from various sources, such as social media, financial records, and personal behaviour, which can then be used to make decisions about an individual's eligibility for loans, jobs, insurance, and other services. **Ethical Data Principles that are violated by social scoring:**
   - **Privacy and Data Protection:** Social scoring involves thorough data collection and analysis, often without individuals' explicit consent. This violates the principle of respecting individuals' privacy and protecting their personal data.
   - **Fairness and Non-Discrimination:** Social scoring can lead to unfair discrimination based on biased data and algorithms. It may perpetuate existing social inequalities and biases, resulting in unjust treatment of certain groups or individuals.
   - **Transparency and Accountability:** Social scoring systems often lack transparency in how scores are calculated and used. This undermines accountability and individuals' ability to understand and challenge decisions made about them.
   - **Human Autonomy:** Social scoring can infringe on individuals' autonomy by using automated systems to make decisions that significantly affect their lives without their input or control.

d. The EU AI Act explicitly prohibits social scoring in Article 5(1)(c): "*(c) The placing on the market, putting into service, or use of AI systems by public authorities or on their behalf for the evaluation or classification of the trustworthiness of natural persons over a certain period based on their social behaviour or known or predicted personal or personality characteristics, with the social score leading to detrimental or unfavourable treatment of certain natural persons or whole groups thereof in social contexts unrelated to the contexts in which the data was originally generated or collected.*" (EU Artificial Intelligence Act)

Question 4

a. Data governance refers to the overall management of the availability, usability, integrity, and security of the data employed in an enterprise. It involves a set of processes, policies, standards, and metrics to ensure the effective and efficient use of information in enabling an organisation to achieve its goals. (Stedman, n.d.)

   In the context of indigenous and Māori data sovereignty, governance means that Indigenous Peoples have the right to control the data that is about them or their lands. This includes developing cultural governance protocols and actively taking part in the stewardship and access of this data, especially in the context of Indigenous Knowledge.

   Data ownership is about who has legal rights and control over the data, like Indigenous communities having rights to their own information. Data management involves the day-to-day tasks of storing, retrieving, and maintaining data. Data governance, on the other hand, sets the rules and policies for how data should be handled, ensuring that both ownership rights and management practices are aligned with ethical standards and respect for stakeholders.

b. Professor Tahu Kukutai defines Māori Data Governance as the process of putting Māori data in Māori hands. The aim is to enable Māori to have control and authority over data related to them, their environments, relationships, culture, and identity. This governance ensures that data is managed in a way that is *tika* (correct and fair), safe and supports the flourishing of Māori as people. The essence of Māori Data Governance lies in putting Māori data into Māori hands and allowing Māori to decide how their data is handled.

   The importance of Māori Data Governance stems from the historical context of colonisation in which data about Māori had been collected and used to serve the agendas, priorities, and needs of others, rather than those of the Māori community. Māori Data Governance looks to reframe this dynamic by placing Māori at the centre of decision-making about their data's usage, protection, and the stories it tells. In doing so, Māori Data Governance is important to rectify the imbalances of the past and ensure that data serves to benefit Māori while respecting their sovereignty and enhancing their well-being.

c. CARE Principles of Indigenous Data Governance:
    - **Collective Benefit:** The CARE principle of Collective Benefit emphasises that data ecosystems should function in ways that allow Indigenous Peoples to derive benefits from the data. By deciding not to share their data openly, Te Hiku Media ensures that the benefits of the data remain within the Māori community. This decision allows them to control how the data is used to support community goals, enhance cultural preservation, and ensure that any derived benefits are directed back to the community.
    - **Authority to Control:** This principle underscores the importance of recognising Indigenous Peoples' rights and interests in their data and empowering their authority to control it. It includes deciding how Indigenous lands, territories, resources, knowledge, and geographical indicators are represented within data. Te Hiku Media's choice reflects a commitment to maintaining control over their data. By not sharing their data with open-source projects, they protect their right to decide how their language and knowledge are represented and used. This safeguards against potential misuse or misrepresentation that could arise if the data were openly accessible without appropriate controls.
    - **Responsibility**: Responsibility involves sharing how data are used to support Indigenous self-determination and collective benefit. It also includes accountability through meaningful and openly available evidence of these efforts. Te Hiku Media takes responsibility for the stewardship of their data, ensuring it is used in ways that directly support the self-determination and collective benefit of the Māori community. By keeping the data within their control, they can more effectively monitor and manage its use, ensuring alignment with community values and goals.
    - **Ethics**: The Ethics principle asserts that the rights and wellbeing of Indigenous Peoples should be the primary concern throughout the data life cycle and across the data ecosystem. Ethical considerations are paramount in Te Hiku Media's decision. Protecting the rights and wellbeing of the Māori community involves ensuring that their data is not exploited or used in ways that could harm their cultural integrity or economic interests. By restricting access, Te Hiku Media prioritises the ethical handling of the data, ensuring it is used in a manner that respects and promotes Māori values and wellbeing.

d. The tension between open data and curated data lies in balancing transparency with security and privacy. Open data is available to anyone, it promotes transparency and innovation but can expose sensitive information if not properly managed. Curated data restricts access to trusted users to protect privacy and ensure data integrity, but this can limit the broader benefits of data availability.

    For example, the health sector, patient data is often curated due to privacy concerns. While researchers may receive help from open access to this data to advance medical research, patient confidentiality must be kept. Thus, only authorised individuals can access sensitive health data to ensure that it is used

ethically and responsibly, highlighting the tension between the need for openness and the need for privacy and control.

Bibliography

1. EU Artificial Intelligence Act. (n.d.). *Article 5: Prohibited Artificial Intelligence Practices* . Retrieved from EU Artificial Intelligence Act: https://artificialintelligenceact.eu/article/5/
2. Greubel, A., Andres, D., & Hennecke, M. (2023, April 28). Analyzing Reporting on Ransomware Incidents: A Case Study . *Social Sciences, 12*(5).
3. Kelly, S., & Resnick-ault, J. (2021, June 9). *One password allowed hackers to disrupt Colonial Pipeline, CEO tells senators* . Retrieved from Reuters: https://www.reuters.com/business/colonial-pipeline-ceo-tells-senate-cyber-defenses-were-compromised-ahead-hack-2021-06-08/
4. Kerner, S. M. (2022, April 26). *Colonial Pipeline hack explained: Everything you need to know*. Retrieved from TechTarget: https://www.techtarget.com/whatis/feature/Colonial-Pipeline-hack-explained-Everything-you-need-to-know
5. Morrison, S. (2021, June 9). *How a major oil pipeline got held for ransom*. Retrieved from Vox: https://www.vox.com/recode/22428774/ransomeware-pipeline-colonial-darkside-gas-prices
6. Office of Public Affairs. (2021, June 7). *Department of Justice Seizes $2.3 Million in Cryptocurrency Paid to the Ransomware Extortionists Darkside* . Retrieved from U.S. Department of Justice: https://www.justice.gov/opa/pr/department-justice-seizes-23-million-cryptocurrency-paid-ransomware-extortionists-darkside
7. Osborne, C. (2021, May 13). *Colonial Pipeline ransomware attack: Everything you need to know*. Retrieved from ZD NET: https://www.zdnet.com/article/colonial-pipeline-ransomware-attack-everything-you-need-to-know/
8. Sanger, D. E., & Perlroth, C. K. (2021, May 13). *Cyberattack Forces a Shutdown of a Top U.S. Pipeline*. Retrieved from The New York Times: https://www.nytimes.com/2021/05/08/us/politics/cyberattack-colonial-pipeline.html
9. WALLIX. (n.d.). *What Happened in the Colonial Pipeline Ransomware Attack*. Retrieved from WALLIX: https://www.wallix.com/what-happened-in-the-colonial-pipeline-ransomware-attack-2/#:~:text=The%20DarkSide%20group%20operated%20through,in%20order%20to%20regain%20access.