Critical Analysis of the Menlo Principles Through the Lens of Normative Ethics

In the bounds of Artificial intelligence-related technologies, the Menlo Principle serves as a guide for navigating the ethical scope of AI creations and implementations. Based on normative ethics, these principles target to guide developers to make sure that AI technologies are developed and used in approaches that follow ethical principles and enhance the welfare of society. This essay will aim to interpret and discuss deeper into each of Menlo's Principles, analyzing the real-life implications and critically evaluating their effectiveness from the perspective of normal ethics.

Safety

The principle of safety emphasizes the constraint of safeguarding that AI systems do not cause detriment to individuals or communities. From a normative ethics standpoint, the principal associates with the deontological principle (an ethical theory that says actions are good or bad according to a clear set of rules) which underlines the significance of completing one's personal ethical obligations. Safety is essential but the attempt to achieve it may sometimes result in situations that lead to grey areas where it is hard to come to a certain ethical decision. Taking into the example of self-driving automobiles, while the primary objective is to prevent accidents and lower the rate of injuries and deaths, the implementation of safety measures such as extensive data collection leads to privacy infringement. According to J. Hunter and Duncan (2023), self-driving cars constantly require the location to pinpoint the vehicle's location. This raises the question of privacy safety if one's personal information regarding travel patterns and routines is accessed by malicious individuals.

Moreover, self-driving cars are designed with numerous cameras and microphones which are persistently recording and storing data on the manufacturer's servers, raising significant concerns about privacy infringement. This issue was highlighted in an article from The Guardians (2023) which reported on Tesla's failure to protect its database resulting in a massive leakage. The report also cited the breach containing 100 gigabytes of confidential data, including sensitive details such as private email addresses, phone numbers, employees' salaries, customers' bank details, and even the social security number of Tesla CEO, Elon Musk. These occurrences stress the importance of data protection measures to avoid the leakage or misuse of sensitive information gathered by AI systems.

Although safety is a fundamental principle, comparing safety with ethical factors like privacy and autonomy necessitates careful consideration and compromises. However, it is essential to acknowledge that before making any informed decisions, empowering users with vital information will allow associated risks that come with AI systems to be recognized.

Transparency

In the process of making AI systems management understandable and concise to users and stakeholders, transparency holds a crucial aspect as it emphasizes honesty and integrity which is also firmly established in virtue ethics. It is quite challenging especially in the realm of complex AI systems such as deep learning neural networks because transparency cultivates reliance and accountability. In algorithmic decision-making procedures, the complexity of the inherent algorithm could hinder meaningful explanations for non-professionals or certain individuals such as learners.

One compelling example of transparency issues would be China's Social Credit System which calculates a social credit score for each citizen based on various factors including social behaviour and financial responsibility. The main object of the system is to achieve social stability and enhance social trust while promoting economic development and improving governance. However, the lack of transparency around how the social credit system assesses a certain individual and determines the score remains a controversy as the information provided by the Government is commonly inaccurate and incomplete. (Lee, 2020) This opacity challenges for people to understand the system which further leads to questions about the algorithm's equity and due process.

Within institutions, the absence of transparency could lead to a lack of trust among stakeholders and hinder the accountability mechanism that holds responsible conduct and decision-making processes. As a result, practical implementation requires a balance between providing clear explanations and preventing information overload to achieve transparency.

Accountability

Accountability is one of the principles that ensures individuals and organizations are held responsible for the outcomes of AI systems. This principle resembles outcome-oriented morals which could be evaluated as actions based on their own consequences and aims to accomplish general welfare.

Promoting this principle is commendable in theory but applying it in practical use could be challenging as the system could unintentionally continue exiting disparities if the automated system was developed under biased data. For instance, in 2018, Facebook faced intensive scrutiny and criticism in the aftermath of the Cambridge Analytica scandal (Confessore, 2018). It was reported a third-party app collected millions of personal data from Facebook users without their consent with the intention to use it for tailored political marketing during the 2016 US presential election. The outrage underlined significant accountability issues as Facebook faced allegations. The organisation failed to protect user data and fell short of defending the break in court. Moreover, concerns arose in the community as there was a failure from Facebook to take responsibility for the misuse of user data by third-party developers and advertisers.

Digital platforms will always have the risk of continued privacy issues and a decline in consumer faith without clear systems of accountability and responsibility for breaches of trust. This proves that tackling accountability issues is crucial for safeguarding the protection of user privacy and maintaining credibility in the digital ecosystem. Despite that the principle of fairness and equity is a noble aspiration, and the motivation is to justify biases and serve greater to society, the realization of achieving them necessities persistent vigilance and proactive actions as determining what constitutes impartiality in different contexts could be subjective and influenced by own values.

Fairness and Equity

The principle of fairness and equity accentuates the importance of inhibiting AI systems from perpetuating or escalating existing disparities. The concept additionally associates with the principle of distributive justice which seeks to equally distribute advantages and disadvantages among all members of the community.

Although fairness and equity are essential for following ethical conduct, it can be ambiguous and complicated to fully establish the principle in the development and deployment of AI systems. For instance, COMPAS, criminal justice systems exploit AI technology to execute algorithmic decision-making. These algorithms are employed to evaluate the risk of recidivism

among individuals who are anticipating trial or sentencing. Then, the assessment system examines the person's criminal history, demographic information, and socioeconomic background to predict the likelihood of committing further offenses. However, concerns arise around the deficiency of transparency and accountability in how the algorithm processes as a study reported that the risk assessment tool was certain to falsely output incorrect labels to black defendants as being at a higher risk of recommitting offenses compared to other defendants (Angwin, Larson, Mattu, & Kirchner, 2016). This builds heavily debated questions about the equality and accuracy of the algorithm and results in doubt about the algorithm's role in the criminal justice system.

The issue in the discussed instance originates from the lack of transparency around the algorithm's evaluation process and the complexity of holding developers and users accountable for potential biases or errors in the system. If there is no clear set of instructions or guidance for oversight and responsibility, there will always be a risk of potential miscalculations and inaccuracies as in this case, faulty judgment in the criminal justice system. Therefore, though fairness and equity are a laudable goal, extensive deliberation and reflection regarding roles and responsibilities are necessary in the implementation of automated systems.

Privacy and Data Governance

Privacy and data governance emphasize safeguarding data handling practices and the significance of protecting one's privacy rights. The theory further relates to the deontological principle where individuals should be regarded as ends in themselves, rather than solely as instruments to achieve an objective.

For maintaining individual independence and dignity, respecting privacy rights is essential, but it could lead to conflict with other factors such as security and public safety. A comprehensive data protection regulation called the General Data Protection Regulation (GDPR), implemented by the European Union (EU), came into effect in May 2018 with the ambition to enhance the privacy rights of individuals within the EU and the European Economic Area (EEA). The regulation imposes strict constraints on organisations that collect, process and store personal data, irrespective of the organisation's reputation or location. The GDPR presents major principles such as transparency, purpose limitation, data minimisation, and accountability which organizations or companies must follow in the process of handling personal data. The developers were also introduced to grant the users with greater control of their personal data including the right to access, rectify or erase their data.

As the highlight of the GDPR's significant implication for business and organisations worldwide emphasize the importance of privacy and data in today's digital world, it proves that developers or organizations should adopt a proactive approach to data protection and privacy, prioritizing the rights and interests of users and maintain high standards of transparency and accountability. Although achieving privacy needs robust legal and technological protections, this leads to increased benefits for society in the context of long-term data and privacy security.

Diversity, Inclusion, and Accessibility

From time to time as the world get more diverse ethnicities and cultures, new technologies and innovations should also advance with a commitment to universal accessibility. Diversity, inclusion, and accessibility should be encouraged to developers or businesses in the development and deployment of AI systems in terms of diverse perspectives. The principle states with the ethics of care which focuses on empathy, compassion, and the recognition of mutual reliance.

While following the principle is essential for new innovations, achieving the desired system in the tech industry remains a challenge as it demands diligent effort and a meticulous approach. There are numerous real-life implications in the context of AI in this case. Firstly, companies like Microsoft have developed AI tools such as Seeing Asi, which assists individuals with visual impairment. The tool exploits the computer's camera and natural language process to recognize and describe certain objects, read text aloud, identify people, and even describe scenery. Similarly, AI-powered speech recognition technology such as Voice-activated virtual assistants like Apple's Siri, Amazon's Alexa, and Google Assistant have been developed to improve access to individuals with mobility impairments or communication disorders. These tools also enable users to interact with devices and access information using voice commands resulting in more accessible and intuitive technology. Moreover, there are AI-powered captioning and transcription services to assist individuals with hearing impairment. These services use machine learning algorithms to generate captions or transcribe spoken language automatically in real-time during video conferences, presentations, or multimedia content. This ensures certain individuals with hearing impairments enjoy equal access to information as any other users.

These examples outline how AI technology can be considered to promote greater independence and information accessibility as it follows the principles of diversity, inclusion, and accessibility. This also shows that a society could be more equitable and accessible for everyone if the automated systems could be implemented in a solution that accommodates a diverse range of users. However, like other principles discussed, diversity, inclusion, and accessibility are laudable goals, and the process of achieving to desired system that follows the principle demands systemic challenges within the tech industry and society.

Conclusion

In conclusion, the Menlo Principle consists of six different principles: Safety, Transparency, Accountability, Fairness and Equity, Privacy and Data Governance, and Diversity, Inclusion, and Accessibility. These principles provide valuable frameworks for addressing the moral implications of AI developments. Based on normative ethics, each principle has been analyzed critically, comparing its strengths and weaknesses while discussing real-life implications. As these principles also reflect moral values. Their applicable implementation involves handling difficult trade-offs and aligning competing objectives. Through continual discourse and introspective analysis, stakeholders should focus on achieving responsible AI technology development and deployment which serve the greater good of the community while following ethical guidelines before coming to an informed decision.

Bibliography

- Anon., 2016. *Machine Bias.* [Online]
  Available at: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing
  [Accessed 1 April 2024].

- Anon., 2018. *Ethics Explainer: Deontology.* [Online]
  Available at: https://ethics.org.au/ethics-explainer-deontology/
  [Accessed 1 April 2024].

- Anon., 2020. [Online]
  Available at: https://www.scmp.com/economy/china-economy/article/3096090/what-chinas-social-credit-system-and-why-it-controversial
  [Accessed 1 April 2024].

- Anon., 2023. *Report: 'massive' Tesla leak reveals data breaches, thousands of safety complaints.* [Online]
  Available at: https://www.theguardian.com/technology/2023/may/26/tesla-data-leak-customers-employees-safety-complaints
  [Accessed 1 April 2024].

- Anon., n.d. *EU General Data Protection Regulation (GDPR).* [Online]
  Available at: https://www.trendmicro.com/vinfo/us/security/definition/eu-general-data-protection-regulation-gdpr
  [Accessed 1 April 2024].

- Confessore, N., 2018. *Cambridge Analytica and Facebook: The Scandal and the Fallout So Far.* [Online]
  Available at: https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html
  [Accessed 1 April 2024].

- Guercio, A., 2024. *Digital accessibility in the era of artificial intelligence—Bibliometric analysis and systematic review.* [Online]
  Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10905618/
  [Accessed 1 April 2024].

- Jayden D, D. D., 2023. *Navigation the roahd ahead: privacy and security risks of self-driving cars.* [Online]
  Available at: https://elevenm.com.au/blog/navigating-the-road-ahead-privacy-and-security-risks-of-self-driving-cars/
  [Accessed 1 April 2024].