AI Friend or Foe? Contextualised in Terms of Computer Safety

1. **Data Privacy and Security**
   AI systems can act as both a safeguard and a threat when it comes to data privacy. On one hand, advanced AI algorithms can detect data breaches, identify vulnerabilities, and help secure sensitive information stored on computers. On the other hand, malicious AI tools, such as those employed in phishing or hacking, can exploit weaknesses in systems, compromise personal data, and breach user confidentiality. The duality of AI highlights the importance of robust cybersecurity measures to curtail its misuse.
2. **Malware Detection vs Creation**
   AI has revolutionised malware detection by identifying threats faster than traditional methods through behavioural analysis and anomaly detection. However, the same technology can be used to create sophisticated and adaptive malware that can evade conventional antivirus programmes. This arms race between protective AI and malicious AI underscores the critical need for continual development of safety protocols and defensive capabilities.
3. **Automated Decision-Making Risks**
   AI's ability to make automated decisions can improve system efficiency, such as by granting or restricting access based on machine learning models. However, poorly designed or inadequately tested AI systems may lead to unintended consequences, such as locking out legitimate users or allowing unauthorised access due to biases or errors in the algorithm. Therefore, ensuring transparency, accountability, and rigorous testing of AI systems is essential to prevent safety flaws.
4. **AI in Social Engineering Attacks**
   AI-powered chatbots and voice synthesis tools are increasingly being used in social engineering attacks, such as impersonating trusted individuals or entities to

extract sensitive information. These tools can deceive even tech-savvy users, posing a significant threat to computer safety. Raising awareness and educating users about recognising AI-driven scams is imperative to mitigate this risk and to foster a safer digital environment.