The Dark Side of AI:
The Misuse of Artificial Intelligence in Online Child Exploitation

Problem

The rapid increase in both quality and accessibility of generative artificial intelligence (AI) has resulted in new vulnerabilities in the context of online child exploitation. The development of realistic image, video, and audio generative technology has provided new opportunities for creating realistic child sexual exploitation material (CSEM).

Aims

The aims of this research were to gain a better understanding of ways in which offenders are using generative AI in the context of online child exploitation, as well as consider whether this new technology facilitates new or unique harms to victims.

Methodology

This thesis thematically analysed 396 news article sources across 49 cases of reported AI misuse, which were retrieved from three AI incident databases: the AI, Algorithmic and Automation Incidents and Controversies (AIAAIC) Repository, the Artificial Intelligence Incident Database (AIID), and the Organisation for Economic Co-operation and Development AI Incidents Monitor (OECD AIM).

Key Findings

Offenders are using generative AI for the **creation** of entirely artificial children, the **revictimisation** of past abuse survivors, and the **exploitation** of children who have not been previously abused.

Whilst online AI-facilitated exploitation has similar, if not the same harms as contact offences for victims and society, it has also made the **creation of realistic material significantly easier** through "nudification" applications. These applications have been used by adult offenders, but also by young people in schools to victimise classmates in acts of bullying and harassment.

The ease of creating AI CSEM has implications for sextortion as offenders can generate realistic nude images from regular, innocent images of real children and use them to **blackmail victims without needing to engage in a grooming process**.

Limitations

This thesis was limited to incidents that had been reported through these databases, and as such, may have missed other incidents or situations that had not been reported to these specific databases. Furthermore, the sources themselves may not have been necessarily true depictions of reality, but rather certain perspectives filtered through the media.

Implications and Recommendations

This research noted that some **offenders are choosing AI to create material as they perceive it to be victimless**, however the sources also stated that the use of AI has lasting impacts and harms on children.

Prevention in this context could involve the **clear criminalisation of any misuse of this technology** in places where it is not already illegal, as well as the specification that regardless of the generative nature of this technology, there are still very real victims being harmed.

Children who are using AI to victimise their fellow students may not understand the extent to which their actions harm others. **Education should focus on the harms of creating these images**, as well as what to do should they find themselves having been victimised.

Whilst the results of this thesis indicate the what and how of this offending, there is a distinct lack of the why, and as such **future research could analyse primary sources** rather than secondary sources by interviewing offenders who utilise AI to produce CSEM **to examine why they choose to use this technology**.