Introduction
Parents frequently struggle to control children's online exposure since they feel they have more control over their offline activities than their online exposure, due to their lack of Internet education. Parents can better comprehend cybersecurity if they are exposed to it early on and are aware of the risks it presents. In order to educate parents about their children's safety online, this study explores the appropriate standards to develop a dataset on child online protection. Deep learning has received a lot of attention recently in the subject of cybersecurity, which includes children's online safety. As a result, it is crucial to explore the permissible standards that can be applied when developing a dataset for children online protection for a deep learning model.

Problem Statement
- Parents frequently struggle to control their children's online exposure, due to lack of Internet knowledge.
- Lack of deep learning models developed to educate parents on protecting children safely online.
- Lack of public dataset that address child online safety.

Objective
Explore the appropriate standards to develop a dataset that is suitable for a deep learning model that can educate parents on ways to protect their children online.

Related Work

<table>
<thead>
<tr>
<th>Cybersecurity Framework/Model/Standard</th>
<th>Framework/Model/Strategic focus area</th>
</tr>
</thead>
<tbody>
<tr>
<td>NIST (Cybersecurity Framework CFRP)</td>
<td>Identify, protect, detect, respond, and recover</td>
</tr>
<tr>
<td>NIST (Cybersecurity Framework CFRP)</td>
<td>Risk identification, vulnerability reduction, threat reduction, consequence management, and enclave cybersecurity outcomes</td>
</tr>
<tr>
<td>Guidelines for Industry on Child Online Protection</td>
<td>Educating children, parents, and educators about children's protection and responsibilities online.</td>
</tr>
<tr>
<td>Guidelines for Parents, Guardians and Educators on Child Online Protection</td>
<td>Personal device safety and security; law, education of parents, guardians, and teachers; education for children and communication</td>
</tr>
<tr>
<td>Standards for measuring child online protection</td>
<td>Risk prone conduct of children (activities and time spent online); cyber threats and incidents; reaction of children to such incidents; and preventive measures</td>
</tr>
<tr>
<td>Child Protection in the Online/Offline Environment</td>
<td>Understand the social and cultural context (online / offline); social networks and abuse; reactions of children to online experiences, and reactions of help and assistance</td>
</tr>
</tbody>
</table>

Proposed Approach

In order to develop a dataset that is adequate for a deep learning model that can educate parents, this study suggests using the Dataset Development Lifecycle by Hutchinson et al. (2021).

Dataset Development
- Yang et al. (2023): The research framework of cement production life cycle inventory dataset for China.
- Khan et al. (2023): A Framework for Dataset Accountability.
- Pauldala et al. (2021): Dataset design and Development.
- Hutchinson et al. (2021): Dataset Development Lifecycle.
- Xu et al. (2020): Contents of the MBD Dataset.
- Pushkarna et al. (2022): Purposeful and Transparent Dataset Documentation for Responsible AI.
- Göbel et al. (2022): A holistic forensic data set synthesis framework.
- Baumgartner et al. (2020): The pushshift telegram dataset.

Conclusion

We explore the necessary standards for developing a dataset suited for a deep learning model that can educate parents in order to protect their children online. The development of the child online protection dataset can be supported by related literature on cybersecurity that highlights risk identification awareness and training, creating a more secure and age-appropriate online environment, children's risk-prone behaviour, online threats and incidents, and children’s reactions to those incidents as well as preventive measures. In addition, the Dataset Development Lifecycle was suggested to develop the dataset.

References

Dataset Development
- Should I use the dataset?
- Does the dataset meet its requirements?
- Is the dataset safe to use?
- Is the dataset likely to have previously unforeseen consequences?
- How will data be collected and handled?
- How will access be controlled?
- How will data be used?
- How will data be stored?