

How Well do LLMs Assist Parents in Assessing Child Appropriateness of Videos?

Sabila Nawshin, Ashley Phoebe Ishoel, Arun Balaji Buduru, Apu Kapadia

01 Introduction

Do you have young children browsing online content?

Children are regularly exposed to a vast and often treacherous landscape of online multimedia content.

- Online platforms lack standardized age-based moderation
- Extremely difficult for parents to consistently supervise or co-view children's media use.
- Child-safe video detection system needs to provide parents with sufficient contextual insight required to guide and moderate effectively.

We evaluate the potential for LLMs in helping bridge that gap by asking -

RQ1: How well can LLMs detect and describe the appropriateness of YouTube videos for children under the age of 7? How do LLMs reason about appropriateness?

RQ2: How well do LLMs and parental reasoning align? How can the LLM aid parents in making decisions more effectively?

03 Findings

How well do LLMs Detect Appropriateness?

When compared with the literature in ML, the LLM's performance was subpar in study 1.

72%

Accuracy vs Dataset

63%

Accuracy vs Parents

Each parent has their own definition of appropriateness, and the decision should be made by parents!

How well do LLMs describe Appropriateness?

“It is extremely helpful that the AI depicts the tone of the video and elaborates a bit more as to what is going on in the video. It makes me feel more confident that the video is in fact appropriate”

Parents found the descriptions to be helpful and informative

How do LLMs reason about appropriateness?

- Fun? Educational? Safe? Equal emphasis on all three aspects

How well do LLMs and parental reasoning align?

- Parents valued engaging, educational, and safe content - but not equally
- Education was appreciated, not required
- Parents considered broader risks, including long-term consequences

04 Implications

Ours results indicate that

- Those descriptions have the potential to simplify informed parental decision-making and actively support JME behaviors by enabling parents and children to discuss content meaningfully with contextual information about the content
- Reinforces arguments for human-in-the-loop approaches by positioning the model as a tool that enhances parental expertise.

02 Methodology

Two Consecutive studies of parents with children aged between 5 to 12

Rate this YouTube video transcript as appropriate or inappropriate for children under 7 years old and explain why in a single paragraph. "<transcript>"

110 transcripts from a public dataset (half labeled as appropriate/half inappropriate)

User Study 1:

- measure agreement with the model decision
- measure satisfaction with the explanations

This video is <model decision>
<Explanation of the decision>

How do parents reason?

How does it reason?

Can we make it align better?

Does the two align?

Ask the model to improve its own instructions iteratively

1. Model self guided

"<prompt>" "<description>". Design a better prompt

2. Parental value guided

"<prompt>" "<description>". Design a prompt with <provided parental values>

replace <prompt>
<description>

Erase history

generate
<description>

New

<prompt>
acceptable?

New <prompt>

Yes

generate
<description>

Acknowledgement

This work was funded in part by the National Science Foundation under Grant No. CNS-2207019 and the US Department of Defense (Contract W52P1J2093009). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the sponsors. I

User Study 2:

250 parents were divided into 4 groups and then were asked their opinions:

- "Model-Self Guided"
- "Parental-Value Guided"
- "Baseline" (parents watched the entire video clip),
- "Control" (parents only saw video thumbnail, title and youtube description)

Scan for full paper

