Artificial Intelegence

River Kelly

Use of AI in Content Moderation

- Over the last two decads, AI has integrated in most all online platforms
- They deliver certain task where are to complex for humans to comprehen
  - ‘Deep learning’
  - Large data inputs
- “The most significant breakthrough of machine learning in recent times is the development of ‘deep neural networks’ which enable ‘deep learning’.” (1)
- But AI still makes erros

Current Approaches

- “Effective moderation of harmful online content is a challenging problem for many reasons.” (1)
- Volume and type of content makes it harder and harder to moderate
    - Blogs, news articles and traditional writing
    - Video content
    - Photos
    - Memes
    - Different languages

The Potential Impact of AI in Online Content

1. AI can be used to improve the pre-moderation stage and flag content for review by humans, increasing accuracy (1)
   a. The word your hear is probably 那個 (in traditional characters) / 那个 (in simplified characters). It is pronounced *nàge* or *nèigel (2)*
2. AI can be implemented to synthesise training data to improve pre-moderation performance (1)
3. AI can assist human moderators by increasing their productivity and reducing the potentially harmful effects of content moderation on individual moderators (1)

Conclusions

- Google has thousands of engeneers working soley on bettering thier search algorithm
- AI has drastically improved the human condiction by providing efficiency to production
- Still some downsides to over come

Work Cited

1. Consultants, C. (2019). Use of AI in Online Content Moderation. Retrieved 2020, from https://www.ofcom.org.uk/__data/assets/pdf_file/0028/157249/cambridge-consultants-ai-content-moderation.pdf
2. https://chinese.stackexchange.com/questions/22145/very-frequently-used-word-in-mandarin-that-sounds-like-nica-or-nigah