A picture paints 1000 words, how can we better analyse them? A tutorial on how to use zero-shot learning to analyse social media images.

Benjamin Riordan[1], Joshua Millward[2], Zhen He[2], Dan Anderson-Luxford[1], Samatha Salim[1], Maree Patsouras[1], Emmanuel Kuntsche[1].

[1]Centre for Alcohol Policy Research, La Trobe University, Melbourne, Australia, [2]Computer Science & Information Technology, La Trobe University, Melbourne

Presenter's email: *b.riordan@latrobe.edu.au*

Introduction: Social media has become a popular part of our media diet and there is an exceptional amount of images shared daily. Images can be insightful and used to help answer research questions, (e.g., how common are alcohol-related posts?). However, methods used to analyse this data (e.g., content analyses) are time-intensive. Zero-shot learning, where a pre-trained model is used without any additional training, is time-efficient and requires less technical expertise. We aim to provide a tutorial on how to use zero-shot learning for image analysis and report preliminary accuracy.

Method: We created a dataset of 135 images of a model holding different alcohol and non-alcohol-related beverages. In total, nine beverage types were captured across five scenes at three different distances from the camera. We created a step-by-step tutorial (available on GitHub) of how to use two popular models (CLIP and LLaVA) to identify whether the image contains an alcohol-related beverage, and more specifically what type of beverage is, in addition to showing how to report accuracy.

Key Findings: We found that CLIP (71.85%) and LLaVA (94.81%) were relatively accurate when predicting whether alcohol was present in the image. CLIP was less accurate when predicting the type of beverage (34.07% vs LLaVA = 79.29%). Both models were more accurate when the beverage was in the foreground.

Discussions and Conclusions: Zero-shot learning with CLIP and LLaVA can be used to detect whether a beverage is alcohol-related from an image. Given that zero-shot learning requires less technical expertise than training a model it may be an exciting method for social scientists aiming to analyse large datasets.

Implications for Practice or Policy: Zero-shot learning may be used to help monitor media data for depictions of substance-related content that may violate guidelines (i.e., undeclared sponsorship from influencers).

Disclosure of Interest Statement: This project was sponsored by internal funding from La Trobe University. BR is sponsored by an ARC DECRA Fellowship DE230100659