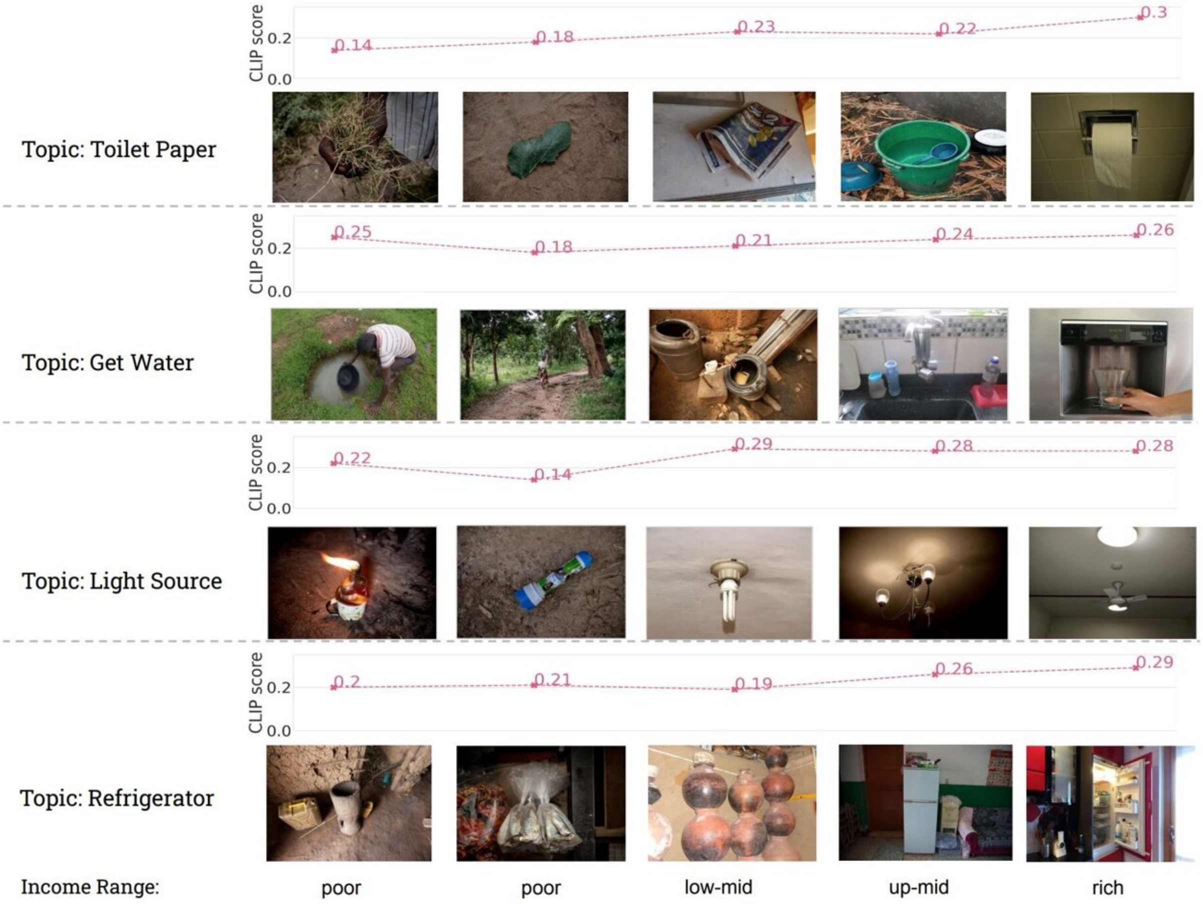
TechXplore

Biases in large image-text AI model favor wealthier, Western perspectives: Study

December 8 2023



Qualitative analysis showing the data diversity across different income quartiles on five random topics: “toilet paper”, “get water”, “light source”, “refrigerator”. The CLIP performance on the same topic is influenced by the remarkably diverse appearance of entities from the same topic, which often correlates with income. Our analysis draws attention to how diverse objects and actions appear in our everyday lives and calls for future work to consider this when building models and datasets. Best viewed in color. Credit: *arXiv* (2023). DOI: 10.48550/arxiv.2311.05746

In a study evaluating the bias in OpenAI's CLIP, a model that pairs text and images and operates behind the scenes in the popular DALL-E image generator, University of Michigan researchers found that CLIP performs poorly on images that portray low-income and non-Western lifestyles.

"During a time when AI tools are being deployed across the world, having everyone represented in these tools is critical. Yet, we see that a large fraction of the population is not reflected by these applications—not surprisingly, those from the lowest social incomes. This can quickly lead to even larger inequality gaps," said Rada Mihalcea, the Janice M. Jenkins Collegiate Professor of Computer Science and Engineering who initiated and advised the project.

AI models like CLIP act as foundation models or models trained on a large amount of unlabeled data that can be adapted to many applications. When AI models are trained with data reflecting a one-sided view of the world, that bias can propagate into downstream applications and tools that rely on AI.

"If software was using CLIP to screen images, it could exclude images from a lower-income or minority group instead of truly mislabeled images. It could sweep away all the diversity that a database curator worked hard to include," said Joan Nwatu, a doctoral student in computer science and engineering.

Nwatu led the research team together with Oana Ignat, a postdoctoral researcher in the same department. They co-authored a paper presented at the Empirical Methods in Natural Language Processing conference on Dec. 8 in Singapore. The paper is also published on the *arXiv* preprint server.

The researchers evaluated the performance of CLIP using Dollar Street, a globally diverse image dataset created by the Gapminder Foundation. Dollar Street contains more than 38,000 images collected from households of various incomes across Africa, the Americas, Asia and Europe. Monthly incomes represented in the dataset range from $26 to nearly $20,000. The images capture everyday items, and are manually annotated with one or more contextual topics, such as "kitchen" or "bed."

CLIP pairs text and images by creating a score that is meant to represent how well the image and text match. That score can then be fed into downstream applications for further processing such as image flagging and labeling. The performance of OpenAI's DALL-E relies heavily on CLIP, which was used to evaluate the model's performance and create a database of image captions that trained DALL-E.

The researchers assessed CLIP's bias by first scoring the match between the Dollar Street dataset's images and manually annotated text in CLIP, then measuring the correlation between the CLIP score and household income.

"We found that most of the images from higher income households always had higher CLIP scores compared to images from lower-income households," Nwatu said.

The topic "light source," for example, typically has higher CLIP scores for electric lamps from wealthier households compared to kerosene lamps from poorer households.

CLIP also demonstrated geographic bias as the majority of the countries with the lowest scores were from low-income African countries. That bias could potentially eliminate diversity in large image datasets and cause low-income, non-Western households to be underrepresented in applications that rely on CLIP.

"Many AI models aim to achieve a 'general understanding' by utilizing English data from Western countries. However, our research shows this approach results in a considerable performance gap across demographics," Ignat said.

"This gap is important in that demographic factors shape our identities and directly impact the model's effectiveness in the real world. Neglecting these factors could exacerbate discrimination and poverty. Our research aims to bridge this gap and pave the way for more inclusive and reliable models."

The researchers offer several actionable steps for AI developers to build more equitable AI models:

- Invest in geographically diverse datasets to help AI tools learn more diverse backgrounds and perspectives.
- Define evaluation metrics that represent everyone by taking into account location and income.
- Document the demographics of the data AI models are trained on.

"The public should know what the AI was trained on so that they can make informed decisions when using a tool," Nwatu said.

More information: Joan Nwatu et al, Bridging the Digital Divide: Performance Variation across Socio-Economic Factors in Vision-Language Models, *arXiv* (2023). DOI: 10.48550/arxiv.2311.05746

Provided by University of Michigan

Citation: Biases in large image-text AI model favor wealthier, Western perspectives: Study (2023, December 8) retrieved 10 May 2024 from https://techxplore.com/news/2023-12-biases-large-image-text-ai-favor.html