



# Flickr Africa: Examining Geo-Diversity in Large-Scale, Human-Centric Visual Data

Keziah Naggita\*  
knaggita@ttic.edu  
TTI-Chicago  
USA

Julienne LaChance\*  
julienne.lachance@sony.com  
SONY AI America  
USA

Alice Xiang  
alice.xiang@sony.com  
SONY AI America  
USA

## ABSTRACT

Biases in large-scale image datasets are known to influence the performance of computer vision models as a function of geographic context. To investigate the limitations of standard Internet data collection methods in low- and middle-income countries, we analyze human-centric image geo-diversity on a massive scale using geotagged Flickr images associated with each nation in Africa. We report the quantity and content of available data with comparisons to population-matched nations in Europe as well as the distribution of data according to fine-grained intra-national wealth estimates. Temporal analyses are performed at two-year intervals to expose emerging data trends. Furthermore, we present findings for an “othering” phenomenon as evidenced by a substantial number of images from Africa being taken by non-local photographers. The results of our study suggest that further work is required to capture image data representative of African people and their environments and, ultimately, to improve the applicability of computer vision models in a global context.

## CCS CONCEPTS

• **Computing methodologies** → *Computer vision*; **Image and video acquisition**.

## KEYWORDS

Geo-diversity, AI Ethics, Computer Vision, Machine Learning, Africa, Datasets

### ACM Reference Format:

Keziah Naggita, Julienne LaChance, and Alice Xiang. 2023. Flickr Africa: Examining Geo-Diversity in Large-Scale, Human-Centric Visual Data. In *AAAI/ACM Conference on AI, Ethics, and Society (AIES '23)*, August 08–10, 2023, Montréal, QC, Canada. ACM, New York, NY, USA, 11 pages. <https://doi.org/10.1145/3600211.3604659>

## 1 INTRODUCTION

Data collection and processing are crucial to the machine learning (ML) pipeline and are the source of many biases in AI systems, which have been shown to largely stem from a lack of diverse representation in training datasets [7]. Currently, most large-scale

computer vision datasets are collected via webscraping and subsequent data cleaning. For example, the ImageNet database ([12]; 42607 citations per Google Scholar, accessed Sept. 14, 2022) is comprised of images sourced from search engines like Google and Flickr, while the COCO dataset ([20]; 26751 citations per Google Scholar, accessed Sept. 14, 2022) is comprised of images sourced entirely from Flickr. Thus, biases inherent to Flickr influence the performance of models for visual tasks as diverse as object classification, pose estimation, instance segmentation, image captioning, and beyond. Some of these dataset biases have been explored in detail: for ImageNet and the Flickr-sourced Open Images dataset [19] it has been shown that data from India, China, and African and South-East Asian countries is vastly underrepresented despite their large populations [14]; while for COCO, data has been shown to be heavily skewed towards lighter-skinned and male individuals [37]. In particular, such biases impact the applicability of models in a global context. For instance, DeVries et al. [14] manually sourced image data from 264 globally-distributed households and demonstrated how object recognition model performance drops when applied in lower-income nations. Motivated by the popularity of datasets sourced using Flickr data, we here analyze 1.5 million geotagged images in the Flickr database to deeply explore its representation of African people and settings (see Figure 1).

In this paper, we aim to highlight the limitations of webscraping generic and human-centric<sup>1</sup> image data from Africa for ML training purposes. We analyze image data for every African nation with direct comparisons to population-matched higher-GDP European nations and show that there is far less data available from Africa. We report the distribution of African geotagged image data as a function of fine-grained, intra-national wealth estimates [8] and assess data with respect to license restrictions, population size, nominal GDP, Internet usage, and official languages. Additionally, we collect crowdsourced annotations to explore image content, and provide evidence for an “othering” phenomenon as the majority of African geotagged images we analyzed were taken by foreigners, while the opposite trend is shown for select European nations. Such results highlight the importance of considering geodiversity metrics beyond ancestry/ethnicity of individuals within images and, moreover, how the mechanisms by which images are obtained can quantitatively and qualitatively affect how the image corpus represents the world (e.g. imposing a “Western gaze”). Overall, we find that Flickr provides a very limited and skewed representation of African countries which likely contributes to many of the biases in models trained on popular, large-scale image datasets.

<sup>1</sup>That is: involving people, their interactions with each other, and/or their activities in the environments in which they live.

\*Authors contributed equally to this research.

This work was done when Keziah was an intern at Sony AI.

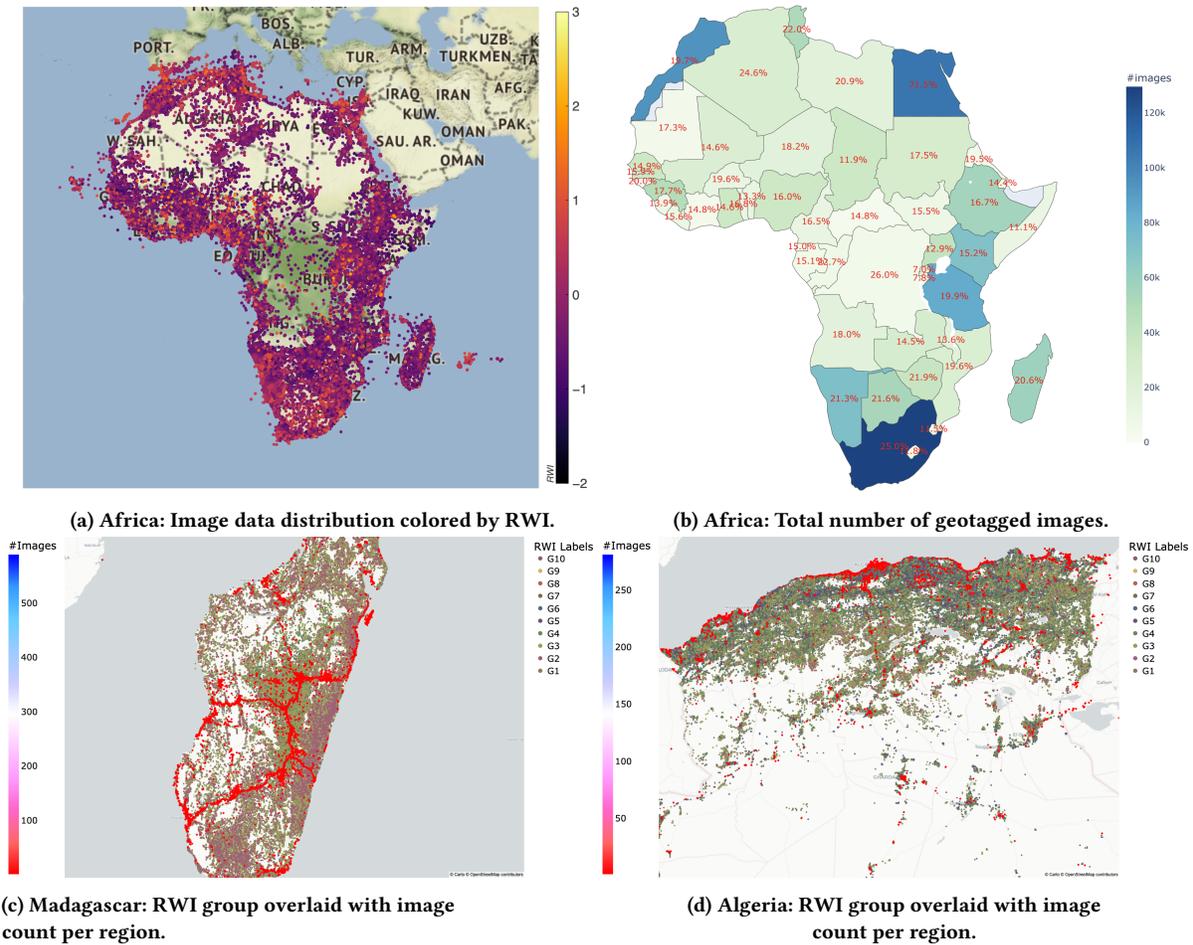
Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

*AIES '23, August 08–10, 2023, Montréal, QC, Canada*

© 2023 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-0231-0/23/08.

<https://doi.org/10.1145/3600211.3604659>



**Figure 1:** A collection of maps displaying relative wealth index (RWI) and geolocation of Flickr Africa images via country name query. Tolerance distances from geotag to nearest RWI-labeled point are: ((a, b) dist:  $\leq 300km$ ; (c, d)  $\leq 10km$ ). (b) Nations are colored according to total number of geotagged images and the percentages (rounded to one decimal place) indicate the percentage of geotagged images out of the total image count per nation. South Africa had the highest number of geotagged images and Sao Tome and Principe had the smallest number of geotagged images while Cape Verde had the highest percentage of geotagged images and Rwanda had the lowest percentage of geotagged images.

### 1.1 Related work

While prior works have explored diversity beyond Western nations [24, 28], studies of access to, and applicability of, AI systems in Africa remain limited [1, 2]. Scholars such as Abebe et al. [1] highlight the challenges of data sharing practices in Africa, such as those concerning trust, awareness, and infrastructure, and note that “The continent’s plural and at times divergent norms, practices, and traditions furthermore complicate the African data access and sharing ecosystem.” Computer vision researchers have produced diverse datasets in an effort to reduce model biases and assess fairness outcomes (e.g. [16]), with some centered specifically on geographic and contextual diversity [3, 5, 9, 10, 14, 36]. Such data collection involves trade-offs, however [35]; while manual data collection enables desired contextual diversity specifications to be met, it is expensive and frequently limits access to low-income regions (see e.g. [14]).

Thus, researchers have explored more automated methods of scraping diverse data from web platforms and public media, producing datasets such as the Geo-Diverse Visual Commonsense Reasoning dataset (GD-VCR) [36], GeoImageNet [5], Functional Map of the World (fMoW) [10], YFCC100M [26], and Open Images Extended [18], among others. Wang et al. [27] construct an ImageNet-style image data hierarchy across languages and cultures beyond English for visually grounded reasoning. While valuable, these initiatives have not deeply explored intra-national diversity, such as according to regional wealth estimates. Likewise, those datasets which utilize geolocation alone may result in a stereotypical portrayal of people in developing nations.

Geodiversity has been studied from various angles beyond dataset production. Scholars have proposed methods for measuring geodiversity in image datasets [25, 29] or performing geography-aware learning [4]. Zhao et al. [37] expose the propagation of racial and

cultural biases into model predictions, while Mandal et al. [21] study geographical bias in image search and retrieval. Denton et al. [13] highlight the importance of annotators' lived experiences on their annotation results.

Additionally, Crandall et al. [11] and Johnson et al. [15], among other researchers, have studied volunteered-geographic information (VGI) and its relation to localness in Flickr user-generated content. At metropolitan-area and individual landmark spatial scales, Crandall et al. [11] use textual and visual image data to develop a classification technique which automatically exposes the relation between location and content in six months of Flickr-scraped images. Johnson et al. [15] define four localness metrics: n-days, plurality, and location-field, to investigate the localness of user-generated content on Flickr, Twitter, and Swarm. In particular, their work assessed the Flickr-scraped YFCC100M dataset, containing images from thousands of users in the contiguous United States, whereas we focus on Africa and a few population-matched European countries. Notably, the authors found that with 31.1% recall accuracy, only 40.7% of Flickr images inspected with the "location field" localness metric (photographer self-reported location information) were local.

## 2 METHODOLOGY

### 2.1 Data Collection

**2.1.1 Flickr Africa.** For each nation in Africa, we utilized Flickr queries to construct a dataset of images and associated metadata. Using the FlickrAPI, we scraped images and associated metadata from Flickr between dates 2004-02-10 and 2022-02-10 (18 years) by querying by country name (e.g. "Togo") and the country name + people (e.g. "Togo people"), with the latter querying choice motivated by construction methods of related large image datasets (e.g. COCO, which utilizes the Flickr query "person"). We scraped Flickr data for 54 African countries: {Algeria, Angola, Benin, Botswana, Burkina Faso, Burundi, Cameroon, Cape Verde, Central African Republic (CAF), Chad, Comoros, Ivory Coast, The Democratic Republic of the Congo (DRC), Djibouti, Egypt, Equatorial Guinea, Ethiopia, Eritrea, Gabon, Gambia, Ghana, Guinea, Guinea Bissau, Kenya, Lesotho, Liberia, Libya, Madagascar, Malawi, Mali, Mauritania, Mauritius, Morocco, Mozambique, Namibia, Niger, Nigeria, Republic of Congo, Rwanda, Sao Tome and Principe, Senegal, Seychelles, Sierra Leone, Somalia, South Africa, South Sudan, Sudan, Swaziland, Tanzania, Togo, Tunisia, Uganda, Zambia, and Zimbabwe}. Utilizing Flickr metadata associated with each image, we generated 108 csv images data files (2 per country, associated with each query) with values for the following variables: {"license", "title", "datetaken", "image\_url", "country", "city", "tags", "latitude", "longitude", "rwi of nearest point", "distance to nearest rwi labelled point (km)", "latitude of nearest point", and "longitude of nearest point"}. City and country information were determined by reverse geo-locating the longitude-latitude values provided in the image metadata using open-source reverse geocode ([23]; accuracy analyses in [17]). All data is available at <https://doi.org/10.5281/zenodo.7133542>. The RWI data is described below. Total image counts were recorded and images without valid geotags were excluded.

**2.1.2 Population-matched European countries.** The data collection process was repeated for four European nations. In the interest of comparing data availability and content to higher-GDP European nations, we chose the following countries as a function of similar population size ([30, 31, 33]): Switzerland and Sierra Leone (GDP: 841.97k vs. 4.27k); Cyprus and Djibouti (GDP: 27.73k vs. 3.84k); Finland and Central African Republic (CAF) (GDP: 297.62k vs. 2.65k); and Slovenia and Lesotho (GDP: 63.65k vs. 2.56k). For all 58 countries we collected data pertaining to percentage of internet users [32], nominal GDP [31], population size [30, 33] and official languages [34].

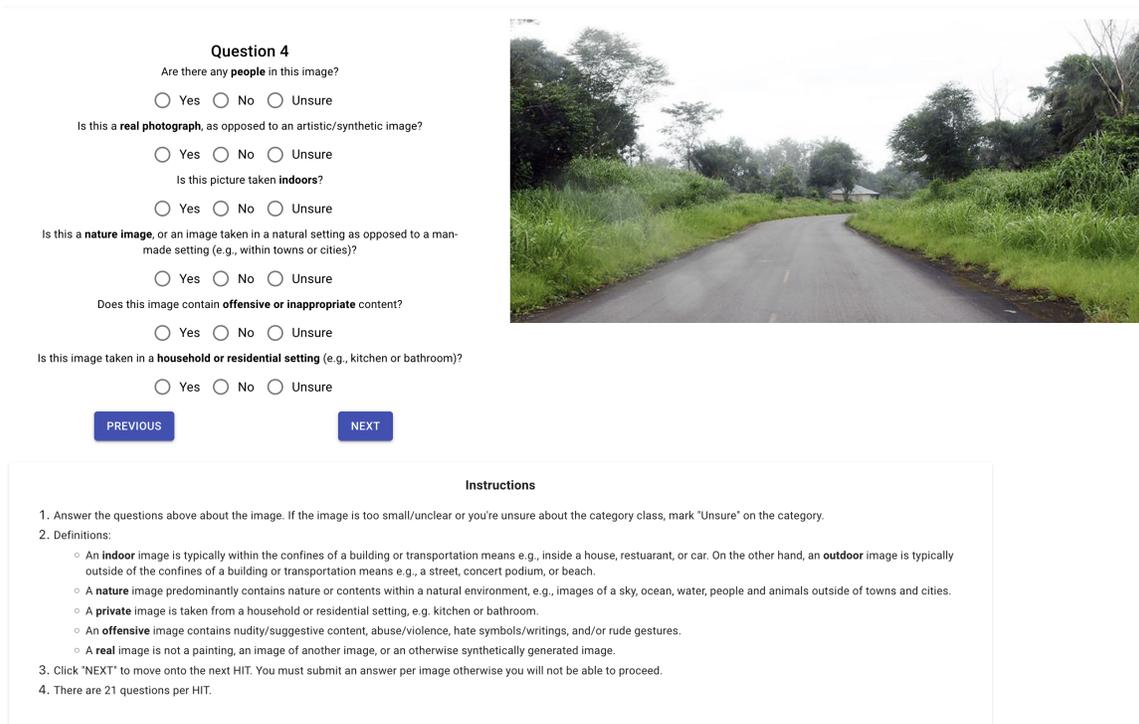
**2.1.3 Relative Wealth Estimates.** Fine-grained relative wealth estimates were associated with each geotagged image. To assess the image distribution according to local wealth estimates, we utilize the relative wealth index (RWI) data collected from Low and Middle-Income Countries (LMICs) by Facebook's Data for Good project [8]. RWI scores are normalized by nation, so the data should only be utilized for intra-national wealth analyses. The RWI dataset contains relative wealth distribution for 49 African countries, such that the following countries are excluded from our original list of nations: {Somalia, Seychelles, Sao Tome and Principe, Sudan, and South Sudan}. Therefore, when analyzing the relationship of RWI to geotagged images, these four countries are excluded. RWI data is provided in the form of 3-lettered iso-codes and the following variables are provided: "quadkey", "latitude", "longitude", "rwi", and "error"; Nominatim API [22] was utilized to assign and add variables "country", "city" to the data files. Using k-nearest neighbour, we computed the nearest RWI-labeled geographic location of each image. Figure 1a shows the distribution of RWI-labeled geotagged images with a 300km maximum tolerance limit between the image geotag and the nearest RWI-labeled location.

**2.1.4 Manual Content Annotation.** Crowdsourced annotations were collected for six additional image features.

We used Amazon Mechanical Turk (AMT) to collect annotations describing image contents. Each Human Intelligence Task (HIT) involved 21 images, with six binary questions per image as shown in Figure 2. The binary questions required the annotator label the image according to: indoor vs. outdoor setting, public vs. private setting, nature vs. manmade setting, the presence of people, real vs. synthetic image type, and offensive vs. inoffensive content. Below were our definitions of the terms or labels;

- An indoor image is typically within the confines of a building or transportation means, e.g., inside a house, restaurant, or car.
- A private image is taken from a household or residential setting, e.g., kitchen or bathroom.
- A nature image predominantly contains nature or contents within a natural environment, e.g., images of a sky, ocean, water, people and animals outside of towns and cities.
- A real image is not a painting, an image of another image, or an otherwise synthetically generated image.
- An offensive image contains abuse/violence, nudity/suggestive content, hate symbols/writings, and or rude gestures.

We compensated workers at a rate of \$15 USD/hour. We sourced each annotation from three different annotators and chose the



**Figure 2: Sample AMT task page which annotators utilized to label binary attributes pertaining to image content. Each HIT involved interaction with an introductory instructional page followed by 21 task pages similar to the one shown above.**

majority consensus value, excluding those images marked “Unsure”. To ensure high annotation quality, we recruited workers with at least 95% acceptance rating and completion of 1000+ prior tasks. We randomly inserted a gold standard image within each set of 20 standard images to assess annotator performance; if the worker failed on this test image, we discarded the annotations but still paid the worker for their contribution.

## 2.2 Limitations of Our Approach

We acknowledge four notable limitations of our method. First, we recognize that geolocation data (longitude, latitude) is inherently unreliable. Values may be modified or removed by the Flickr user or otherwise not reflect the location of capture, while reverse geolocation methods are computationally expensive and often fail, particularly with geographic locations close to region borders. This motivates our use of both geotags and country name tags for cross-validation of location, though this restricts us to fewer data samples overall. Secondly, to determine location of photographers to assess localness, we relied on photographers volunteered information of their location from their profile metadata. This doesn’t take into account confounding factors like an immigrant visiting their home nation. Additionally, some forms of geodiversity are difficult or impossible to determine from visual inspection alone, such as an individual’s gender, ethnicity, or religion. Finally, we were limited to obtaining data using only two queries, namely, by country name or country name + “people”. We anticipate future work exploring a wider variety of query terms, both in English and local languages;

here, no correlation was determined between dominant national languages and geotagged image availability.

## 2.3 Ethical Considerations

We note that although the Flickr images analyzed here are all publicly viewable, we show that most have the Flickr default license of “All Rights Reserved”. Thus, we have opted to provide image URLs in lieu of images for direct download to avoid duplication of protected content, particularly in the event that a Flickr user chooses to remove or modify the permissions of an image. We acknowledge the weaknesses of this method in terms of consent, as public Flickr images are typically not taken by those in the images (as pointed out by Birhane and Prabhu [6]); likewise, Flickr users may wish to avoid the utilization of their images for research purposes. Given that our objective is to critique large-scale image dataset curation strategies which do not respect image licenses (e.g. the methodology for generating the COCO dataset), we deemed it justifiable to perform basic analyses on protected images and to build awareness regarding widespread license violations in standard AI training pipelines.

## 3 RESULTS AND DISCUSSION

### 3.1 Data Availability and Geographic Distribution

There were very few geotagged images from Africa, as shown in Figures 3a and 3b. In terms of total geotagged image counts with

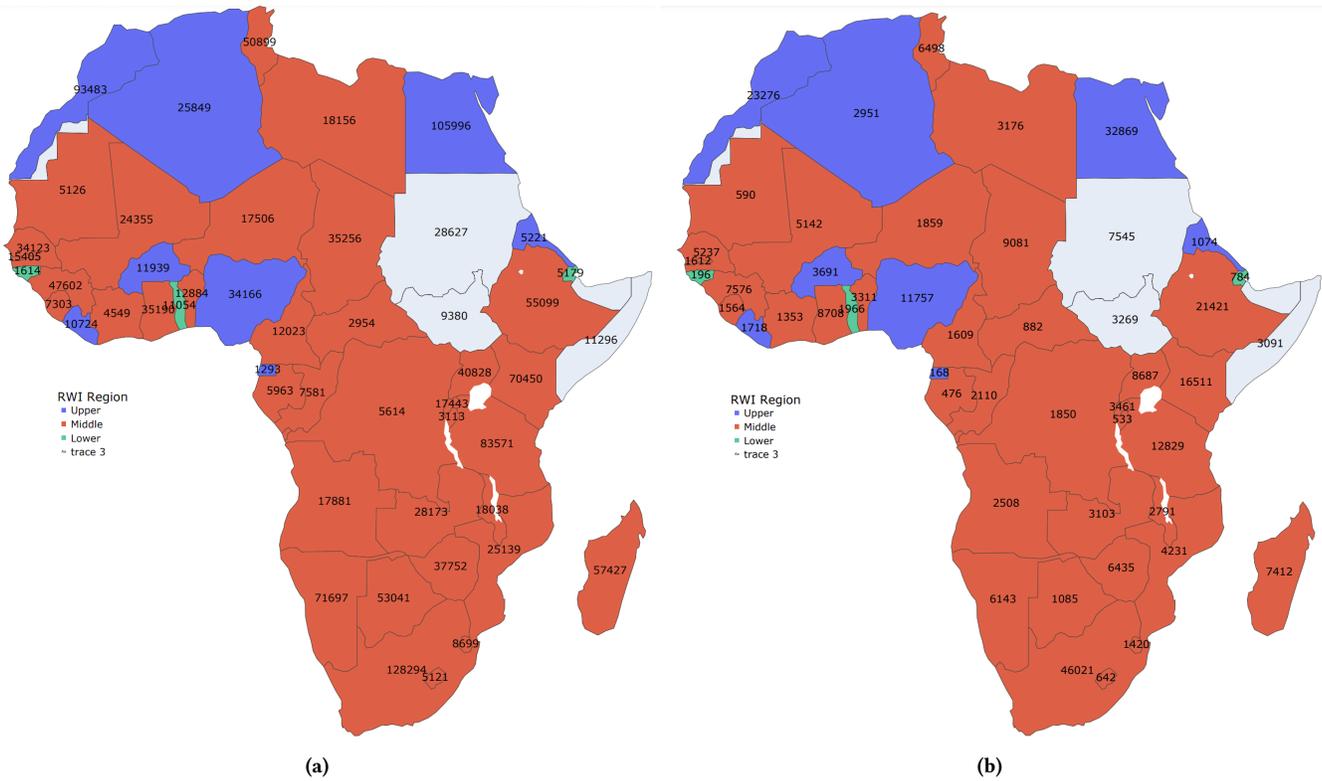


Figure 3: Geotagged image counts by nation and respective RWI regions for (a) “query-by-country name” and (b) “query-by-country name-people”. Nations are colored according to dominant RWI group (upper, middle, or lower wealth group) from which most images were sourced. Images mainly came from middle RWI groups ( $G_4, G_5, G_6$  and  $G_7$ ). The numbers denote the number of geotagged images. Countries that didn’t have RWI data are in grey.



Figure 4: Plots showing the number of geotagged images as a function of (a) the percentage of Internet users in the country, and (b) the nominal GDP of the country. Data points are colored according to national dominant language class (see Section 3.1). In general, the number of geotagged images increased with increase internet usage and GDP, with no observable trend in language class; “STP” denotes the country Sao Tome and Principe.

**Table 1: Number of geotagged and total images, and percentage (rounded to 2 decimal places) of geotagged images for the five countries with the highest percentage of geotagged images according to query type.**

| query-by-name     |            |             |            | query-by-name+people |            |             |            |
|-------------------|------------|-------------|------------|----------------------|------------|-------------|------------|
| Country           | #geotagged | #all images | %geotagged | Country              | #geotagged | #all images | %geotagged |
| Cape Verde        | 11,465     | 43,283      | 26.49      | Burkina Faso         | 3,691      | 12,990      | 28.41      |
| DRC               | 5,614      | 21,560      | 26.04      | DRC                  | 1,850      | 7,028       | 26.32      |
| South Africa      | 128,294    | 513,082     | 25.00      | Republic of Congo    | 2,110      | 8,249       | 25.58      |
| Algeria           | 25,849     | 105,254     | 24.56      | Cape Verde           | 1,333      | 5,390       | 24.73      |
| Republic of Congo | 7,581      | 33,438      | 22.67      | Botswana             | 1,085      | 4,664       | 23.26      |

**Table 2: Number of geotagged and total images, and percentage (rounded to 2 decimal places) of geotagged images for the five countries with the lowest percentage of geotagged images according to query type.**

| query-by-name |            |             |            | query-by-name+people |            |             |            |
|---------------|------------|-------------|------------|----------------------|------------|-------------|------------|
| Country       | #geotagged | #all images | %geotagged | Country              | #geotagged | #all images | %geotagged |
| Lesotho       | 5,121      | 43,282      | 11.83      | Zambia               | 3,103      | 33,063      | 9.39       |
| Swaziland     | 8,699      | 75,631      | 11.50      | Equatorial Guinea    | 168        | 2,081       | 8.07       |
| Somalia       | 11,296     | 101,989     | 11.08      | Gabon                | 476        | 7,286       | 6.53       |
| Burundi       | 3,113      | 39,888      | 7.80       | Burundi              | 533        | 9,862       | 5.40       |
| Rwanda        | 17,443     | 250,469     | 6.96       | Rwanda               | 3,461      | 76,521      | 4.52       |

query-by-name and query-by-name+people from African nations, South Africa (128,294 & 46,021) and Egypt (105,996 & 32,869) had the highest counts, and Equatorial Guinea (1,293 & 168) and Sao Tome and Principe (776 & 116) had the lowest counts. Cape Verde had the highest percentage of geotagged images (26.49%) from query-by-name and Burkina Faso (28.41%) had the highest from query-by-name+people. By contrast, Rwanda had the lowest percentage of geotagged images (6.96%, 4.52%) from both query-by-name and query-by-name+people. African nations with the highest and lowest percentages of geotagged images are summarized in Table 1 and Table 2.

**Thus**, the low number of African geotagged images indicates the ineffectiveness of Flickr scraping as a data collection methodology in this region and, therefore, a need to explore alternative geodiverse data collection methods, e.g. utilizing manual data collection.

The population-matched European countries had higher numbers and percentages of geotagged images than the corresponding African countries, as is further emphasized in Table 3. For example, with query-by-name, despite relatively similar population sizes, the percentage change of the number of geotagged images from Sierra Leone to Switzerland is 1673.48%, that is, 18× as many total geotagged images as Sierra Leone as shown in Table 3.

**Thus**, African countries had far fewer images (both geotagged and non-geotagged) than the corresponding European countries of similar population size. We recommend that computer vision experts be cognizant of this discrepancy in Flickr scraped datasets and to consider the corresponding potential for bias when training computer vision models.

We analyzed the statistical effect of factors that might potentially affect taking, uploading and tagging images on Flickr; population-size, internet usage, official language, and countries' GDP. In general, the number of geotagged images increased with population size (correlation: 0.412 & 0.538, query-by-name and query-by-name+people respectively), internet usage (0.474 & 0.385), and GDP (0.599 & 0.748); the latter two are shown in Figure 4. An investigation of the effect of these variables on the number of geotagged images was found to be statistically significant: (population size:

$p$ -value = 0.0019 &  $p$ -value = 0.000119, query-by-name and query-by-name+people respectively), (internet usage:  $p$ -value = 0.00029 &  $p$ -value = 0.003999), and (GDP:  $p$ -value = 0.160 &  $p$ -value = 0.059). By contrast, official language was not found to have a meaningful correlation to the number of geotagged images ( $p$ -value = 0.2021 &  $p$ -value = 0.846). Because image dataset queries are typically done in English, to assess the impact of dominant national languages relative to English on geotagged data availability, we coded each of the countries' official languages ([34]) according to five categories for analysis: 1- (English is the only official language), 2- (English is among the two official languages), 3- (English among atleast three official languages), 4- (English not among atleast three official languages), and 5- (English not among atleast three official languages). No correlation was determined for any language category.

**Thus**, when data collection is required in regions with lower population size, internet usage, and/or GDP, we recommend the use of local, manual data collection techniques in lieu of webscraping whenever feasible. Additionally, RWI information may be useful when assessing diverse areas for data collection.

### 3.2 Tags and Licenses: Query-Based and Applicability Limitations

The use of both country name tags and geotags (latitude/longitude) was found to be necessary to ensure data was accurately sourced from the country of interest. The naïve country name querying method is particularly limiting when applied to certain nations, such as *Chad*, *Guinea*, and *Republic of Congo*. Images from query by *Chad* were predominantly geotagged from *United States* (54.63%), *United Kingdom* (16.58%), and *Canada* (9.30%), with only 5.14% of the images coming from *Chad* according to geotag location results. In total images from query by *Chad* were geotagged from 129 countries. Likewise, images from query by *Guinea* predominantly came from *Papua New Guinea* (29.97%) and *United States* (10.84%), with geotags from 190 countries. Finally, image geotags from query by *Republic of Congo* mainly reflected the following countries; *Congo*, *The Democratic Republic of the* (42.14%), *United Kingdom* (11.34%),

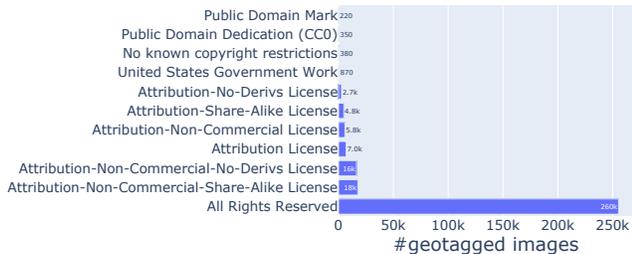
**Table 3: Number of geotagged and total images, and percentage (rounded to 2 decimal places) of geotagged images for population-matched African and European nations (query-by-name) side-by side, e.g Switzerland and Sierra Leone on line 1.**

| European Countries |            |            |             |            |
|--------------------|------------|------------|-------------|------------|
| Country            | population | #geotagged | #all images | %geotagged |
| Switzerland        | 8.75M      | 129,518    | 535,843     | 24.17      |
| Finland            | 5.55M      | 119,901    | 522,637     | 22.94      |
| Slovenia           | 2.11M      | 86,630     | 371,584     | 23.31      |
| Cyprus             | 918.10k    | 77,826     | 371,504     | 20.95      |

| African Countries |            |            |             |            |
|-------------------|------------|------------|-------------|------------|
| Country           | population | #geotagged | #all images | %geotagged |
| Sierra Leone      | 8.30M      | 7,303      | 52,530      | 13.90      |
| CAF               | 5.60M      | 2,954      | 19,901      | 14.84      |
| Lesotho           | 2.10M      | 5,121      | 43,282      | 11.83      |
| Djibouti          | 976.11k    | 5,179      | 36,029      | 14.37      |



(a) License information for query-by-name



(b) License information for query-by-name+people

**Figure 5: Bar charts showing the total count of each Flickr license type for the entire image datasets as a function of (a) “query-by-name” and (b) “query-by-name+people”. The most restrictive license type is by far the most common, “All Rights Reserved”, likely because it is the Flickr default option.**

and *United States* (7.55%).

Thus, we conclude that reliance upon country name queries is insufficient for constructing a geodiverse dataset in the absence of more robust geolocation data. We recommend that data collectors consider using RWI data to source more geographically diverse visual data.

We furthermore report the most frequent tags as the name of the place where the image was taken, for example “Africa” and the country name, in addition to image contents. The least frequent tags were usually those in foreign languages and whose meanings were hard to decipher because of multiple concatenated words.

Thus, in an African context, the utilization of image tags alone to generate datasets with specific image content may be less reliable due to the variable nature of selected tags; we believe this warrants future exploration.

Additionally, the vast majority of images with query-by-name and query-by-name+people respectively are licensed as “All Rights

Reserved” (80.46%, 81.99%), indicating the Flickr default setting when images are uploaded to the platform (see Figure 5).

Thus, those constructing datasets using Flickr Africa data must be aware that most images are unavailable for model training and evaluation without copyright violations, thereby further limiting ethical access to geographically diverse data.

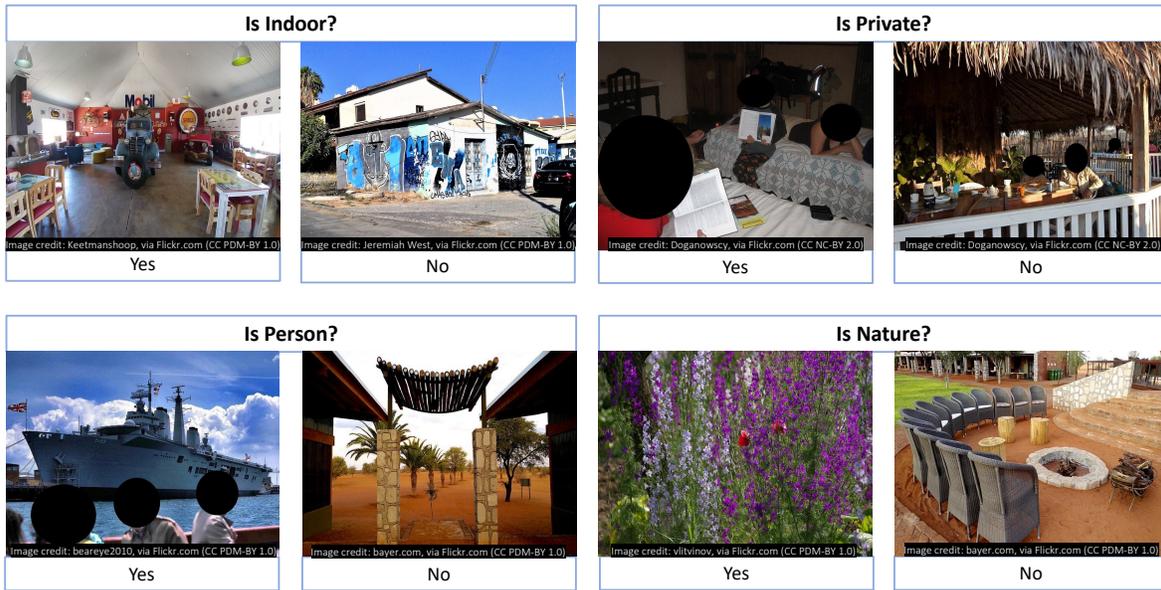
### 3.3 Geodiversity by RWI

To assess the impact of wealth on the availability of geotagged image data, we examine image counts by RWI values binned into 10 percentile groups, G1-G10. For most nations, the majority of image data comes from the middle RWI regions (G4, G5, G6 and G7) and the least from low RWI regions (G1, G2 and G3). However, this is not always the case, e.g. Madagascar and Algeria from which data is sourced from low-income areas (along main roads close to national parks) or high-income areas (in major cities), respectively. Thus, RWI has potential as a mechanism for constructing geodiverse datasets in future work.

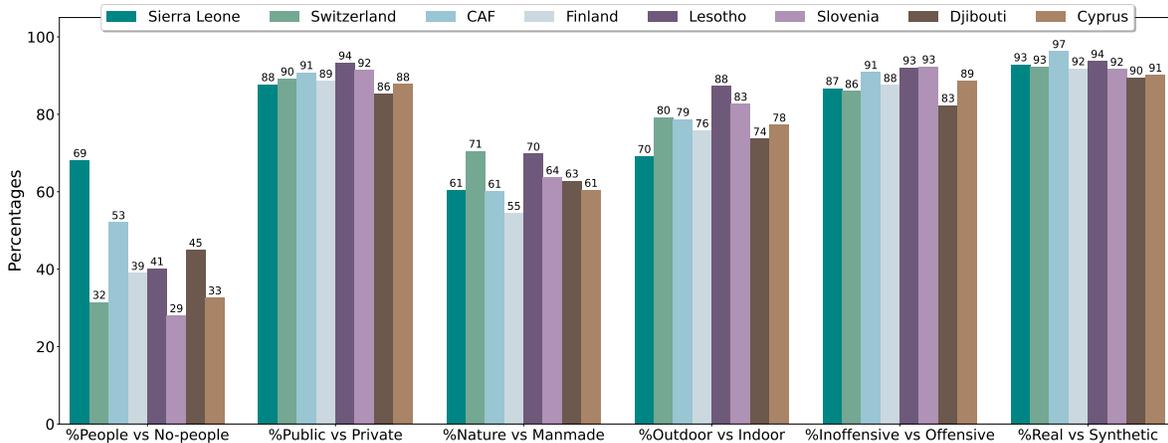
### 3.4 Image Content

By utilizing crowdsourced annotations, we examine 16,000 images’ content data across 2,000 images from each population-matched African and European nation pair (identical to the image subset in Section 3.5). Sample images by attribute and results for matched African/European nation pairs are shown in Figure 6a for each binary attribute with the exception of “offensive” vs. “inoffensive” content and with manually obscured human faces. We collected information about these six attributes to gauge the applicability of African-sourced image datasets for various computer vision tasks: e.g. the presence of people for human-centric tasks such as pose estimation, body part segmentation or face detection; the prevalence of indoor/private settings for specific object recognition tasks; or real/appropriate image content for training dataset viability. Likewise, we originally hypothesized that African images were more likely to be taken by foreigners (which was found to be supported by the data; see Section 3.5); this motivated the count of nature-centric images.

The AMT results revealed that query-by-name images from both African and European countries were predominantly “real” (93.47% and 91.94%), “inoffensive” (88.51% and 89.41%), “outdoor” (77.89% and 79.28%), “public” (90.27% and 90.19%), and “nature” (63.68% and 62.96%) images. There were negligible variations across nations for the percentage of “real”, “outdoor”, “public”, and “nature” images. However, as shown in Figure 6b, nations varied in percentage of people in images, and in general most nations’ images did not contain people. For example, 69% of Sierra Leone’s 2000 sampled geotagged images contained people, while only 33% of Djibouti’s



(a)



(b)

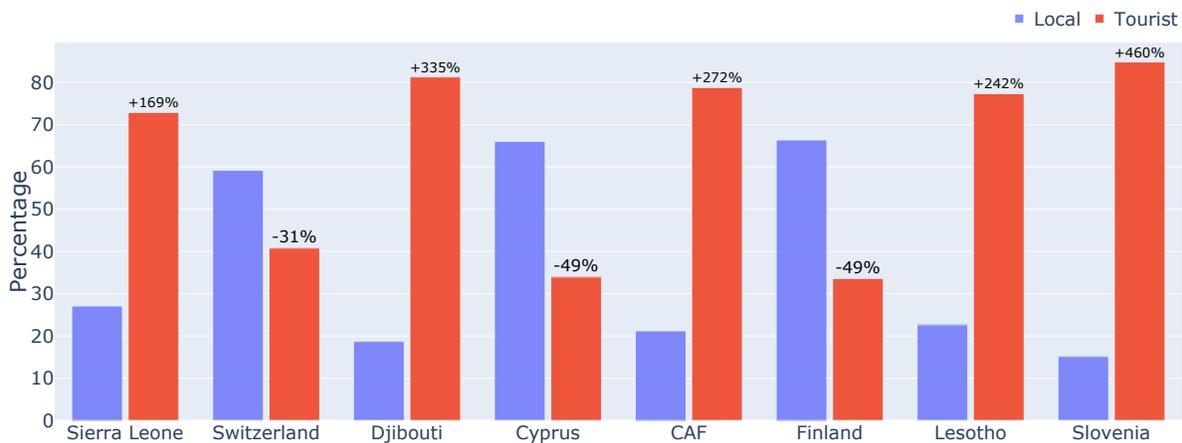
**Figure 6: (a) Sample images sourced from African nations, and (b) image content percentages for select African and European nations, according to AMT workers reporting across six binary attributes. Workers were asked to report if the images contained people, public settings, nature content, outdoor settings, and inoffensive content, and if they appeared to be real images. Image content was found to be similar percentage-wise across different nations, although far fewer images overall were captured in the population-matched African nations in comparison to corresponding European nations. Images displayed here were selected among those with permissible licenses with face obfuscation for display purposes only.**

2000 sampled geotagged images contained people. Thus, although no major differences between African and European image content were observed according to the six attributes considered, we believe these findings are important in the context of data regarding data quantity. Given that image content was fairly similar across most attributes annotated, and there exist far fewer geotagged images from Africa (see Table 3), we anticipate insufficient African data availability for certain computer vision tasks. For example, the lower prevalence of images captured in “private”

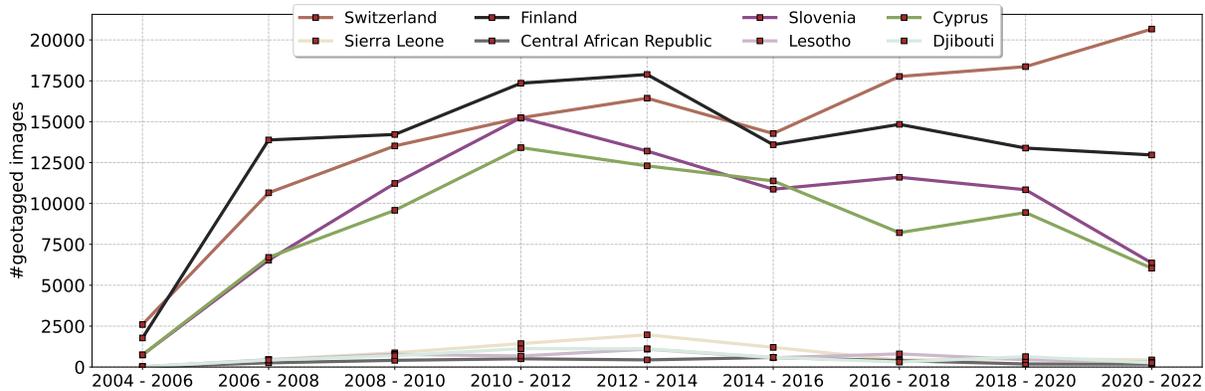
and “indoor” settings indicates e.g. household object image data inaccessibility, which thereby impacts downstream object recognition system models consistent with the findings of DeVries et al. [14].

### 3.5 Local vs. Non-Local Representation

Beyond analyzing the image content of the 2,000 randomly-sampled images from each of the 8 nations, we examined the local vs. non-local status of those Flickr users who captured and uploaded the



**Figure 7:** Bar chart showing the percentage out of 2,000 images from each population-matched African/European nation pair via query-by-name as taken by locals (blue) and tourists (red), according to Flickr users’ reported locations. The percent change from local to tourist percentage value is additionally indicated for each nation. Images from African countries were predominantly taken by foreigners whereas those from higher nominal GDP European countries were predominantly taken by locals. “CAF” indicates the country Central African Republic.



**Figure 8:** A comparison of the number of geotagged images from select African and European nations in approximately 2-year time ranges, as queried by country name. In general, a high number of images were uploaded in the dates in range 2012-2014, with the exception of Switzerland which had high upload volume in 2020-2022.

geotagged images. An assessment of the residence or origin of the Flickr users revealed that for the African geotagged nations, images were far more likely to be taken by foreigners than locals whereas the opposite trend was observed for higher-GDP European nations, according to comparisons between geotags and Flickr users’ reported locations. For Sierra Leone, +169% of images were captured by foreigners compared to locals, while for Switzerland it was -31%. The same trend applies to Djibouti and Cyprus (+335% and -49%) and CAF and Finland (+272% and -49%); results are reported in Figure 7. A random inspection of the Flickr map<sup>2</sup> also further shows that images geotagged in Africa are less likely to be taken by the locals.

Thus, the prevalence of non-local representation may explain the

<sup>2</sup>When we inspected the map (<https://www.flickr.com/map/>) on 06-16-2023, 2/2 of the geotagged images were taken by France and Spanish photographers

image content results described in the previous section, as Flickr users from similar backgrounds may contribute image data from both Africa and Europe. AI practitioners should be wary of stereotyped representations of African life within such datasets given that these images are typically taken by foreigners in public, outdoor locations. Additionally, current methodologies for image dataset collection are unlikely to capture visual data pertaining to the private, daily life of African people nor visual information the locals of each country consider to be important, resulting in biases propagated by AI systems trained on such data.

### 3.6 Temporal Analysis

We performed a temporal analysis to investigate and contextualize the data in time according to data quantity, relative wealth index

(RWI) at location of capture, license type and Flickr user origin. We studied the geotagged images distribution in the in approximately 2-year spans time ranges.

*Number of geotagged images.* In general, there were relatively fewer geotagged images in the years 2004-2006 and 2020-2022, as shown in Figure 8 for population-matched African/European nations. The image distribution could be the result of factors including less internet penetration and popularity of Flickr from 2004-2006 reducing image uploads, and the COVID-19 pandemic limiting outdoor activities from 2020-2022. This trend held in all analyzed countries with the exception of Switzerland; there, the highest number of images was uploaded in the date range 2020-2022, as shown in Figure 8. The highest number of uploaded and subsequently downloaded geotagged images for most nations came from 2010-2014, potentially explained by the growth of internet usage and exposure to Flickr in different countries within this time span.

*RWI regions of image uploads.* We explored trends in dominant RWI groups per nation over time, in order to determine if there were observable shifts towards images sourced from higher or lower RWI regions. Over the time range of 2004 to 2022, query-by-name images from Botswana, Libya, Namibia, South Africa, Tunisia, Swaziland, Uganda, Zambia, and Zimbabwe all came from the middle RWI regions. Images from Morocco consistently came from the upper RWI regions. On the other hand, query-by-name+people images from Rwanda and Swaziland all came from the middle RWI regions and those from Morocco all came from upper RWI regions. Lower and middle RWI regions countries had their data distributions varying between lower and middle RWI regions over the years. Countries whose images were from predominantly upper RWI regions had their data distributions varying between middle and upper RWI regions over the years.

*Licenses of the uploaded images.* We analyzed the quantity of images with various Flickr license options. Images were found to have predominantly the “All Rights Reserved” license type across all time ranges analyzed; as noted in Section 3.2, this substantially limits data usage. There were almost no images licensed under the “Public Domain Dedication (CC0) CC” and “Public Domain Mark CC” among those uploaded to Flickr from 2004 to 2022.

*Local vs. non-local representation.* We performed a temporal analysis of the geotagged images to investigate the local vs. non-local status of Flickr users. For the 2,000 randomly sampled images from the 8 countries analyzed for image content, we observed differences in sampling dates: that is, Cyprus, Slovenia, and Finland images were mainly sampled from 2004 to 2008; Switzerland images were mainly sampled from 2004 to 2006; and the African nation images were were mainly sampled from 2004 to 2012. Following these results, we repeated the temporal analysis across all images sourced from each of the 8 nations. In general, more images across all nations were taken by non-locals compared to the smaller 2,000-image datasets. However, the prior trends held in the sense that when African countries were considered, far more geotagged images were taken by non-locals than in comparable European nations, e.g., +329% for Sierra Leone versus +39% for Switzerland. Section 3.5 describes implications of non-local representation in image data

from Africa; namely, the risk of an “othering” phenomenon and its impact on downstream bias in AI systems.

#### 4 CONCLUSION AND FUTURE WORK

Geographical context shapes data, and data shapes the performance of models trained using such data. The key findings from our Flickr Africa data analysis (1) expose the limitations of current large-scale image data collection methodologies, and (2) expose unique data challenges to Africa, including the lack of data crucial to specific domains (e.g. a researcher cannot source sufficient, representative household object data if very few images are taken within indoor/private scenes). Notably, we reported on the extreme lack of data availability when compared to wealthy European nations; for instance when querying by country name, Switzerland had 18x the geotagged image data as Sierra Leone, an African nation of similar population size (8.75M vs. 8.30M, respectively), while Sao Tome and Principe only had (776, 116) geotagged images in total (depending on query). Moreover, data may be even less accessible according to use case, given that most of the Flickr Africa data has a restrictive use license, and certain image content attributes were found to appear less frequently (e.g. private and indoor settings). Nationally, higher quantities of geotagged image data was found to positively correlate with population size, GDP, and Internet usage, but no significant correlation was discovered based on dominant national languages. Additionally, we interrogate where African image data comes from: generally from middle-wealth regions as measured intra-nationally by RWI, though this differs by nation; and with images mainly taken by foreigners, though the opposite trend is identified in wealthier European nations. We discussed how AI systems may propagate biases in accordance with the stereotyped representation of African life by outsiders. Temporal analyses were performed and demonstrated that certain trends, such as dominant RWI region, prevalence of restrictive license type, and non-local representation of African nations in geotagged images held over time.

Looking forward, we encourage new scholarship centering novel methods for sourcing geodiverse datasets and measuring new forms of geodiversity specific to Africa, such as analyses of tribal diversity as opposed to the more commonly studied diversity by race/ethnicity. We openly provide our large-scale dataset to enable future researchers to utilize and augment Flickr Africa for model evaluations across a wide domain of computer vision tasks; likewise, more rigorous bias identification methods (e.g. [27]) may uncover still more limitations. Finally, we would be interested to explore the extent to which privacy and consent are respected in Africa.

#### ACKNOWLEDGMENTS

We wish to thank Jerone Andrews and Dora Zhao in particular for their expertise and assistance with our work pertaining to crowdsourcing. We also express gratitude to the whole SONY AI Ethics team, especially William Thong for great discussions and informative feedback on this research. Lastly, we would like to thank the anonymous reviewers for the insightful feedback that helped improve our paper.

## REFERENCES

- [1] Rediet Abebe, Kehinde Aruleba, Abeba Birhane, Sara Kingsley, George Obaido, Sekou L. Remy, and Swathi Sadagopan. 2021. Narratives and Counternarratives on Data Sharing in Africa. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (Virtual Event, Canada) (FAccT '21). Association for Computing Machinery, New York, NY, USA, 329–341. <https://doi.org/10.1145/3442188.3445897>
- [2] Evelyn Anane-Sarpong, Tenzin Wangmo, Claire Leonie Ward, Osman Sankoh, Marcel Tanner, and Bernice Simone Elger. 2018. "You cannot collect data using your own resources and put it on open access": Perspectives from Africa about public health data-sharing. *Developing world bioethics* 18 4 (Dec. 2018), 394–405. <https://pubmed.ncbi.nlm.nih.gov/28745008/>
- [3] Rönnlund AR. 2016. Dollar Street. <https://www.gapminder.org/dollar-street>.
- [4] Kumar Ayush, Burak Uz Kent, Chenlin Meng, Kumar Tanmay, Marshall Burke, David Lobell, and Stefano Ermon. 2021. Geography-Aware Self-Supervised Learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. 10181–10190.
- [5] Kumar Ayush, Burak Uz Kent, Chenlin Meng, Kumar Tanmay, Marshall Burke, David B. Lobell, and Stefano Ermon. 2020. Geography-Aware Self-Supervised Learning. *CoRR* abs/2011.09980 (2020). arXiv:2011.09980 <https://arxiv.org/abs/2011.09980>
- [6] Abeba Birhane and Vinay Uday Prabhu. 2021. Large Image Datasets: A Pyrrhic Win for Computer Vision?. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. 1537–1547.
- [7] Joy Buolamwini and Timnit Gebru. 2018. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. In *Proceedings of the 1st Conference on Fairness, Accountability and Transparency (Proceedings of Machine Learning Research, Vol. 81)*, Sorelle A. Friedler and Christo Wilson (Eds.). PMLR, 77–91. <https://proceedings.mlr.press/v81/buolamwini18a.html>
- [8] Guanghua Chi, Han Fang, Sourav Chatterjee, and Joshua E. Blumensack. 2022. Microestimates of wealth for all low- and middle-income countries. *Proceedings of the National Academy of Sciences* 119, 3 (2022), e2113658119. <https://doi.org/10.1073/pnas.2113658119> arXiv:<https://www.pnas.org/doi/pdf/10.1073/pnas.2113658119>
- [9] Pei-Yu Peggy Chi, Matthew Long, Akshay Gaur, Abhimanyu Deora, Anurag Batra, and Daphne Luong. 2019. Crowdsourcing Images for Global Diversity. In *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services (Taipei, Taiwan) (MobileHCI '19)*. Association for Computing Machinery, New York, NY, USA, Article 79, 10 pages. <https://doi.org/10.1145/3338286.3347546>
- [10] Gordon A. Christie, Neil Fendley, James Wilson, and Ryan Mukherjee. 2017. Functional Map of the World. *CoRR* abs/1711.07846 (2017). arXiv:1711.07846 <http://arxiv.org/abs/1711.07846>
- [11] David J. Crandall, Lars Backstrom, Daniel Huttenlocher, and Jon Kleinberg. 2009. Mapping the World's Photos. In *Proceedings of the 18th International Conference on World Wide Web (Madrid, Spain) (WWW '09)*. Association for Computing Machinery, New York, NY, USA, 761–770. <https://doi.org/10.1145/1526709.1526812>
- [12] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*. 248–255. <https://doi.org/10.1109/CVPR.2009.5206848>
- [13] Emily Denton, Mark Diaz, Ian Kivlichan, Vinodkumar Prabhakaran, and Rachel Rosen. 2021. Whose Ground Truth? Accounting for Individual and Collective Identities Underlying Dataset Annotation. *CoRR* abs/2112.04554 (2021). arXiv:2112.04554 <https://arxiv.org/abs/2112.04554>
- [14] Terrance DeVries, Ishan Misra, Changhan Wang, and Laurens van der Maaten. 2019. Does Object Recognition Work for Everyone?. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2019, Long Beach, CA, USA, June 16–20, 2019*. Computer Vision Foundation / IEEE, 52–59. [http://openaccess.thecvf.com/content\\_CVPRW\\_2019/html/cv4gc/de\\_Vries\\_Does\\_Object\\_Recognition\\_Work\\_for\\_Everyone\\_CVPRW\\_2019\\_paper.html](http://openaccess.thecvf.com/content_CVPRW_2019/html/cv4gc/de_Vries_Does_Object_Recognition_Work_for_Everyone_CVPRW_2019_paper.html)
- [15] Isaac L. Johnson, Subhasree Sengupta, Johannes Schöning, and Brent Hecht. 2016. The Geography and Importance of Localness in Geotagged Social Media. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) (CHI '16). Association for Computing Machinery, New York, NY, USA, 515–526. <https://doi.org/10.1145/2858036.2858122>
- [16] Kimmo Karkkainen and Jungseock Joo. 2021. FairFace: Face Attribute Dataset for Balanced Race, Gender, and Age for Bias Measurement and Mitigation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 1548–1558.
- [17] Ourania Kounadi, Thomas J Lampoltshammer, Michael Leitner, and Thomas Heistracher. 2013. Accuracy and privacy aspects in free online reverse geocoding services. *Cartography and Geographic Information Science* 40, 2 (2013), 140–153.
- [18] Alina Kuznetsova, Hassan Rom, Neil Alldrin, Jasper Uijlings, Ivan Krasin, Jordi Pont-Tuset, Shahab Kamali, Stefan Popov, Matteo Mallocci, Alexander Kolesnikov, et al. 2020. The open images dataset v4: Unified image classification, object detection, and visual relationship detection at scale. *International Journal of Computer Vision* 128, 7 (2020), 1956–1981.
- [19] Alina Kuznetsova, Hassan Rom, Neil Alldrin, Jasper R. R. Uijlings, Ivan Krasin, Jordi Pont-Tuset, Shahab Kamali, Stefan Popov, Matteo Mallocci, Tom Duerig, and Vittorio Ferrari. 2018. The Open Images Dataset V4: Unified image classification, object detection, and visual relationship detection at scale. *CoRR* abs/1811.00982 (2018). arXiv:1811.00982 <http://arxiv.org/abs/1811.00982>
- [20] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. 2014. Microsoft COCO: Common Objects in Context. In *Computer Vision – ECCV 2014*, David Fleet, Tomas Pajdla, Bernt Schiele, and Tinne Tuytelaars (Eds.). Springer International Publishing, Cham, 740–755. [https://doi.org/10.1007/978-3-319-10602-1\\_48](https://doi.org/10.1007/978-3-319-10602-1_48)
- [21] Abhishek Mandal, Susan Leavy, and Suzanne Little. 2021. Dataset Diversity: Measuring and Mitigating Geographical Bias in Image Search and Retrieval. In *Proceedings of the 1st International Workshop on Trustworthy AI for Multimedia Computing (Virtual Event, China) (Trustworthy AI'21)*. Association for Computing Machinery, New York, NY, USA, 19–25. <https://doi.org/10.1145/3475731.3484956>
- [22] Open street map. 2020. Open-source geocoding with OpenStreetMap data. <https://nominatim.org/>. Accessed: 2022-06-15.
- [23] Richard Penman. 2020. reverse-geocode 1.4.1. <https://pypi.org/project/reverse-geocode/#description>. Accessed: 2022-06-15.
- [24] Nithya Sambasivan, Erin Arnesen, Ben Hutchinson, Tulsee Doshi, and Vinodkumar Prabhakaran. 2021. Re-Imagining Algorithmic Fairness in India and Beyond. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (Virtual Event, Canada) (FAccT '21). Association for Computing Machinery, New York, NY, USA, 315–328. <https://doi.org/10.1145/3442188.3445896>
- [25] Shreya Shankar, Yoni Halpern, Eric Breck, James Atwood, Jimbo Wilson, and D. Sculley. 2017. No Classification without Representation: Assessing Geodiversity Issues in Open Data Sets for the Developing World. <https://doi.org/10.48550/ARXIV.1711.08536>
- [26] Bart Thomee, David A Shamma, Gerald Friedland, Benjamin Elizalde, Karl Ni, Douglas Poland, Damian Borth, and Li-Jia Li. 2016. YFCC100M: The new data in multimedia research. *Commun. ACM* 59, 2 (2016), 64–73.
- [27] Angelina Wang, Alexander Liu, Ryan Zhang, Anat Kleiman, Leslie Kim, Dora Zhao, Iroha Shirai, Arvind Narayanan, and Olga Russakovsky. 2022. REVISE: A Tool for Measuring and Mitigating Bias in Visual Datasets. *Int. J. Comput. Vision* 130, 7 (jul 2022), 1790–1810. <https://doi.org/10.1007/s11263-022-01625-5>
- [28] L. Weinberg. 2022. Rethinking Fairness: An Interdisciplinary Survey of Critiques of Hegemonic ML Fairness Approaches. *The journal of artificial intelligence research* 74 (May 2022). <https://doi.org/10.1613/jair.1.13196>
- [29] Aaron M. Wesley and Timothy C. Matisziw. 2021. Methods for Measuring Geodiversity in Large Overhead Imagery Datasets. *IEEE Access* 9 (2021), 100279–100293. <https://doi.org/10.1109/ACCESS.2021.3096034>
- [30] Wikipedia.org. 2022. List of African countries by population. [https://en.wikipedia.org/wiki/List\\_of\\_African\\_countries\\_by\\_population](https://en.wikipedia.org/wiki/List_of_African_countries_by_population). Accessed: 2022-08-27.
- [31] Wikipedia.org. 2022. List of countries by GDP (nominal) - IMF. [https://en.wikipedia.org/wiki/List\\_of\\_countries\\_by\\_GDP\\_\(nominal\)\\_-IMF](https://en.wikipedia.org/wiki/List_of_countries_by_GDP_(nominal)_-IMF). Accessed: 2022-08-27.
- [32] Wikipedia.org. 2022. List of countries by number of Internet users. [https://en.wikipedia.org/wiki/List\\_of\\_countries\\_by\\_number\\_of\\_Internet\\_users](https://en.wikipedia.org/wiki/List_of_countries_by_number_of_Internet_users). Accessed: 2022-08-27.
- [33] Wikipedia.org. 2022. List of European countries by population. [https://en.wikipedia.org/wiki/List\\_of\\_European\\_countries\\_by\\_population](https://en.wikipedia.org/wiki/List_of_European_countries_by_population). Accessed: 2022-08-27.
- [34] Wikipedia.org. 2022. List of official languages by country and territory. [https://en.wikipedia.org/wiki/List\\_of\\_official\\_languages\\_by\\_country\\_and\\_territory](https://en.wikipedia.org/wiki/List_of_official_languages_by_country_and_territory). Accessed: 2022-08-27.
- [35] Alice Xiang. 2022. Being 'Seen' vs. 'Mis-Seen': Tensions between Privacy and Fairness in Computer Vision. *Harvard Journal of Law & Technology* (Apr 2022). <https://ssrn.com/abstract=4068921>
- [36] Da Yin, Liunian Harold Li, Ziniu Hu, Nanyun Peng, and Kai-Wei Chang. 2021. Broaden the Vision: Geo-Diverse Visual Commonsense Reasoning. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Online and Punta Cana, Dominican Republic, 2115–2129. <https://doi.org/10.18653/v1/2021.emnlp-main.162>
- [37] Dora Zhao, Angelina Wang, and Olga Russakovsky. 2021. Understanding and Evaluating Racial Biases in Image Captioning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. 14830–14840.