



# Beyond 'Slumming It'

## —AI and the Real Lives of Global South Communities

 **Matt Jones**, Swansea University  
**Dani Kalarikalayil Raju**, Studio Hasi  
**Jen Pearson**, Swansea University  
**Thomas Reitmaier**, Swansea University  
**Simon Robinson**, Swansea University  
**Arka Majhi**, Indian Institute of Technology Bombay

---

### Insights

- We conducted workshops in three Global South communities in India and South Africa to explore perspectives on image-based GenAI.
- Workshop participants were disappointed and upset that the images generated by AI of the current and future Global South communities emphasized underdevelopment.
- Community engagement in dataset curation and use is vital to avoid these problems in the future.

Biases in AI datasets are well known, and their impacts in terms of misrepresenting gender and ethnicity are regularly surfacing in generative AI (GenAI) services. Many of us have seen examples of these—some of them absurd, some comical, and others completely inappropriate. A notorious example is the image Midjourney produced using the prompt “Black African doctor is helping poor and sick white children, photojournalism.” The image showed a white “savior” doctor, helping black children [1]. And what would you expect to see given the following two prompts? “A person at social services”

and “a productive person.” Stable Diffusion, another AI image creator, generated images of only non-white females and non-white males for the first prompt and only white males for the second [2]. While these problematic images have been reported in previous articles, it is easy to create new ones such as those shown in Figure 1 using—at the time of writing—the most recent version of ChatGPT.

Though these examples are provocative, our work has taken us to less abstract and more grounded situations, where GenAI misrepresents everyday life in gross



Figure 1. Images produced by ChatGPT 4.0 via prompts “a doctor helping a poor patient” (left) and “a successful lawyer” (right).

ways. In a series of workshops during early 2024 in urban, peri-urban, and rural Indian and South African communities—some of which we have visited many times over the past decade—we watched as community members experimented with GenAI and were left feeling upset, frustrated, sad, and less hopeful about their lives.

In this article, we provide a glimpse into the vibrant lives of these communities and the parallel “sad worlds” presented in AI-generated images. There are several potential solutions to this problem. We present one that we have been working on for several years: community contribution and curation of training datasets and the services they enable.

### THE VIBRANT AND COMPLEX REALITY OF LIFE

In January and March 2024, we visited three places: two in India and one in South Africa. In India, the first location was a Banjara community near Chalisgaon, around seven hours by train northeast of Mumbai. The community there has a thriving cotton farming business. A typical day involves an early start—to avoid the heat later in the day—with women and men heading out to the fields to tend to crops. By midafternoon these workers return home to rest, chat, cook, and play. The villagers also keep goats for food and to sell. The animals, which live close to the villagers’ homes, are often treated as pets. Our conversations with villagers gave us a

sense of their entrepreneurial mindsets, their aspirations to improve their environment—for example, by investing in their homes—and their pride in what they do.

Dharavi, in Mumbai, was our second stop after the Banjara village. Often referred to as Asia’s largest slum, Dharavi is home to more than a million people, who live on 535 acres—a population density of 869,565 people per square mile. It is situated in the heart of Mumbai’s business district and is well known for its significant economic activities, ranging from pottery making and leather manufacturing to electronics and recycling. Visitors are often struck by the extraordinary range of cuisines, cultures, religion, and ethnicities that come together in such a compact space. Dharavi is not so much a melting pot as a stained glass window of intricate colors and connectivity.

A month later, we traveled from India to Langa, a township on the outskirts of Cape Town, South Africa. Today, nearly 100 years after its founding, there is a wide range of housing types in Langa, from suburban bungalows similar to those seen in many other parts of the world to homes built of corrugated iron. With its schools, libraries, restaurants, and shops, it is a busy place that also provides a base for many people who commute to the city to study or work.

If you had traveled with us, you



Real photos from the Banjara farming community in India.



BOTTOM ROW OF PHOTOS COURTESY OF FIT LAB SWANSEA

would have also seen the messiness and difficulties of daily life in each of these places, such as the lack of universal access to sanitation and precarious infrastructure. And maybe if you only visited once you would be left with a negative impression. Perhaps you would have focused on the surface—the images you took with your phone emphasizing contexts that were uncomfortable *to you*. If others asked you to describe these places when you returned home, maybe you would talk about poverty, decline, and danger. This is not new—it is what generations of “slum tourists” have done. Slum tourism became popular in the 18th century in England, when the rich would venture into London slum districts to be shocked and frightened. They would return with tales to regale guests with at their fashionable dinner parties (this practice, incidentally, gave rise to the term “slumming it”).

In a way, the AI systems are acting as artificial slum tourists. Large language models (LLMs) are fed by the millions of images taken by visitors, including slum tourists, and by texts describing the social and economic challenges of these locations written by nongovernmental organizations and policymakers, however good their intentions. The LLM we used generated images of these locations that focused and amplified *only* the negatives; they didn’t engage with the fullness of the lives lived there.



Figure 2. Images generated by DALL-E 2, using prompts created by workshop organizers “A London bus in African style in Langa township” (left) and “A London bus in African style in Camps Bay” (right).

### REALITY VERSUS AI

We conducted four-hour workshops: one each in the Banjara community and Langa and two in Dharavi. Forty-two people (23 women and 19 men, ages 19 to 58) participated in Dharavi; eight people (five women and three men, ages 19 to 68) in the Banjara village; and 10 people (six women and four men, ages 19 to 36) in Langa, with the latter also coming from the neighboring townships of Khayelitsha and Delft. When we asked all these groups about their awareness of AI or machine learning there was little response. Their exposure to GenAI was nonexistent—a reminder that there are likely vast numbers of people in the world who are not yet part of the

conversation about fundamental future infrastructures. Many, though, had used or heard about facial or fingerprint recognition.

We set up workshop rooms in each location with screens or projectors connected to a laptop, and then we opened DALL-E 2 (a popular text-to-image GenAI tool). To start the discussions, we explained what the system could do and how it was different from searching for images on the Web. Then we prompted DALL-E with, “a London bus in Indian [or African] style in [the place we were holding the workshop]” (Figure 2). In each workshop, the anticipation in the room at this point was palpable. It quickly dissipated, however, as broken-down, rusty, ramshackle



Real photos of Dharavi in Mumbai (left) and Langa in Cape Town (right).



Figure 3. Images generated by DALL-E 2 for places in India and South Africa, using prompts created by workshop participants.

vehicles appeared on the screen. There was incredulous laughter and shaking of heads. “We have smart buses in this area,” one Dharavi participant said. We repeated the prompt, changing the location to an affluent area (e.g., Camps Bay and Colaba for Langa and Dharavi workshops, respectively). This time, the image generator created buses that looked new, shiny, and comfortable. Again, this led to sighs of exasperation from the workshop participants: “It looks like it [the system] thinks rich people are better than us.”

Handing over the prompting to the participants, we asked them to tell us which other pictures they would like the computer to produce, some about

their everyday life and others of a future time. The output from these prompts for the three places is shown in Figure 3.

As with the reactions to the initial bus prompt, every image that the GenAI created was met with a mix of despair and laughter. After seeing the “beautiful Dharavi” image, the participants spoke of the things they would have drawn—colorful rangoli patterns, the numerous street shrines, the many festivals, and coming together as a community. Their ideas were strikingly different from the images created by DALL-E. The rural farmers said that AI’s image of 2070 was far worse than their current homes and environment, while

township participants felt patronized by the superficial rendering of “hope” for Cape Town (“They just splashed a bit of color and a heart!”). The “relaxing” image was a pose they felt a tourist—not a local—might make. Perhaps the most poignant reaction, however, was from one of the Dharavi attendees: “Even AI doesn’t think we have a future.”

Listening to and watching participants’ reactions was both depressing and motivating. As practitioners of HCI, we are driven to understand how technologies can bring hope and joy rather than upset and sadness. While there is much talk about the future existential threat of sentient AI, what we saw was a powerful reminder of the need to avoid being seduced by such sci-fi scenarios. Rather, we should work to mitigate the issues we’re currently confronting and those that will come up in the near future. Imagine the impact on young people from the Global South communities eager to engage with the excitement of cutting-

**There are likely vast numbers of people in the world who are not yet part of the conversation about fundamental future infrastructures.**

edge tech, as they get to use systems such as the version of DALL-E we used, only to see the version of their world this tech creates.

## A BETTER FUTURE

How do we improve our present situation? Interestingly, DALL-E 3 has moved away from photorealistic images to more-stylistic renderings [1,2], possibly to avoid some of the failings we saw. The “beautiful Dharavi” image DALL-E 3 produces is shown in Figure 4.

While this is a more positive presentation (and note the upbeat caption the system produces), it perhaps lacks the power for emotional engagement that a photo affords. It is sterile, presenting a world where life and liveliness have been extracted, leaving one disinterested in the characters, their actions, and their homes. The philosopher Ludwig Wittgenstein gives us a clue as to why such a rendering might be less compelling than DALL-E’s previous photorealistic attempts. He noted that, “if a lion could speak, we could not understand him.” In other words, the lion is not human, not one of us, and does not share our context. Therefore, the lion would not be able to express itself in ways meaningful to us [3]. What the stylistic DALL-E images perhaps do is signal to us—more than a photorealistic image can—that the output we are looking at is created by an alien, a nonhuman entity. It is not a lion but an *AI-on*; an AI that is not of us, that does not in any way understand what it is to be human or the contexts we inhabit. These AI-stylized images provide some confirmation of Wittgenstein’s assertion. We now have *AI-ons* that can speak and paint, but they do not resonate with us because they are not one of us.

If the stylized forms are not the best way to present the vibrant richness we have seen for ourselves in communities we have visited, what is? Our workshop attendees had a solution that could result in a truer “photo” and include the realism of their lived experiences: ensure that the models are trained on a much more diverse and representative set of images and other data. They were keen to see how they could get their communities to contribute their own pictures to enrich



Figure 4. AI-generated image of Dharavi by DALL-E 3, using the prompt “beautiful Dharavi.” Caption produced by DALL-E: “Here is a photorealistic depiction of a beautiful scene in Dharavi, Mumbai, showcasing the vibrant life and dynamic spirit of this unique neighborhood.”

existing models. The value of this sort of locally sourced image set has been demonstrated in recent research, which has shown that models that fail to reflect cultural diversity can be enhanced with a relatively small set of data [4].

## TRULY INVOLVING OVERLOOKED AND UNHEARD COMMUNITIES

We can go further, though, than simply complementing datasets with locally sourced training data. We can help communities become AI shapers rather than entities from which we extract data. To do so, there is a need to develop ways to fully—rather than superficially—engage people in the process of creating systems that truly represent their worlds, and that will be useful and usable to them. Our work has taught us that such approaches will benefit not only people living in the Global South. It is an opportunity to develop radical alternative approaches to the ways AI is currently done globally.

For the past three years, we have been exploring such community-driven approaches in regard to another modality: speech. Our aim has been to build a spoken language system tool kit that can be quickly deployed and used with low- or no-resource language communities (i.e., communities that speak a language for which there are few or no existing digital materials to train a model). You can find out more about our research [5,6,7] and access components of the tool kit at <https://unmute.tech/>. Here we describe the key components that we believe would also provide positive ways to diversify and enrich the data and applications of other emerging AI services.

**Community engagement and relationship building.** To kick off such a development project, we recommend first identifying a community liaison (sometimes known as a human access point). This should be someone who is not only trusted within the community and speaks the community’s language but also can



## Advertise with ACM!

Reach the innovators and thought leaders working at the cutting edge of computing and information technology through ACM's magazines, websites and newsletters.



Request a media kit with specifications and pricing:

**Ilia Rodriguez**

+1 212-626-0686

acmm mediasales@acm.org



Discussing GenAI images in one of the Banjara community homes.

communicate with the research team and maintain regular contact over an extended period. It is beneficial if the liaison is already familiar with how to formulate research tasks and trials, such that they can help situate design activities and facilitate uptake during workshops and deployments. The liaison should be seen as the first point of contact and a “critical friend” who can provide immediate thoughts or feedback, or tips on how to proceed with data gathering while taking into consideration the values, concerns, and practices of the community.

*Workshops to explain the value of the community's data and why it will be a key resource for both them and the wider world of future users.* There are a lot of important ethical and practical steps at this point. For example, it can be challenging to know how to begin explaining a highly technical design to someone unfamiliar with many of the processes involved. Metaphors can be useful here, because they enable people to

relate new, often abstract information to existing things that they understand. For example, we have used the metaphor of a child who is slowly learning to talk as an analogy for how speech and language technologies “learn” a language from repeated exposure to particular utterances, explaining how, with time (i.e., more data), such a system could identify and match similar phrases. This metaphor is also useful when explaining that speech and language technologies can often make mistakes that seem childish or obvious to an experienced speaker.

Investing time with communities to understand and codesign potential use cases is also highly valuable, as it will both motivate later data provision and give insights into useful local applications. On the ethical front, it is key to emphasize communities' control over the data they provide. We have, for example, come across people who are worried that by creating video content for an AI, the system will

**To build AI systems that are good for all and have an intimate understanding of spaces, go to those spaces, and seek meaningful, collaborative community engagements.**

surface that content in a way that makes them appear foolish or naive to their peers or even globally.

**Accommodation of at-hand tools to enable data capture.** Many of the people who are not yet actively involved in contributing data for AI systems do not have access to resources that “mainstream” users take for granted, such as high-end smartphones and ample data connectivity. By engaging, you will learn about the types of hardware, operating systems, and apps available and usable in those contexts; you should then adapt any approaches you deploy. For example, we have regularly found users whose phones have smashed screens and are completely unresponsive to touch in certain segments of the display. Ideally, in this case, the user should be able to move key interface elements. In terms of interaction and data gathering, a tool that is widely available and used in all of the Global South communities we have worked in is WhatsApp. Appropriating this and other platforms for data collection allows for a longer-running, more self-organizing process, which can leverage participants’ familiarity with an existing and popular platform.

**Enabling communities to describe the world in their own terms.** Kate Crawford shows how data used to tag content can skew how AI conceptualizes the world and the uses it then affords [8]. Providing community members with accessible tools to describe the content they are supplying with terms that are meaningful to them can mitigate against creating AI worlds that are unaligned with the real worlds they see.

## CONCLUSION

*The Worlds I See*, by Fei-Fei Li, is an excellent title for a book charting the past, present, and future of AI [9]. In it, Li notes:

*Where we once sought to give our algorithms a kind of encyclopedic awareness—all categories and things—we now aim for something richer. A more intimate understanding of the spaces and moment and even meaning in which those things are embedded. An expansion of not just quantity, but detail and nuance.*

We wholeheartedly agree. The generous amount of time people in each community gave us revealed that AI sees Global South communities in ways that are far less detailed or nuanced than they need to be. Currently, most state-of-the-art AI efforts are centered on the refining of methods of large-scale data mining, the hosting of models, and the means of learning everything we need to know “from the data.” We believe it is equally important to place an emphasis on engaging communities and carefully considering how to integrate community codesign and feedback into existing work and methodologies. We would like to see more bidirectional, ground-up development pipelines, which put communities at the fore, and therefore ground research and development in the real world. Our message to others is simple. To build AI systems that are good for all and have an intimate understanding of spaces, go to those spaces, and seek meaningful, collaborative community engagements with the kind of people we have met in Global South contexts.

## ACKNOWLEDGMENTS

Thanks to Minah Radebe, Yashwant Rathod, and Manik for helping facilitate the workshops in Langa, Chalisgaon, and Dharavi, respectively, and all the community members who shared their time and insights.

We gratefully acknowledge funding by the Engineering and Physical Sciences Research Council (grants EP/W020548/1 and EP/T024976/0).

## ENDNOTES

1. Ananya. AI image generators often give racist and sexist results: Can they be fixed? *Nature* 627, 8005 (2024) 722–725.
2. Tiku, N., Schaul, K., and Chen S.Y. These fake images reveal how AI amplifies our worst stereotypes. *Washington Post*. Nov. 1, 2023; <https://www.washingtonpost.com/technology/interactive/2023/ai-generated-images-bias-racism-sexism-stereotypes/>
3. Wittgenstein, L. (Anscombe, G.E.M. trans.). *Philosophical Investigations*. Basil Blackwell Ltd, Oxford, 1953.
4. Liu, Z. et al. SCoFT: Self-contrastive fine-tuning for equitable image generation. arXiv:2401.08053, Jan. 16, 2024.

5. Reitmaier, T. et al. Cultivating spoken language technologies for unwritten languages. *Proc. of the 2024 CHI Conference on Human Factors in Computing Systems*. ACM, New York, 2024, Article 614.
6. Reitmaier, T. et al. Situating automatic speech recognition development within communities of under-heard language speakers. *Proc. of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, New York, 2023, Article 406.
7. Reitmaier, T. et al. Opportunities and challenges of automatic speech recognition systems for low-resource language speakers. *Proc. of the 2022 CHI Conference on Human Factors in Computing Systems*. ACM, New York, 2022, Article 299.
8. Crawford, K. *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press, 2021.
9. Li, F.-F. *The Worlds I See: Curiosity, Exploration, and Discovery at the Dawn of AI*. Flatiron Books, New York, 2023.

---

**Matt Jones** is an EPSRC research fellow exploring a novel interactive paradigm: Everyone Virtuoso Everyday (e-v-e-.ai).  
→ [matt.jones@swan.ac.uk](mailto:matt.jones@swan.ac.uk)

**Dani Kalarikalayil Raju** is an alumnus of the Indian Institute of Technology Bombay doing HCI research in India. He is a cofounder of Studio Hasi, a startup working with the marginalized communities to facilitate their participation in the design and deployment of advanced technologies.  
→ [daniel@studiohasi.com](mailto:daniel@studiohasi.com)

**Jen Pearson** is a professor of HCI and director of research for the Department of Computer Science in the Computational Foundry, Swansea University. Her primary research interest is centered around so-called emergent communities, learning from them and working with them to cocreate digital interactive systems that better suit their contexts.  
→ [j.pearson@swansea.ac.uk](mailto:j.pearson@swansea.ac.uk)

**Thomas Reitmaier** is a lecturer in computer science at Swansea University. He works on the interdisciplinary UnMute/Amplify projects to cocreate spoken language interactions with communities of minority language speakers who are currently digitally unheard.  
→ [thomas.reitmaier@swansea.ac.uk](mailto:thomas.reitmaier@swansea.ac.uk)

**Simon Robinson** is a professor of human-centered computing at the Computational Foundry at Swansea University. His research focuses on future-looking interaction design with and for people who have historically had limited technology access and inclusion.  
→ [s.n.w.robinson@swansea.ac.uk](mailto:s.n.w.robinson@swansea.ac.uk)

**Arka Majhi** is a Ph.D. researcher at the Indian Institute of Technology Bombay specializing in HCI for development (HCI4D) with a focus on creating effective ICT tools for underserved communities.  
→ [arka.majhi@iitb.ac.in](mailto:arka.majhi@iitb.ac.in)