Transforming Social Media Signals into Accurate Population Maps for Crisis Response

Google NotebookLM/WorldPop, 4 February 2026

Transcript

Speaker 1: Welcome to the Deep Dive. Before we get into the data, we need to be transparent about who we are.

Speaker 2: That's right.

Speaker 1: We are not real people. We are AI-derived voices generated from source material uploaded by WorldPop.

Speaker 2: And it is also important to state clearly that the audio has been edited, checked, and validated by experts at WorldPop to ensure accuracy.

Speaker 1: Okay, with that covered, let's jump in. Imagine a nightmare scenario for a humanitarian aid worker. A typhoon hits.

Speaker 2: Right.

Speaker 1: You have resources, you have food, you have medical teams. But you have one massive blind spot. You don't actually know where the people are.

Speaker 2: It sounds impossible in the age of GPS and smartphones, doesn't it? But, you know, aid organisations usually rely on census data.

Speaker 1: And that's static. It's old.

Speaker 2: Exactly. It tells you where people lived 3, 5, maybe even 10 years ago. It doesn't tell you where they are today, right now.

Speaker 1: So today we're looking at a solution to that lag. We're covering a new research pre-print co-authored by WorldPop Principal Research Fellow, Dr. Shengji Lai that establishes a statistical framework for transforming biased and anonymised social media signals into high-resolution population estimates.

Speaker 2: It's about figuring out how to take messy, noisy Facebook location data. and clean it up enough to actually save lives in real time.

Speaker 1: But messy feels like a huge understatement.

Speaker 2: Yeah.

Speaker 1: I mean, you can't just count Facebook users and assume that's everyone. “My grandmother isn't on Facebook”. And what about places with no signal?

Speaker 2: You're right, but there's actually an even bigger hurdle than your grandmother not having an account. It's privacy. Specifically, something called differential privacy.

Speaker 1: This is the censoring issue the paper talks about.

Speaker 2: That's the one. Meta, Facebook's parent company has this really strict rule. They divide the world into these little map tiles, about 4.8 kilometres square. And if a specific tile has fewer than 10 users, they don't just show a low number, they censor it. They show nothing at all. Zero.

Speaker 1: Which makes sense for privacy, I guess. You don't want to be able to identify the one person living on a remote farm.

Speaker 2: Right, but for a disaster model, it creates these huge data gaps. And unfortunately, those gaps disproportionately erase rural and low population areas. So, the very people who might need the most help are suddenly invisible.

Speaker 1: So, the team had to figure out how to fill in those black holes.

Speaker 2: And they used what's called a Bayesian imputation approach. Think of it like a puzzle where you're missing some pieces. You don't know exactly what the missing piece looks like, but based on the colours and the lines of all the pieces around it, you can calculate the probability of what fits there.

Speaker 1: So, they used the patterns in the visible data to mathematically reconstruct the hidden data.

Speaker 2: And it worked. By doing this, they restored data coverage for 5.5% of rural areas. Areas that were before this completely invisible in the data set.

Speaker 1: That's potentially thousands of people appearing back on the radar. But that only solves one problem, the privacy gap. You still have to link those Facebook users to the actual total population.

Speaker 2: And this is where the model gets really clever. They found they could bridge that gap using three key predictors. First, the level of urbanisation. Pretty straightforward. Second, the demographic makeup, specifically the number of working age adults, since they're most likely to have phones.

Speaker 1: And the third one was nighttime lights.

Speaker 2: Yeah, it's a brilliant proxy for human activity. The radiance of lights at night correlates so strongly with human presence and, you know, technology usage.

Speaker 1: So even if someone isn't on Facebook, if they live in a town that glows at night, the model can infer a population density.

Speaker 2: You got it.

Speaker 1: So, they have this complex model math, demographics, satellite lights, Facebook pings. How did they prove it actually works in the real world.

Speaker 2: They looked at the Philippines; it's a perfect testing ground.

Speaker 1: Why there?

Speaker 2: Well, about 78% of the population is on Facebook, which is huge, and the country faces a very high risk of natural disasters like typhoons. But the real stroke of genius was when they tested it.

Speaker 1: When was that?

Speaker 2: May 4th, 2020.

Speaker 1: Wait, May 2020. That's peak COVID lockdown.

Speaker 2: Bingo. Usually social media data is frantic, right? People are moving around.

Speaker 1: But on that day, everyone was stuck at home.

Speaker 2: Exactly. Their night time location on Facebook matched their actual census residents almost perfectly.

Speaker 1: So, it gave them a rare ground truth, a perfect baseline to check their math against.

Speaker 2: It did. And the error rates were remarkably low. We're talking about 19% for urban areas and around 24% for rural. In the chaotic world of disaster statistics, having a real-time signal that's that accurate is a massive step forward.

Speaker 1: Because unlike the census, this isn't on a 10-year cycle.

Speaker 2: No, it's dynamic. If the nighttime lights shift or if demographics in an area change, the estimates update. It moves from a static snapshot to a live feed of humanity.

Speaker 1: To read the full pre-print, follow the link below.