Global CENSUS Archive

SPOTLIGHT

November 2026

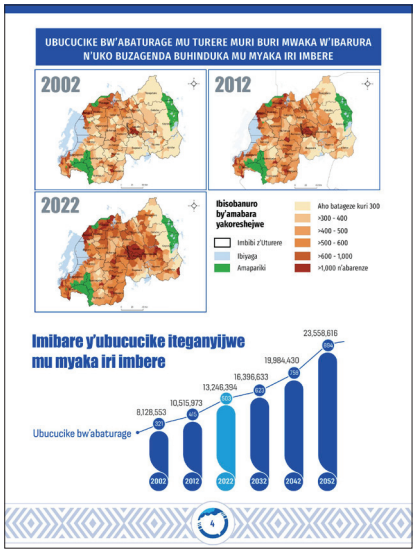
A newsletter from East View highlighting census data from around the world

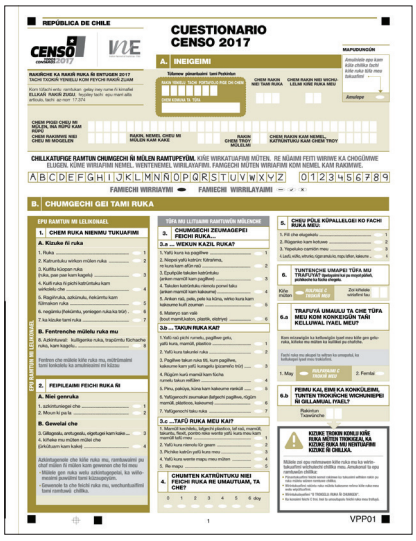
CENSUSES AS PRIMARY SOURCES IN THE AGE OF ARTIFICIAL INTELLIGENCE

This year's Charleston conference (November 2–6) was marked by discussion and exploration of the effects of AI in library collections and workflows. A variety of sessions investigated its practical applications and impact on scholarly communications. A recurring theme was that libraries need to play an active role in shaping AI development, ensuring that AI-assisted research remains grounded in trusted content, transparent sourcing, and appropriate access rights. East View sponsored session, "From Scattered Sources to Lasting Collections: Censuses, Maps, and Research" reinforced the importance of primary sources as the foundation of research.

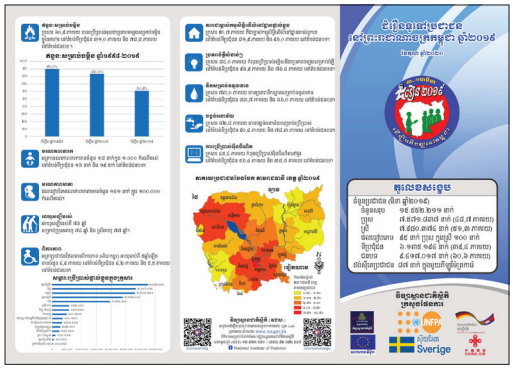
> **As AI tools increasingly mediate how researchers discover and synthesize information, the quality and contextual richness of the underlying source material become critical.**

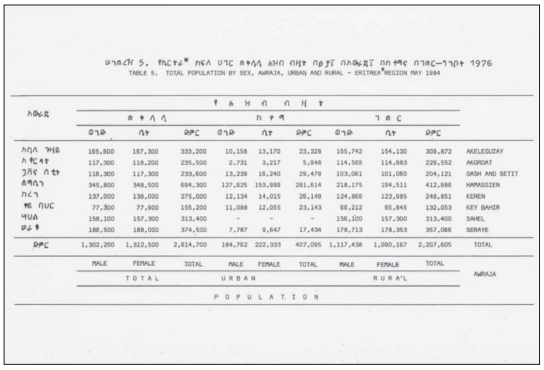
Researchers might assume that historical or contemporary census data could be accessed or summarized through AI agents. But that is not actually the case. For example, an independent evaluation by the UK fact-checking organization Full Fact found that ChatGPT and Gemini repeatedly misrepresented 2021 UK Census figures about English-language proficiency. Several responses attributed approximately 138,000 people living in households where no one spoke English to the foreign-born population. Although the models identified a real figure (137,876), they misclassified it; the actual foreign-born subgroup numbered 129,440 among people aged 16 and older. Census publications frequently contain dense tables, footnotes, subtotals, and multilevel headings. There is research suggesting that the LLMs have limitations seeking information from tables, especially ones that have complicated formats. A 2025 IMF study found that ChatGPT and Gemini frequently misreported official economic statistics, often producing plausible but incorrect figures, highlighting that while GenAI excels at synthesizing information, it remains unreliable for retrieving precise official data.





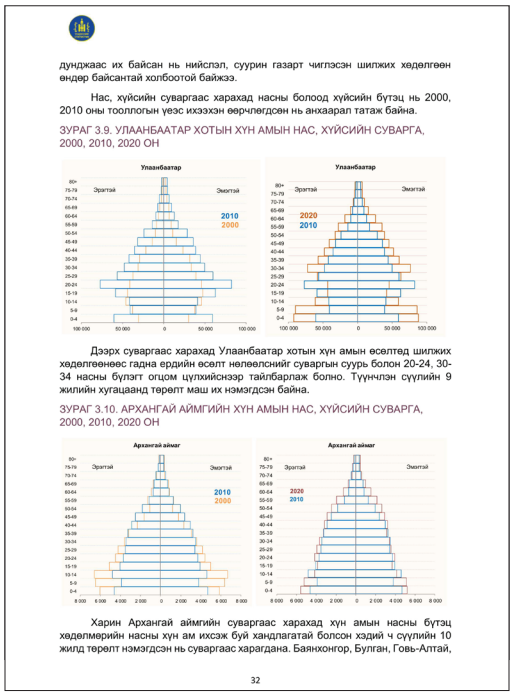
Top: Rwanda, 2022. Census Results. Booklet. Kinyarwanda Language, p. 5. https://gca.eastview.com/CENSUS-2685394BO.1.5
Bottom: Chile, 2017. Census Questionnaire. Mapudungun Language, p. 1. https://gca.eastview.com/CENSUS-2363884BO.1.1





Left: Cambodia, 2019. Census Report Leaflet. Khmer Language, p. 1.
https://gca.eastview.com/CENSUS-2678241BO.1.1

Right: Ethiopia, 1984. Preliminary Report. v01 no01, p. 55.
https://gca.eastview.com/CENSUS-2539992BO.1.55



дунджаас их байсан нь нийслэл, суурин газарт чиглэсэн шилжих хөдөлгөөн өндөр байсантай холбоотой байжээ.

Нас, хүйсийн суваргаас харахад насны болоод хүйсийн бүтэц нь 2000, 2010 оны тооллогын үеэс ихээхэн өөрчлөгдсөн нь анхаарал татаж байна.

ЗУРАГ 3.9. УЛААНБААТАР ХОТЫН ХҮН АМЫН НАС, ХҮЙСИЙН СУВАРГА, 2000, 2010, 2020 ОН

Улаанбаатар

Эрэгтэй Эмэгтэй

2010
2000

Улаанбаатар

Эрэгтэй Эмэгтэй

2020
2010

Дээрх суваргаас харахад Улаанбаатар хотын хүн амын өсөлтөд шилжих хөдөлгөөнөөс гадна ердийн өсөлт нөлөөлснийг суваргын суурь болон 20-24, 30-34 насны бүлэгт огцом цүлхийснээр тайлбарлаж болно. Түүнчлэн суулийн 9 жилийн хугацаанд төрөлт маш их нэмэгдсэн байна.

ЗУРАГ 3.10. АРХАНГАЙ АЙМГИЙН ХҮН АМЫН НАС, ХҮЙСИЙН СУВАРГА, 2000, 2010, 2020 ОН

Архангай аймаг

Эрэгтэй Эмэгтэй

2010
2000

Архангай аймаг

Эрэгтэй Эмэгтэй

2020
2010

Харин Архангай аймгийн суваргаас харахад хүн амын насны бүтэц хөдөлмөрийн насны хүн ам ихсэж буй хандлагатай болсон хэдий ч суулийн 10 жилд төрөлт нэмэгдсэн нь суваргаас харагдана. Баянхонгор, Булган, Говь-Алтай,

32

Mongolia, 2020. Thematic Study. v03. Gender Equality, p. 32.
https://gca.eastview.com/CENSUS-2698939BO.1.33

These models might also confuse original historical figures with later revisions. Furthermore, census categories change over time. Without an actual human analysis, categories like race or ethnicity could be used without taking the historical context into account. Administrative boundaries within countries also frequently change. Current LLMs confuse boundaries and treat changing boundaries as permanent across different census years. In other words, a number can be accurately extracted from a census publication but assigned to the wrong geographic unit, making the resulting analysis unreliable. And finally, not all census publications are published in languages that play a big role in LLM training.

As AI tools increasingly mediate how researchers discover and synthesize information, the quality and contextual richness of the underlying source material become critical. Digitized historical documents, official census publications, and structured historical datasets are particularly valuable because they provide researchers with verifiable evidence while making otherwise difficult-to-access materials discoverable and usable in new AI-driven research environments. The Global Census Archive supports this work by bringing together historical census publications and structured GIS datasets from around the world, providing researchers with authoritative primary sources and geographically detailed data.

Further reading on artificial intelligence and censuses:

- "AI Chatbots Benchmarking Report: An Independent Assessment," Full Fact: https://fullfact.org/ai-benchmark/report-one/
- "StatGPT: AI for Official Statistics," IMF: https://www.imf.org/en/publications/departmental-papers-policy-papers/issues/2026/03/10/statgpt-ai-for-official-statistics-573514

Global CENSUS Archive

The East View **Global Census Archive**® (GCA) is an innovative program to collect official data and publications issued by the national census authorities of more than 200 countries, from the early 19th century to the present day, providing a key resource for scholars and analysts to find current and historical census data to support critical research into our changing world. GCA provides a single platform experience to explore the world's census publications: all countries and regions, the complete depth of published census materials, including original census publications in e-book form, as well as modern tabular data in Excel format and GIS files for advanced users.