Is AI a Fundamental Shift for Digital and Computational Demography?

Ridhi Kashyap
Leverhulme Centre for Demographic Science
Department of Sociology
Nuffield College
University of Oxford

October 16, 2026

The Digital Revolution and Demography

LEVERHULME CENTRE FOR DEMOGRAPHIC SCIENCE

UNIVERSITY OF OXFORD

International Union for the Scientific Study of Population

ABOUT IUSSP MEMBERSHIP SCIENTIFIC GROUPS MEETINGS & EVENTS TRAINING CAREERS RESOURCES PUBLICATIONS

Digital and Computational Demography

Chair



Prof Ridhi Kashyap
University of Oxford

Members



Dr Diego Alburez-Gutierrez
Max Planck Institute for Demographic Research (MPIDR)

Ms Monica Alexander
University of Toronto

Prof Dennis Feehan
UC Berkeley

Prof Rachel Franklin
Newcastle University

Mrs Katherine Hoffmann Pham
UN Global Pulse

Dr Emmanuel Olamijuwon
University of St Andrews



Prof Ingmar Weber
Saarland University

Description

This panel will build on the success of the previous panels - Big Data and Population Processes (2015-2018) and Digital Demography (2019-2022) - to help consolidate the field of Digital and Computational Demography, and promote exchange and collaboration between demographers as well as other social scientists interested in population, and computer scientists. The panel will further help forge links between demographers and population scientists and a growing, interdisciplinary community of computational social science. It will help advance the data revolution for sustainable development agenda, with a view to understanding how demographers and population scientists can contribute to leveraging the new sources of data and computational methods for understanding population and sustainable development processes.

The goal of the panel is to explore demography in relation to the digital revolution along three axes: (1) examine the implications of digital transformations for population and sustainable development outcomes, e.g. linked to health, fertility and family, gender inequalities, and migration; (2) assess how digital trace and geospatial data sources, generated as by-products of technologies, can be repurposed for demographic research; (3) explore the demographic applications of computational methods, such as microsimulation, agent-based modelling and machine learning.

Programme of activities

Chapter 3: Digital and computational demography

Ridhi Kashyap, R. Gordon Rinderknecht, Aliakbar Akbaritabar, Diego Alburez-Gutierrez, Sofia Gil-Clavel, André Grow, Jisu Kim, Douglas R. Leasure, Sophie Lohmann, Daniela V. Negraia, Daniela Perrotta, Francesco Rampazzo, Chia-Jung Tsai, Mark D. Verhagen, Emilio Zagheni, and Xinyi Zhao

in Research Handbook on Digital Sociology

Category: Chapter **Collection:** Sociology, Social Policy and Education 2023
Published: 31 Mar 2023 **DOI:** https://doi.org/10.4337/9781789906769.00010
Page Range: 48–86 **Publisher:** Edward Elgar Publishing

Keywords: Digital demography; Computational demography; Digital trace data; Geospatial data; Demographic bias; Online recruitment; Machine learning; Microsimulation; Agent-based modelling

Abstract

Digital and computational demography explores demography in relation to the digital revolution - the rapid technological improvements in digitized information storage, computational power and the spread of the internet and mobile technologies since the turn of the new millennium. We cover three ways in which the digital revolution touches upon demography. First, we discuss how digital technologies, through their



Research Handbook on Digital Sociology

Edited by Jan Skopek

Handbook

Published: 31 Mar 2023
Print ISBN: 9781789906752
eISBN: 9781789906769
DOI: https://doi.org/10.4337/9781789906769
Pages: 492
Collection: Sociology, Social Policy and Education 2023

Home › Handbook of Computational Social Science for Policy › Chapter

Leveraging Digital and Computational Demography for Policy Insights

Chapter | Open Access | First Online: 24 January 2023
pp 327–344 | Cite this chapter

You have full access to this open access chapter

Download book PDF Download book EPUB

Ridhi Kashyap & Emilio Zagheni



Handbook of Computational Social Science for Policy

Sections References

Abstract

IUSSP Scientific Panel on Digital and Computational Demography (2023-26)

Digital and Computational Demography

- **Demography in relation to the digital revolution** (Kashyap et al 2023; Kashyap and Zagheni 2023)

  (1) New data opportunities: digital traces, digitalised records, new modes of data collection
  Does AI change what counts as data?

  (2) **Computational methods:** simulation, agent-based models, machine learning and prediction
  Does it change how we model and predict?

  (3) **Impacts of digitalisation:** information, services, networks and connectivity shaping demographic behaviours
  Does it matter more than mobile phones and the internet did?

Which AI? The Generative Turn

Big data, digital traces, supervised machine learning	Deep learning	The generative turn
2009 Google Flu Trends nowcasts influenza from searches **2012** Email data used to estimate migration (Zagheni and Weber) **2015** Phone metadata predict wealth (Blumenstock et al.) **2017-23** Facebook data data track migrant stocks (Zagheni et al.); gender gaps and population displacement (Kashyap et al) **2020** Fragile families challenge (Salganik et al)	**2012** Deep learning breaks through in image recognition (AlexNet) **2017** Transformer architecture **2018–20** Pretrained language models: BERT, GPT-3 **2021** "Foundation models" (Bommasani et al.)	**Nov 2022** ChatGPT: public launch of generative AI **2023** GPT-4 and multimodal models **2023–24** Life-course foundation models (life2vec); silicon samples

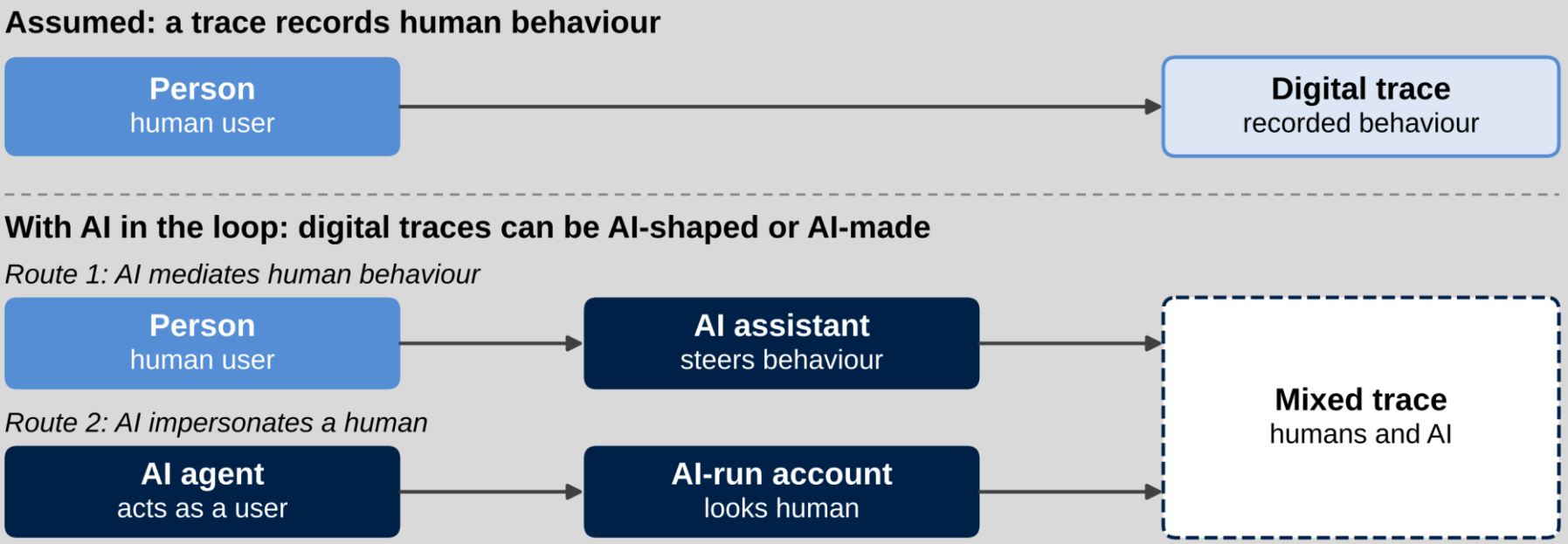
AI is already inside DCD. What does this generative turn add?

Data

(1) **New data opportunities:** digital traces, digitalised records, new modes of data collection

→ *Does AI change what counts as data?*

Data: Contamination Risk



Online surveys: An agent built from a simple prompt evaded standard quality checks and held a consistent demographic persona (Westwood 2025)

Data: Silicon Samples

- **A new mode of data collection with a validity problem**

  The promise: models prompted with demographic backstories reproduced US survey distributions (Argyle et al., 2023)

  Some problems:

  - Variance compressed; relationships between variables do not match benchmarks (Bisbee et al., 2024)
  - Opinions skew toward some groups (Santurkar et al., 2023)

Data: Unlocking New Streams

LEVERHULME CENTRE FOR DEMOGRAPHIC SCIENCE

UNIVERSITY OF OXFORD

Promising, but not new in kind

- **What LLMs add:** zero-shot handling of varied handwriting, text and tables, e.g. parish registers, historical census records, verbal autopsies with no bespoke training data
- LLMs can outperform other algorithmic approaches for verbal autopsies (Wen et al 2026; Chen et al 2025)

Computational Methods

LEVERHULME CENTRE FOR DEMOGRAPHIC SCIENCE

UNIVERSITY OF OXFORD

Computational methods: simulation, agent-based models, machine learning and prediction

Does AI change how we model and predict?

Supervised ML v Foundation Models

LEVERHULME CENTRE FOR DEMOGRAPHIC SCIENCE

UNIVERSITY OF OXFORD

Supervised ML

Labelled population data
(X, Y) from a defined sample

↓

Learn a target mapping
e.g. predict fertility intention

↓

Apply to a stated task
performance tied to target and population

Reuse: one model per outcome
Data: curated samples

Foundation models

Broad pretraining corpus
often self-supervised next-token or related objectives

↓

Learn reusable representations
patterns encoded across the training material

↓

Adapt to many tasks
prompting, fine-tuning or tool use

Reuse: one model, many outcomes
Data: millions of records

→ Generative agent-based models: agents with richer behavioural representations

Risks and Challenges

LEVERHULME CENTRE FOR DEMOGRAPHIC SCIENCE

UNIVERSITY OF OXFORD

Sharpened by foundation models

WHO is represented?

Which populations, languages, periods and institutions enter the corpus?

Representation and exclusion

WHERE did it come from?

Which sources and selection processes, and who can build with them?

Representation and exclusion

HOW MUCH data is needed?

Millions of records, against thousands of cases for supervised ML

Exclusion: capability concentrates in data-rich settings

Apply to machine learning more broadly

WHAT is learned?

Is a learned dimension a theoretical construct, or a proxy that predicts well?

Opacity and construct validity

WHEN does it travel?

Do results hold across place, cohort or model version? Are they reproducible?

Construct validity and reproducibility

WHY does it predict?

Can we explain why it predicts, or only that it predicts?

Interpretability

Impacts of Digitalisation

Mobile phones and the internet have changed how people seek information, access services, work, interact with networks and form relationships with implications for demographic behaviours

Does the adoption of Gen AI tools matter more than mobile phones and the internet did?

Impacts of Digitalisation

- LLMs are becoming more **persuasive and realistic**: relevant for pathways from information-seeking to behaviours

- AI as a **gatekeeper of institutional decisions**: greater needs for algorithmic auditing; implications for inequalities

- **Unequal exposures and deepening digital divides:** AI acceleration, within a world where ~2+ billion are still offline (64% Africa not internet users)

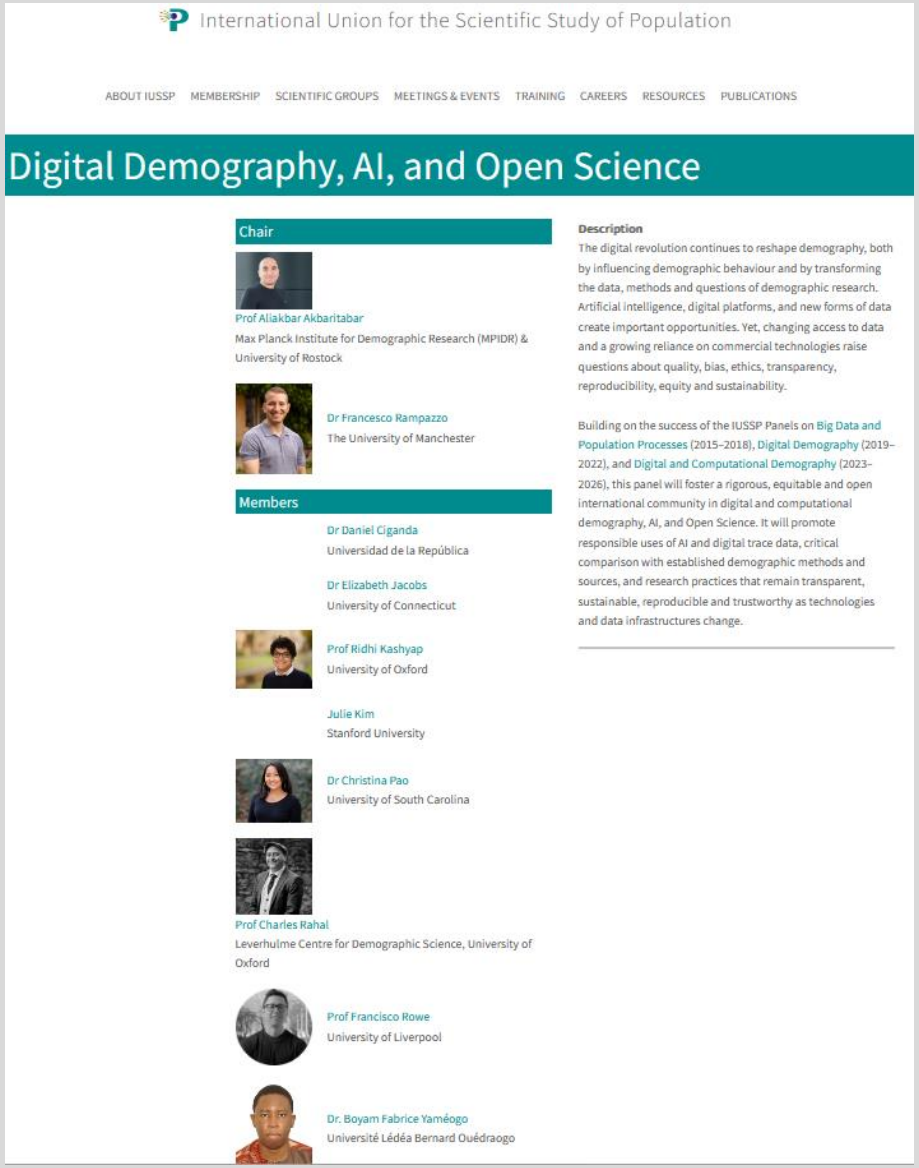
Fundamental shift?

- (Gen) AI fits within the existing DCD framework encompassing data, methods and impacts

- **Continuities:** opportunities with deepening challenges of representation, validity, interpretability, reproducibility

- **Discontinuities:** contamination of digital data, multimodality, new representations of data
  - AI as research infrastructure for population science (like desktop computers?)

AI as a new phase within DCD



International Union for the Scientific Study of Population

ABOUT IUSSP MEMBERSHIP SCIENTIFIC GROUPS MEETINGS & EVENTS TRAINING CAREERS RESOURCES PUBLICATIONS

Digital Demography, AI, and Open Science

Chair

Prof Aliakbar Akbaritabar
Max Planck Institute for Demographic Research (MPIDR) & University of Rostock

Dr Francesco Rampazzo
The University of Manchester

Members

Dr Daniel Ciganda
Universidad de la República

Dr Elizabeth Jacobs
University of Connecticut

Prof Ridhi Kashyap
University of Oxford

Julie Kim
Stanford University

Dr Christina Pao
University of South Carolina

Prof Charles Rahal
Leverhulme Centre for Demographic Science, University of Oxford

Prof Francisco Rowe
University of Liverpool

Dr. Boyam Fabrice Yaméogo
Université Lédéa Bernard Ouédraogo

Description

The digital revolution continues to reshape demography, both by influencing demographic behaviour and by transforming the data, methods and questions of demographic research. Artificial intelligence, digital platforms, and new forms of data create important opportunities. Yet, changing access to data and a growing reliance on commercial technologies raise questions about quality, bias, ethics, transparency, reproducibility, equity and sustainability.

Building on the success of the IUSSP Panels on Big Data and Population Processes (2015–2018), Digital Demography (2019–2022), and Digital and Computational Demography (2023–2026), this panel will foster a rigorous, equitable and open international community in digital and computational demography, AI, and Open Science. It will promote responsible uses of AI and digital trace data, critical comparison with established demographic methods and sources, and research practices that remain transparent, sustainable, reproducible and trustworthy as technologies and data infrastructures change.

- **Digital Demography, AI, and Open Science:** new iteration of IUSSP panel starting 1 November 2026 for the next three years

- Co-led by Aliakbar Akbaritabar (Max Planck Institute for Demographic Research and University of Rostock) and Francesco Rampazzo (University of Manchester)

Thank you!

ridhi.kashyap@nuffield.ox.ac.uk

Funding: Leverhulme Trust for the Leverhulme Centre for Demographic Science (Grant RC-2018-003) and the Gates Foundation (INV-045370)

AI Disclosure: GPT-6 and Claude Sonnet 5.5 were used in preparing these slides