

Study: The Infrastructure Paradox of AI Development in LMICs

Data Availability Statement

The dataset supporting the findings of this study is publicly available in the Harvard Dataverse repository at <https://doi.org/10.7910/DVN/VJY5QV>. They comprise the following:

1. Dataset and variable documentation

The repository contains the raw survey responses as exported from Google Forms and the cleaned analytical dataset used for all reported results. Variables cover country, respondent role and sector, income-group classification, electricity infrastructure, internet infrastructure, hardware access and cost, cloud access and cost, training data access and quality, AI policy status, AI interventions and investments, and AI innovation strength. A variable codebook and data dictionary (PDF) documents all variable definitions, coding schemes, and transformations.

Key transformations include: standardisation of country names; mapping of countries to World Bank income groups (low-income, lower-middle-income, upper-middle-income); recoding of categorical survey responses into ordinal numeric variables for affordability, reliability, accessibility, and quality measures; grouping of multi-response items into broader analytical categories; and construction of derived binary indicators for AI policy presence and AI innovation strength.

2. Analysis code

The repository includes a Python notebook and a Stata do-file reproducing the full analytical workflow. The Python notebook covers data import, variable renaming, standardisation, income-group assignment, recoding, descriptive summaries, and figure generation. The Stata do-file covers value labelling, construction of derived variables, cross-tabulations, chi-square tests of association stratified by income group, and export of the analytical tables reported in the manuscript. The main inferential procedure was the chi-square test of independence, examining associations between infrastructure variables and three outcomes: existence of AI policy, existence of AI interventions or investments, and AI innovation strength. All percentage distributions reported in the tables were generated from within-group cross-tabulations.

3. Survey instrument

The full survey questionnaire, including the informed consent disclosure presented to participants, is available as a supplementary PDF document attached..