README for Reproducibility Packages

Contents

Overview

This package includes data and code to reproduce figures and tables in the paper "Who on earth is using generative AI? Global trends and shifts in 2025".

Data Availability

This section will outline where and how the data supporting the findings of the study can be accessed and used. This is crucial, especially for replicating the results, as only the same data will be able to produce consistent results. Make sure to list all the datasets used and categorize them as follows:

- ☐ All data are publicly available.
- ☒ Some data cannot be made publicly available.
- ☐ No data can be made publicly available.

Data Sources

The data is proprietary and cannot be made public or used for any other purposes. Main data is from Semrush, a website traffic data provider. The data is saved in "data" folder.

Semrush datasets (proprietary)

- GenAI_by_country_202101_202505.csv
    - Description: Monthly visits, unique users, and engagement metrics (bounce rate, average visit duration, pages per visit) for 60 generative AI tools by country.
    - Version/Access date: Extracted in June 2025, covering January 2021 to May 2025.
    - Access: Requires an institutional subscription and API credentials from Semrush (https://www.semrush.com/projects/). Data were retrieved via Python API calls and exported to CSV. Replicators must set up an account and

obtain API access; instructions are available from Semrush's API documentation.

- Benchmark_by_country_202101_202505.csv
    - Description: Monthly visits and users for a set of benchmark websites (e.g., Google, YouTube) by country. Used for comparison with generative AI tools.
    - Version/Access date: Extracted in June 2025, covering January 2021 to May 2025.
    - Access: Same as above. Provided here as CSV exports, since replication requires licensed Semrush API access.
- genai_list2506.xlsx
    - Description: Full list of 60 generative AI tools used in the paper, including tool name, target domain, category (e.g., Chatbot, Image, Video & Audio, Productivity), and additional descriptive fields (traffic snapshots, producing country). This file defines the scope of tools analyzed.
    - Version/Access date: Assembled June 2025, last updated with Semrush traffic values on June 3, 2025.
    - Access: Internal reference file created by the World Bank Digital Development team. Cannot be publicly released because it includes licensed Semrush traffic values that are confidential and subject to subscription restrictions.
    - Notes: The analysis in the code does not rely on the manually entered traffic values in this file (those are for reference only). All quantitative results are computed directly from the Semrush raw CSV datasets.

Other internal datasets (derived, not public)

- 2024crosscountry_reg.dta
    - Description: Prepared dataset combining Semrush outcomes with country-level covariates for April 2024, used in regression analysis.
    - Citation: Liu, Yan and He Wang (2024). Who on Earth is using generative AI? World Bank, Washington, DC, USA. Available at: https://blogs.worldbank.org/en/digital-development/who-on-earth-is-using-generative-ai-
    - Access: Internal dataset prepared by the World Bank Digital Development team. Cannot be publicly released as it incorporates licensed Semrush data, which is confidential and subject to subscription restrictions.
    - Contact. He Wang hwang21@worldbank.org

Public supporting data can be included in the package

- **dt2025.dta**
  Description. Country level digital development indicators used as covariates. The file has been reduced to two public variables:
    - using_int_p Individuals using the Internet percent of population.
      - WDI indicator code. **IT.NET.USER.ZS**
      - Source page. World Development Indicators at the World Bank Open Data portal https://data.worldbank.org/indicator/IT.NET.USER.ZS
      - Example programmatic access in Stata. wbopendata, indicator("IT.NET.USER.ZS") language(en) long clear
    - Fixed_mbps Fixed network speed from Ookla Open Data.
      - Project page. Ookla Open Data https://github.com/teamookla/ookla-open-data
      - Download options. Follow the repository instructions to obtain the fixed performance parquet files from the public AWS S3 bucket or the BigQuery public dataset. Typical steps are
        i go to the repository page above and read the Data Access section
        ii if using AWS, list the fixed performance path s3://ookla-open-data/parquet/performance/type=fixed/ and copy the quarter directories for the required years
        iii if using BigQuery, add the public project shown in the repository and query the fixed performance tables, then export a country level extract.
      - Licensing. The Ookla Open Data license permits redistribution of derived aggregates with attribution. This file contains a country level fixed speed measure and can be included in the package with citation to Ookla Open Data.
- **countryname_concordance_FY26.dta**
    - Description: Concordance that maps ISO codes to World Bank regions and income classifications for FY26 and adds alternative spellings of country names. The table draws on public World Bank country metadata and manual edits.
    - Variables from WDI metadata: “countryname_inwdi”, “wbname”, “iso2”, “iso3”, “cen_lat”, “cen_lon”, “wbregion_full”, “wbregion_name_full”, “region_wdi”, “regionname_wdi”, “adminregion_wdi”, “adminregionname_wdi”, “income_full_fy26”, “income_name_full_fy26”, “lendingtype_wdi”, “lendingtypename_wdi”.
    - Manual edits: countryname (harmonized spellings), region_2ssa, region_withhic, wbregion_client, wbclient, subregion_ce, plus group flags

(eu27, eu28, oecd, asean5, lac6, fcs, smallstates, g20, g7, eu, eurozone, apec, brics5).

Statement about Rights

- ☒ I certify that the author(s) of the manuscript have legitimate access to and permission to use the data used in this manuscript.
- ☐ I certify that the author(s) of the manuscript have documented permission to redistribute/publish the data contained within this replication package. Appropriate permission are documented in the LICENSE.txt file.

Instructions for Replicators

- All figures and tables in the paper can be replicated with the two do-files in the "code" folder.

List of Exhibits

The provided code reproduces:

- ☐ All numbers provided in text in the paper
- ☒ All tables and figures in the paper
- ☐ Selected tables and figures in the paper, as explained and justified below

Requirements

Computational Requirements

No particular requirements.

Software Requirements

- **Stata version 15**
- **Excel**

Memory and Runtime and Storage Requirements

Run time is within 30 minutes.