

ReadMe: Reproducibility Package for “How is Government using AI?”

1. Overview

This README provides instructions for reproducing the results presented in *How Is Government Using AI?*, a background paper prepared for the World Bank’s *World Development Report (WDR) 2026: The Promises of Artificial Intelligence*.

The reproducibility package contains the datasets and code required to reproduce the analyses and figures presented in the paper. Replicators should download the full contents of the package and preserve the existing folder structure.

The main replication file, `AI Background Paper Reproducibility.Rmd`, imports and processes the required datasets and reproduces the analyses and figures presented in the paper. Before running the code, replicators will need to update the author-specific working directories and file paths to reflect the location of the reproducibility package on their computer. Detailed instructions are provided in Section 3: Instructions for Replicators.

All figures generated by the replication code are saved in the 3 - Output folder.

2. Data availability

The primary data source for this paper is the AI & Data for Better Governance Survey, designed and administered by a DEC-PR team as part of the analytical work for the *World Development Report 2026*. The survey provides a comprehensive view of government AI adoption from both whole-of-government and sectoral perspectives.

The survey comprises 13 questionnaires: one Agency questionnaire, completed by the country’s central digital authority or equivalent institution, which covers AI adoption and the broader enabling environment; and two questionnaires for each of six core government sectors—public finance, taxation, procurement, civil service, health, and education. At the sector level, the Systems questionnaire captures AI applications using data from management information systems, while the Manager questionnaire examines AI adoption, governance, and enabling conditions within the relevant ministry or agency.

In addition, the analysis uses the World Bank’s official FY2026 country income classifications to group economies by income level (Table 1).

Table 1. List of datasets used for the analysis

Dataset	Description	Access policy	URL
Agency Questionnaire_WIDE.csv	AI & Data for Better Governance Survey: Whole-of-government questionnaire	Public	https://microdatalib.worldbank.org/index.php/catalog/20124
Dfbg_managers_datasets: Education Manager Survey.csv, Health Manager Survey.csv, Civil Service Manager Survey.csv, Taxation Manager Survey.csv, Public Finance Manager Survey.csv, Public Procurement Manager Survey.csv	AI & Data for Better Governance Survey: Managers from 6 core sectors (tax, public finance, procurement, education, health, civil service) questionnaires	Public	https://microdatalib.worldbank.org/index.php/catalog/20124
Dfbg_systems_datasets : Systems Questionnaire Education_WIDE.csv, Systems Questionnaire Health_WIDE.csv, Systems Questionnaire Civil Service_WIDE.csv, Systems Questionnaire Tax_WIDE.csv, Systems Questionnaire Public Finance_WIDE.csv, Systems Questionnaire Procurement_WIDE.csv	AI & Data for Better Governance Survey: Systems Experts from 6 core sectors (tax, public finance, procurement, education, health, civil service) questionnaires	Public	https://microdatalib.worldbank.org/index.php/catalog/20124
OGHIST_2026_07_01.csv	World Bank GNI per capita Operational Guidelines &	Public	https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups

	Analytical Classification		
--	------------------------------	--	--

3. Instructions for replicators

The analysis is contained in a single R Markdown file, AI Background Paper Reproducibility.Rmd. The file imports and cleans the survey datasets, merges World Bank income classifications, constructs the variables and indices used in the analysis, and reproduces the figures presented in the paper.

To reproduce the results:

1. Open RStudio.
2. Install the R packages required by the replication code: dplyr, tidyr, readr, purrr, tibble, ggplot2, scales, forcats, ggrepel, stringr, knitr, janitor, and readxl. The code also uses functions from additional packages, including ggthemes, so replicators should ensure these are installed before running the complete file.
3. Download the complete reproducibility package and preserve the following folder structure:

Reproducibility Package/

```
|
|
|— 1 - Data/
|— 2 - Code/
|  └─ AI Background Paper Reproducibility.Rmd
└─ 3 - Output/
```

4. Open AI Background Paper Reproducibility.Rmd in RStudio.
5. Update the author-specific directory paths at the beginning of the script to correspond to the location of the reproducibility package on the replicator's computer. In particular, replicators should update:
 - the working directory (setwd) to point to the 1 - Data folder;
 - output_dir to point to the 3 - Output folder;
 - INPUT_MANAGER to point to dfbg_managers_dataset.csv; and

- any remaining absolute paths beginning with C:/Users/ so that they point to the corresponding files within 1 - Data.
6. Run the R Markdown file from beginning to end. The order of execution should not be changed because objects created in earlier sections are used in subsequent analyses.
 7. The script reads the Agency, Manager, and Systems questionnaire datasets from the 1 - Data folder. For the sector-level analysis, separate Manager and Systems files are loaded for taxation, public procurement, health, public finance, civil service, and education.
 8. The script also uses OGHIST_2026_07_01_new.xlsx to assign economies to the World Bank FY2026 income classifications.
 9. All final figures generated using ggsave() are saved in the 3 - Output folder.
 10. The .Rmd can also be knitted to HTML to produce a single document containing the code-generated analyses, figures, and supporting tables.

Table 2 below summarizes all the exhibits that appear in the paper.

Table 2. List of Exhibits

Exhibit	Description	Output file
Figure 1*	Key actions for responsible AI implementation	AI_BACKGROUND_1.JPEG
Figure 2	Most needed form of support for AI adoption across sectors	AI_BACKGROUND_2.JPEG
Figure 3	Data sources leveraged by government AI projects, by income level	AI_BACKGROUND_3.JPEG
Figure 4	Internal AI use in government, by income level	AI_BACKGROUND_4.JPEG
Figure 5	Areas of AI use in government, by income level	AI_BACKGROUND_5.JPEG

Exhibit	Description	Output file
Figure 6	External AI use in government, by income level	AI_BACKGROUND_6.JPEG
Figure 7	Internal AI use in government, by sector	AI_BACKGROUND_7.JPEG
Figure 8	Provider model of flagship AI applications, by income level	AI_BACKGROUND_8.JPEG
Figure 9	Distribution of agency questionnaire indices, by income level	AI_BACKGROUND_9.JPEG
Figure 10	Correlations between government AI maturity indices	AI_BACKGROUND_10.JPEG
Figure 11	Most widely reported barriers to AI adoption, by income level	AI_BACKGROUND_11.JPEG
Figure 12	External barriers to AI adoption, by income level	AI_BACKGROUND_12.JPEG
Figure 13	Key barriers to AI adoption, by sector	AI_BACKGROUND_13_1.JPEG, AI_BACKGROUND_13_2.JPEG
Figure 14	Stage of adoption of a national AI strategy, by income level	AI_BACKGROUND_14.JPEG
Figure 15	Government adoption of GenAI guidelines, by income level	AI_BACKGROUND_15.JPEG
Figure 16	Level of AI performance evaluation, by income level	AI_BACKGROUND_16.JPEG
Figure 17	Map of AI adoption vs. governance: four archetypes	AI_BACKGROUND_17.JPEG
Figure 18	Return on investment of AI projects, by income level	AI_BACKGROUND_18.JPEG

Exhibit	Description	Output file
Figure 19	Top AI use cases in management information systems, by sector	AI_BACKGROUND_19_1.JPEG, AI_BACKGROUND_19_2.JPEG

*Note on Figure 1

The open-ended responses to Question 34 were categorized into thematic groups using an AI-assisted classification step (CLAUDE). The replication code documents the exact classification prompt and categories used. The resulting classifications are provided in `data_export (10).xlsx`, which is used as an input to reproduce Figure 1.

This means replicators do not need to rerun the AI classification to reproduce the published figure; they use the provided classified dataset.

4. Requirements

The analysis was conducted in R using an R Markdown (.Rmd) file. Replicators should have R and RStudio installed to run the replication code and knit the complete reproducibility document.

5. 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 permissions are documented in the LICENSE.txt file.