Using GenAI to Enhance Statistical Reporting

UNICEF Data & Analytics Section

CIS Region on the Path to Achieving the Sustainable Development Goals

May 21-23, 2025, Bishkek, Kyrgyz Republic

AI Challenge / AI Possibility

National Statistical Offices have long served as a **key enabler for evidence-based policy** and reliable government decision making.

But the landscape is shifting:

- AI technology is evolving faster than our ability to make use of it.
- The public sphere is increasingly influenced by non-traditional data sources.
- Trust in data, even official data, is fragile.

Our shared challenge: How do we **safeguard the credibility of official data and statistics**, while **embracing new tools** that allow us to deliver faster, focused, ever more relevant insights?

GenAI: A Flexible Tool for the Data Lifecycle

While uncertainty remains, the promise of GenAI for official data and statistics is immense. Across the data lifecycle, new capabilities are emerging. A few examples:

- **Survey Design**: LLMs can test question clarity, simulate interviews, and propose cultural adaptations.
- **Data Processing**: Natural language interfaces accelerate pipeline debugging, cleaning, and recoding.
- **Analysis**: AI can support fast exploratory analysis, anomaly detection, and summarization.
- **Reporting**: GenAI enables tailored narratives, visual summaries, and multi-audience outputs.
- **Localization**: AI agents can generate district-specific narratives in multiple languages.

Frontier Data Technology Unit | Yves Jaques yjaques@unicef.org

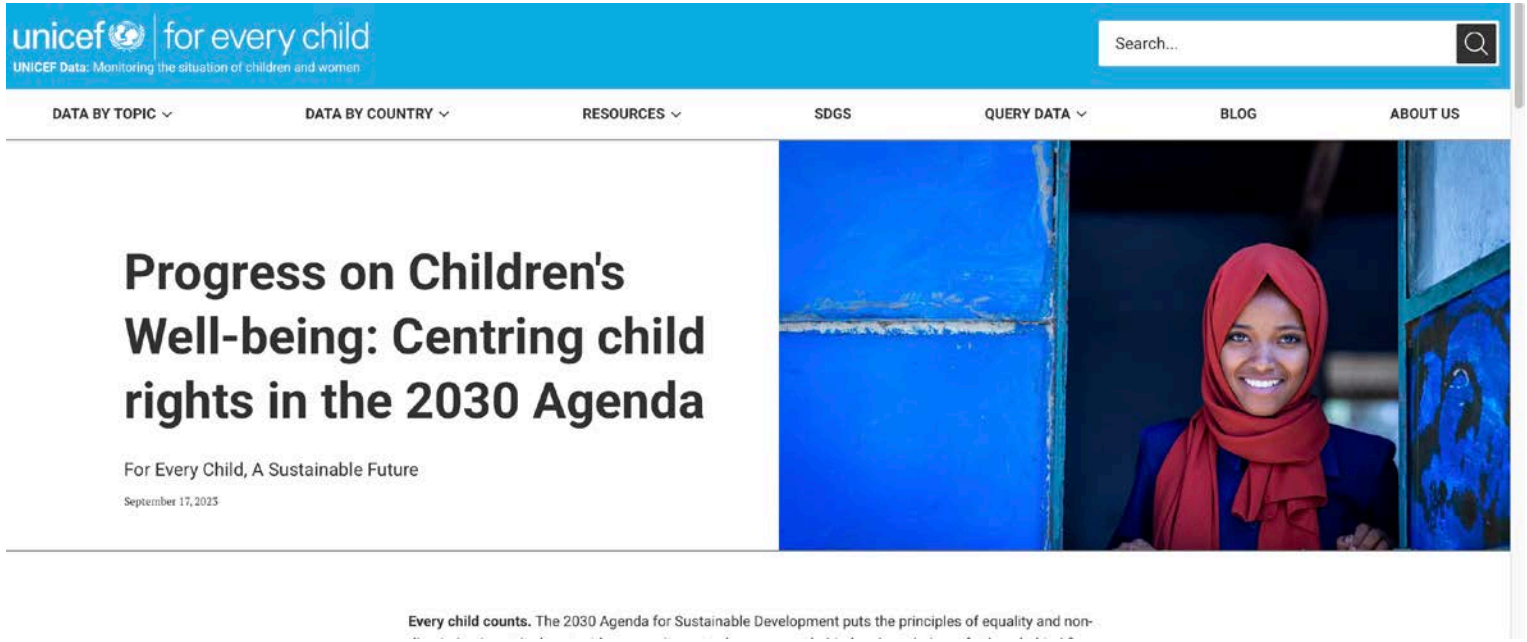
Our Approach: Move Fast ~~and~~ *do not* Break Things

UNICEF is exploring GenAI guided by principles of:

- **Guardrails**: Limiting scope to reduce hallucination risk.
- **Augmentation**: Supporting, not replacing, human analysts.
- **Test Test Test**: Logging usage, using disclaimers, and building in strong QA workflows.

Three current field-tested experiments are helping us assess **feasibility, safety, and value**!

Test Case 1 – SDG Report Chatbot
Data Communications Team



Test Case 1 – SDG Report Chatbot
Data Communications Team

Challenge: surveying our users told us that our 98-page SDG report was difficult to navigate.

Solution: A chatbot built over a single flagship PDF using GPT-4o-mini.

- Scoped to only answer questions grounded in the PDF.
- Out-of-scope queries receive standard disclaimers.
- Human-in-the-loop QA to thoroughly test quality.
- Cost controls and user feedback mechanisms implemented.

Impact: Improved access and understanding without sacrificing statistical rigor.

Test Case 2 – AHEAD: Localized Narratives for Health

Health Analytics Team

Frontier Data Technology Unit | Yves Jaques yjaques@unicef.org

Test Case 2 – AHEAD: Localized Narratives for Health

Health Analytics Team

Challenge: contextualizing public health administrative data across hundreds of districts.

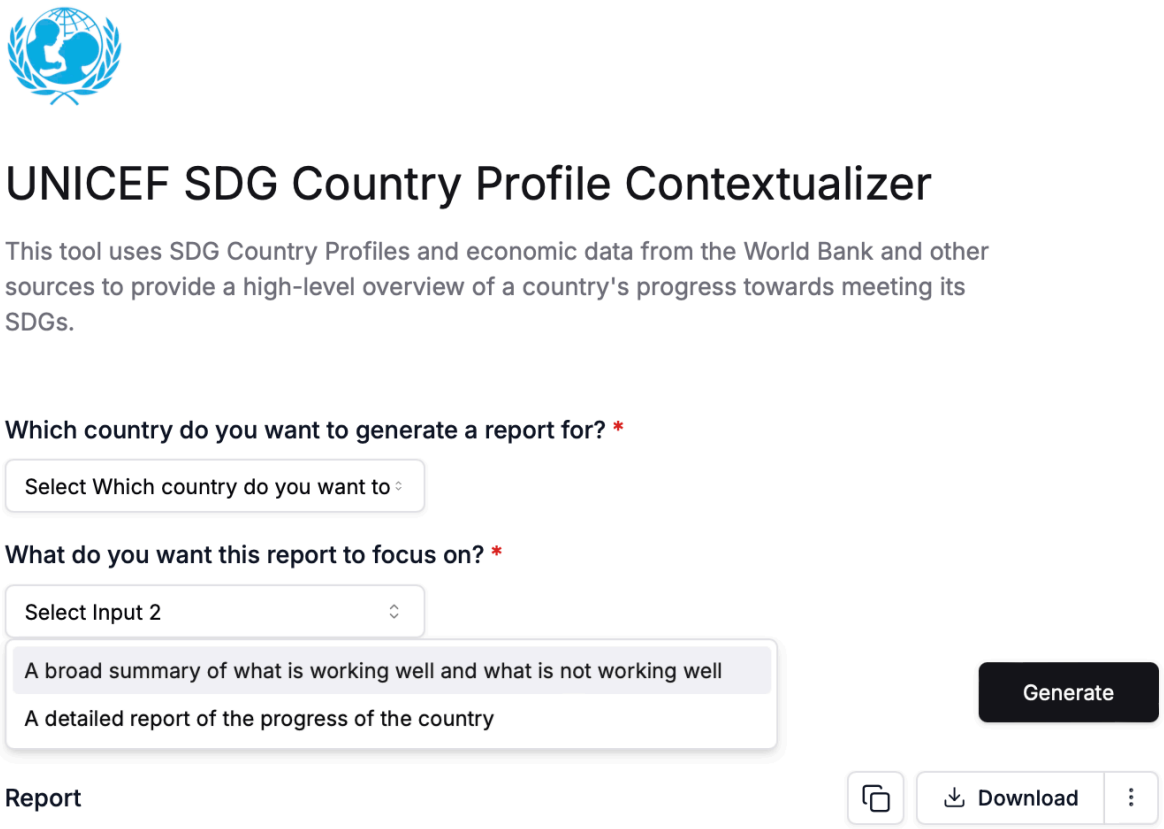
Solution: R-generated analytics from primary admin data with vector database of health reports passed into AI prompts for narrative output.

- Narrative generation **based on actual policies and local context** with vector data base built on consistently updated national and local reports.
- **District-specific outputs** with clear recommendations based on local reporting.
- Mixture of “traditional” **scripted data generation** and visualization, connected with well-engineered **data pipelines**, and then **enhanced with GenAI narration** capabilities.
- **Challenge areas**: Calibrating AI narration with precise prompt engineering, controlling hallucinations, ensuring output transparency.

Impact: Hyper localized and actionable data-driven narratives generated at scale.

Case 3 – GEN-SDG: Enhancing SDG Country Briefs

Frontier Data Network



Frontier Data Technology Unit | Yves Jaques yjaques@unicef.org

Case 3 – GEN-SDG: Enhancing SDG Country Briefs

Frontier Data Network

Challenge: New SDG benchmarking methods are rigorous but complex to interpret.

Solution: Retrieval-Augmented Generation (RAG) prototype built with MIT Sloan.

- AI-generated narratives over SDG benchmarking PDFs using various role-based summaries.
- Live integration of additional World Bank data for economic context.
- Evaluated for accuracy and tailored messaging.

Impact: Role-based prompts with GenAI can facilitate analytic outputs tailored to diverse audiences

Lessons: GenAI parsing of Data visualization remains problematic, scripted outputs of structured data into visualizations still more reliable in this area. Hybrid approach best!

Where Next? Responsible GenAI in Official Statistics

Where do we see opportunities for further integration of GenAI?

- **Role-aware outputs**: clear value in generating custom formats for user groups: directors, researchers, private sector, general public.
- **Data-grounded generation**: further testing to anchor AI on structured official data, eliminating (minimizing?) hallucination risk .
- **Shared QA infrastructure**: collaboration across the official data and statistics community: prompt libraries, evaluation methods, results.
- **Field-first design**: deploying responsible tools that are usable by local staff requiring little engineer and analyst intervention.
- **Transparent methods**: test where and how we can produce result reproducibility where source data fed to models is opaque.

Frontier Data Technology Unit | Yves Jaques yjaques@unicef.org

Final Reflections

GenAI will either foster or impede the future of how we communicate official data and statistics. It is up to us to drive its responsible use to enhance our delivery. Our experience at UNICEF has shown us that:

- Statisticians, Data Scientists, Data and Software Engineers must **lead the way together**.
- Collectively **we have the experience over the data lifecycle** to design robust implementations that build strong **quality assurance** into AI innovation.
- **Probabilistic tools like GenAI can coexist** with our rigorous official methodologies when leveraged where their value is high (narrative augmentation, summarization, localization, contextualization, personalization)
- Trust is built through **process transparency** and **rigorous testing**.

We are cautiously hopeful that **GenAI can reinforce our ability** to generate timely and relevant analytics that **build public trust** in data in a world starved for credible, actionable information.