



AI Era Challenges and Opportunities: Reaching Users of Official Data and Statistics

57th Session of the United Nations Statistical Commission
Friday Seminar on Emerging Issues: AI-Readiness for Official Data and Statistics
Keynote Address Delivered by **Dr. Koh Eng Chuan, Chief Statistician**, Singapore Department of Statistics
Acknowledgements due to Andy Chiang, Malcolm Cai, Priyadharshini N, Tess Tan, Tay Jun Wen, and Tow Joon Han for their valuable inputs

Introduction

Government statisticians today operate in the midst of a rapidly evolving Artificial Intelligence (AI) landscape. They are already using AI in their data work, in the way they process and code, how they scrape data and generate new insights.

The public is turning to AI systems to search for information. Increasingly, AI influences how policy debates are framed and how decisions are made. Many no longer visit official statistical websites as they look for easy access to answers to their queries. Answers are increasingly intermediated through third-party platforms aided by algorithms — search engines, chatbots, dashboards, and automated policy tools. AI engines and applications continue to evolve and progress to become more powerful, faster, and sophisticated.

This AI development raises a central question:

How do government statistical offices remain the authoritative source of official statistics if they are no longer the primary interface through which users encounter them?

Exploring this question requires examining how users now access statistics, the risks this creates, and the opportunities it opens for rethinking the role and approaches of a statistical office.

Today's Users and Access Methods

Understanding who uses official statistics today and how they access them provides a good context for these challenges.

Traditionally, users of official statistics were well defined. They included government officials and policymakers, researchers and analysts, the media, businesses, and members of the public. These users ranged from sophisticated analysts to those looking at headlines. Much of the dissemination model was built around these users — structured publications, tables for standard and customised requests, press releases, and official government websites.

Today, however, there is a rapid rise of **new and indirect users**. AI systems—large language models, search engines, and automated decision tools—are now major consumers of official statistics. People increasingly receive statistical information not as raw data, but as AI summaries, explanations, or answers to questions.

Consider how many people now search for information. Instead of navigating to a statistical office's website, they type a question into Google, ChatGPT, Claude, or Microsoft Copilot:

What is inflation this year?

Is unemployment rising?

How much do statisticians earn compared to data scientists?

Google used to surface the link to the website of the statistical office, but now they may provide an AI answer as their first response. Google and other AI platforms provide different responses to data search depending on prompts and timing. The key shift being witnessed is:

From direct consumption of official statistics to machine-mediated interpretation of them.

This shift has profound implications for the visibility, authority, and relevance of a government statistical office.

Core Challenges from AI-Mediated Data

Government statistical offices face four interrelated challenges from AI-mediated data access.

Diminished Visibility and Authority

AI responses often omit sources altogether. Google's AI search has recently introduced links citing where its results were referenced, but other AI search engines may not provide such information.



Even when sources are mentioned, official and unofficial data are frequently blended without distinction. From the user's perspective, authority shifts away from the statistical office and towards the platform providing the answer.

The risk is subtle: official statistics disappear into background infrastructure. Trust accrues to the interface – the search engine or AI assistant – rather than the institution that produced the data with care, independence, and public accountability.

Over time, this erodes recognition of the unique role of official statistics and weakens the social contract that underpins the work of government statisticians.

AI Hallucination and Misinterpretation

AI systems are very fluent, but they do not 'understand' statistics in the way trained statisticians do. As a result, common and recurring errors occur, such as:



- Using any available data, outdated data, or from non-official sources
- Ignoring revisions, rebasing, or seasonal adjustment
- Overlooking methodological breaks or changes in definitions
- Comparing data from different sources with varying scope and coverage

The AI responses often sound technically plausible but are statistically incorrect.

These subtle errors are particularly dangerous. They can mislead policymakers, journalists, and the public without triggering immediate scepticism.

Accessibility and Usability

The third challenge concerns the relative ease of access and usability of AI-mediated answers compared to official statistics.



Expert users can leverage their domain knowledge to effectively interpret official statistics – understanding the terminology, classifications, and metadata. However, those who do not have such expertise face substantial accessibility challenges. They have to invest significant time and effort to learn technical concepts and methodological nuances to make sense of statistical outputs.

By contrast, AI systems provide information in plain language, framed as direct answers with confident explanations. Accuracy largely depends on the prompts entered.

This creates a real tension between methodological rigor and accessibility. If official statistics are technically correct but difficult to understand, whilst AI summaries are easy to consume but imperfect, many users are likely to choose convenience over precision.

Timeliness Versus Statistical Quality

The age-old issue of timeliness versus quality presents another challenge.



Users expect answers immediately rather than waiting weeks or months. AI-driven environments create expectations of instant, latest figures. They may pull data from any available sources to provide users with the most timely information, ignoring statistical quality.

Official statistics, however, operate within necessary constraints – validation and accuracy, confidentiality protection, governance processes, and professional review.

The risk is that users turn to faster, lower-quality sources – over which statistical offices have little or no control – leading to poor analysis and weak evidence for decision-making. When users become indifferent to data sources, the relevance of official statistics may be threatened.

Case Study: When AI Gets Official Statistics Wrong

These challenges are illustrated through an example involving AI responses on the youth unemployment rate for Singapore. It cited figures that appear credible.

Table 1: Comparison of Responses from Various AI Platforms on Different Days

Question: What is the youth unemployment rate of Singapore?				
Platform	15 Jan 2026		20 Jan 2026	
	Response	Source Cited	Response	Source Cited
ChatGPT	6.8% Y2024	Ministry of Manpower Singapore	7.8% Y2024 (Prioritised) 6.8% Y2024	Trading Economics, World Bank data (Prioritised) Ministry of Manpower Singapore
Google AI	7.8% Y2024	Statista.com	7.8% Y2024	Statista.com (Source hidden to non-logged in users)
Claude	5.7% 2Q2025	The Online Citizen article*	6.8% Y2024 (Prioritised) 5.7% Y2024	Vulcan Post (indirectly quoting the Ministry of Manpower) The Online Citizen article
perplexity	6.8% Y2024	Ministry of Manpower Singapore	7.8% Y2024 (Prioritised) 6.8% Y2024	Various websites all pointing to World Bank data; Trading Economics, the global economy, YCharts (Prioritised) Ministry of Manpower Singapore
deepseek	4.4% Y2023	Not indicated	5.7% Y2023	Not indicated
Gemini	6.8% Y2024	Not indicated	6.8% Y2024	Mothership, indirectly quoting the Singapore Department of Statistics

● Wrong Responses
 ● Partially Correct Responses
 ● Correct Responses

* Data provided by Claude on 15 Jan 2026 was correct, but it refers to youths under 30 years. The definition of youths used by the Ministry of Manpower Singapore and the International Labour Organisation for unemployment refers to those aged 15-24.

However, closer inspection revealed significant problems. The responses used different definitions or scope on labour force, cited sources that are not official, and the answers change within just five days (Table 1). Some responses are partially correct, whilst some are entirely wrong. Yet they all appear genuine, usually providing more information and explanation than requested, giving them an expert appearance. For non-expert or convenience-seeking users, these results would likely be accepted without question.

This case study highlights key lessons for government statisticians:

AI errors are often subtle, not absurd

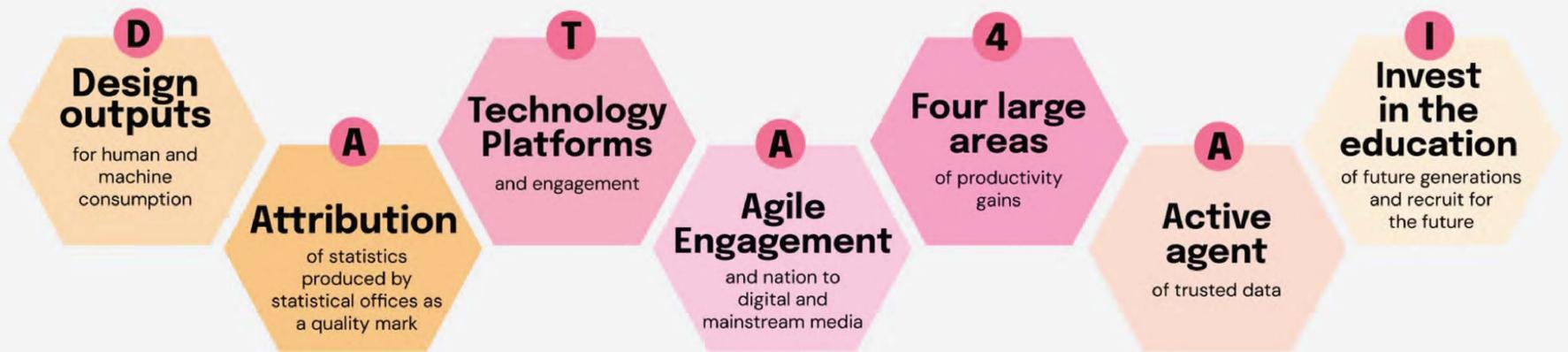
AI errors propagate quickly and at scale

Silence from official sources allows misinterpretation and misinformation to become the dominant narrative, particularly since investigating how errors occur requires considerable effort

Opportunities and Strategic Responses for Government Statisticians

The challenges outlined present real opportunities for government statisticians to respond strategically and quickly. These opportunities can be organised around a **DATA4AI strategic framework** (Figure 1).

Figure 1: DATA4AI Strategic Framework



Design Outputs for Human and Machine Consumption

Government statistical offices must design outputs for both machines and humans. Search engine optimisation (SEO) is no longer optional. If AI systems cannot easily find official statistics, they effectively do not exist in the digital information ecosystem.

Statistical offices should evaluate if their data are truly AI-ready. AI systems require specific data formats, structural frameworks, and technical interfaces to accurately access and utilise official statistics. This calls for investment in rich metadata, accessible APIs and interoperability, clear standards and classifications, and version controls and timestamps.

D Can agentic AI bots talk to generative AI bots or find data through standardised data formats such as statistical data and metadata exchange (SDMX), open APIs, or model context protocols (MCP)? The Singapore Department of Statistics (DOS) exemplifies this approach by promoting use of APIs, and developing AI-ready websites by working on SEO. DOS's AI chatbot SANDRA, currently in beta, uses semantic search capabilities to understand the context of users' data enquiries and uses large language model to retrieve data via APIs from an e-Service, the SingStat Table Builder. DOS is also in the process of including API documentation in OpenAPI specifications to enable search engines and AI chatbots to discover DOS's data and information. A future consideration is MCP, which provides a standardised mechanism to facilitate integration and interoperability with other AI applications.

Statistical offices have to recognise that machines are among their primary users and design accordingly.

Attribution of Statistics Produced by Statistical Offices as a Quality Mark

Strong attributions mechanisms are essential to protect and clearly signal official statistics.

This includes clear guidance on citation, persistent identifiers for key indicators, digital markers of authenticity, and ensuring official website links appear in AI responses. Metadata has to be clear with websites open to all, so that the AI systems can identify and link to official sources. In Singapore, DOS is working on improving metadata with clear meta titles and concise meta descriptions to improve discoverability. Structured data and sitemap.xml files are included to help search engines better understand the information offered on DOS's platforms and for crawlers to discover webpages and datasets. Additionally, DOS plans to improve core web vitals on how quickly and smoothly users interact with the webpages – a factor that search engines such as Google consider when ranking search results.

Users and AI systems must be able to recognise what is official, current, and authoritative.

Technology Platforms and Engagement

Proactive Engagement with technology platforms and the AI ecosystem is crucial.

This includes dialogue with search engines and AI developers, with potential coordination through international bodies such as the United Nations Statistics Division or the World Bank, to engage major platforms in a systematic way for their AI bots to use official data as first responses.

However, security and trust remain paramount. Systems containing large amounts of data need to stay secured. AI bots that reach into official websites will have to be carefully scrutinised or managed to prevent data breaches.

Participation in data and AI governance standards is therefore critical.

Agile Engagement and Explanation to Digital and Mainstream Media

Statistical offices must shift from pure publishing to agile engagement and explanation.

Traditional tables and PDF reports alone are no longer sufficient. This requires working nimbly with mainstream and social media to ensure official statistics continue to be referenced.

A

Agility means producing plain-language explanations that anticipate common questions and likely AI misinterpretations. Statistical offices may consider publishing accessible information nuggets and analysis that can be easily reproduced by both mainstream media and social media users. Visualisations have to evolve beyond traditional bar charts to include tree maps, box and whiskers plot, Sankey diagrams, network and nodes graphs, and heatmaps. New visualisation tools are available for such uses, such as Power BI, Tableau, and Qlik.

The role of a statistical office increasingly includes interpretation, not just dissemination, requiring innovative approaches to attract and engage users.

Four Large Areas of Productivity Gains

AI offers substantial opportunities and internal productivity improvements across four large areas of the data value chain:

1. Data collection using web AI tools
2. AI/ ML tools for processing, quality checks, drafting and analysis
3. AI systems for data dissemination by statistical offices and data retrieval by users
4. AI-generated and curated new analytical insights

4

Singapore's experience includes SANDRA and DICE. SANDRA helps users with data queries – on a simple prompt, it charts data and provides metadata and other related links. DICE is used in data processing, which is a machine learning (ML)-powered tool that recommends the most appropriate Singapore Standard Classification codes based on textual descriptions.

Using web scrapers and other AI/ ML tools, the potential for creating new knowledge is great. Aside from price captures, DOS has created experimental indicators such as e-commerce, green economy, innovative companies, and enterprise characteristics.

These tools enable statisticians to focus on higher-value work – analysis, stewardship, engagement, and keep in pace with AI developments.

Active Agent of Trusted Data

Government statisticians must accept their **evolving role**:

A

- From data producer to **data steward** – step up to set data governance standards and usage across government agencies and the administrative data they generate.
- From publisher to **interpreter and consultant** – work with agencies and the larger public to discuss and educate on statistical interpretation and usage for analysis to answer hypothesis and provide key insights and understanding to the economy and society.
- From neutral passive provider to **active agent of trusted data** – transform statistical offices as the official authority of data, data stewardship and data usage across government, academia and businesses by providing advisories and insights.

This evolution strengthens rather than compromises independence, positioning statistical offices as the official source for quality, objective statistics amid AI-mediated information.

Invest in the Education of Future Generations and Recruit for the Future

Investment in statistical literacy for the AI age is essential. Collaboration with educational institutions enables the younger generations of data consumers to understand the importance of quality official statistics.

I

Users need to be equipped with skills to question AI-mediated statistics, understand uncertainty, and recognise the importance of methodology, definitions and context. New statisticians should be AI-trained, proficient in programming languages such as Python and R, and understand AI operations for effective engagement.

This is not just a technical task – it is a public education and recruitment mission.

Conclusion: A Public-Sector Call to Action

Official statistics cannot compete with the speed of AI, but statistical offices should understand and leverage AI effectively. Official statistics remain essential for unbiased and open understanding of the society, and for evidence-based policies. The task is ensuring that official statistics are easily discoverable by AI systems, engaged beyond official portals, and explained in clear language. This approach protects the public value of official statistics in the age of AI.

References

UNSC (2026) – Presentation of statistical data and metadata, dissemination of statistics through the Statistics Division and its publications and open data, 57th UNSC Agenda item 3(m) – E/CN.3/2026/15, New York, 3–6 March 2026. [2026-15-DataPresentationDisseminationOpenData-E](#)

World Bank (2025) – Fostering AI-Readiness and Responsible Redistribution of Official Statistics, Room Document for 56th UNSC, New York, 4–7 March 2025. https://unstats.un.org/UNSDWebsite/statcom/session_56/documents/Room_Document-WorldBank-E.pdf

World Bank (2026) – AI-Readiness and (Re)Distribution of Official Statistics, Opportunities, Challenges, and Recommendations, prepared by the World Bank with input from CCSA members, 57th UNSC Agenda item 5(h) – E/CN.3/2026/31, New York, 3–6 March 2026. [BG-5h-CCSA_AI_Readiness_Official_Statistics_v2-E](#)