

# Generative AI for official statistics

InKyung Choi

Statistical Management and Modernisation Unit

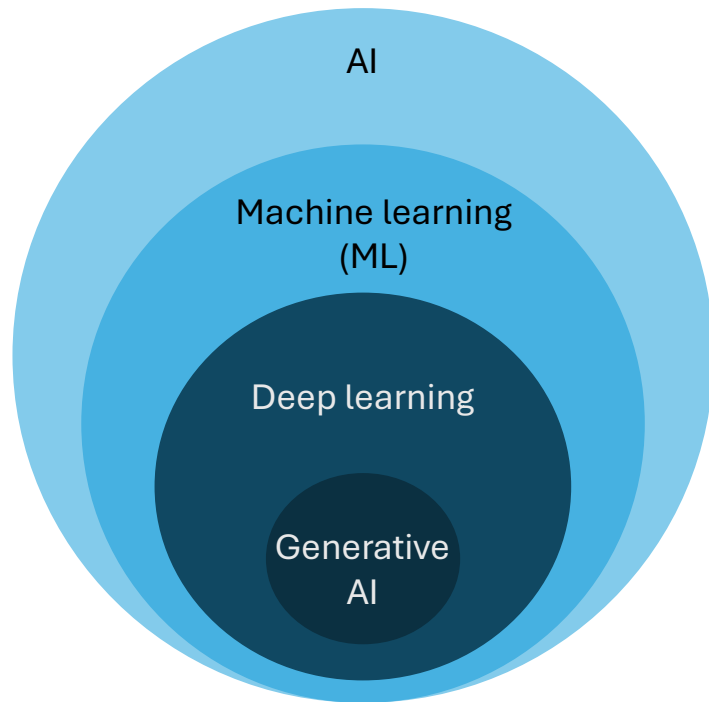
United Nations Economic Commission for Europe (UNECE)

@2<sup>nd</sup> Sprint on AI and Data Science for Economic Statistics (Dec 12, 2024)



**1. What    2. Why    3. Where    4. How    5. Open  
questions**

# What?



- **Artificial Intelligence (AI)** is a broad field of computer science that focuses on creating systems and machines capable of performing tasks that typically require human intelligence.
- **Machine Learning (ML)** is a subset of AI that involves the use of algorithms and statistical models to enable computers to improve their performance on a specific task through learning from data, without being explicitly programmed.
- **Generative AI (Gen AI)** refers to AI systems that can generate new content or data that is not explicitly derived from existing examples.

# What? – ML, DL and Gen AI



## Machine Learning (ML)

Images of cats and dogs

**Data**

+

guidance on which features to  
look for (e.g., ear, nose, size)

**Feature**

=>

the pointier ears are, the smaller the  
size is, the more likely to be a cat

**Prediction**



# What? – ML, DL and Gen AI

Lights?

Stroke?



Vermeer



Rembrandt

Subject?

Colour?

Position?

## Machine Learning (ML)

Images of cats and dogs

+

guidance on which features to look for (e.g., ear, nose, size)

=>

the pointier ears are, the smaller the size is, the more likely to be a cat

**Data**

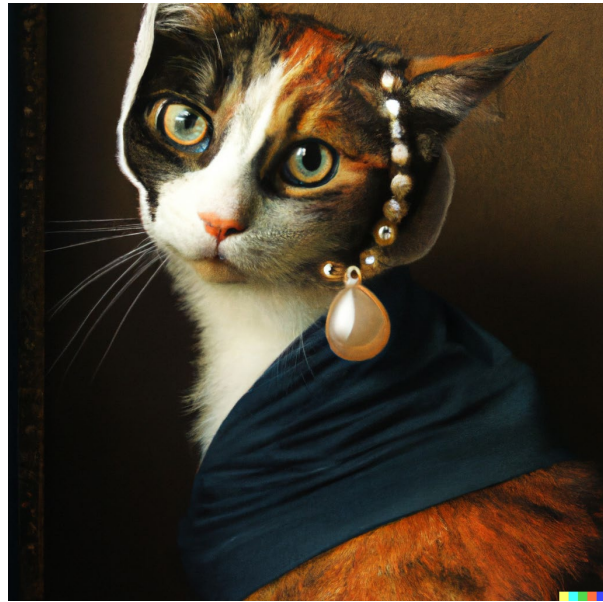
**Feature**

?

**Prediction**

# What? – ML, DL and **Gen AI**

“cat with a  
pearl earring  
painted in  
Vermeer style”



**Inference / estimation**

**Prediction**

**Generation**



“golden  
retriever  
portrait in  
Rembrandt  
painting style”

# Why?

Gen AI offers unprecedented capabilities

As statistical organizations strive to

- Expanding data and services
- Engaging more and better with citizens
- Modernizing internal operations



# Why?

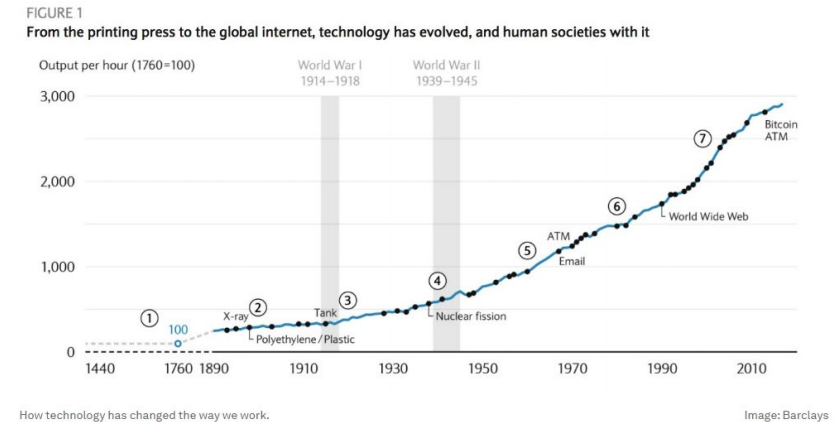
Gen AI offers unprecedented capabilities

As statistical organizations strive to

- Expanding data and services
- Engaging more and better with citizens
- Modernizing internal operations

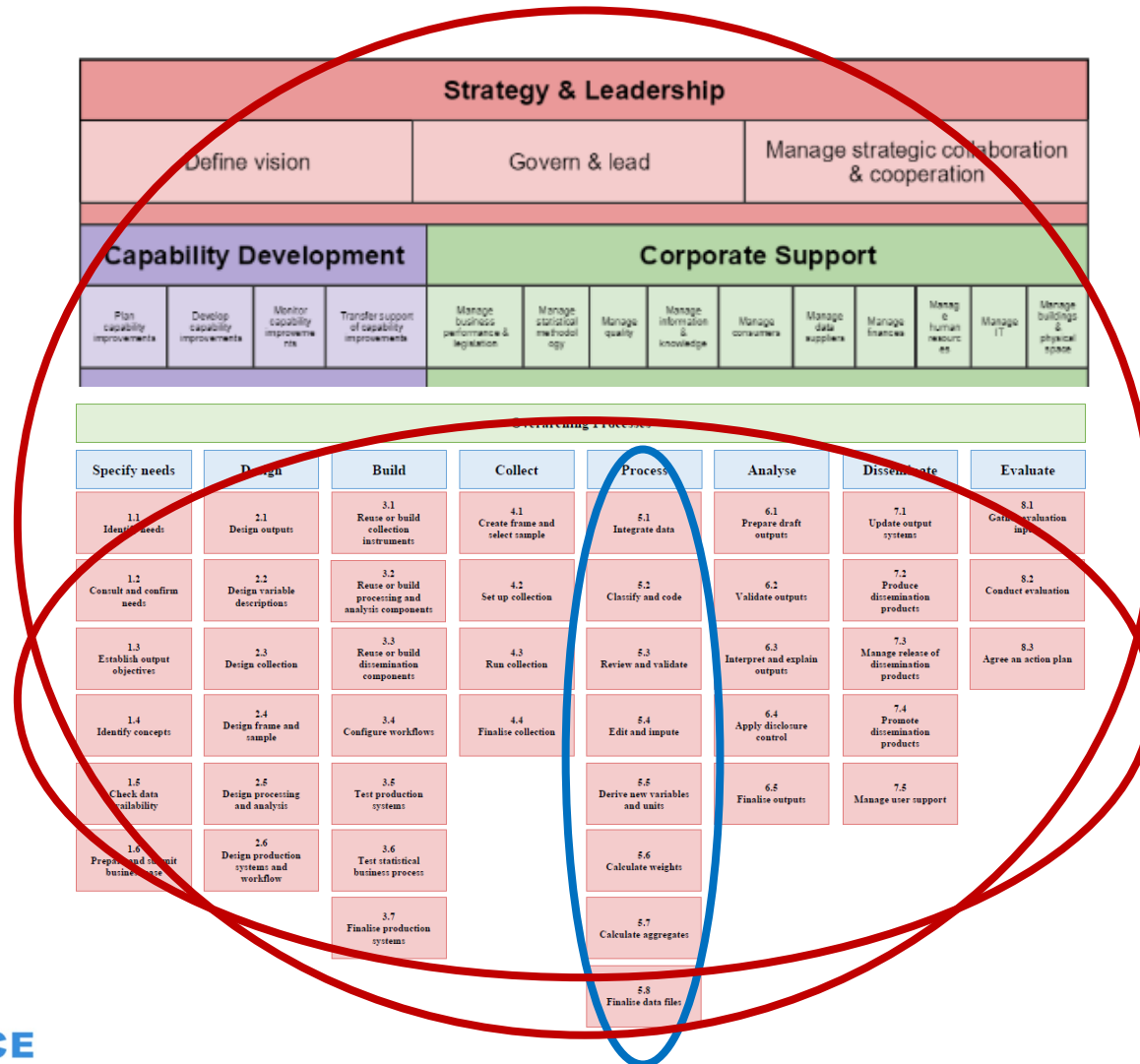
➡ Increase efficiency and productivity

➡ Improve service delivery





# Where?



Other activities in statistical organizations

Statistical production and dissemination

# Where?

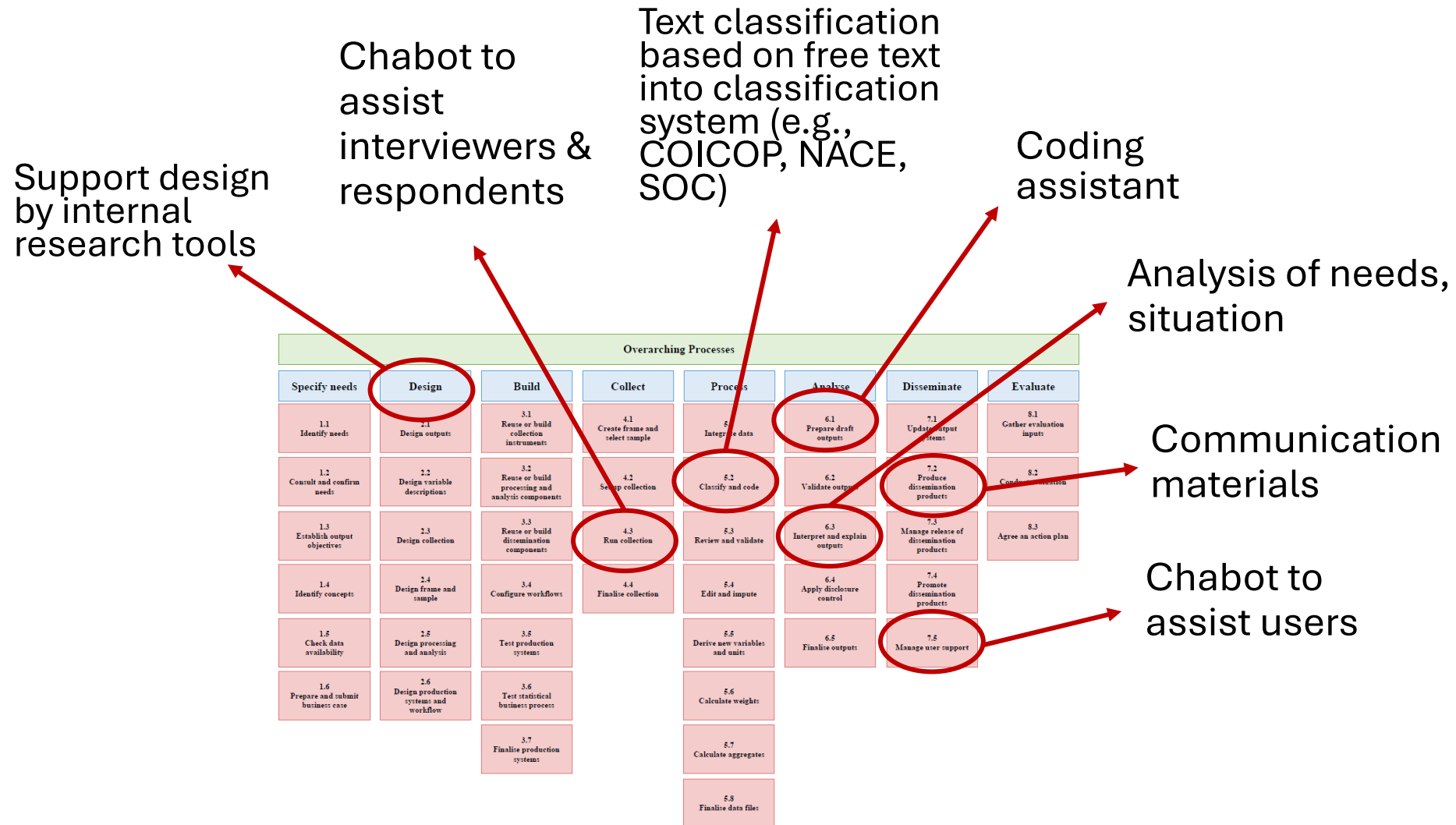
Level of impact generative AI has or is expected to have on the following work areas in the works of statistical organizations next 2-3 years ([CES Survey on the use of generative AI in statistical organizations \(July 2024\)](#))

	Highly impactful	Moderately impactful	Slightly impactful	Not impactful at all	Not sure	Average score
Data collection and processing	6	17	15	1	2	2,72
Data analysis	8	17	13	3	0	2,73
Dissemination and communication	13	16	9	2	0	3,0
Coding and IT development	21	15	4	0	1	3,43
Other administrative tasks	8	14	12	3	4	2,73

\* Average score is calculated using following scale: highly impactful – 4, moderately impactful – 3, slightly impactful – 2 and not impactful at all – 1 ("not sure" is excluded).

# Examples to be presented in Session 3 & 4

## Where?



# How?

## Level of concern regarding risks associated with use generative AI in your organizations (CES Survey on the use of generative AI in statistical organizations (July 2024))

	Very concerned	Moderately concerned	Slightly concerned	Not concerned at all	Average score
Ethical concerns	19	9	12	1	3,12
Accuracy and lack of validation mechanism	25	13	3	0	3,54
Confidentiality, privacy and security	27	12	2	0	3,61
Copyright and legal issues	16	13	8	4	3,0
Negative public perception	10	17	12	2	2,85
Dependency to external providers	12	13	13	3	2,83
Overuse by staff	3	10	20	8	2,20
Misuse (e.g., neglecting human oversight)	14	21	6	0	3,20
Lack of reproducibility	9	22	8	2	2,93

\* Average score is calculated using following scale: very concerned – 4, moderately concerned – 3, slightly concerned – 2 and not concerned at all – 1

# How?

Level of concern regarding risks associated with use generative AI in your organizations [\(CES Survey on the use of generative AI in statistical organizations \(July 2024\)\)](#)

	Very concerned	Moderately concerned	Slightly concerned	Not concerned at all	Average score
Ethical concerns	19	9	12	1	3,12
Accuracy and lack of validation mechanism	25	13	3	0	3,54
Confidentiality, privacy and security	27	12	2	0	3,61

- ... compounded by being **national statistics organisations**

Higher stake for error

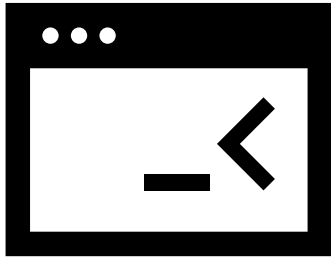
How to ensure quality?

In-house or not?  
Change management



# How?

## Technical guardrails



- Self-hosting with open-source models
- RAG, multi-agent

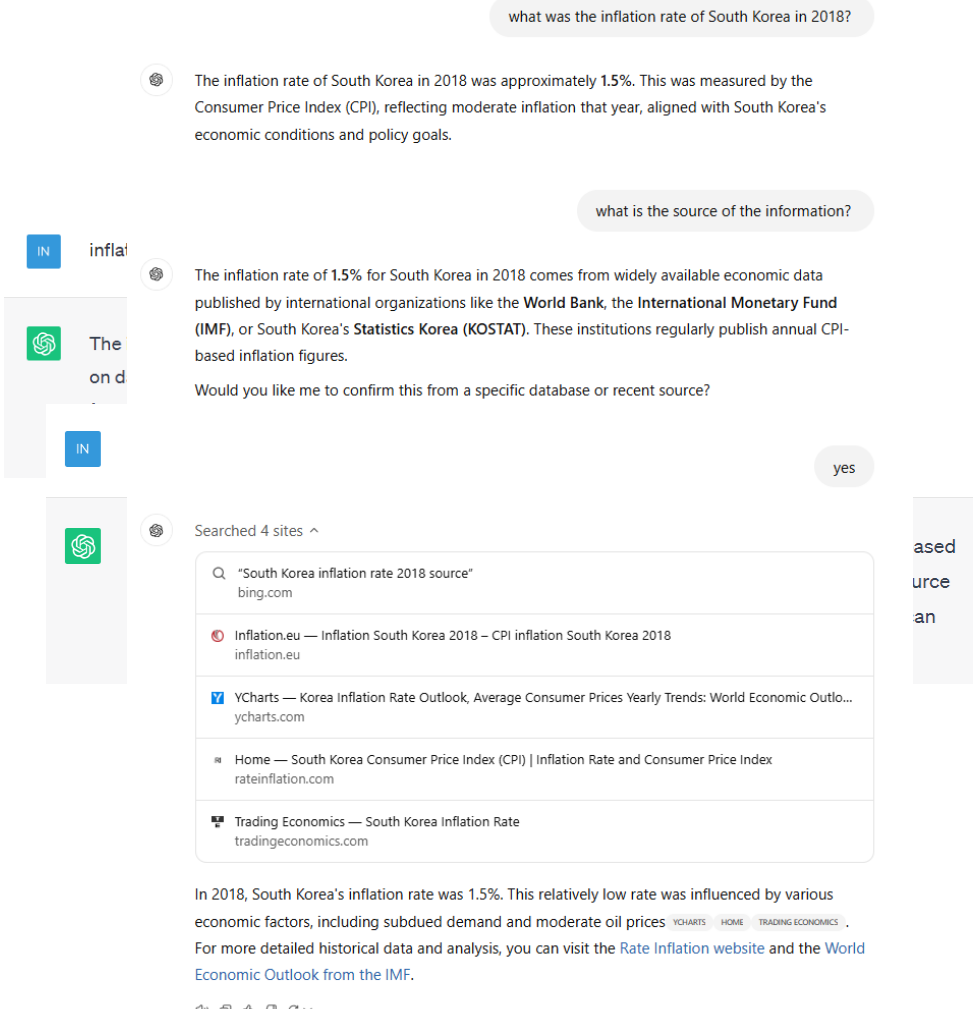
## Governance measures



- Human in the loop
- Governance framework & guidelines

# Open questions

- Fundamental role of NSOs to provide quality statistics for decision making & public
- What if decision makers and public don't get their information from NSOs?
- ...but this already happened (e.g., google search, SNS)
- Gen AI brings in new challenges
- How to address?



what was the inflation rate of South Korea in 2018?

The inflation rate of South Korea in 2018 was approximately 1.5%. This was measured by the Consumer Price Index (CPI), reflecting moderate inflation that year, aligned with South Korea's economic conditions and policy goals.

what is the source of the information?

The inflation rate of 1.5% for South Korea in 2018 comes from widely available economic data published by international organizations like the **World Bank**, the **International Monetary Fund (IMF)**, or South Korea's **Statistics Korea (KOSTAT)**. These institutions regularly publish annual CPI-based inflation figures.

Would you like me to confirm this from a specific database or recent source?

yes

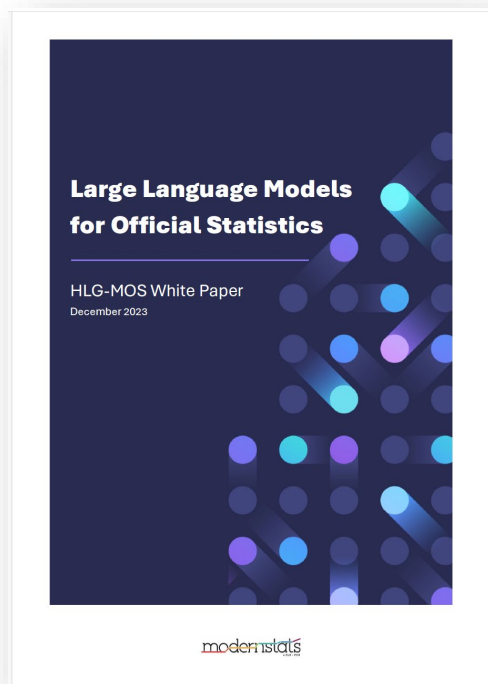
Searched 4 sites ^

Q	"South Korea inflation rate 2018 source"
bing.com	
inflation.eu	Inflation.eu — Inflation South Korea 2018 – CPI inflation South Korea 2018
ycharts.com	YCharts — Korea Inflation Rate Outlook, Average Consumer Prices Yearly Trends: World Economic Outlo...
rateinflation.com	Home — South Korea Consumer Price Index (CPI)   Inflation Rate and Consumer Price Index
tradingeconomics.com	Trading Economics — South Korea Inflation Rate

In 2018, South Korea's inflation rate was 1.5%. This relatively low rate was influenced by various economic factors, including subdued demand and moderate oil prices. For more detailed historical data and analysis, you can visit the [Rate Inflation website](#) and the [World Economic Outlook from the IMF](#).

# Thank you

# Resources on Gen AI for official statistics



[LLMs for Official Statistics \(2023\)](#)

**Generative AI use cases in official statistics**

This page contains generative AI use cases in national and international statistical organizations presented at various relevant meetings. If you have any use case / research you wish to add to this list, please submit it via [pull request to this repo](#) or [this form](#).

**List of use cases**

(See data file in JS)

Use case type	Title	Research type	Organization	Date	AI models used	Approach	Tools	Code availability
1	Code and IT development	Paper	CSO Ireland	2023-12	GPT-3.5			
2	Code and IT development	Presentation	France Insee	2023-11	GPT-3.5			
3	Text generation for analysis	Paper	Statistics Canada	2023-11	GPT-3.5			
4	Text generation for analysis	Presentation	Statistics Norway	2024-05	GPT-4	Prompt engineering	Chat GPT	N/A
5	Text generation for data processing (APIs)	Presentation	Bank of International Settlements	2024-03	GPT-3.5	Prompt engineering, RAG	OpenAI AI Assistant	N/A
6	Text generation for data processing (APIs)	Presentation	Bank of International Settlements	2024-03	GPT-3.5	Prompt engineering, RAG	OpenAI AI Assistant	N/A
7	Communication - technical materials	Presentation	Indonesia BIS	2024-03	GPT-3.5	Prompt engineering		N/A
8	Communication - images in videos (avatars)	Presentation	ESKP/STAT	2024-04		Prompt engineering	CHARLS, ChatGPT, video tools, image creation, Midjourney	N/A
9	Data dissemination and information search	Paper	UNCTAD	2024-05	Midjourney	RAG	ChatGPT, BingGPT, Perplexity, Sonnet	Yes



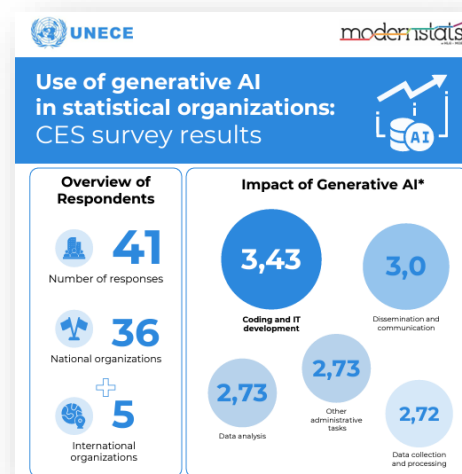
[Github repo on generative AI use cases \(2024\)](#)

Coming up



HLG-MOS Generative AI for Official Statistics Project (2024-2025)

Workshop on Generative AI (May 2025)



[CES survey on the use of gen AI \(2024\)](#)