



PARiS  
21



# Unlocking AI

## AI and Official Statistics in Rwanda

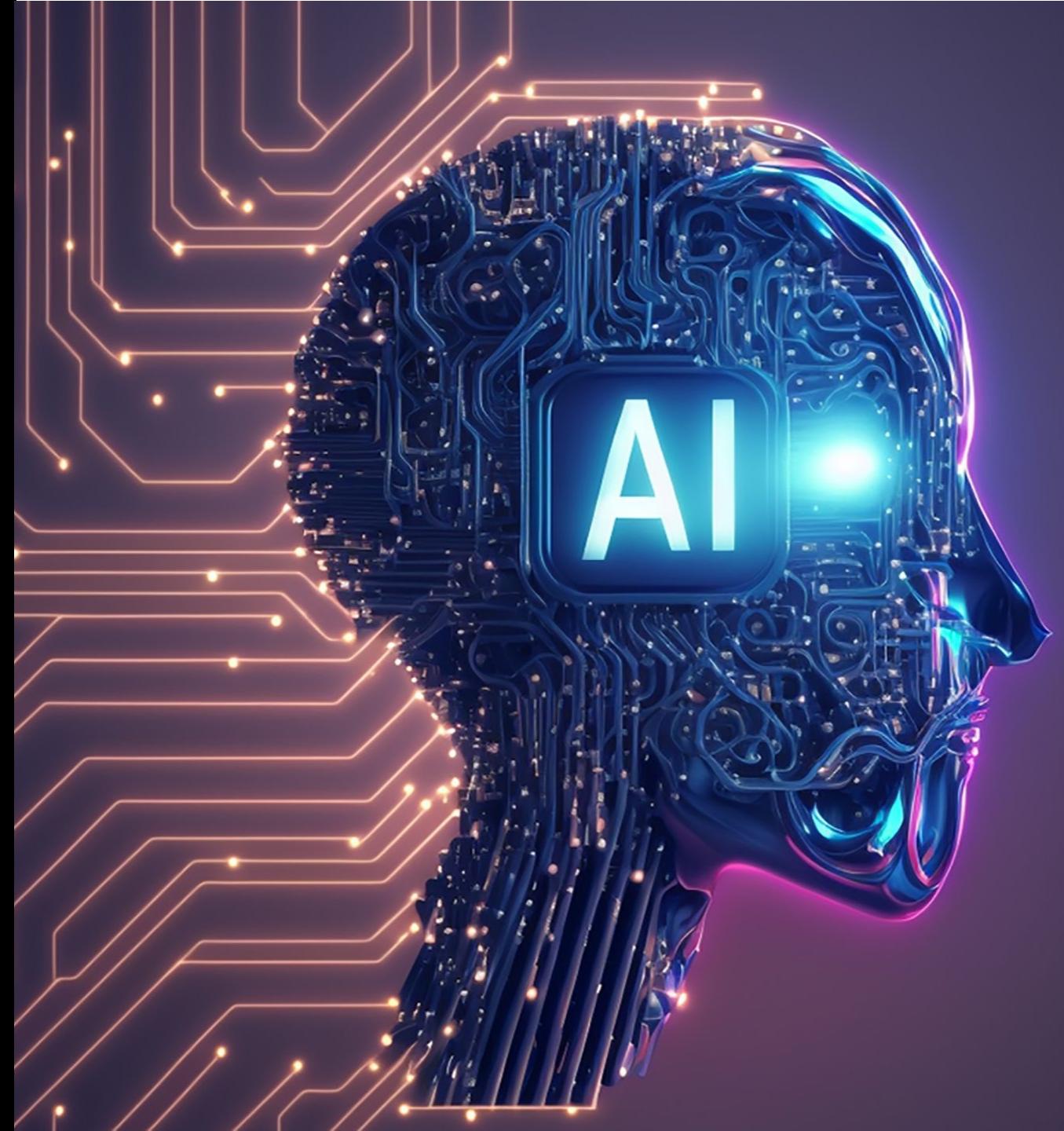
PARIS21 and SJTU

Aug 5-8, 2025

Kigali, Rwanda

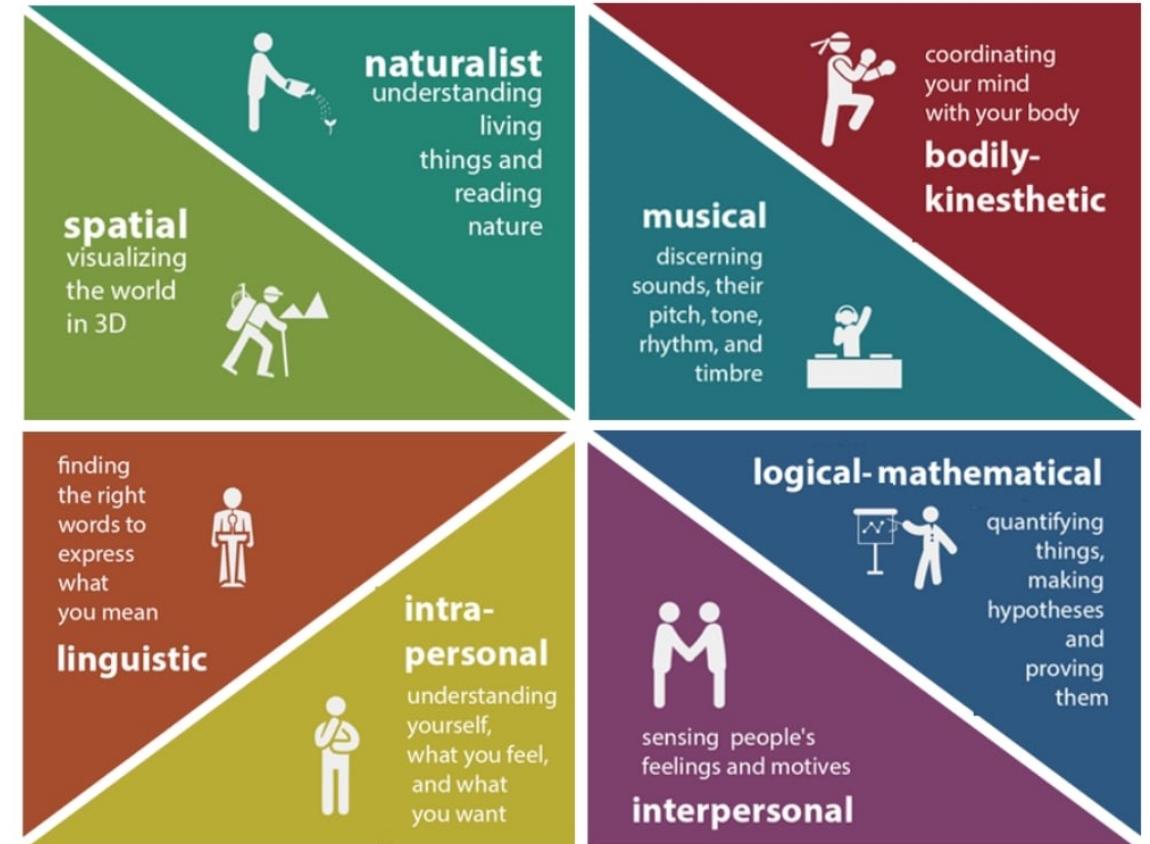
# What is AI?

- Definitions
- Foundations



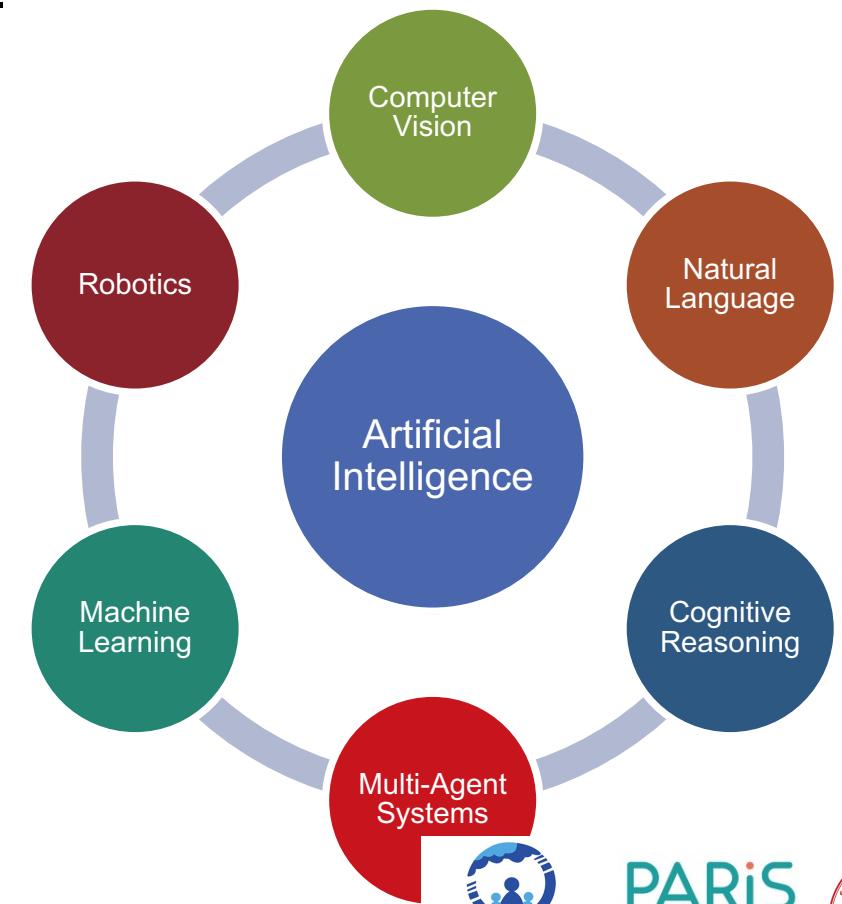
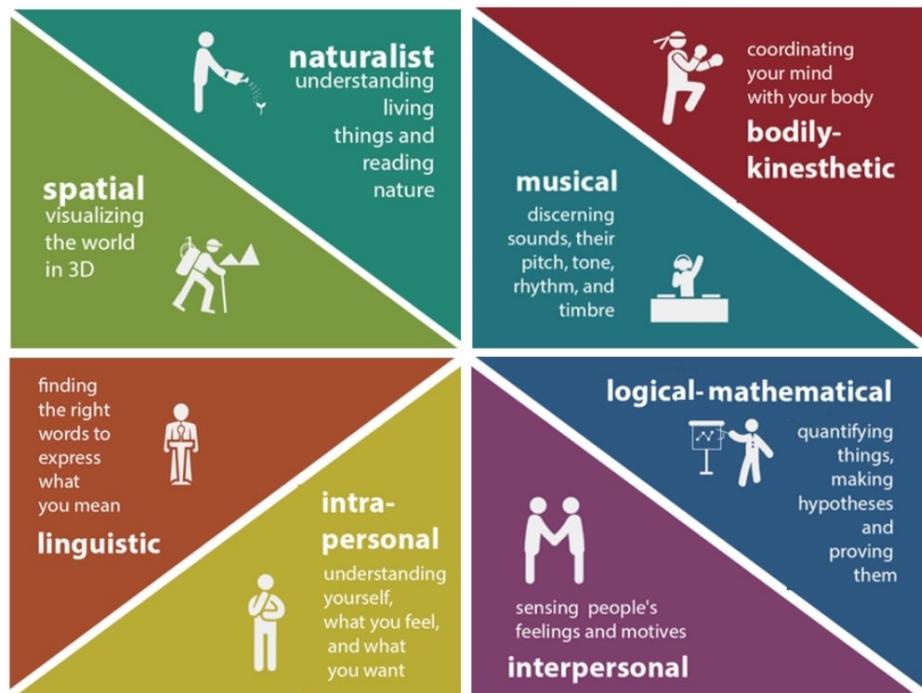
# Starting with Human Intelligence

- Human intelligence
- *Homo sapiens* -- “Man the wise”
- Gardner's Theory: Multiple Intelligences



# Starting with Human Intelligence

- AI can be seen as the scientific endeavor to **simulate**, **extend**, and **expand** some of these human intellectual capabilities using machines.

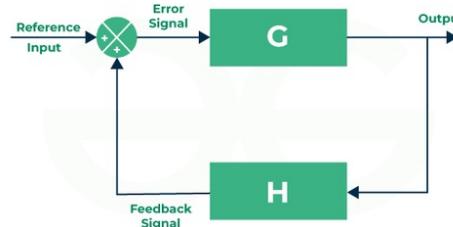
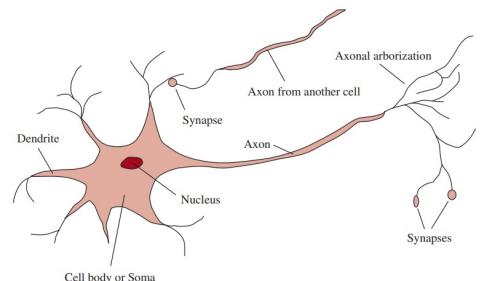
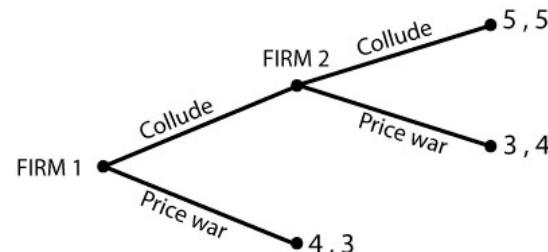


# The Foundations of AI

AI is an interdisciplinary field that integrates

- Philosophy (since ~400 BCE)
  - Mathematics (since ~800 CE)
  - Economics (since ~1776)
  - Neuroscience (since ~1861)
  - Psychology (since ~1879)
  - Linguistics (since ~1957)
  - Computer Engineering (since ~1940)
  - Control Theory (since ~1948)

to design systems capable of performing tasks that typically require human intelligence.



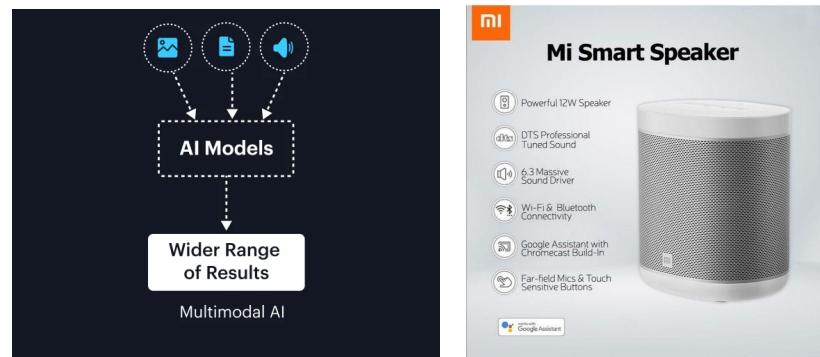
# Different Types of AI

- NLP & How it works
- Computer Vision
- Multimodality
- Cognitive Reasoning
- AI + X



# Types of AI

- Computer Vision and Automated Driving
- Multimodality: **text, images, audio, video**
- Cognitive Reasoning & Strategy
- Large Language Models



PARIS  
21!



# Natural Language Processing (NLP)



- Natural Language Processing, or NLP, is the capability of a computer program to understand, interpret, and generate human language.
- The **core** technology behind **the generative AI tools**
- Perhaps, the most directly applicable technology for your daily work

## HOW does NLP work?

# Natural Language Processing (NLP): How it works

How does a machine “understand” and “write” language?

- The Core Principle: A Super-Powered **“Next-Word Predictor”**
- An incredibly powerful system for predicting the next most likely word in a sentence.
- TRAINING: It has been **trained** on a vast library of text and books from the internet.
  - The training set is of utmost importance!



# Natural Language Processing (NLP): How it works

- Plausible, Not Necessarily Factual

- AI's goal is to create a statistically probable sequence of words >> plausible
- It is designed to *sound* correct and fluent.

**However, the AI has no internal database of facts!**

**Hallucinate:** AI can invent facts, figures, and sources that are completely false

# AI for Official Statistics

- Use cases in other NSOs
  - Risks & benefits
- Additional considerations



# AI is a transformative tool reshaping the production and use of statistics

It makes data exploration more intuitive and accessible, even for non-experts!



Allows us to draw insights and valuable data and information in seconds from new types of sources (images, text, paper records, etc.)



Enables the extraction of new information from existing sources (forecasting, nowcasting) and improve their quality and granularity



Facilitates visualising complex datasets and generates tailored reports, turning raw data into actionable insights

# How other NSOs are using AI

Country	Challenge	Solution	Lessons learned
Mongolia	Data collection for Mongolia's agricultural livestock census is costly and time-intensive.	Tested deep learning with satellite images to detect animals and yurts, supporting livestock tracking and population mapping.	AI can work well, but high-resolution images are costly and critical; strong partnerships are needed to access quality data and scale solutions.
United Kingdom	It's hard to know how official data are used in decision making and political debates.	The ONS built ParliAI, an AI tool that scans UK parliamentary records to track mentions of ONS data using large language models.	AI can reveal how data is used in public discussions; open-sourcing tools like ParliAI helps others adapt and apply similar solutions.
Canada	Manually coding millions of write-in responses in the census is costly and time-consuming.	Statistics Canada used FastText, an NLP algorithm, to automatically code about 7 million responses across 31 questions.	AI can greatly reduce the need for manual coding in large surveys, saving time and resources while maintaining accuracy.
Colombia	Poverty data was not detailed enough to support local policy decisions.	The NSO used AI and daytime and nighttime satellite images to estimate poverty across the entire country, for a 70X increase in data points.	AI models can greatly increase data granularity, helping governments target policies more effectively across regions.
Ireland	Transitioning from SAS to R can be slow and complex for users unfamiliar with both languages.	The CSO created a generative AI tool that automatically translates SAS code into R and helps users understand the original code.	Generative AI can support modernisation by speeding up code migration and helping staff learn new tools more easily.

# Generative AI Risks for NSOs

While AI offers benefits, it also presents risks that must be addressed.



**Opacity**

**Complexity**

**Bias**

**Autonomous behavior**

It is essential to adopt a balanced approach and take action to ensure transparency and accountability of AI systems. This involves developing legislative measures that protect citizens' rights and promote trust in technology.

## **Recommendations for AI developers:**

- Develop own ethical policies for generative AI.
- Ensure training data is representative and unbiased.
- Consider the confidentiality of user information and personal data.

## **Recommendations for AI users:**

- Do not input confidential, sensitive, or proprietary information.

# Input Risks: The Foundation of AI's Actions

## Data Privacy & Security

### ■ *Data Privacy Policies of Major AI Models (latest versions, by 2025.8)*

– Will my data be collected? **YES.**

– Is my data safe with them? **Maybe not.**

#### 1. Personal Data we collect

**ChatGPT**

We collect personal data relating to you (“Personal Data”) as follows:

**Personal Data You Provide:** We collect Personal Data if you create an account to use our Services or communicate with us as follows:

- **Account Information:** When you create an account with us, we will collect information associated with your account, including your name, contact information, account credentials, date of birth, payment information, and transaction history, (collectively, “Account Information”).
- **User Content:** We collect Personal Data that you provide in the input to our Services (“Content”), including your prompts and other content you upload, such as files, images, and audio, depending on the features you use.
- **Communication Information:** If you communicate with us, such as via email or our pages on social media sites, we may collect Personal Data like your name, contact information, and the contents of the messages you send (“Communication Information”).
- **Other Information You Provide:** We collect other information that you may provide to us, such as when you participate in our events or surveys or provide us with information to establish your identity or age (collectively, “Other Information You Provide”).

**ChatGPT**

#### 2. How we use Personal Data

We may use Personal Data for the following purposes:

- To provide, analyze, and maintain our Services, for example to respond to your questions for ChatGPT;
- To improve and develop our Services and conduct research, for example to develop new product features;
- To communicate with you, including to send you information about our Services and events, for example about changes or improvements to the Services;
- To prevent fraud, illegal activity, or misuses of our Services, and to protect the security of our systems and Services;
- To comply with legal obligations and to protect the rights, privacy, safety, or property of our users, OpenAI, or third parties.

#### 3. Disclosure of Personal Data

We may disclose your Personal Data in the following circumstances:



# Mitigating the risks of (generative) AI

Ethical  
Frameworks

Data  
readiness

Transparency

Governance

Collaboration

Capacity  
Building



# Effectively using generative AI in the NSO

- Free vs paid tools
  - If the tool is free, any inputs you are giving it are generally being used to train the model
- Data privacy
  - All major genAI platforms are collecting your data, and there is no guarantee it will be kept safe
- Prioritising where AI can add the most value
  - It should be seen as a means to an end, not the end itself
- Effective prompting
  - CLEAR technique (Context, Limitations, Examples, Adaptation, Feedback)

# Additional resources

- AI through the lens of official statistics and the SDGs: What are the benefits and risks? (2024) by Dilek Fraisl, Linda See, François Fonteneau and Johannes Jütting ([source](#))
- OECD AI Principles (updated 2024) ([source](#))
- UNECE Generative AI for Official Statistics (2025) ([source](#))
- PARIS21 Catalogue of Generative AI Use Cases (2025) ([source](#))
- Data science and official statistics: Toward a new data culture (2021) by Stefan Schweinfest and Ronald Jansen ([source](#))
- Collecting, generating, and analyzing national statistics with AI: What benefits and costs? (2023) by Maria Rim and Youngsun Kwon ([source](#))
- The impact of Artificial Intelligence on Official Statistics (2023) ([source](#))
- The use of data science in a national statistical office (2022) ([source](#))

# Recap of this week's workshop on genAI

- Discussions about AI for official statistics, AI risks and ethics, prompt engineering, and more
- Robust testing of the Gender Data chatbot
- Evaluation mechanism developed and applied to the Gender Data chatbot
  - Can be applied to future chatbots as well
- Checklist of risks and suggested mitigation strategies, for the chatbot
- Guide for effective prompting

