



Large Language Models for Official Statistics

Contributors
(20 people)



Statistics
Canada

Statistique
Canada

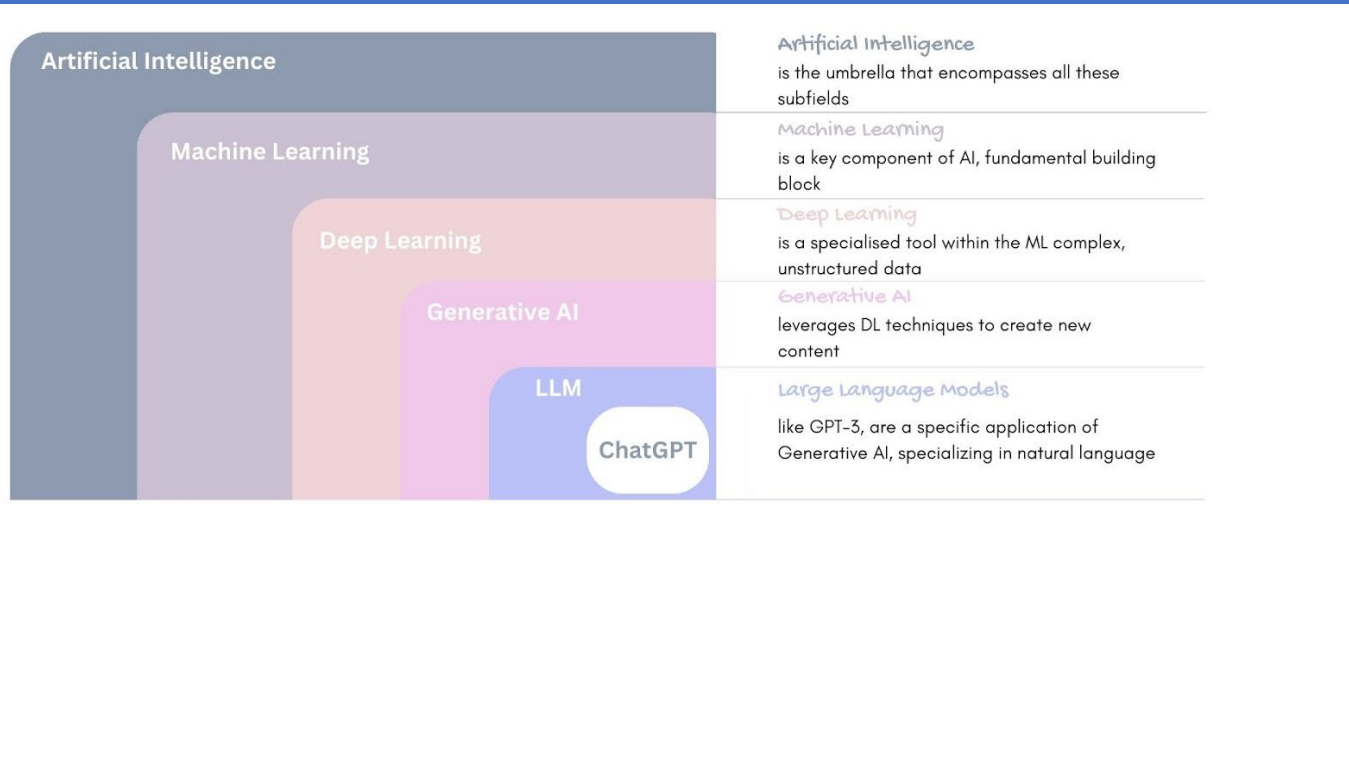


Overall structure of document

1. Introduction on LLM
2. Implications and opportunities for Official Statistics
3. Use cases for Statistical Organisations
4. Risks and mitigation measures
5. Considerations as Statistical Organisations move forward with LLMs

Section 1 - Introduction Section

Introduction Section



- What are large language models?
- How do they work?
- Foundation Models
- Fine Tuning Models
- Role of Open Source

Section 2 - Implications and Opportunities for Official Statistics

Section 2 - Implications and Opportunities for Official Statistics

Workplace Tasks -

- Communications
- Idea Generation
- Project Management & Planning
- Translation of (spoken) languages
- Presentations
- Educational Outcomes
- Image Generation

Statistical Production -

- Survey Design
- Data classification & coding
- Validation & editing data
- Coding Helper/Coding translator
- Updating statistical standards
- Synthetic data generation
- Customised delivery and searching of statistics

Section 3 – Use Cases For Statistical Organisations

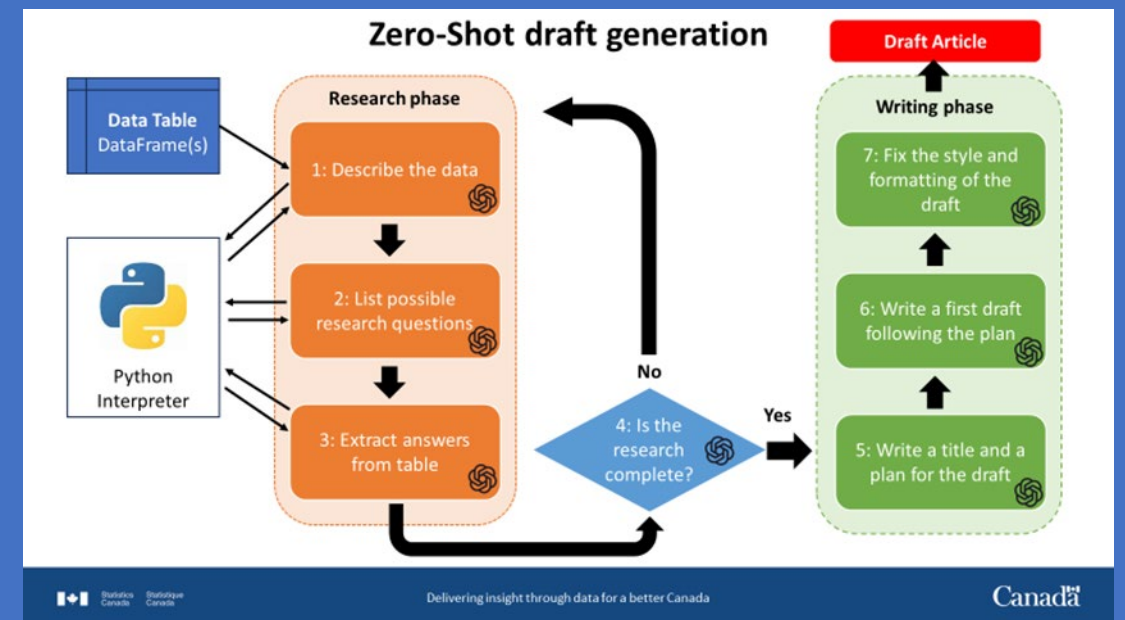
Section 3 – Use Cases For Statistical Organisations

ABS - Updating Statistical Classification Definitions

CSO Ireland - Code Translation and Explanation (SAS to R) Using LLMs

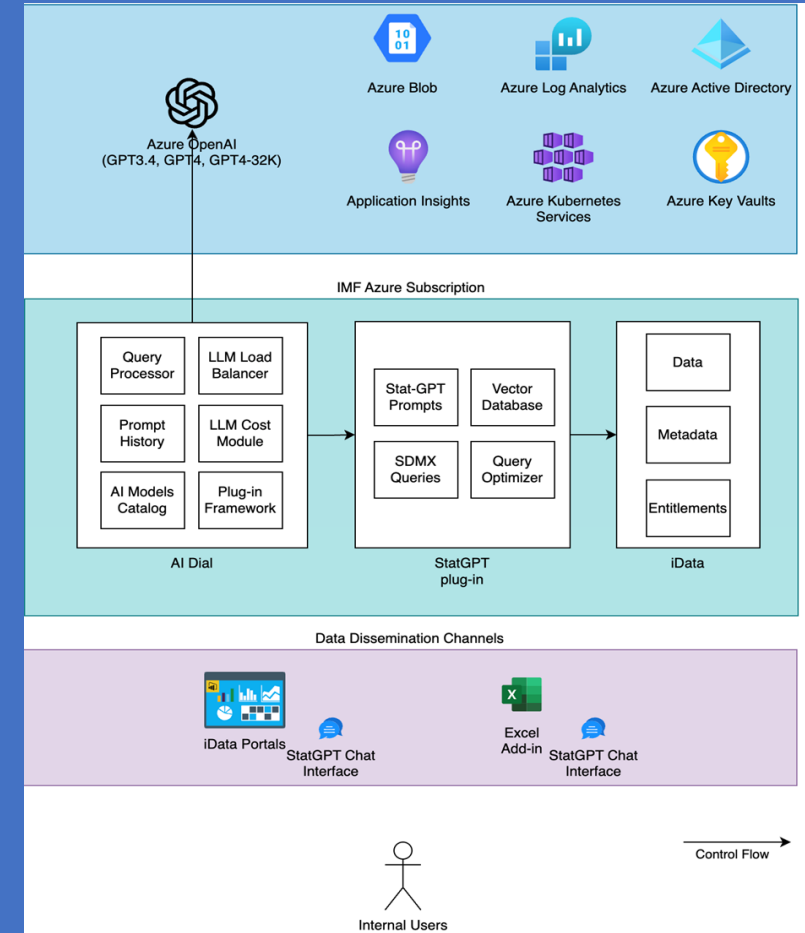
IMF – StatGPT

Statistics Canada - Report Generation Using LLMs



Section 3 – Use Cases For Statistical Organisations -Coverage

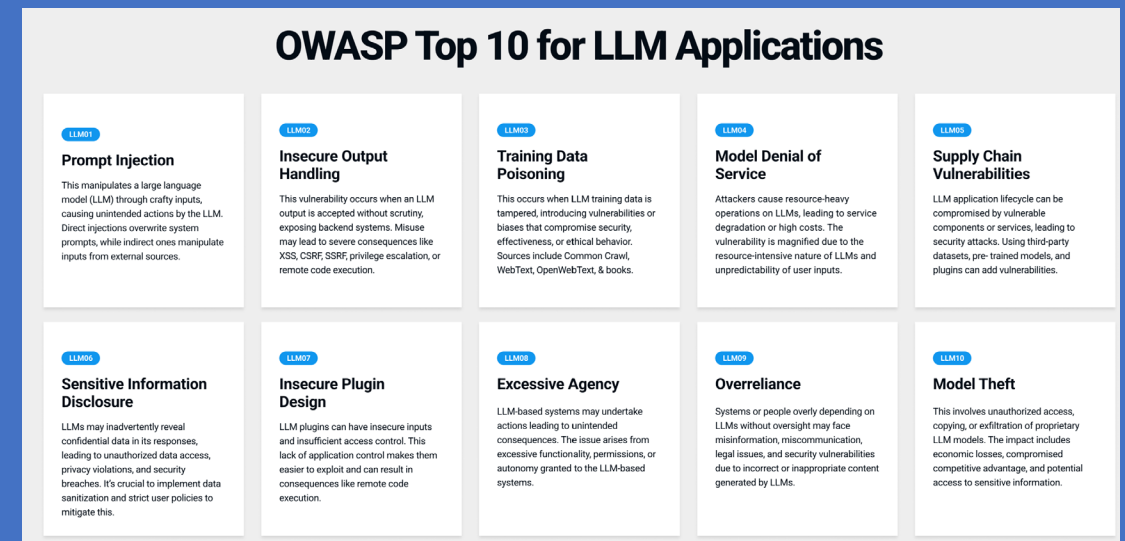
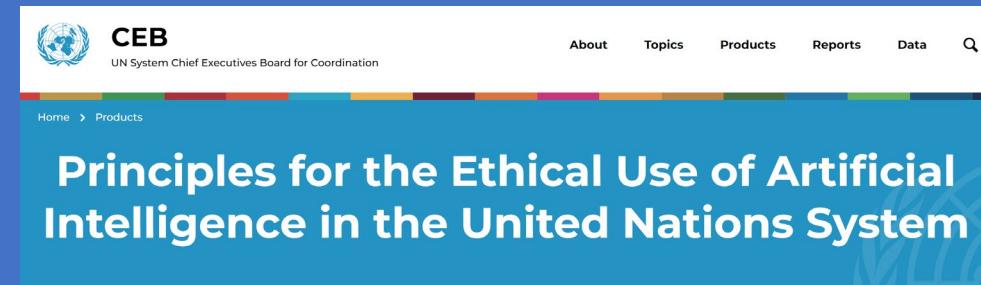
- Business Problem and Value Addition
- Preliminary Results
- IT and HR Requirements
- Quality Metrics
- Stakeholder Response
- Barriers
- Risks and Mitigations
- Next Steps & Other Potential Use Cases



Section 4 – Risks And Mitigation Measures

Section 4 – Risks And Mitigation Measures

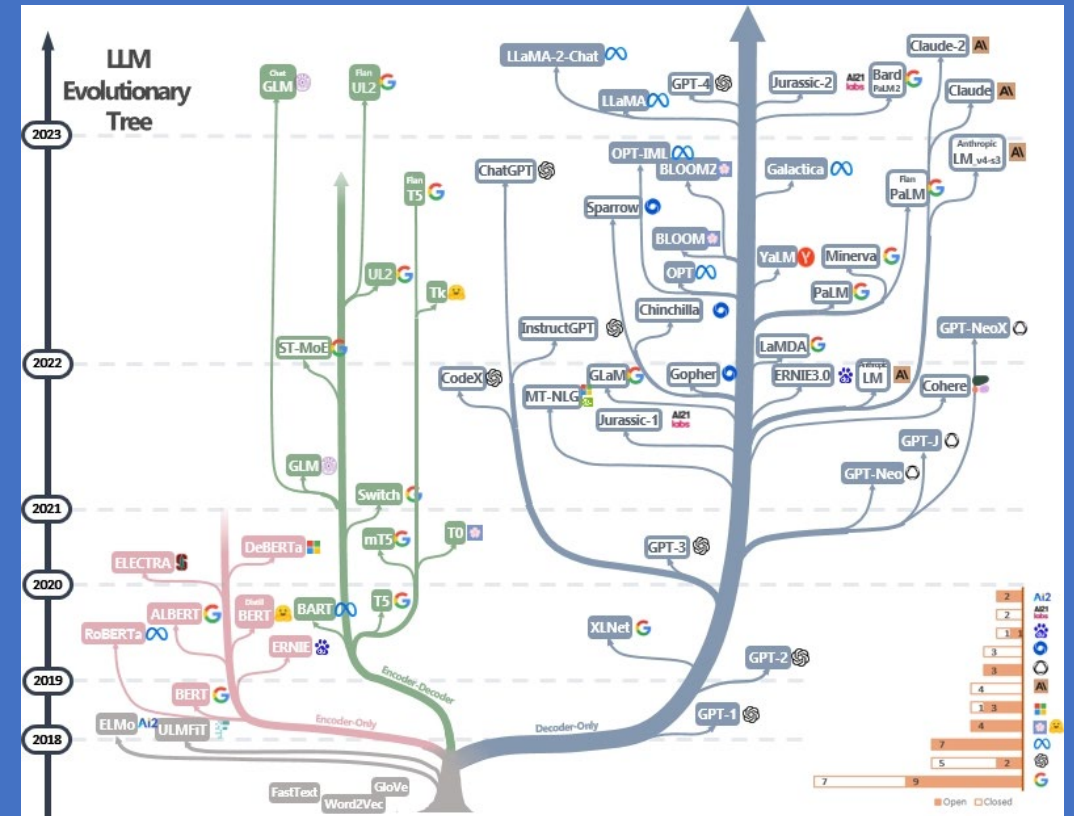
- Ethics - Application of UN AI Principles to LLMs
- Accuracy and Lack Of Validation
- Privacy & Security Concerns
- Copyright & Legal Issues
- Lack Of Literacy & Understanding



Section 5 - Considerations as Statistical Organisations Move forward with LLMs

Considerations

- Governance
- Engagement with tech companies
- Considerations around open access
- Communications with the public
- Future Works & Suggestions



Closing Considerations

Closing Considerations



Still 'Phase 1' of LLMs

Analogy of mobile phones.

Early days of mobile phones -

- Considered for conversation only
- Move from 'big' models to small models with same/better performance
- Consolidation (e.g. Apple/Android)
- Underestimation of future implications
- For statistical organisations, decisions likely to move from 'how can we use' to 'how much impact' will it have for us....