

Preserving and Accessing: Secure AI Solutions for Sensitive Information Systems

Presented by

Georgina Cosma

Professor of AI & Data Science
Loughborough University

About the Presenter

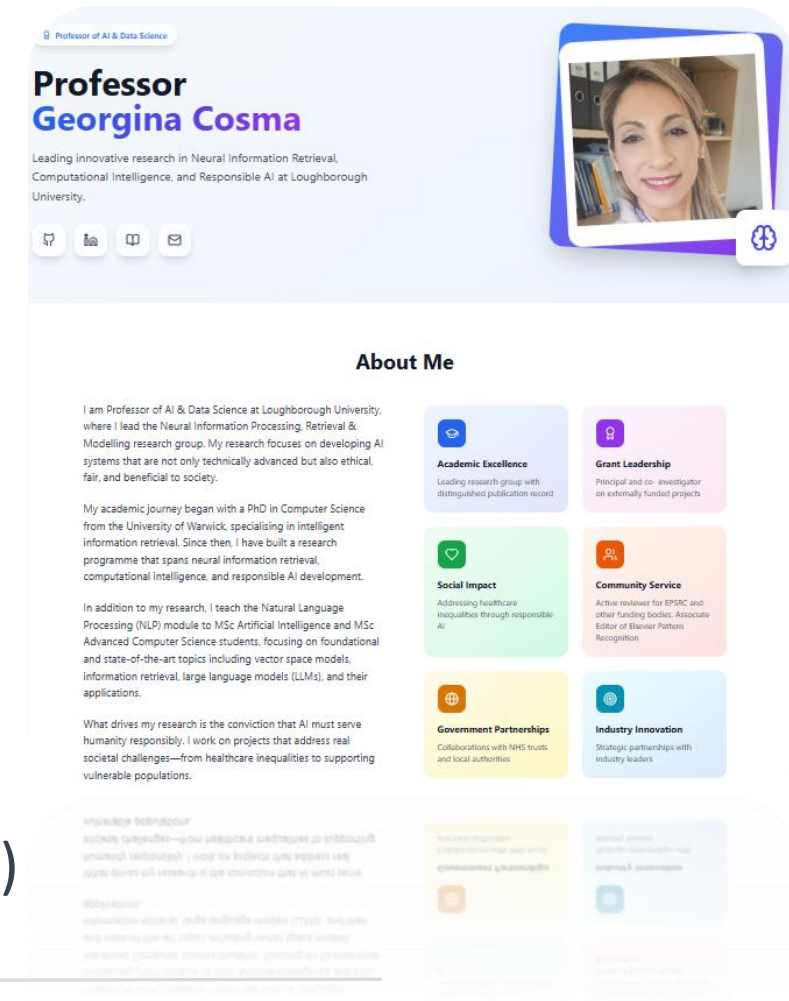
Academic Position

- Professor of AI & Data Science, Loughborough University
- PhD in Computer Science, University of Warwick (2008)

Research Focus

- Natural Language Processing
- Information retrieval (text, multi-modal)
- Information processing and modelling
- Continual learning and unlearning
- Responsible AI innovation (bias management and mitigation)

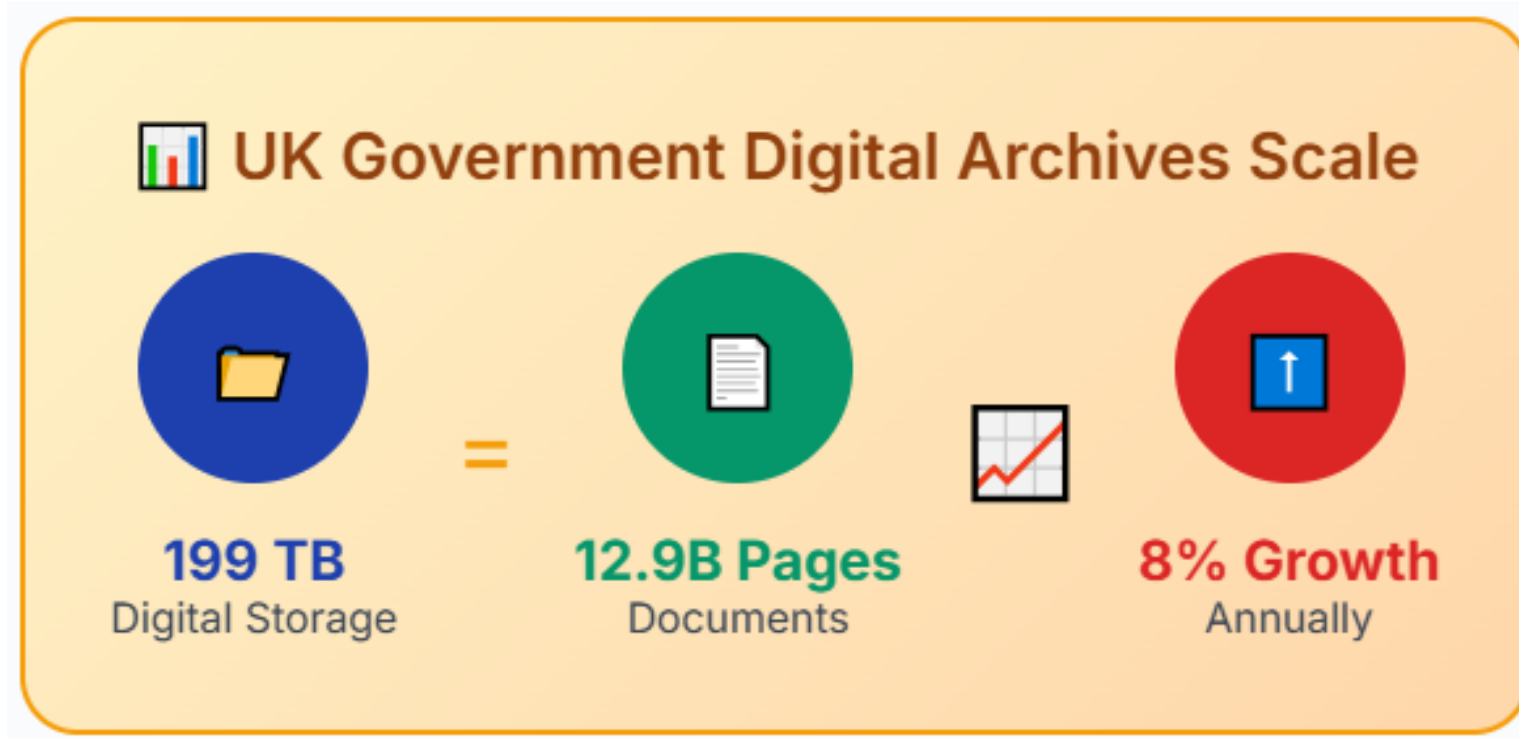
Contact: Georgina Cosma
g.cosma@lboro.ac.uk
[@gcosma1](https://www.gcosma1.com)
www.cosmos-ai.com



Contents

- Information retrieval foundations and limitations
- Retrieval Augmented Generation (RAG) systems
- Advanced agentic AI for multi-step reasoning
- Custom training approaches for sensitive data

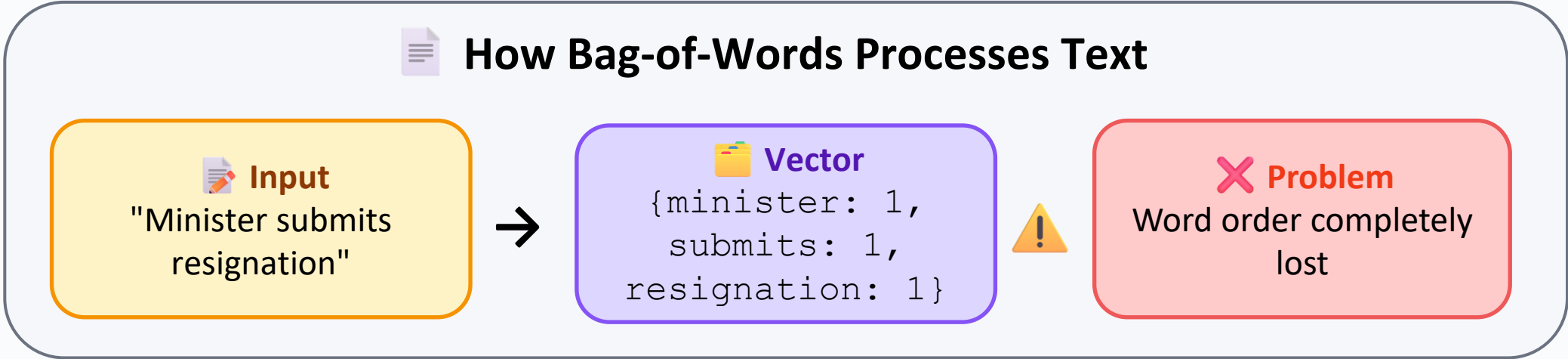
Digital Archives Scale



When reviewers read Cabinet Office records to judge their potential historic value they are looking at:

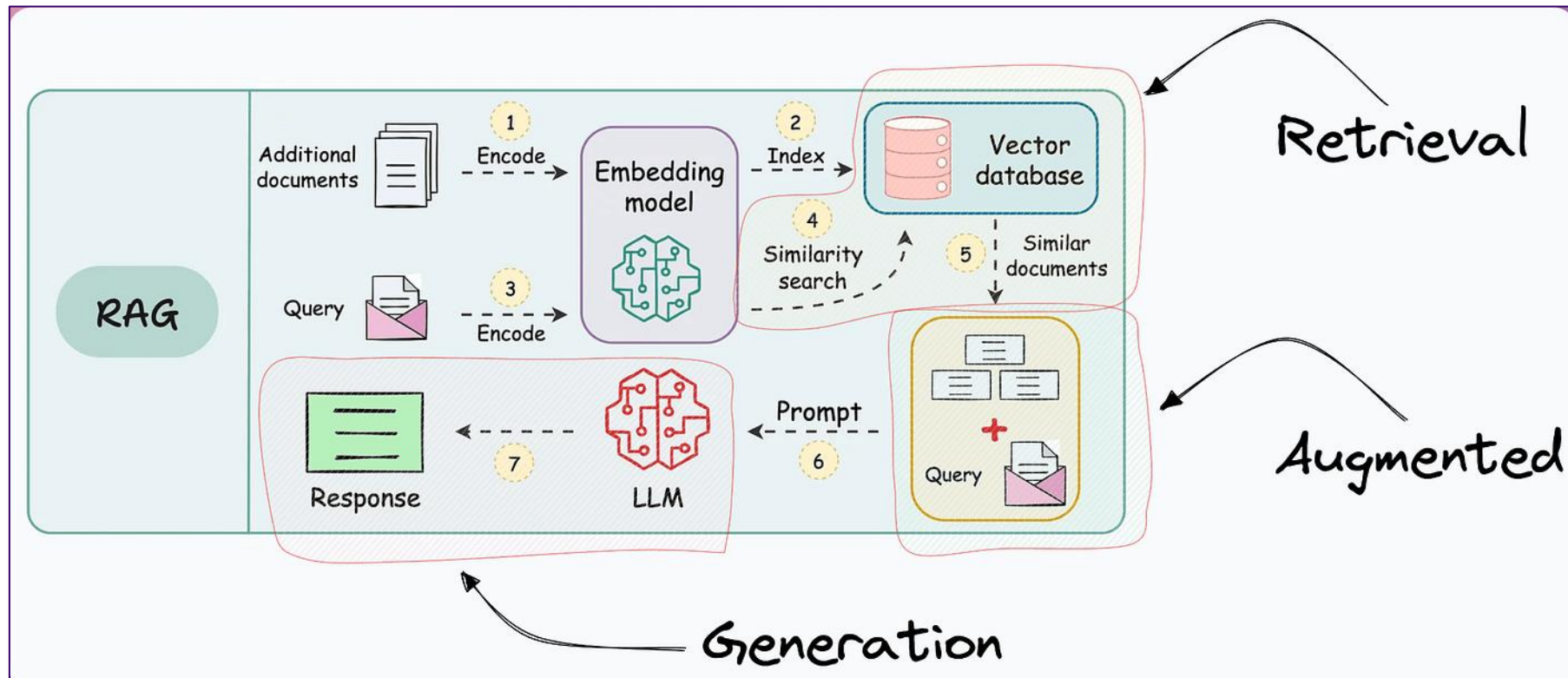
- What is the information about, is it for example international diplomacy or some internal administrative issue?
- Who is involved, who is mentioned in the text, who created it and who was it received by?
- Format, as this provides clues as to why the record was created and its function.

Bag-of-Words: The Foundation



Solving the Scale Problem: What is RAG?

Retrieval Augmented Generation (RAG): AI technique that combines **information retrieval** with **text generation**.



Secure government applications with purpose-built RAG architectures using only verified internal data sources and with no external connectivity.

Solving the Scale Problem with RAG systems

Example:

Input: "Cabinet infrastructure spending Q2 2024"

RAG Processing Pipeline:

1. Semantic vector search across document corpus
2. Cluster related policy documents by spending theme
3. Extract key financial decisions and commitments
4. Link ministers, departments, implementation timelines
5. Generate executive summary with source attribution

Sample Output:

"Based on analysis of [X] relevant documents, three major infrastructure investments were identified totalling £[Y]B across transport, digital, and energy sectors..."

Features: Full citations • Confidence scoring • Processing time: sub-3 seconds



Semantic Search & Clustering



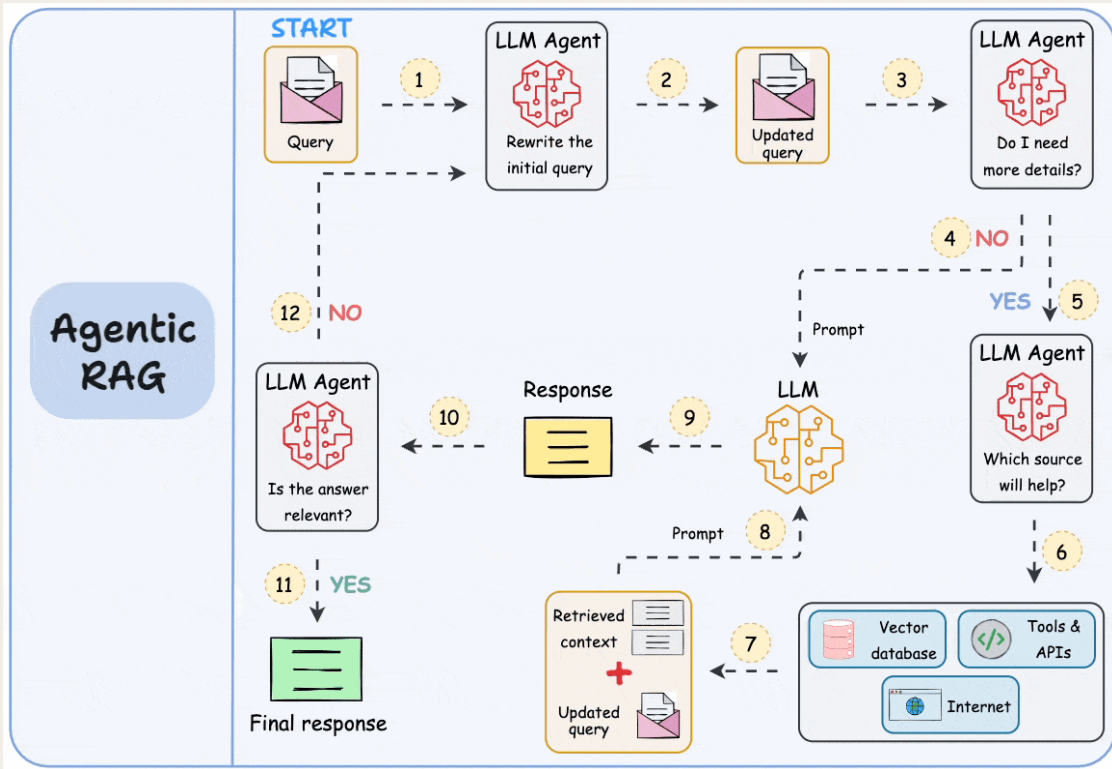
Real-time Source Attribution



Adaptive Output Formats

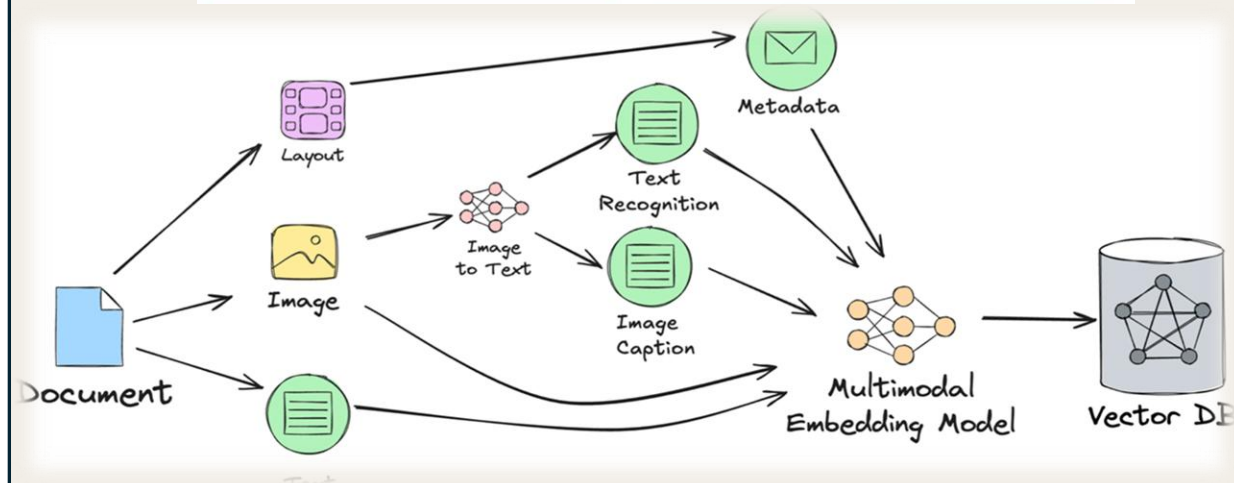
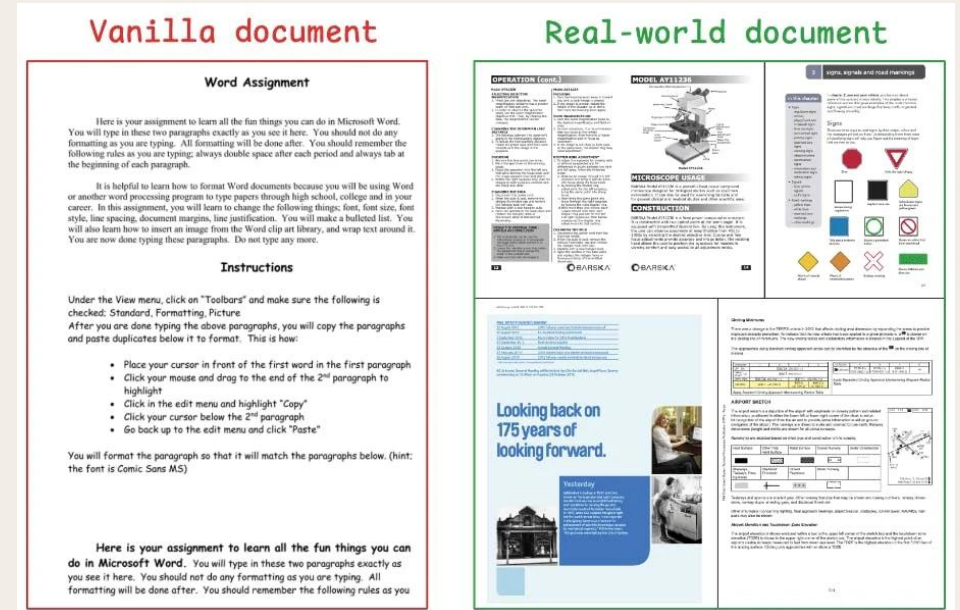
Agentic RAG systems

Agentic RAG



"Agents function as intelligent systems capable of independent reasoning—analysing requirements, developing strategies, and refining approaches until they deliver optimal outcomes, rather than merely executing predetermined workflows. LLMs provide the cognitive foundation for this autonomous decision-making."

Multi-modal RAG



Standard and Agentic RAG Capabilities

Standard RAG Capabilities

Information retrieval & search

Question answering

Summarisation

Concept annotation

Document classification

Information extraction

Content analysis

Report generation

Document comparison

Entity recognition

Sentiment analysis

Citation & attribution

Agentic RAG Capabilities

Trend analysis

Compliance checking

Multi-step reasoning

Strategic planning

Dynamic search refinement

Multi-document synthesis/Cross-document
Relationship mapping

Multi-modal integration

Iterative investigation

Complex workflow execution

Adaptive response generation

Verification & validation

Conditional/rule-based reasoning

RAG and Hallucination

Secure RAG Minimises Hallucination

- Trained from scratch on internal documents only
- Responses grounded in verified internal documents
- No external data contamination
- Source attribution enables verification

Remaining Risks ⚠️

- May misinterpret retrieved information
- Could inappropriately combine information from different retrieved documents

Government Safeguards

- **Source Attribution** - Direct links to original documents
- **Confidence Scores** - Response reliability indicators
- **Human Verification** - Manual verification of AI outputs

Experience in Secure AI Applications



Leicestershire County Council

- Ethical explainable AI for early intervention systems
- Transparent decision-making with full audit trails



Healthcare Services Safety Investigation Branch

- LLM concept annotation for identifying factors leading to adverse maternity outcomes
- Secure processing of sensitive healthcare data
- RAG, Concept annotation, Summarisation, Clustering, Thematic Analysis



SVGC Knowledge Transfer Partnership

- Document deduplication algorithms for government digital systems
- Document clustering and draft/version identification.



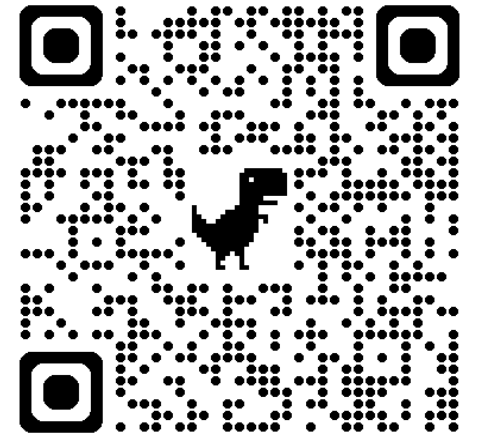
Opportunities for collaboration?

Let's Discuss Your AI Requirements

Contact:

Professor Georgina Cosma
Department of Computer Science
School of Science
Loughborough University

Email: g.cosma@lboro.ac.uk
X: @gcosma1
Website: cosmos-ai.com



Preserving and Accessing: Secure AI Solutions for Sensitive Information Systems

Abstract:

This presentation explores secure artificial intelligence (AI) techniques for government and sensitive information systems. Traditional AI approaches rely on external APIs and unverified training data, creating security vulnerabilities. Key techniques include purpose-built Retrieval Augmented Generation (RAG) systems using only curated, internal data sources with zero external dependencies. These secure techniques address the considerable challenge of processing vast government archives through semantic search capabilities that understand context beyond simple keyword matching. The presentation covers advanced approaches including agentic RAG for autonomous decision-making and multi-modal processing techniques. We also examine the complex challenge of "unlearning" - removing specific information from AI systems. These techniques demonstrate how AI can be implemented safely within sensitive environments whilst maintaining transparency and accountability.