AT2 Part A: Group Project  
GHG Digital Consultant

short line

Daniela Valencia Pineda: 24901098

Brayam Pineda Miranda:

Santiago Cardenas Carmona:

Juan Diego Balsero Leon:

Camilo Uribe Guerra:

36118 Applied Natural Language Processing - Autumn 2025

Master of Data Science and Innovation

University of Technology of Sydney

Table of Contents

[1. Project overview and scope 2](#_Toc195123329)

[2. Project phases and team contributions 2](#_Toc195123330)

[3. Data collection and sources 2](#_Toc195123331)

[4. Approaches and methodology 2](#_Toc195123332)

[5. Key Findings and insights 2](#_Toc195123333)

[6. Project outcomes and business value 3](#_Toc195123334)

[7. Challenges encountered 3](#_Toc195123335)

[8. Recommendations and future work 3](#_Toc195123336)

[9. References 3](#_Toc195123337)

[Appendix 3](#_Toc195123338)

# Project overview and scope

1. Key Objectives
2. Purpose: Define the core aim of your chatbot: to assist stakeholders in understanding and developing regulations around Greenhouse Gas (GHG) emissions in Australia.
3. Target users: Policymakers, environmental analysts, legal advisors, or researchers involved in sustainability and climate regulation.
4. Scope:
   1. Focus on Australian regulations.
   2. Reference international regulations to inspire local policy improvements.
   3. Provide accurate and contextualized answers using RAG + LLM.
   4. Limitations: language scope, depth of legal interpretation, domain constraints.

* XYZ Project objectives and scope

# Project phases and team contributions

1. Break the development timeline into distinct phases. Example:
   1. **Phase 1 – Problem Framing & Planning**
   2. **Phase 2 – Data Collection & Preprocessing**
   3. **Phase 3 – RAG Pipeline Implementation**
   4. **Phase 4 – API Integration & App Development**
   5. **Phase 5 – Evaluation & Refinement**
2. For each phase, list **who did what**:
   1. [Name] – Collected AU regulatory documents, annotated question samples
   2. [Name] – Built the RAG pipeline
   3. [Name] – Integrated API with front-end
   4. [Name] – Managed report writing and model evaluation
3. Summarize the **overall system architecture**:
4. Web-based chatbot
5. RAG pipeline (document retrieval + LLM response generation)
6. Use of API to access a powerful LLM
7. Outline NLP techniques used:
8. Document chunking
9. Embedding generation
10. Semantic search
11. Prompt engineering

short dash

# Data collection and sources

1. Australian regulations (e.g., National Greenhouse and Energy Reporting (NGER) Scheme, Safeguard Mechanism, state-specific rules)
2. International regulations (e.g., EU ETS, U.S. EPA GHG rules, UNFCCC)
3. Document formats: PDFs, scraped websites, .docx files
4. Storage and retrieval: Explain how documents are stored (e.g., chunked, embedded) and queried using RAG.

short dash

# Approaches and methodology

* NLP methods and techniques used, including justification for choices made

short dash

# Key Findings and insights

1. Summarize:
   1. Accuracy and usefulness of responses.
   2. How well the chatbot matched user questions to relevant regulatory content.
   3. Gaps or inconsistencies observed (e.g., international docs using different terminology).
2. Optional: Include user feedback or testing cases.
3. What your chatbot achieves:
   1. Enables faster access to GHG regulatory knowledge.
   2. Supports cross-comparison of policies (AU vs. global).
   3. Educates users with limited legal or environmental expertise.
4. Stakeholder impact:
   1. Government, climate policy analysts, startups in green tech.

short dash

# Project outcomes and business value

* Outcomes and value added

short dash

# Challenges encountered

Challenge: Ambiguity in user questions.

* Solution: Fine-tuned prompt templates and chunk filtering.

Challenge: Varying formats of regulatory documents.

* Solution: Unified preprocessing and content chunking pipeline.

Challenge: Latency in API responses.

* Solution: Streamlined chunk retrieval, reduced context window.

# Conclusions

short dash

# Recommendations and future work

1. Recap what was achieved.
2. List improvements for future development:
3. Fine-tuning on domain-specific data
4. UI/UX enhancements
5. Adding feedback collection and analytics

short dash

# References

* Include a list of references used throughout the project report.

short dash

# Appendix