Appendix A

The Watsonx Prototype

This Appendix section provides a visual for the prototype; as well as serving as a guide.

A.1 Watsonx.ai Models

We can select a user project and model for inferencing as shown in Figure A.1. We can also view information about the model here.

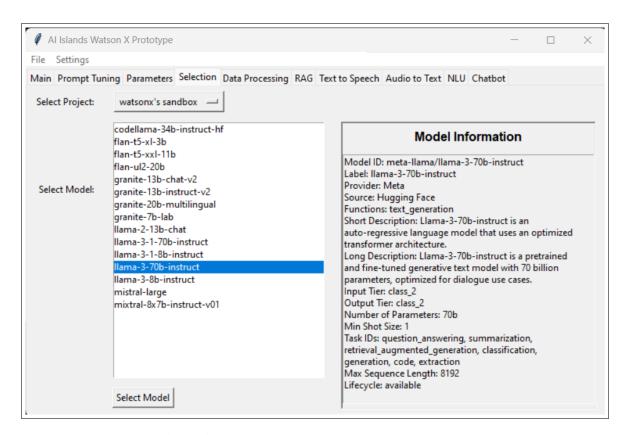


Figure A.1: The Watsonx Prototype - Model Selector

We can set parameters for the model, as shown in Figure A.2.

APPENDIX A. THE WATSONX PROTOTYPE

Al Islands Watson X Prototype	_	×
File Settings		
Main Prompt Tuning Parameters Selection Data Processing RAG Text to Speech Audio to Text NLU	Chatbot	
Temperature: 0.7		
Top P: 1.0		
Top K: 50		
Max New Tokens: 1000		
Min New Tokens: 1		
Repetition Penalty: 1.0		
Random Seed: 42		
Stop Sequences: None		

Figure A.2: The Watsonx Prototype - Parameter Configuration

We can inference the model on the *Main* tab, as shown in Figure A.3. It is currently configured to accept a dataset, as shown in Figure A.4.

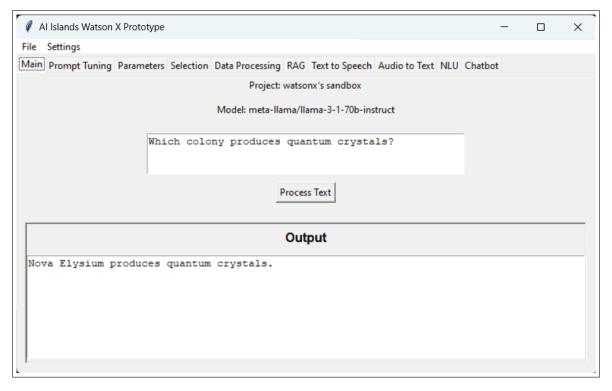


Figure A.3: The Watsonx Prototype - Model Inference With RAG Setting On

APPENDIX A. THE WATSONX PROTOTYPE

We can select a dataset after it has been processed within the *Data Processing* tab and configure the model to use the dataset by checking the 'Use Dataset' option, as shown in Figure A.4.

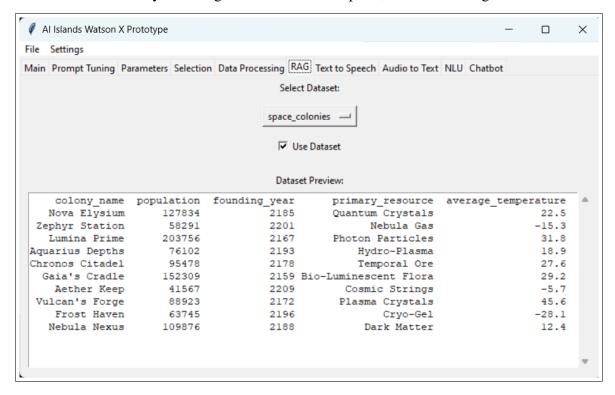


Figure A.4: The Watsonx Prototype - RAG Dataset Selector

We can preserve chat history and use a chatbot instead of single inferencing, as shown in Figure A.5.

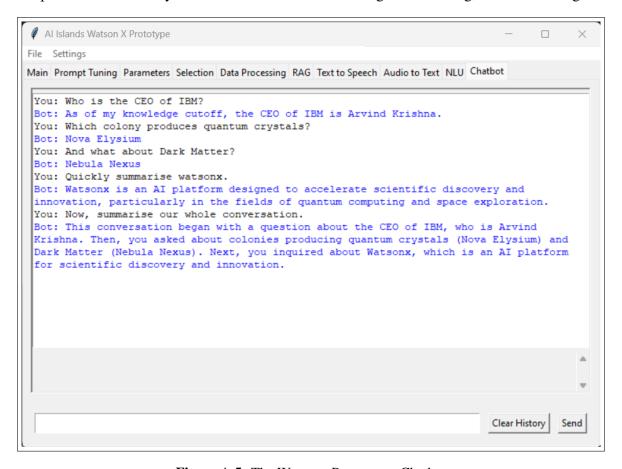


Figure A.5: The Watsonx Prototype - Chatbot

APPENDIX A. THE WATSONX PROTOTYPE

A.2 Watson Services

The prototype has the NLU, TTS and STT services available. Figure A.6 shows the NLU being inferenced, with all features.

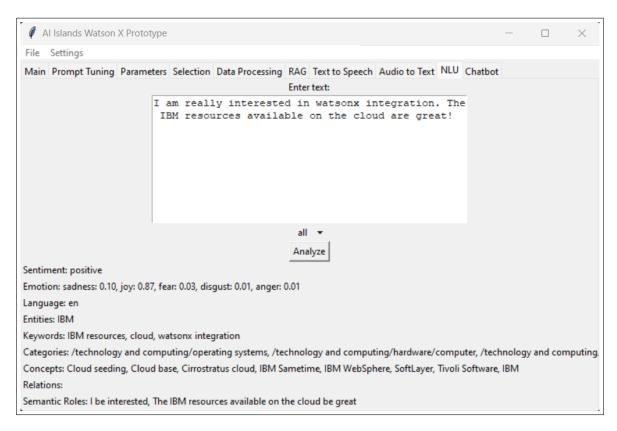


Figure A.6: The Watsonx Prototype - NLU

The method used to authenticate the services is the unified authentication method, as discussed in the main report, Chapter 6.