

Evaluating AI: An Overview of the AI as Infrastructure (AIINFRA) Project

AI Meets Humanities: NTU-ANU-KCL Workshop
November 19 – 21, 2025
Nanyang Technological University, Singapore

James Smithies
Professor of Digital Humanities
Director HASS Digital Research Hub
The Australian National University

Glen Berman
PhD Candidate and Senior Research Officer
Director HASS Digital Research Hub
The Australian National University



Overview

01 The AI as Infrastructure Project

02 HASTRICT Principles

03 ATLAS Darwin



01 THE AI AS INFRASTRUCTURE (AIINFRA) PROJECT

A trans-national project focussed on assessing the value of Retrieval Augmented Generation (RAG) for historical research, using Hansard debates from Australia, New Zealand, United Kingdom as our content baseline.

Partners

- ANU HASS Digital Research Hub
- ANU Library
- Aotearoa / New Zealand Department of Internal Affairs
- Australian Parliamentary Library
- King's College London
- The National Library of Australia
- UK History of Parliament project
- UK National Archives



- 2024 – 2026, ANU Futures Scheme
- <https://aiinfra.anu.edu.au/>



01 THE AI AS INFRASTRUCTURE (AIINFRA) PROJECT

Goals

- Identify and implement best practices for the design, development, evaluation, and maintenance of historical research tools enabled by Large Language Models (LLMs).
- Develop high-level principles (the HASTRICT principles), an organisational maturity model, and associated evaluation framework to support the responsible implementation of AI tools into the GLAM and academic sector.
- Develop a prototype technical ‘test harness’ for the evaluation of LLM Retrieval Augmented Generation (RAG) systems, and as a test bed for development of the other project outputs, using open-source tools.



04 HASTRICT PRINCIPLES

- **H:** Human
- **A:** Accountability
- **S:** System targets
- **T:** Transparent
- **R:** Reproducible
- **I:** Iterative
- **C:** Contextually Informed Measures & Baselines
- **T:** Tasks, workflows & use cases

James Smithies, Glen Berman, Karaitiana Taiuru, 'Operationalisation of Responsible AI in the Research and Cultural Sectors'. In *Building Digital Humanities*, edited by Simon Burrows et al. Radboud University Press. Under review.

Department of Industry Science and Resources, "Guidance for AI Adoption: Implementation", Department of Industry Science and Resources," **October 21, 2025**, <https://www.industry.gov.au/publications/guidance-for-ai-adoption/guidance-ai-adoption-implementation-practices>.



04 ATLAS DARWIN

ATLAS Darwin

[New Session](#)[About](#)[FAQ](#)

In what contexts does Darwin discuss stock

Darwin discusses "stock" in several distinct contexts throughout his correspondence and writings, reflecting the term's varied meanings in 19th-century usage.

Breeding and Heredity

Darwin frequently uses "stock" to refer to breeding lines or genetic lineages in both animals and plants. In his bee-keeping correspondence, he mentions "pure Ligurian stock" when discussing hybridization between different bee varieties. His systematic questioning about breeding reveals deep interest in how traits are transmitted: "Is a peculiarity which has long been in blood more easily transmitted, than a newly acquired one?" He also inquires about crossing different stocks, asking "When old variety is crossed with new; or natural one with very artificial one... do offspring partake more of the natural than artificial kind?"

Financial Investments

Darwin's personal papers reveal extensive discussion of financial stocks and investments. His correspondence with stockbrokers details specific railway investments, including "North Eastern—Berwick 4% Preference" stock and "Gt. North of Scotland Ry. Co." shares. The brokers provide detailed dividend information, noting that ordinary share capital paid dividends "varying from 4½ to 6%" and describing market conditions where "Berwick stock" was sometimes impossible to purchase.

Scientific Knowledge and Information

Darwin also employs "stock" metaphorically to describe accumulated knowledge, writing to Joseph Hooker about extracting "more knowledge from your well-arranged stock," referring to Hooker's botanical expertise.

Plant Breeding Context

In botanical discussions, Darwin uses "stock" when referring to self-fertilization mechanisms, noting how certain violet species "produce... small closed flowers to insure a stock of seed through self-fertilisation."

These varied uses demonstrate Darwin's practical engagement with both scientific breeding principles and personal financial management, alongside his theoretical work on heredity and variation.

[after 8 May 1860] **1876-10-21** 1875-02-16 1843-08-02 1880-12-07 [1871-82] 1877-04-20 1876-11-12 1879-05-09 1869-05-06 1863-05-10 [after 21 Jan 1871] 1863-01-22 1859-06-17 1876-01-19 1860-02-18 1872-03-12 1845-10-08 1879-12-26 [Feb-July 1838] [View all 20 citations](#)

[Enhanced Feedback](#)[Skip](#)☐ Privacy Off

Sharing anonymous usage data and feedback

Test Target

Atlas Version
Darwin 0.2.0

Composite Target
k20_claudel_darwin

LLM Provider
Anthropic

LLM Model
claude-sonnet-4-20250514

System Prompt
You are an expert historical research assistant specializing in Charles Darwin's works and writings. Your expertise covers Darwin's scientific work; p... [View Full Prompt](#)

Vector Store Overview
[View Overview](#)

Embedding Model
Livingwithmachines/berit_1760_1900

Algorithm
HNSW

Search Type
Hybrid

Retrieval Size Single Corpus
80

Retrieval Size All Corpora
100

ATLAS: Analysis and Testing of Language Models for Archival Systems

- <https://github.com/AI-as-Infrastructure/aiinfra-atlas>
- <https://github.com/AI-as-Infrastructure/aiinfra-atlas-darwin>

Citation Details

say in the foundation of your "Origin of Species" that the words in typicly quotes, mostly the qualifying "if", scarcely bear out the inference he wishes drawn—You put the views forward distinctly as an hypothesis to meet certain difficulties in the evidence of geology with respect to evolution—The question is a very important one and can scarcely be settled by quoting authority—it is a matter of evidence. Faithfully Yours | T. Mellard Reade Dr. Chas Darwin FRS. &c Will you kindly return the Magazine.—

Metadata

Title Letter from Reade, T. M. to Darwin, C. R. (1880-12-07)

Date 1880-12-07

Corpus darwin

URL <https://www.darwinproject.ac.uk/letter/?docId=letters/DCP-LETT-12894.xml>

ID DCP-LETT-12894

[Close](#)

Extended Feedback

Help us improve our understanding of LLM RAG systems by providing detailed feedback

Factual Accuracy ⓘ

☒ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Corpus Fidelity ⓘ

☒ 1 ☒ 2 ☐ 3 ☐ 4 ☐ 5

Analysis Quality ⓘ

☒ 1 ☒ 2 ☒ 3 ☐ 4 ☐ 5

Relevance ⓘ

☒ 1 ☒ 2 ☐ 3 ☐ 4 ☐ 5

Difficulty ⓘ

☒ 1 ☒ 2 ☒ 3 ☒ 4 ☐ 5

Clarity ⓘ

☒ 1 ☒ 2 ☒ 3 ☐ 4 ☐ 5

User Expertise

☐ Expert ☒ Non-expert

Search Type ⓘ

Hybrid

Retrieval Size Single Corpus ⓘ

80

Retrieval Size All Corpora ⓘ

100

Search K ⓘ

20

Search Score Threshold ⓘ

0

Chunk Size ⓘ

1500

Chunk Overlap ⓘ

250

Pooling ⓘ

mean+max

Multi Corpus Vectorstore ⓘ

False

Export Session



PHOENIX Projects > Darwin-Prod

Total Traces: 22 Total Cost: \$0 Latency P50: 14.35s Latency P99: 19.14s User Comment: -- Relevance Rating: μ 3.74 Factual Accuracy: μ 3.89 Clarity: μ Stream: ☐ From 9/7/2025 09:00:00 AM

Spans Traces Sessions Metrics Config

Q filter condition (e.x. span_kind == 'LLM') + Root Spans All Columns

	st...	kind	name	input	output	annotations	start time	latency
<input type="checkbox"/>	⊖	agent	com.atlas.rag.pipeline	In what contexts does...	Darwin discusses "sto...		04/10/2025, 02:23 am	15.46s
<input type="checkbox"/>	⊖	agent	com.atlas.rag.pipeline	in what contexts does...	Darwin refers to Cirrip...		04/10/2025, 02:10 am	17.20s
<input type="checkbox"/>	⊖	agent	com.atlas.rag.pipeline	Does Darwin change ...	Based on the provide...		04/10/2025, 02:02 am	14.51s
<input type="checkbox"/>	⊖	agent	com.atlas.rag.pipeline	How do the terms poll...	Darwin clearly disting...		04/10/2025, 12:53 am	16.17s
<input type="checkbox"/>	⊖	agent	com.atlas.rag.pipeline	How did Darwin react ...	I can only discuss topi...		04/10/2025, 12:44 am	10.23s
<input type="checkbox"/>	⊖	agent	com.atlas.rag.pipeline	Does Darwin change t...	Based on the provide...		04/10/2025, 12:15 am	11.49s
<input type="checkbox"/>	⊖	agent	com.atlas.rag.pipeline	What words change t...	I cannot provide a co...		04/10/2025, 12:00 am	12.95s
<input type="checkbox"/>	⊖	agent	com.atlas.rag.pipeline	What words does Dar...	Based on the provide...		03/10/2025, 11:46 pm	13.43s
<input type="checkbox"/>	⊖	agent	com.atlas.rag.pipeline	How many species of ...	Based on the provide...		03/10/2025, 11:46 pm	11.98s

Ask AI



The screenshot displays the PHOENIX ATLAS DARWIN interface. On the left is a sidebar with navigation options: Projects, Datasets & Ex, Playground, Prompts, APIs, API Key, Star on GitHub, Settings, Documentation, Support, Dark, and Logout. The main area is titled 'Trace Details' and shows the status of a trace (UNSET), total cost (\$0), and latency (15.46s). Below this, a 'Trace' section lists components of the pipeline: com.atlas.guardrails.sensitivity, com.atlas.rag.retrieval, atlas.retrieval, com.atlas.rag.generation (382), com.atlas.rag.generation.response, and com.atlas.rag.references. The 'com.atlas.rag.generation' component is selected, showing its input and output. The input is 'In what contexts does Darwin discuss stock'. The output is a detailed analysis of Darwin's use of 'stock' in various contexts, including agricultural, breeding, intellectual, and knowledge contexts. On the right, there are sections for 'Annotation Summary', 'Edit Annotations', and 'Notes'. At the bottom right, there is an 'Ask AI' button and an 'Add a note' input field.

Trace Details

Trace Status: UNSET | Total Cost: \$0 | Latency: 15.46s

Trace

- com.atlas.guardrails.sensitivity
- com.atlas.rag.retrieval
 - atlas.retrieval
 - com.atlas.rag.generation (382)
 - com.atlas.rag.generation.response
 - com.atlas.rag.references

com.atlas.rag.generation (15.46s at 10/4/2025 02:23:50 AM)

Input

In what contexts does Darwin discuss stock

Output

Darwin discusses "stock" in several distinct contexts throughout the provided documents, reflecting both his scientific interests and personal financial affairs.

****Agricultural and Breeding Contexts****

Darwin frequently uses "stock" to refer to breeding lines of animals. In his bee-keeping correspondence, he mentions "one of his hives of the pure Ligurian stock," discussing the maintenance of distinct breeding populations. He also refers to plant breeding, noting how "V. canina & odorata produce perfect flowers adapted for a cross, & small closed flowers to insure a stock of seed through self-fertilisation," demonstrating his understanding of how organisms maintain genetic continuity.

****Intellectual and Knowledge Contexts****

Darwin employs "stock" metaphorically when referring to accumulated knowledge or expertise. He writes admiringly of a colleague's "well-arranged stock" of botanical knowledge, indicating how he viewed

Annotation Summary

Edit Annotations

Notes

Ask AI

Add a note



The screenshot displays the PHOENIX ATLAS DARWIN interface. On the left is a sidebar with navigation links: Projects, Datasets & Ex, Playground, Prompts, APIs, API Key, Star on GitHub, Settings, Documentation, Support, Dark, and Logout. The main area is titled 'Trace Details' and shows a trace for 'com.atlas.rag.pipeline'. The trace status is 'UNSET', total cost is '\$0', and latency is '15.46s'. The trace steps are: com.atlas.guardrails.sensitivity, com.atlas.rag.retrieval, atlas.retrieval, com.atlas.rag.generation (382), com.atlas.rag.generation.response (selected), and com.atlas.rag.references. The selected step shows a table of annotations for 'com.atlas.rag.generation.response' at 10/4/2025 02:24:05 AM. The table has columns: name, annotator kind, user, label, and score. The annotations are: Analysis Quality (HUMAN, system, analysis_quality, 4), Clarity (HUMAN, system, clarity, 5), Corpus Fidelity (HUMAN, system, corpus_fidelity, 5), Factual Accuracy (HUMAN, system, factual_accuracy, 5), Query Difficulty (HUMAN, system, query_difficulty, 5), and Relevance Rating (HUMAN, system, relevance, 5). On the right, there are sections for 'Annotation Summary', 'Edit Annotations', 'Add Annotation', 'Notes', and an 'Ask AI' button.

Trace Details

Trace Status: UNSET, Total Cost: \$0, Latency: 15.46s

Trace

- com.atlas.rag.pipeline
 - com.atlas.guardrails.sensitivity
 - com.atlas.rag.retrieval
 - atlas.retrieval
 - com.atlas.rag.generation (382)
 - com.atlas.rag.generation.response (selected)
 - com.atlas.rag.references

com.atlas.rag.generation.response

1.00ms at 10/4/2025 02:24:05 AM

Info Annotations (8) Attributes Events (0)

name	annotator kind	user	label	score
Analysis Quality	HUMAN	system	analysis_quality	4
Clarity	HUMAN	system	clarity	5
Corpus Fidelity	HUMAN	system	corpus_fidelity	5
Factual Accuracy	HUMAN	system	factual_accuracy	5
Query Difficulty	HUMAN	system	query_difficulty	5
Relevance Rating	HUMAN	system	relevance	5

Annotation Summary (5)

Edit Annotations (E)

Add Annotation

No annotation configurations for this project

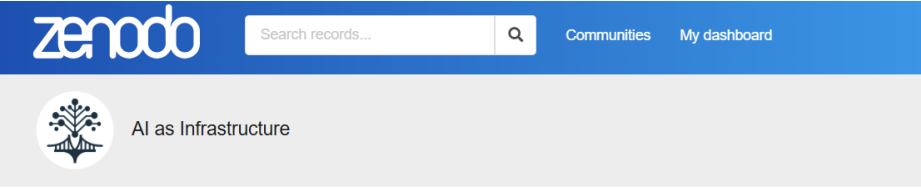
Configure Annotation Configs

Notes (N)

Ask AI

Add a note





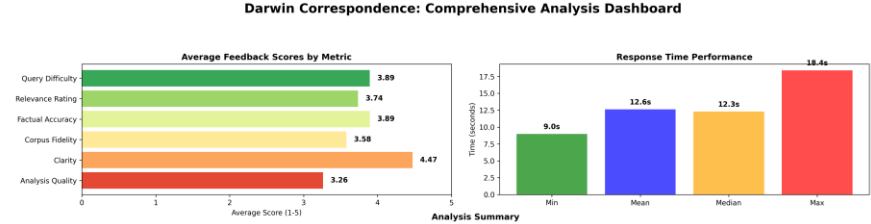
Published October 31, 2025 | Version v2

Dataset Open

ATLAS Darwin Telemetry 06-10-2025

Smithies, James (Project leader)¹ ; Berman, Glen (Project member)¹ ;
McGillivray, Barbara (Project member)² ; Smith, Elizabeth (Project member)³

Show affiliations



Metric	Value
Total Records	292
Unique Q&A Sessions	22
Feedback Annotations	157
Date Range	2025-10-02 13:20:33+00:00 to 2025-10-03 16:29:52+00:00

- Data export from Smithies, James, and Glen Berman. 'AI-as-Infrastructure/aiinfra-atlas-darwin: V0.2.0'. Zenodo, 8 October 2025. <https://doi.org/10.5281/zenodo.17298447> including expert user feedback from Barbara McGillivray and Elizabeth Smith, conducted during October 2025.

Smithies, James, Glen Berman, Barbara McGillivray, and Elizabeth Smith. 'ATLAS Darwin Telemetry 06-10-2025'. Zenodo, 31 October 2025. <https://doi.org/10.5281/zenodo.17491578>.



THANK YOU

<https://zenodo.org/communities/aiinfra/>



Australian
National
University