



Playing Heaven

Remapping Early Modern Neo-Confucian Worlds with AI



Javier Cha
PRESENTING



Donghyeok Choi



Eric Chow



Chien-Cho Chan



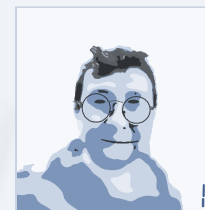
The University of Hong Kong



Yan Hon
Michael Chung
PRESENTING



Miguel
Escobar Varela



Florin-Stefan
Morar



The Hong Kong University of
Science and Technology



National University of Singapore

1

Four Challenges

01

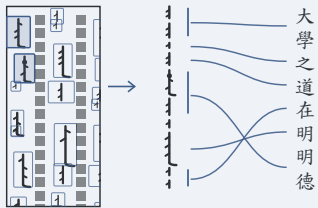
Multilingual Intellectual History

Challenge

Historical accounts tend to privilege a dominant and prestigious written language, even though multilingualism has long been an ordinary condition of human life. In early modern East Asia, Neo-Confucian scholarship circulated principally in Literary Sinitic, but its dissemination and reception involved multiple linguistic communities, literary forms, and sociopolitical settings.

Our Approach

The multilingual capabilities of LLMs extend well beyond the reach of any individual historian, allowing us to recover, connect, and compare historical sources in Manchu, vernacular Chinese, various forms of literary Sinitic, including those practised in Korea. This work provides a basis for wider perspectives involving Japan, Okinawa, Vietnam, and Jesuit scholarship.



03

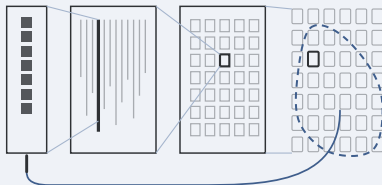
From Individual Texts to Intellectual Formations

Challenge

Intellectual history requires moving from individual passages and debates to larger formations such as schools of thought and literary and aesthetic movements. We also need to factor in that some scholars also pursued forms of semiotics that questioned the a priori existence of fixed concepts. A further challenge is accounting for the contributions of thousands of less prominent scholars who participated in the formation, evolution, and dissemination of these intellectual and cultural traditions.

Our Approach

We organize specialized AI agents in a research pipeline our team has developed, which we call **agentic knowledge distillation**. The agents analyze passages, synthesize higher-order interpretations of essays, and simulate debates among scholars. Each successive layer remains linked to the underlying records, allowing historians to examine the evidence at every interpretive level.



02

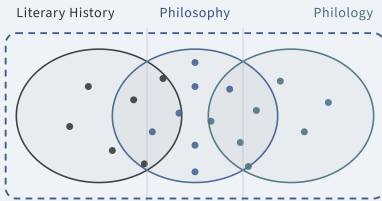
The Boundaries of Modern Disciplines

Challenge

In early modern East Asia, literature, moral philosophy, and philology were overlapping scholarly pursuits, with individual scholars placing different emphases on each. Modern academic divisions, however, treat them as distinct fields of inquiry.

Our Approach

Our computational research environments, Pivot 機 and Traces 跡, employ specialized AI agents to bring together literary, philosophical, and philological developments into an integrated field of view.



04

Understanding and Evaluating Machine Interpretation

Challenge

LLMs can now produce interpretation-like readings of primary sources and secondary scholarship, extending the possibilities of computational hermeneutics beyond earlier approaches centred on word frequencies and semantic drift. The tradeoff is that we know little about how models generate these sophisticated outputs. A further challenge involves evaluating model performance, since interpretive tasks involve a range of plausible responses rather than a single right or wrong answer.

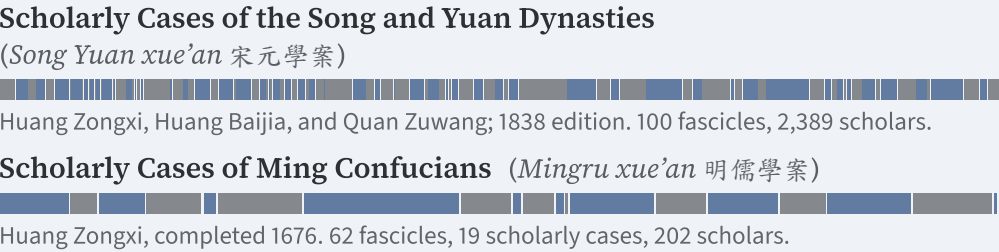
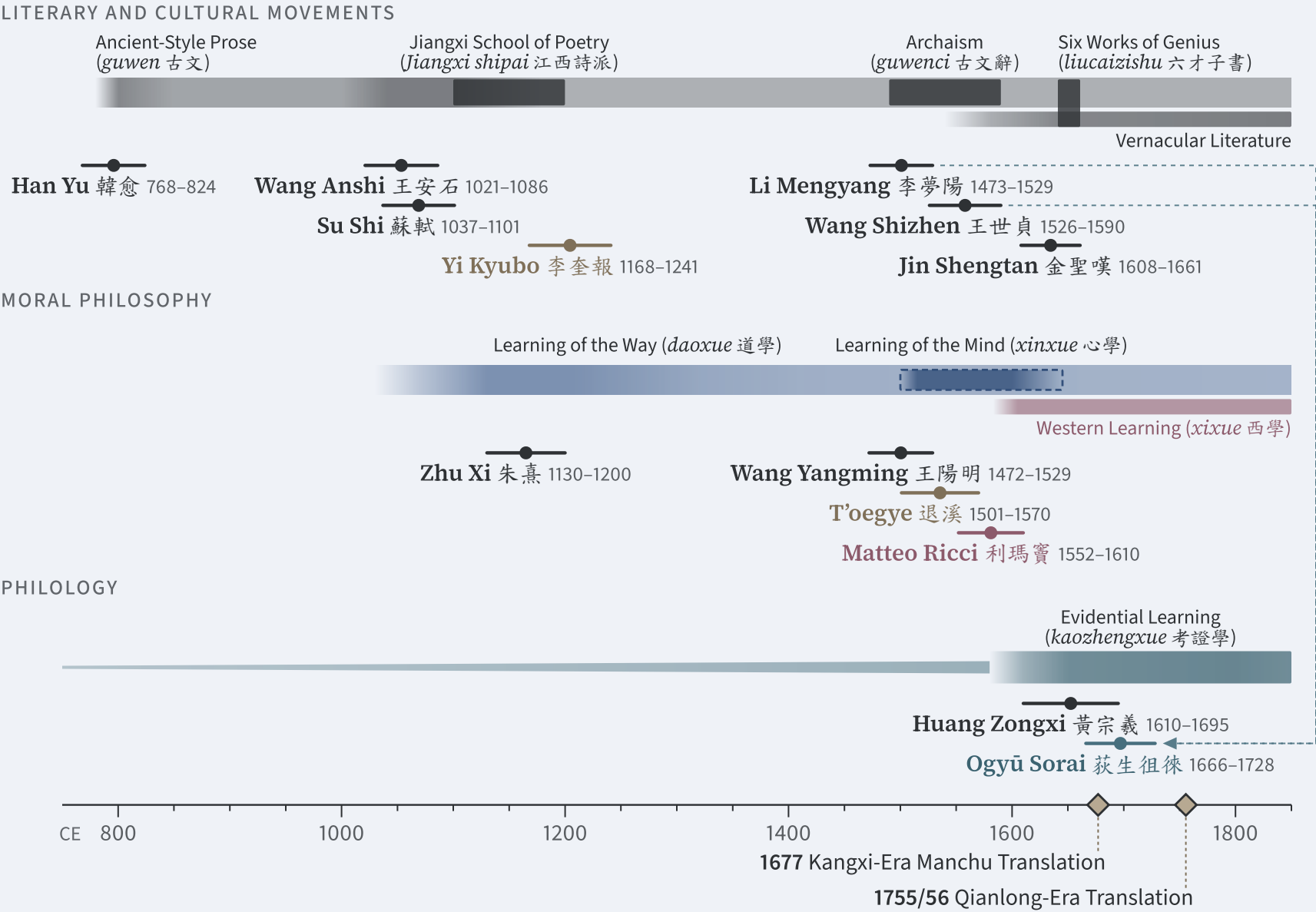
Our Approach

We have experimented with persona-vector steering and sparse autoencoders, and are considering contrastive embedding models for literary-style analysis. We are also exploring evaluation methods involving expert focus groups and comparisons of outputs generated by multiple models, with particular attention to whether their interpretations are plausible, supported by historical evidence, and defensible against alternative readings.

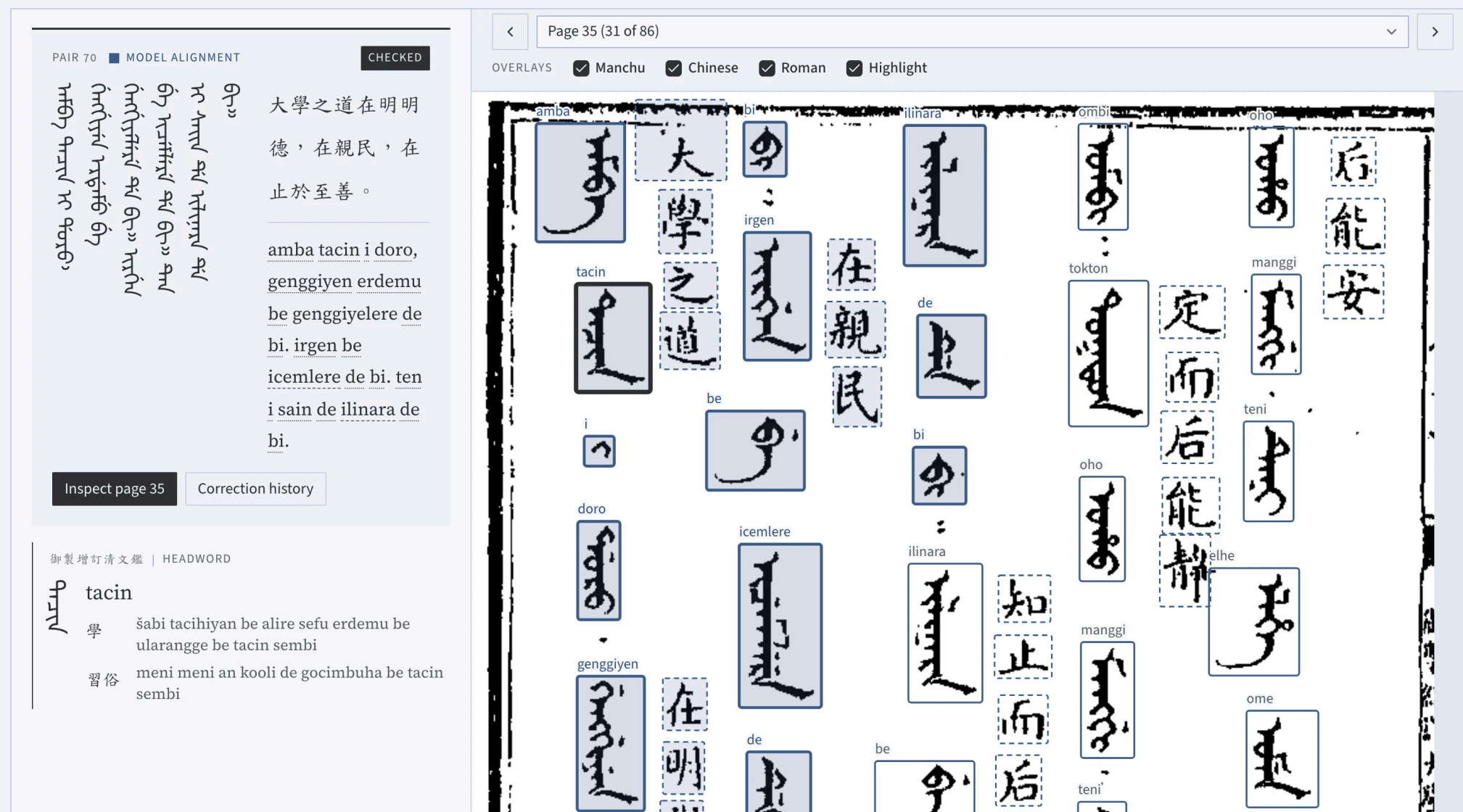


2 Neo-Confucianism and Related Intellectual and Cultural Movements

For more than a millennium, Neo-Confucianism and related intellectual traditions informed literary and aesthetic practices, moral philosophy, and philological inquiry in East Asia and beyond. These traditions developed in close interaction, giving rise to related and sometimes competing schools of thought and cultural movements. Thousands of scholars across China, Korea, Japan, and Vietnam, as well as Jesuit missionaries, participated in these exchanges. Their writings, intellectual affiliations, scholarly fellowships, and cultural practices reflected considerable diversity within and across linguistic and political boundaries. Playing Heaven investigates how these activities and relationships contributed to larger intellectual, social, and cultural developments.



3 Bringing Manchu Sources into Focus



We fine-tuned a VLM for historical Manchu transcription and then integrated word detection to further improve recognition. Building on this research, HKUST and the Staatsbibliothek zu Berlin signed an agreement to advance OCR research on the library's extensive Manchu holdings, managed by its CrossAsia division.

For Playing Heaven, our digital research environment brings Manchu transcriptions, corresponding Chinese passages, and original page images together for comparative investigation. Scholars can inspect and correct individual readings, compare accounts across languages, and pursue historical questions arising from differences in their perspectives and representations.

[CrossAsia — Manchu OCR Announcement ↗](#)

96.28%

Word Accuracy
Real Handwritten Manchu Documents

Llama-3.2-11B-Vision
Fine-Tuned for Historical Manchu
Yan Hon Michael Chung and Donghyeok Choi

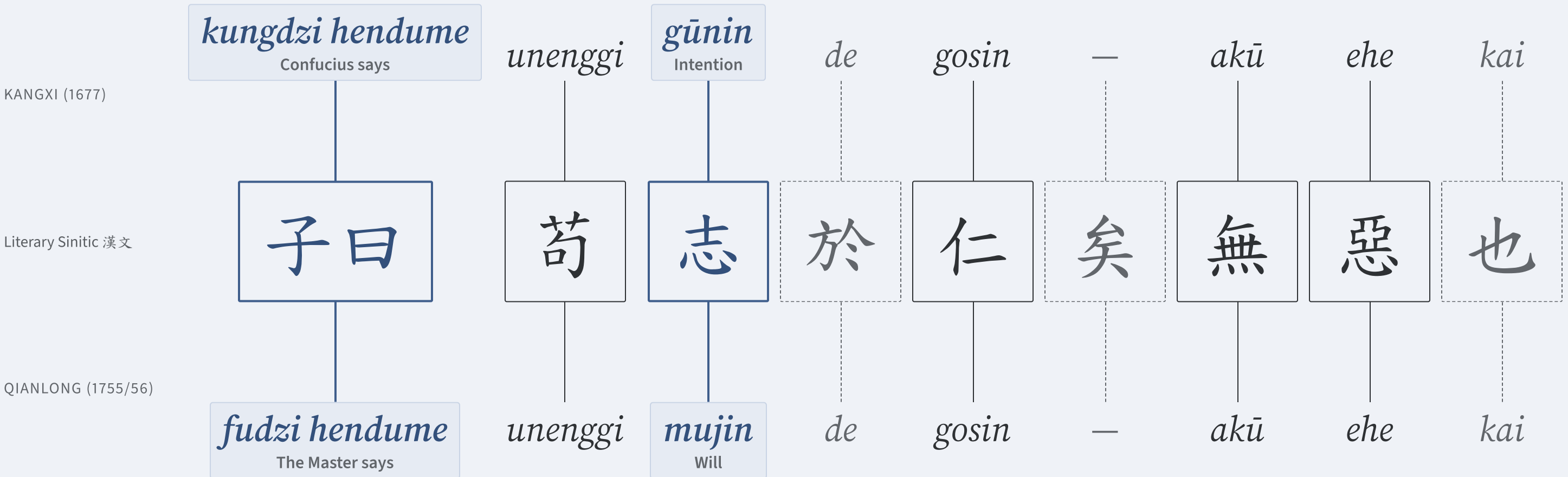
4

We align successive Manchu translations of the Four Books and Five Classics (*Sishu Wujing* 四書五經), the classical canon redefined by Neo-Confucian scholars, to examine how the Kangxi and Qianlong courts rendered and revised classical terminology. Comparisons across editions allow us to identify recurring

translation choices and investigate the editorial decisions of the two imperial courts as they rendered the highly literary language of the Chinese classics into vernacular Manchu.

Analects of Confucius 4.4 | Kangxi (1677) and Qianlong (1755/56) Editions

子曰：苟志於仁矣，無惡也。 The Master said: if the will is set on humaneness, there is no wickedness.



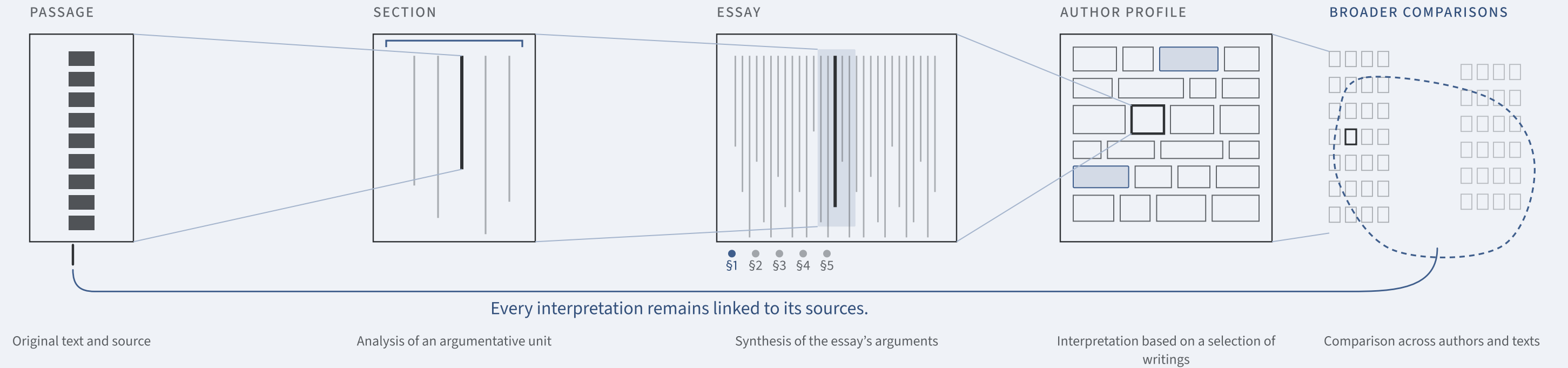
— Unchanged Rendering — Changed Rendering ----- Grammatical Particle or Unaligned Element

5

Pivot 機: Agentic Knowledge Distillation

Pivot organizes specialized AI agents to analyse historical texts and construct successive layers of interpretation, from individual passages to essays, author profiles, and comparative corpora.

Agent skills define the analytical tasks, while Pivot maintains the resulting interpretations as source-linked records that historians can inspect, correct, and revise.



1 Select and Segment the Text

文與可畫篋簞谷偃竹記, line 4 of 29

HISTORICAL SOURCE

故畫竹，必先得成竹於胸中

Therefore, to paint bamboo one must first have the complete bamboo in the breast.

3 Construct a Section Interpretation

RECORDED MODEL OUTPUT

Artistic creation requires a complete, pre-conceived mental image of the whole bamboo prior to applying the brush. Piecemeal construction (adding segment by segment, leaf by leaf) destroys organic integrity. Execution must be immediate and swift to match mental vision.

4 Examine the Interpretive Dimensions

INTERPRETIVE DIMENSIONS

- C1 Semiotics and Epistemology 名實
How artistic knowledge relates to the representation of its subject.
- C9 Aesthetics and the Way 文 / 道
How artistic practice relates to a larger account of the Way.

6

Pivot Debate: Wang Anshi and Su Shi on Dao

Pivot’s Debate feature uses AI agents to compare interpretations of Wang Anshi’s and Su Shi’s writings, examining how each conceived of Dao and its relationship to language, literature, and government. By drawing on their respective essays and

broader intellectual positions, the agents identify fundamental philosophical disagreements and contrasting worldviews that their superficially shared classical vocabulary might otherwise obscure.

MODEL

Gemini 3.8 Flash (Medium)

QUESTION

What are 王安石 (Wang Anshi) and 蘇軾 (Su Shi)’s positions on what is 道 (dao)?

PARTICIPANTS

First王安石 (Wang Anshi)

Second蘇軾 (Su Shi)

positions read from the dossier by the model, not the authors speaking

ROUNDS2 | Comparison between interpretations

DOSSIER

PV.ESS.000678與祖擇之書, Wang Anshi

PV.PRF.000007Profile, Wang Anshi

PV.ESS.000363答張文潛書, Su Shi

PV.SRC.000843Bol, *This Culture of Ours*, excerpt

Begin

Model selected → Question submitted → Dossier assembled

→ Debate begun: PV.DEB.000002, run PV.RUN.002528

RECORDED DEBATE OUTPUT

POSITION READ FOR 王安石 (WANG ANSHI)

Wang Anshi conceives of Dao (道) as grounded in practical governance, statecraft, and administrative order (「治教政令」、「實用和實踐」). True literature (「聖人之所謂文」) is ontologically identical with sage statecraft, policy decrees, and institutional execution [...]
[PV.SRC.000843] [PV.ESS.000678] [PV.PRF.000007]

POSITION READ FOR 蘇軾 (SU SHI)

Su Shi conceives of Dao (道) not as an authoritarian monopoly or rigid statutory code, but as a living, immanent dynamic of nature and balance (「中庸」) [...] literature is co-extensive and contemporaneous with Dao (「與道俱」), emerging organically from moral character rather than enforced conformity.
[PV.SRC.000843] [PV.ESS.000363]

THEY DIFFER ONThe relationship between Literature (文) and Dao (道)

Wang Wang views true literature (「聖人之所謂文」) as coterminous with the actual administrative machinery and moral self-admonition of statecraft, subordinating rhetorical composition to governance [PV.SRC.000843] [PV.ESS.000678]

Su Su views literature as co-extensive and unfolding with the Way (「與道俱」), maintaining that creative, unforced literary engagement allows individuals to independently attain sagehood [PV.SRC.000843]

SOURCE RETURN

PV.ESS.000678與祖擇之書
Wang Anshi’s letter, as its essay synthesis reads it

PV.ESS.000363答張文潛書
Su Shi’s letter, as its essay synthesis reads it

PV.PRF.000007Profile · Wang Anshi
the author profile across his selected essays

Debate outcome PV.MDB.000002, a provisional synthesis written by Gemini 3.8 Flash (Medium): draft, 28 September 2026; Pivot’s validation left it requires review; stale since 1 October, when a new text version changed sections it rests on. The participants’ own models are not in the public record.

Open Debate Outcome in Pivot ↗

7 How Do Language Models Construct Interpretations?

LLMs can produce sophisticated readings of historical texts, but how do they arrive at these interpretations? We examine their internal representations and test how interventions affect their responses, without fine-tuning the underlying language models.

01 Persona-Vector Steering (Intervention)

We introduced activation-based interventions to test whether models could reproduce the documented intellectual positions of Wang Anshi and Su Shi. Steering altered tone and register but did not reliably change models’ interpretations of those positions.

Tone and Register CHANGED

Intellectual Position NOT RELIABLY REPRODUCED

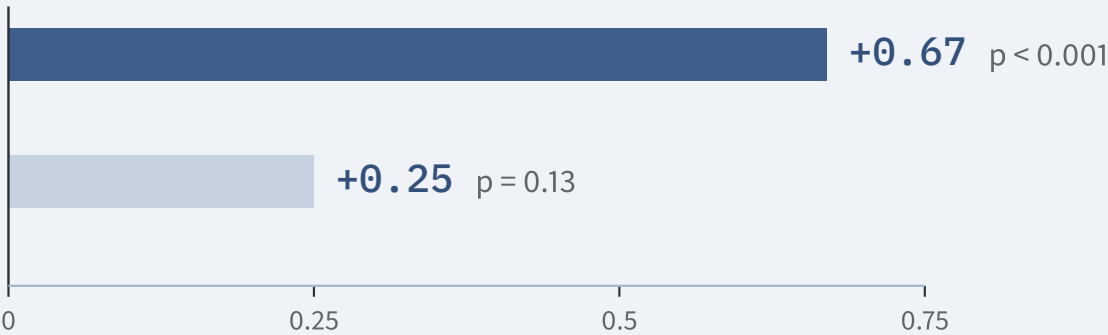
02 Internal Representations (Mechanistic Probing)

We used sparse autoencoders to examine internal features associated with different readings. Initial experiments found that the model’s internal states differed more between passages expressing opposing intellectual positions than between those expressing similar positions when answering questions about the texts, although the effect remained inconclusive on held-out comparisons.

STANCE GAP, STANDARD DEVIATIONS
Answering a question on the issue; layer 60, Gemma 3 27B

Initial Analysis
all 863 pairs

Held-Out Test
210 held-out pairs



Intellectual differences were more apparent in the model’s internal states during question answering. However, the same effect was weaker and statistically inconclusive when tested on a separate set of passage pairs withheld from the initial analysis.

8 Next Steps for Playing Heaven

- 01

Traces 跡: Rethinking Intellectual Formations

Traces reassesses inherited schools, lineages, and intellectual classifications against the surviving writings and documented relationships of thousands of lesser-known scholars. Beginning with retrospective internal histories of Confucian learning, it brings greater attention to scholars whose intellectual and cultural activities have received comparatively little scholarly attention.

[Open Traces 跡 ↗](#)

- 03

Evaluating Machine-Generated Interpretations

As agents produce larger collections of interpretive records, we need formal methods to assess their quality. We plan to combine domain-expert focus groups with comparisons among LLM agents to establish the range of historically defensible interpretations and distinguish plausible readings from those contradicted by the evidence.

- 02

Literary Styles and Aesthetic Practices

Develop specialized embedding models and classifiers using contrastive learning to identify literary styles, rhetorical conventions, and aesthetic features. Examine how these practices circulated among scholars and contributed to literary movements and intellectual affiliations.

- 04

LONGER-TERM DIRECTION

Transregional Intellectual History

Building on our research into intellectual connections between China and Korea and our work on Manchu sources, we hope to extend these investigations to a wider range of linguistic and political settings. This would allow us to examine the diversity of intellectual and cultural traditions within China, beyond the southeastern regions that have dominated scholarship, and their development elsewhere in East Asia and beyond. This next stage would follow the current three-year project, subject to further validation and support.