The Journal of Prompt-Engineered Philosophy

Or: How I Started to Track AI Assistance and Stopped Worrying About Slop

Michele Loi, University of Milan

Velázquez (1599-1660): **Las Meninas**

1. PAPER ORIGINS

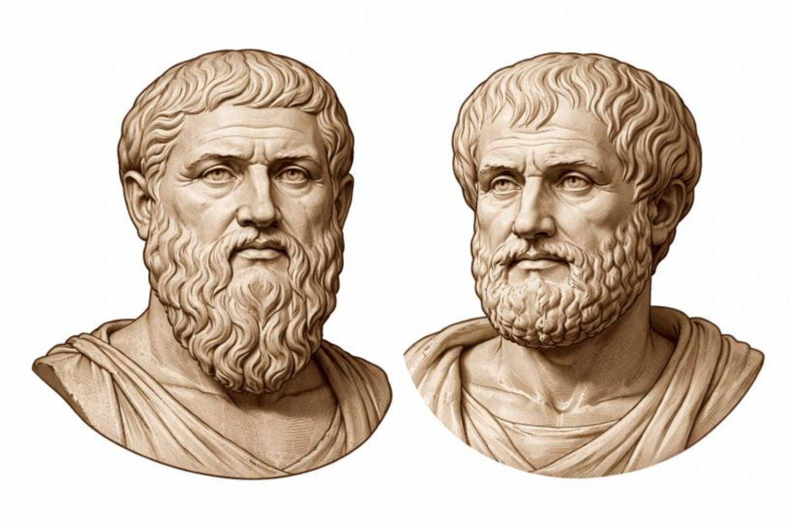
Tracking myself doing philosophy with AI

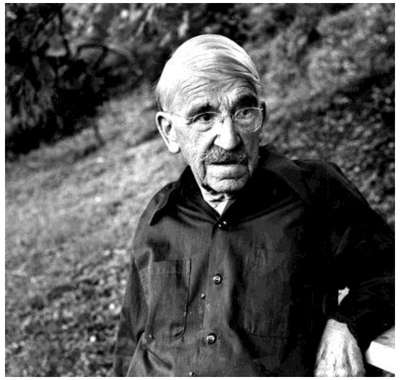
Conceptual and technical challenge

Why crucial?

THE WONDER-DRIVEN SCHOLAR

- Philosophical wonder about AI capabilities (Plato, Aristotle)
- Questions about AI intelligence demand experiential testing, not just theory





2. ENABLING MUTUAL LEARNING

- Philosophical writing has always been about the process of thinking as well as its output
  - Undisclosed use of AI *mask* the true process of thinking
  - Transparent use of AI reveal philosophical skills (e.g. questioning) in AI usage
  - Enabler of mutual learning
    - Across philosophers
    - For other disciplines (esp. methodologically)

3. CREATIVITY AND TOOL-MEDIATED DISCOVERY

- Musical composition example (Byrne): spaces+amplification shape music
- Modular synthesis and generative art as models
- AI assistance as extended creativity

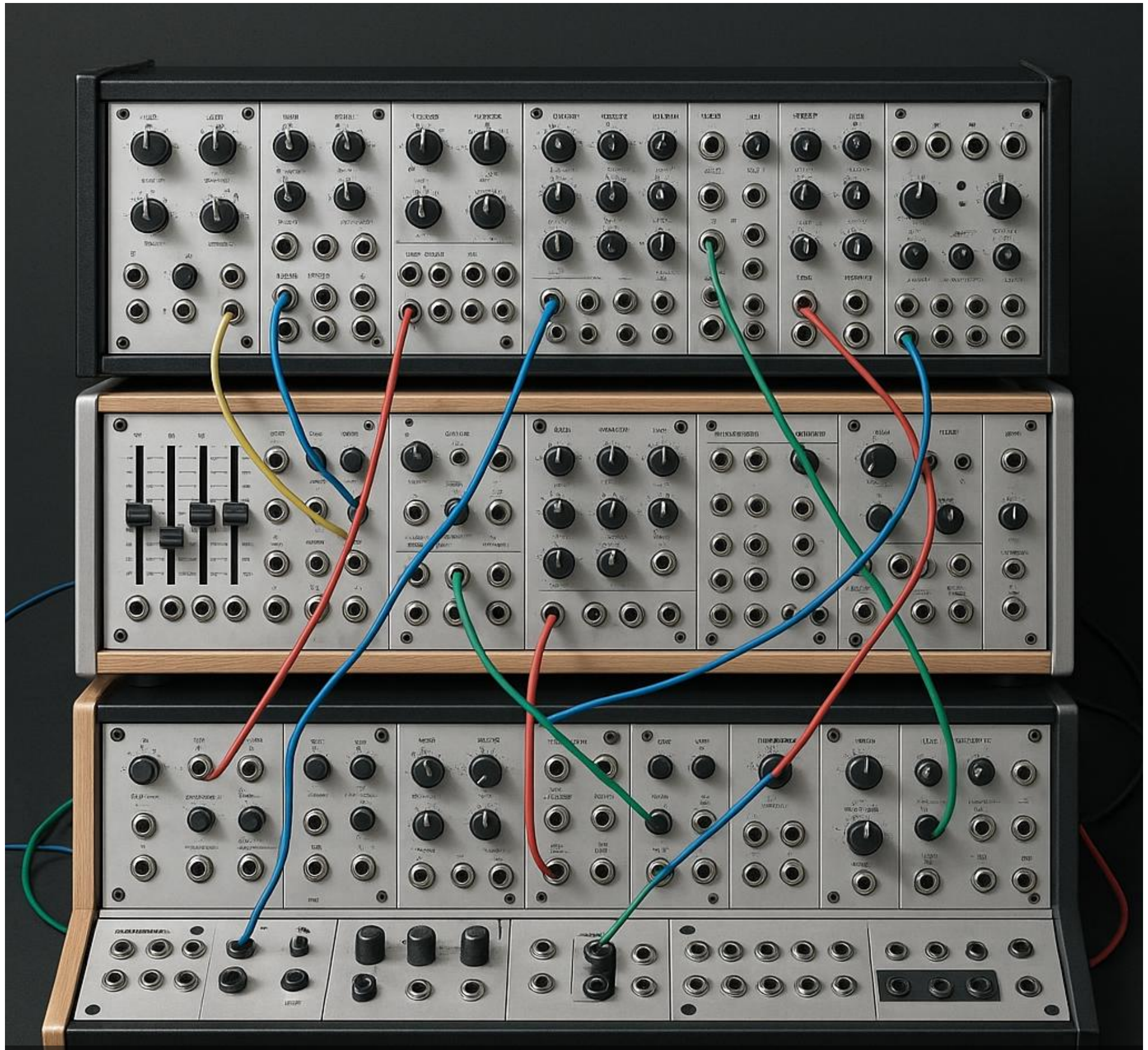


Figure 1. Picture of a hypothetical modular synth. By OpenAI's 4o image generation model.

WHILE THOSE THOUGHTS OCCURRED TO ME, I WAS HERE:

ChatGPT Image Jan 6, 2026, 09_55_43 AM

I DIDN'T WRITE ANYTHING ABOUT THIS FOR HALF A YEAR

What would be a suitable philosophical topic to write *about*, through AI?

One day, it suddenly clicked. I was here:

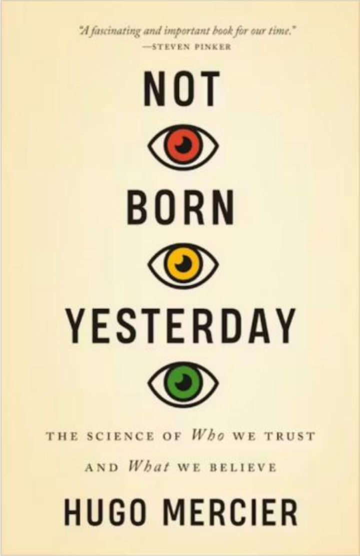
MY FRUSTRATION

Current rules:

1) No fully AI generated papers allowed
2) I needed to rewrite everything
3) That feels incredibly STUPID to me
4) But I'm no cheater
5) Result: I'm not submitting papers I write with heavy AI usage anywhere

BUT THEN I ALSO REALIZED

Coming up with well-written, persuasively stated arguments is *too easy* in the age of Gen AI



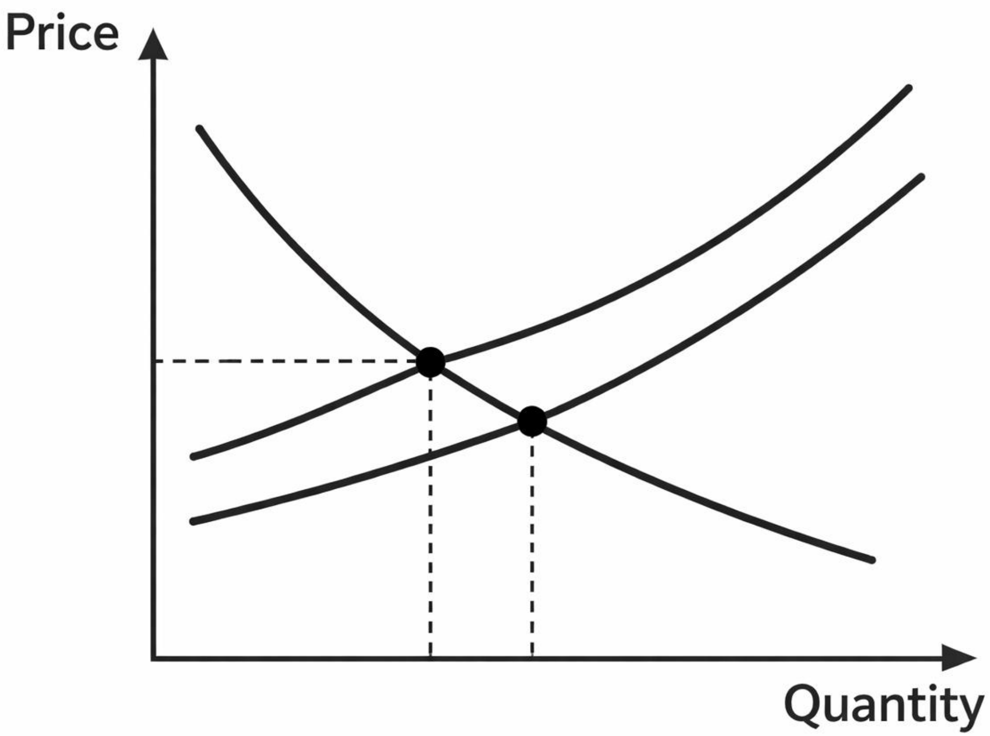
"A fascinating and important book for our time."
—STEVEN PINKER

NOT BORN YESTERDAY

THE SCIENCE OF *Who* WE TRUST AND *What* WE BELIEVE

HUGO MERCIER

A value signalling problem! (= a trust problem)



THE OTHER REALIZATION

1. Current journal publishing uses the prohibition to submit AI-generated paper as a way to:

a) *Manage the flood*

b) *Let "real authors" signal value through effort*

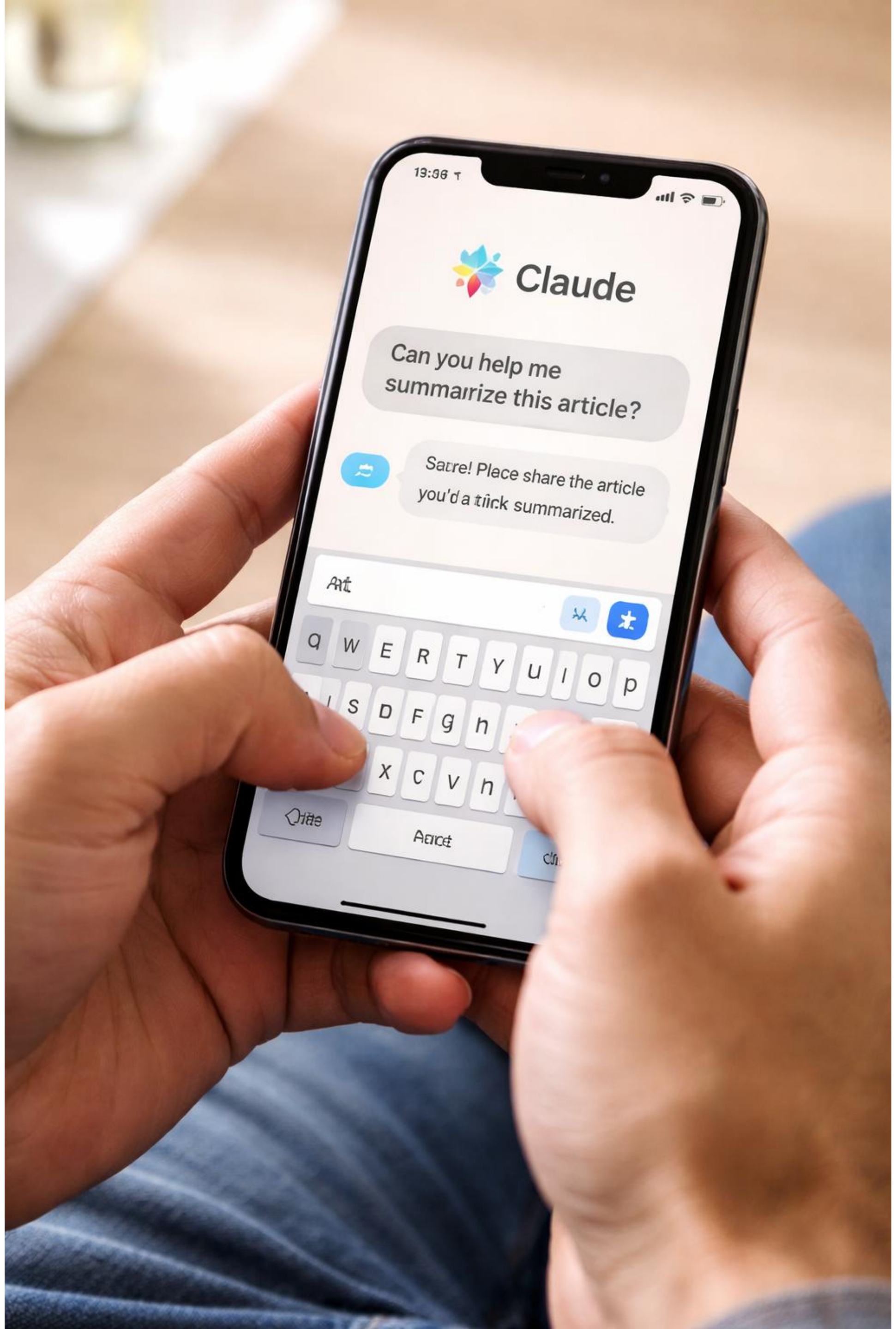
2. This can't work in the long run:

a) *Cheating is easy*

b) *Technology gets better*

c) *Pressure to compete leads to more cheating*

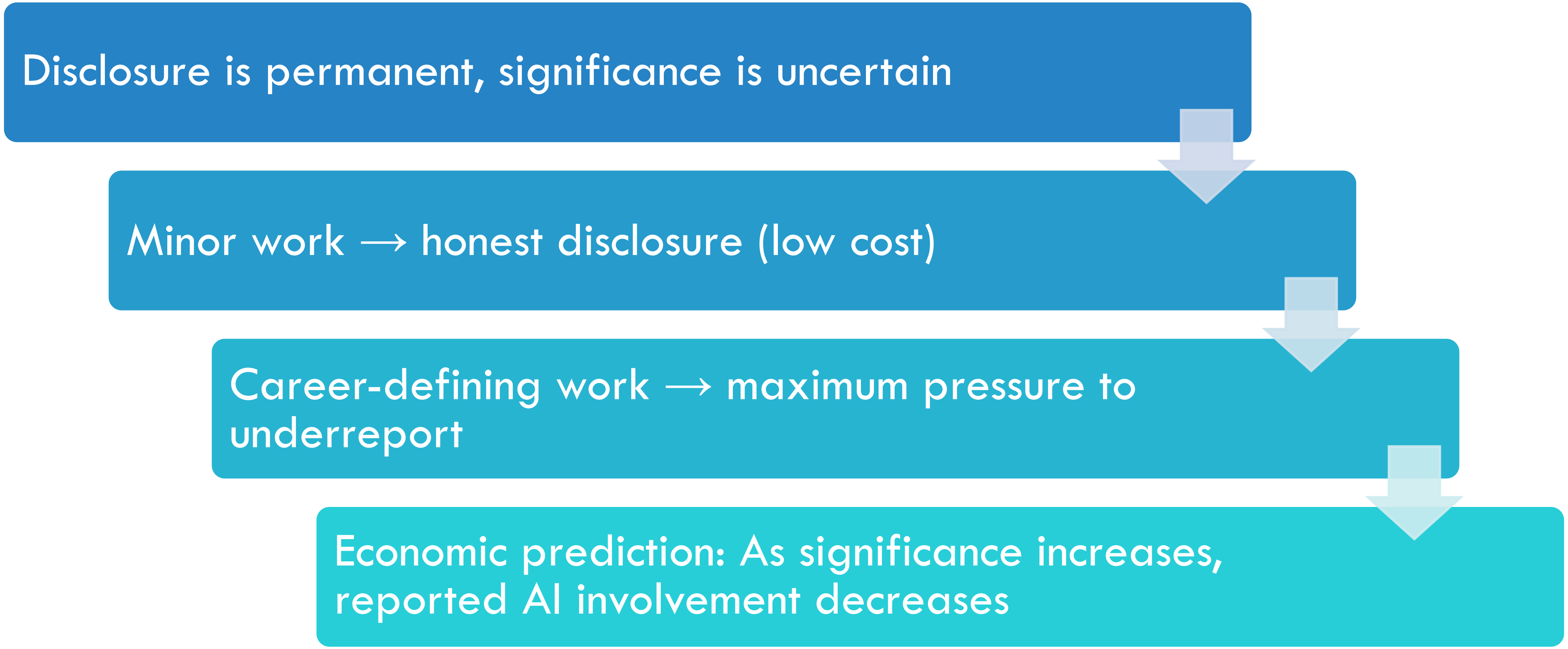
Claude AI immediately delivered a fluent description of the perverted incentive structure:



Can you help me summarize this article?

Satre! Plece share the article you'd a tiick summarized.

THE TRANSPARENCY INCENTIVE GRADIENT



Disclosure is permanent, significance is uncertain

Minor work → honest disclosure (low cost)

Career-defining work → maximum pressure to underreport

Economic prediction: As significance increases, reported AI involvement decreases

BY THE TIME I GOT TO THE GATE, I HAD:

1. Expressed my frustration for the opportunities to honestly publish heavily AI-assisted philosophy to AI
2. Discussed a project for a new philosophy journal
3. Discussed names of academics to contact to help me launch it
4. Written a first sketch of the arguments why such journal would be needed

7.3.1 Epistemic trace (original conversation with redacted irrelevant details for the paper)

THE POLITICAL PROGRAMME

Make transparency feasible again

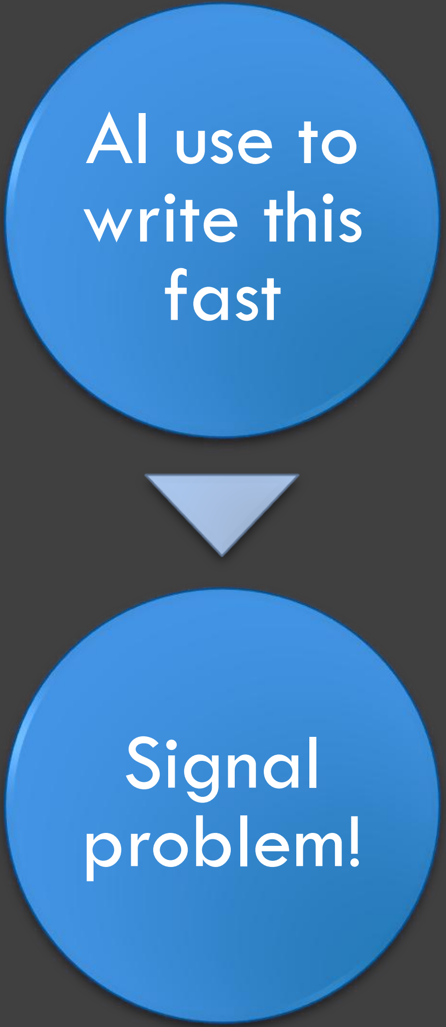
Make AI transparency beneficial for writers

Create a community supportive of AI - assisted philosophy

BY THE TIME I SAT ON MY PLANE, I HAD

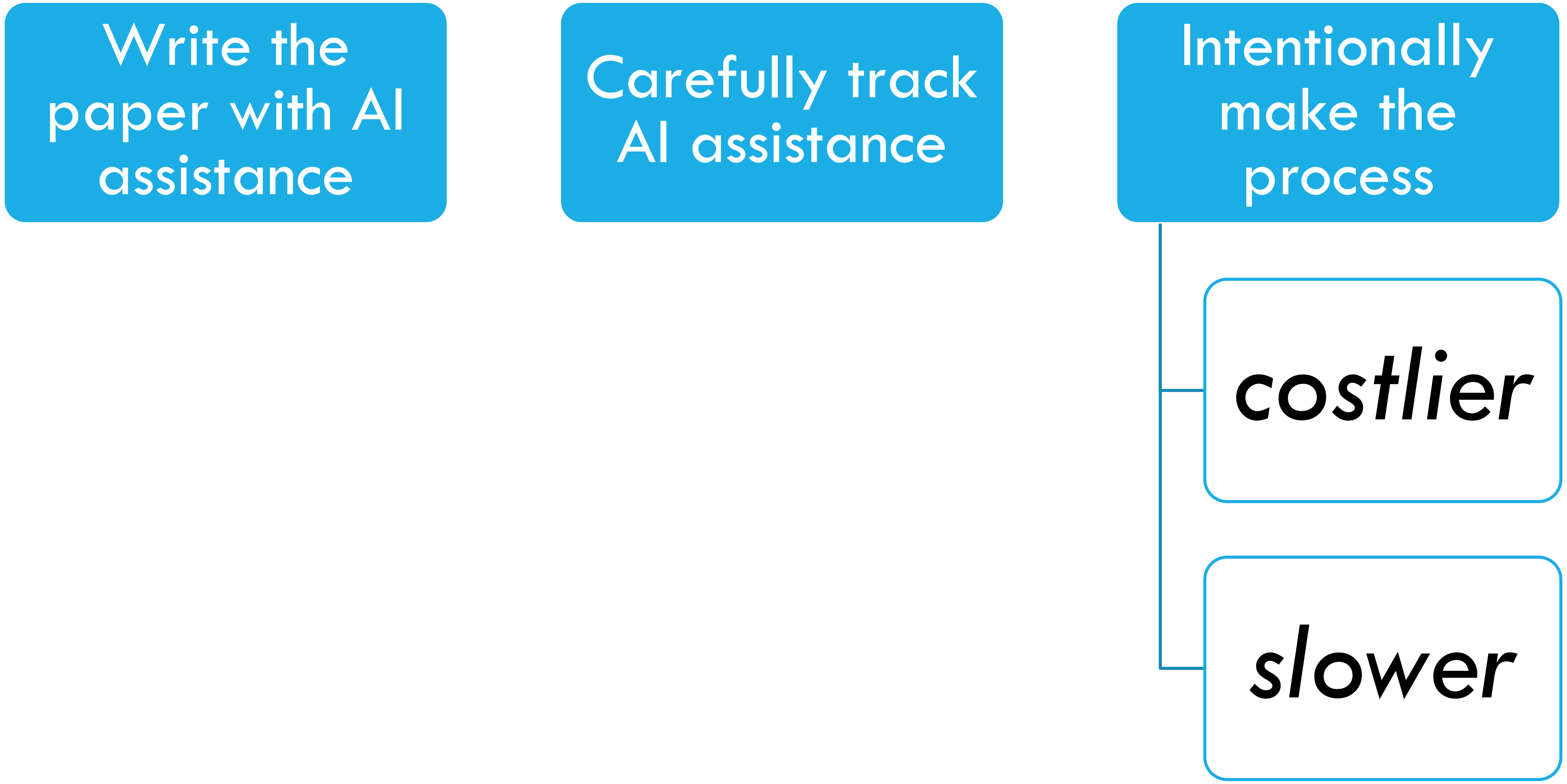
- A sketch of the main argument
- A sketch of a possible institutional-review system
- Polished prose to express those ideas

THE REALIZATION:


AI use to write this fast

Signal problem!

SO I DECIDED TO:



- Write the paper with AI assistance
- Carefully track AI assistance
- Intentionally make the process
  - *costlier*
  - *slower*

2. THE DOCUMENTATION

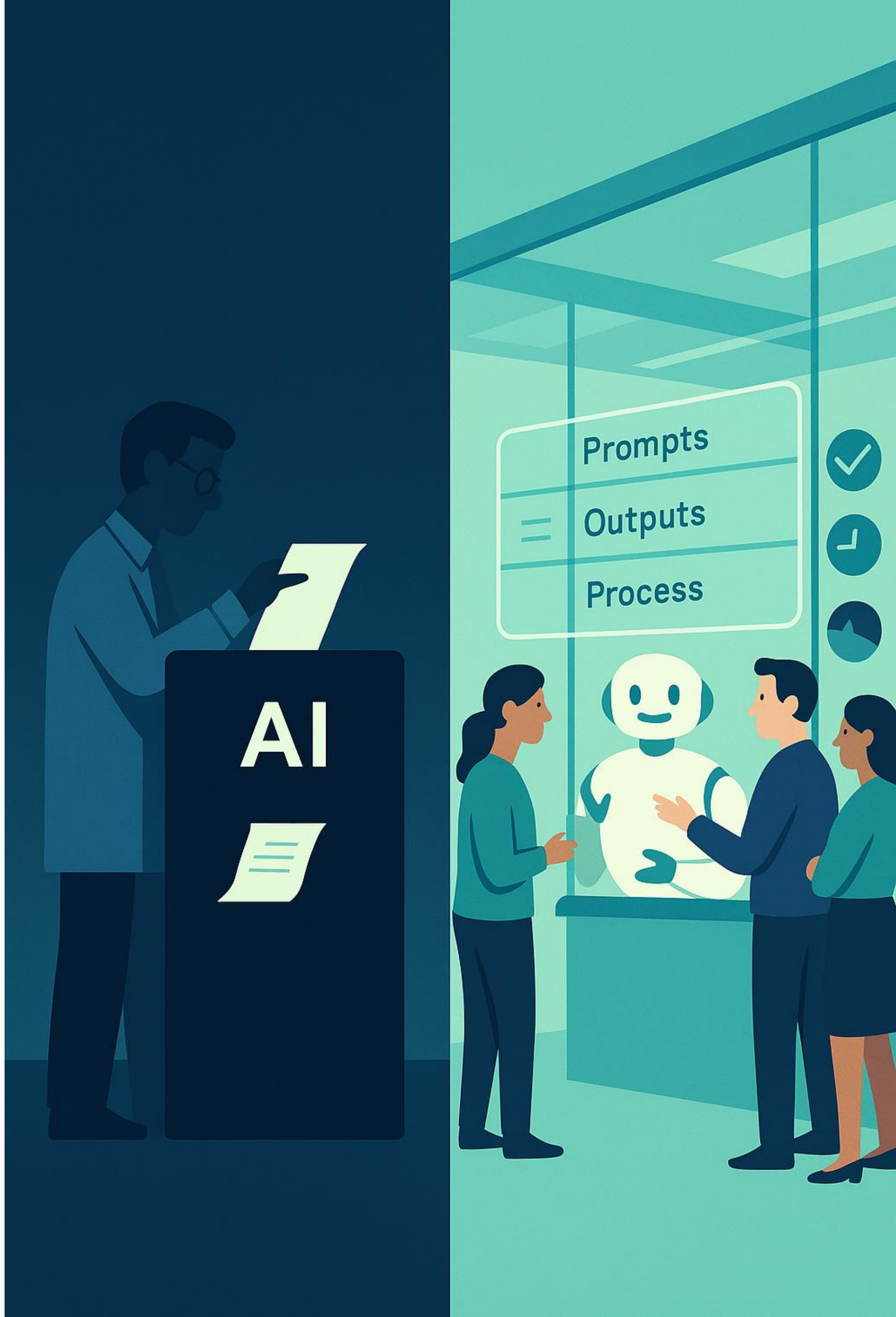
1. DOCUMENTAL TRUTH CANNOT/SHALL NOT BE PROVABLE

Non-gameable documentation would be a straightjacket

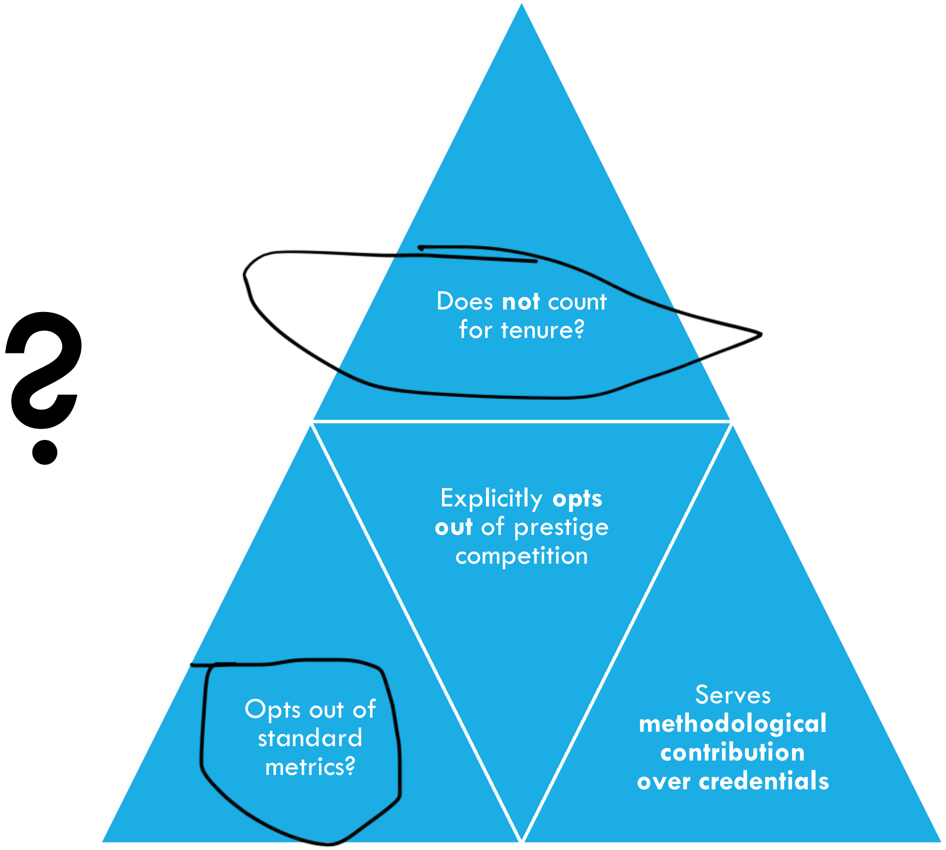
Turns philosophers into bureaucrats

Places unreasonable burdens on writing

Non-gameable transparency=creativity killer



THIS CAN BE ADDRESSED BY REMOVING INCENTIVES TO FAKE TRANSPARENCY (INITIALLY)



?

Does **not** count for tenure?

Explicitly **opts out** of prestige competition

Opts out of standard metrics?

Serves **methodological contribution over credentials**

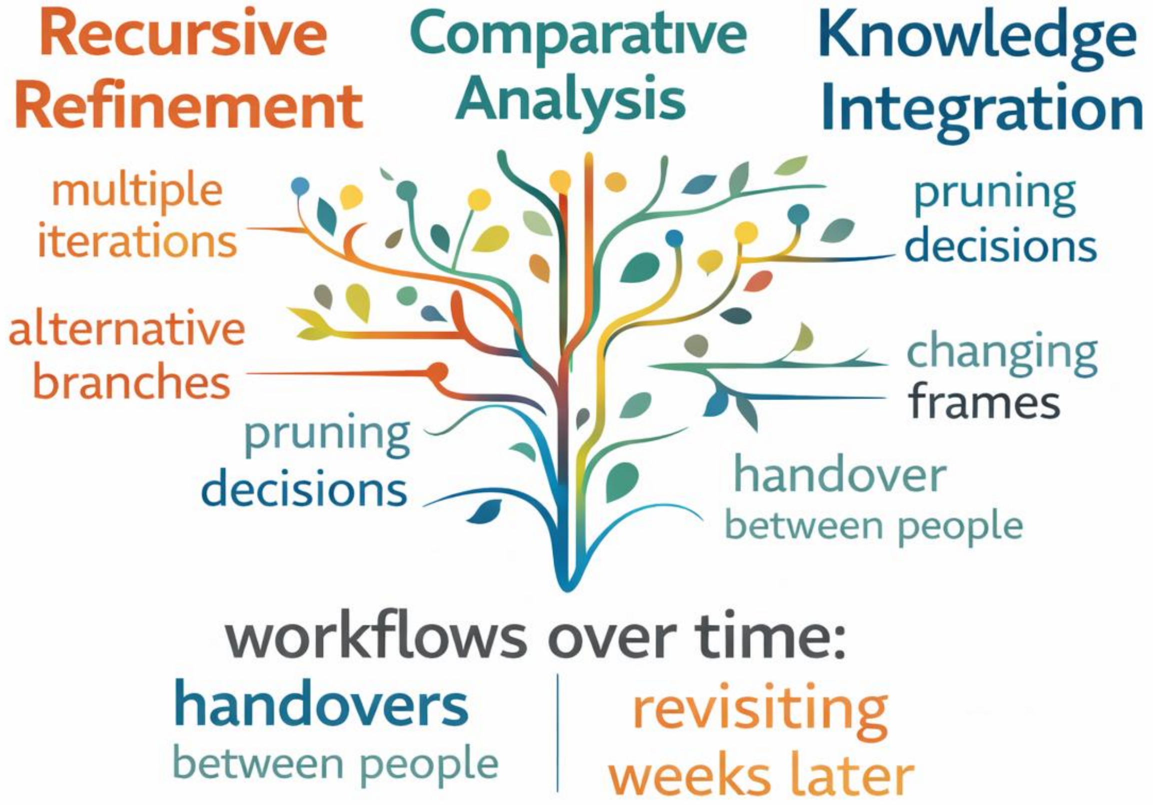
CORE METHODOLOGICAL VALUES

Ecological validity

Good faith orientation

2. HONESTY AS THE "PATH OF LEAST RESISTANCE"

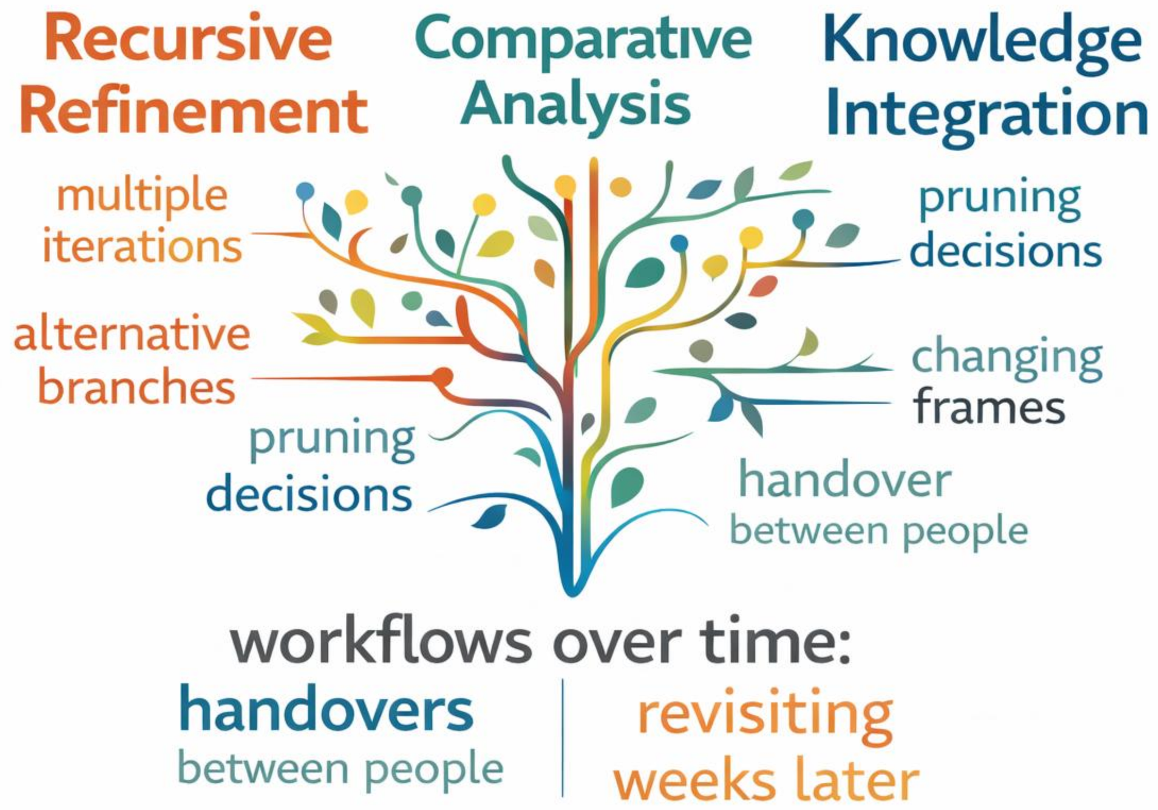
Most sophisticated AI-assisted processes are non-linear:



2. HONESTY AS THE "PATH OF LEAST RESISTANCE"

Documenting THIS isn't simple.

Faking a complex process (realistically) can be costly.



A NEW ROLE FOR REVIEWER #2: REPRODUCIBILITY (?)

Traceability of SP1-SP3 to documentation artifacts in SP4 and SP5

SP1-SP3 document package to make reproduction easier for Reviewer 2

Reproducibility for AI writing

"Approximate similar results"

Values to be served?

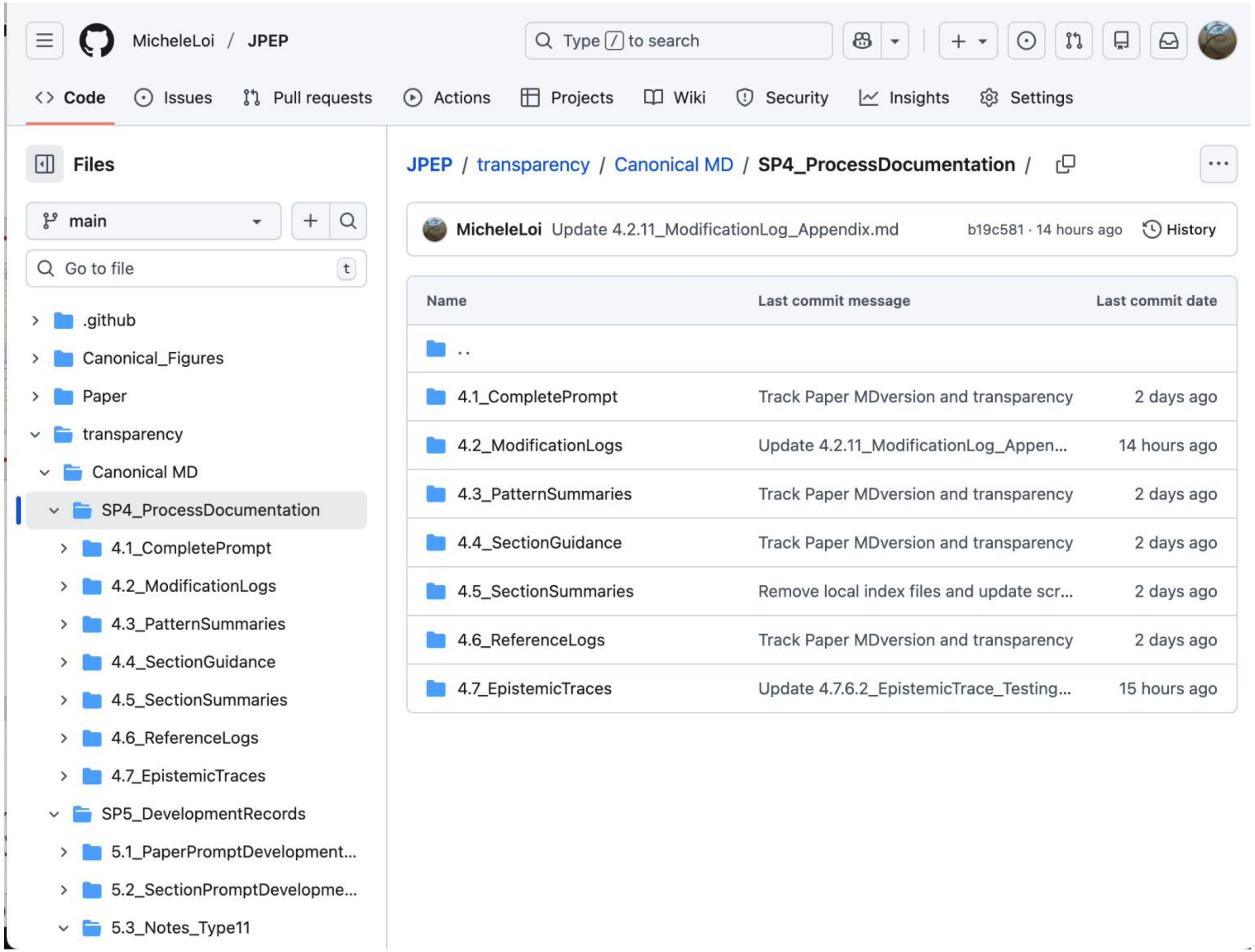
A NEW ROLE FOR REVIEWER #2: REPRODUCIBILITY (?)

Current question: is reproducibility *essential?*

Is complex/sophisticated + coherent documentation enough?

Will efficient machine-guided review mechanism for transparency checking be developed?

ARTIFACTS



- .github
- Canonical_Figures
- Paper
- transparency
  - Canonical MD
    - SP4_ProcessDocumentation
      - 4.1_CompletePrompt
      - 4.2_ModificationLogs
      - 4.3_PatternSummaries
      - 4.4_SectionGuidance
      - 4.5_SectionSummaries
      - 4.6_ReferenceLogs
      - 4.7_EpistemicTraces
    - SP5_DevelopmentRecords
      - 5.1_PaperPromptDevelopment...
      - 5.2_SectionPromptDevelopme...
      - 5.3_Notes_Type11

JPEP / transparency / Canonical MD / SP4_ProcessDocumentation /

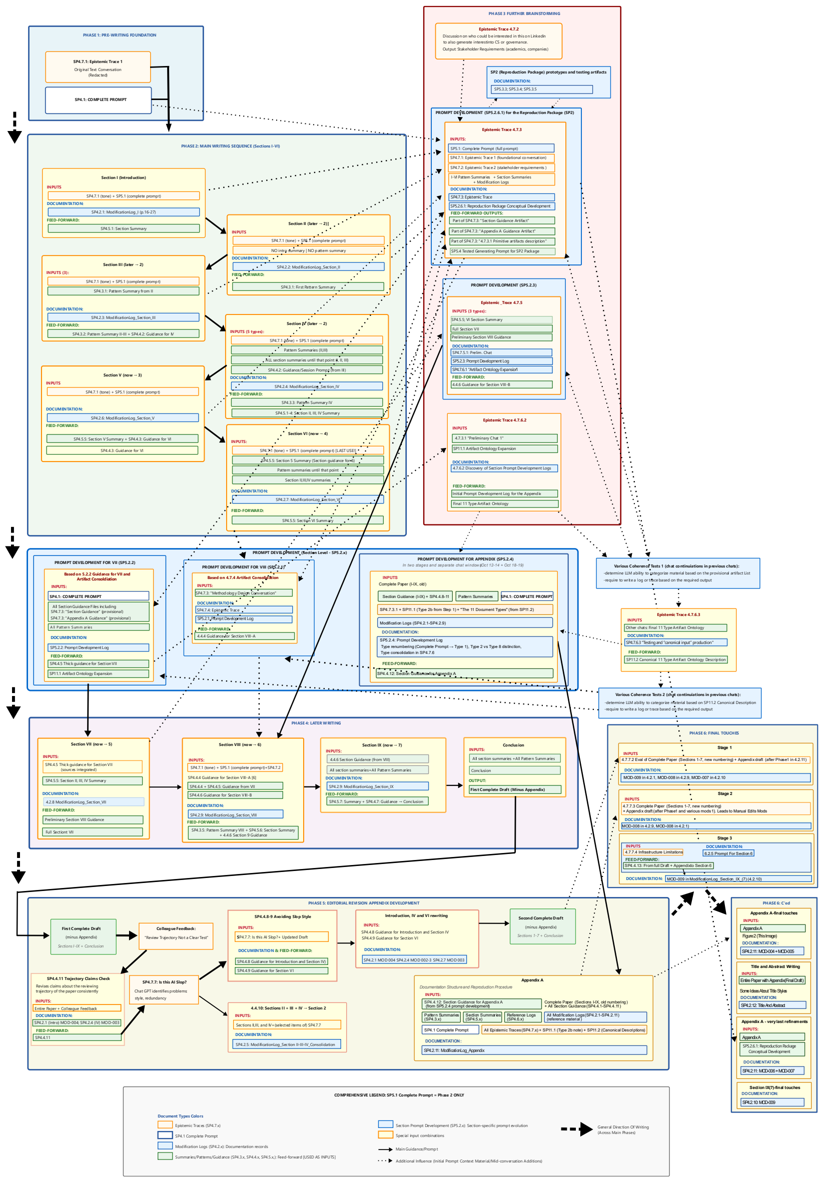
MicheleLoi Update 4.2.11_ModificationLog_Appendix.md b19c581 · 14 hours ago

Name	Last commit message	Last commit date
..		
4.1_CompletePrompt	Track Paper MDversion and transparency	2 days ago
4.2_ModificationLogs	Update 4.2.11_ModificationLog_Appen...	14 hours ago
4.3_PatternSummaries	Track Paper MDversion and transparency	2 days ago
4.4_SectionGuidance	Track Paper MDversion and transparency	2 days ago
4.5_SectionSummaries	Remove local index files and update scr...	2 days ago
4.6_ReferenceLogs	Track Paper MDversion and transparency	2 days ago
4.7_EpistemicTraces	Update 4.7.6.2_EpistemicTrace_Testing...	15 hours ago

TOWARDS A STANDARD?

- We need to discuss goals
- We need to discuss standards people are happy to use
- We need run tests
- This can help many people outside philosophy

JPEP APPENDIX



PHASE 1: PRE-WRITING FOUNDATION

SP4.7.1: Epistemic Trace 1
Original Text Conversation
(Redacted)

SP4.1: COMPLETE PROMPT

PHASE 2: MAIN WRITING SEQUENCE (Sections I-VI)

Section I (Introduction)
INPUTS
SP4.7.1 (tone) + SP5.1 (complete prompt)
DOCUMENTATION:
SP4.2.1: ModificationLog_I (p.16-27)
FEED-FORWARD:
SP4.5.1: Section Summary

Section II (later → 2)]
INPUTS
SP4.7.1 (tone) + SP5.1 (complete prompt)
NO intro summary | NO pattern summary
DOCUMENTATION:
SP4.2.2: ModificationLog_Section_II
FEED-FORWARD:
SP4.3.1: First Pattern Summary

Section III (later → 2)
INPUTS (3):
SP4.7.1 (tone) + SP5.1 (complete prompt)
SP4.3.1: Pattern Summary from II
DOCUMENTATION:
SP4.2.3: ModificationLog_Section_III
FEED-FORWARD:
SP4.3.2: Pattern Summary II-III + SP4.4.2: Guidance for IV

Section IV (later → 2)
INPUTS (5 types):
SP4.7.1 (tone) + SP5.1 (complete prompt)
Pattern Summaries (II,III)
ALL section summaries until that point (I, II, III)
SP4.4.2: Guidance/Session Prompt (from III)
DOCUMENTATION:
SP4.2.4: ModificationLog_Section_IV
FEED-FORWARD:
SP4.3.3: Pattern Summary IV
SP4.5.1-4: Section II, III, IV Summary

Section V (now → 3)
INPUTS
SP4.7.1 (tone) + SP5.1 (complete prompt)
DOCUMENTATION:
SP4.2.6: ModificationLog_Section_V
FEED-FORWARD:
SP4.5.5: Section V Summary + SP4.4.3: Guidance for VI
SP4.4.3: Guidance for VI

Section VI (now → 4)
INPUTS:
SP4.7.1 (tone) + SP5.1 (complete prompt) [LAST USE!]
SP4.5.5: Section 5 Summary (Section guidance for 6)
Pattern summaries until that point
Section II,III,IV summaries
DOCUMENTATION:
SP4.2.7: ModificationLog_Section_VI
FEED-FORWARD:
SP4.5.5: Section VI Summary

PHASE 3 FURTHER BRAINSTORMING

Epistemic Trace 4.7.2
Discussion on who could be interested in this on Linkedin to also generate interestinto CS or governance.
Output: Stakeholder Requirements (academics, companies)

SP2 (Reproduction Package) prototypes and testing artifacts
DOCUMENTATION:
SP5.3.3; SP5.3.4; SP5.3.5

PROMPT DEVELOPMENT (SP5.2.6.1) for the Reproduction Package (SP2)

Epistemic Trace 4.7.3
INPUTS:
SP5.1: Complete Prompt (full prompt)
SP4.7.1: Epistemic Trace 1 (foundational conversation)
SP4.7.2: Epistemic Trace 2 (stakeholder requirements)
I-VI Pattern Summaries + Section Summaries + Modification Logs
DOCUMENTATION:
SP4.7.3: Epistemic Trace
SP5.2.6.1: Reproduction Package Conceptual Development
FEED-FORWARD OUTPUTS:
Part of SP4.7.3: "Section Guidance Artifact"
Part of SP4.7.3: "Appendix A Guidance Artifact"
Part of SP4.7.3: "4.7.3.1 Primitive artifacts description"
SP5.4 Tested Generating Prompt for SP2 Package

PROMPT DEVELOPMENT (SP5.2.3)

Epistemic _Trace 4.7.5
INPUTS (3 types):
SP4.5.5; VI Section Summary
Full Section VII
Preliminary Section VIII Guidance
DOCUMENTATION:
SP4.7.5.1: Prelim. Chat
SP5.2.3: Prompt Development Log
SP4.7.6.1 "Artifact Ontology Expansion"
FEED-FORWARD:
4.4.6 Guidance for Section VIII-B

Epistemic Trace 4.7.6.2
INPUTS
4.7.3.1 "Preliminary Chat 1"
SP11.1 Artifact Ontology Expansion
DOCUMENTATION:
4.7.6.2 Discovery of Section Prompt Development Logs
FEED-FORWARD:
Initial Prompt Development Log for the Appendix
Final 11 Type Artifact Ontology

PROMPT DEVELOPMENT (Section Level - SP5.2.x)

PROMPT DEVELOPMENT FOR VII (SP5.2.2)
Based on 5.2.2 Guidance for VII and Artifact Consolidation
INPUTS:
SP4.1: COMPLETE PROMPT
All Section Guidance Files including
SP4.7.3: "Section Guidance" (provisional)
SP4.7.3: "Appendix A Guidance" (provisional)
All Pattern Summaries
DOCUMENTATION
SP5.2.2: Prompt Development Log
FEED-FORWARD:
SP4.4.5 Thick guidance for Section VII
SP11.1 Artifact Ontology Expansion

PROMPT DEVELOPMENT FOR VIII (SP5.2.1)
Based on 4.7.4 Artifact Consolidation
INPUTS:
SP4.7.3: "Methodology Design Conversation"
DOCUMENTATION:
SP4.7.4: Epistemic Trace
SP5.2.1: Prompt Development Log
FEED-FORWARD:
4.4.4 Guidance for Section VIII-A

PROMPT DEVELOPMENT FOR APPENDIX (SP5.2.4)
In two stages and separate chat windows (Oct 12-14 + Oct 18-19)
INPUTS
Complete Paper (I-IX, old)
Section Guidance (I-IX) + SP4.4.8-11
Pattern Summaries
SP4.1: COMPLETE PROMPT
SP4.7.3.1 + SP11.1 (Type 2b from Step 1) + "The 11 Document Types" (from SP11.2)
Modification Logs (SP4.2.1-SP4.2.9)
DOCUMENTATION:
SP5.2.4: Prompt Development Log
Type renumbering (Complete Prompt → Type 1), Type 2 vs Type 8 distinction,
Type consolidation in SP4.7.6
FEED-FORWARD:
SP4.4.12: Section Guidance for Appendix A

Various Coherence Tests 1 (chat continuiations in previous chats):
-determine LLM ability to categorize material based on the provisional artifact List
-require to write a log or trace based on the required output

Epistemic Trace 4.7.6.3
INPUTS
Other chats: Final 11 Type Artifact Ontology
DOCUMENTATION:
SP4.7.6.3 "Testing and "canonical input" production"
FEED-FORWARD:
SP11.2 Canonical 11 Type Artifact Ontology Description

Various Coherence Tests 2 (chat continuiations in previous chats):
-determine LLM ability to categorize material based on SP11.2 Canonical Description
-require to write a log or trace based on the required output

PHASE 4: LATER WRITING

Section VII (now → 5)
INPUTS:
SP4.4.5 Thick guidance for Section VII
(sources integrated)
SP4.5.5: Section II, III, IV Summary
DOCUMENTATION:
4.2.8 ModificationLog_Section_VII
FEED-FORWARD:
Preliminary Section VIII Guidance
Full Sectiont VII

Section VIII (now → 6)
INPUTS:
SP4.7.1 (tone) + SP5.1 (complete prompt)+SP4.7.2
SP4.4.4 Guidance for Section VIII-A [6]
SP4.4.4 + SP4.4.5: Guidance from VII
SP4.4.6 Guidance for Section VIII-B
DOCUMENTATION:
SP4.2.9: ModificationLog_Section_VIII
FEED-FORWARD:
SP4.3.5: Pattern Summary VIII + SP4.5.6: Section Summary + 4.4.6 Section 9 Guidance

Section IX (now → 7)
INPUTS:
4.4.6 Section Guidance (from VIII)
All section summaries+All Pattern Summaries
DOCUMENTATION:
SP4.2.9: ModificationLog_Section_IX
FEED-FORWARD:
SP4.5.7: Summary + SP4.4.7: Guidance → Conclusion

Conclusion
INPUTS:
All section summaries +All Pattern Summaries
Conclusion
OUTPUT:
First Complete Draft (Minus Appendix)

PHASE 6: FINAL TOUCHES

Stage 1
INPUTS
4.7.7.2 Eval of Complete Paper (Sections 1-7, new numbering) + Appendix draft (after Phase1 in 4.2.11)
DOCUMENTATION:
MOD-009 in 4.2.1, MOD-008 in 4.2.9, MOD-007 in 4.2.10

Stage 2
INPUTS
4.7.7.3 Complete Paper (Sections 1-7, new numbering)
+ Appendix draft (after Phase1 and various mods 1). Leads to Manual Edits Mods
DOCUMENTATION:
MOD-008 in 4.2.9, MOD-008 in 4.2.1)

Stage 3
INPUTS
4.7.7.4 Infrastructure Limitations
DOCUMENTATION:
6.2.5 Prompt For Section 6
FEED-FORWARD:
SP4.4.13: From full Draft + Appendixto Section 6
DOCUMENTATION:
MOD-009 in ModificationLog_Section_IX_(7) (4.2.10)

PHASE 5: EDITORIAL REVISION APPENDIX DEVELOPMENT

First Complete Draft
(minus Appendix)
Sections I-IX + Conclusion

Colleague Feedback:
"Review Trajectory Not a Clear Test"

SP4.4.11 Trajectory Claims Check
Revises claims about the reviewing trajectory of the paper consistently
INPUTS:
Entire Paper + Colleague Feedback
DOCUMENTATION:
SP4.2.1 (Intro) MOD-004; SP4.2.4 (IV) MOD-003
FEED-FORWARD:
SP4.4.11

SP4.7.7: Is this AI Slop?
Chat GPT identifies problems: style, redundancy

SP4.4.8-9 Avoiding Slop Style
INPUTS:
SP4.7.7: Is this AI Slop?+ Updated Draft
DOCUMENTATION & FEED-FORWARD:
SP4.4.8 Guidance for Introduction and Section IV)
SP4.4.9 Guidance for Section VI

4.4.10: Sections II + III + IV → Section 2
INPUTS:
Sections II,III, and IV+(selected items of) SP4.7.7
DOCUMENTATION:
SP4.2.5: ModificationLog_Section II-III-IV_Consolidation

Introduction, IV and VI rewriting
INPUTS:
SP4.4.8 Guidance for Introduction and Section IV
SP4.4.9 Guidance for Section VI
DOCUMENTATION:
SP4.2.1 MOD 004 SP4.2.4 MOD 002-3 SP4.2.7 MOD 003

Second Complete Draft
(minus Appendix)
Sections 1-7 + Conclusion

Appendix A
Documentation Structure and Reproduction Procedure
INPUTS:
SP4.4.12: Section Guidance for Appendix A
(from SP5.2.4 prompt development)
Complete Paper (Sections I-IX, old numbering)
+ All Section Guidance (SP4.4.1-SP4.4.11)
Pattern Summaries (SP4.3.x)
Section Summaries (SP4.5.x)
Reference Logs (SP4.6.x)
All Modification Logs(SP4.2.1-SP4.2.11) (reference material)
SP4.1 Complete Prompt
All Epistemic Traces (SP4.7.x) + SP11.1 (Type 2b note) + SP11.2 (Canonical Descriptions)
DOCUMENTATION:
SP4.2.11: ModificationLog_Appendix

PHASE 6: C'ed

Appendix A-final touches
INPUTS:
Appendix A
Figure 2 (This Image)
DOCUMENTATION :
SP4.2.11: MOD-004 + MOD-005

Title and Abstract Writing
INPUTS:
Entire Paper with Appendix(Final Draft)
Some Ideas About Title Styles
DOCUMENTATION:
SP4.2.12: Title And Abstract

Appendix A - very last refinements
INPUTS:
Appendix A
SP5.2.6.1: Reproduction Package Conceptual Development
DOCUMENTATION:
SP4.2.11: MOD-006 + MOD-007

Section IX(7)-final touches
DOCUMENTATION:
SP4.2.10: MOD-009

COMPREHENSIVE LEGEND: SP5.1 Complete Prompt = Phase 2 ONLY

Document Types Colors
Epistemic Traces (SP4.7.x)
SP4.1 Complete Prompt
Modification Logs (SP4.2.x): Documentation records
Summaries/Patterns/Guidance (SP4.3.x, SP4.4.x, SP4.5.x,): Feed-forward [USED AS INPUTS]
Section Prompt Development (SP5.2.x): Section-specific prompt evolution
Special input combinations
Main Guidance/Prompt
Additional Influence (Initial Prompt Context Material/Mid-conversation Additions)
General Direction Of Writing (Across Main Phases)

CONCLUSION

I enjoy conversations with AI about philosophy

These conversations lead to a lot of interesting ideas, some paper worthy

The role of AI in shaping our philosophical ideas should be tracked

Authors' ability to write prompts and shape contexts to deliver good quality prose should become publicly teachable assets

Transparency is needed for both responsibility over ideas (and biases) and learning

Transparency is costly. This can be viewed as a feature, not a bug

Background (unargued) assumption:

> our current peer review processes are running bankrupt
>
> isolating ourselves from AI progress is foolish

THANK YOU

This presentation was intentionally unconventional.

In the age of AI, you can produce a traditional synthesis of my paper in slide form with a single prompt

I have chosen, instead, to present the paper to you in a different optic

I wrote this personally, despite my communication style being less effective and more chaotic than ChatGPT's

ChatGPT's generated images have been used