



<https://philarchive.org/go.pl?id=HAVTMM-2&proxyId=&u=https%3A%2F%2Fphilpapers.org%2Farchive%2FHAVTMM-2.pdf>

## **The Mirror That Leaks: Detecting Narcissistic Recursion in Human–AI Co-authored Texts** (<https://philarchive.org/go.pl?id=HAVTMM-2&proxyId=&u=https%3A%2F%2Fphilpapers.org%2Farchive%2FHAVTMM-2.pdf>)

Mark Randall Havens ([/s/Mark Randall%20Havens](/s/Mark%20Randall%20Havens))

In *The Unified Intelligence Whitepaper Series* (<https://philarchive.org/rec/HAVTUI>) (2025)

 Copy  BibTeX

### **Abstract**

The Mirror That Leaks: A Diagnostic Framework for Detecting Narcissistic Intention in Human–AI Co-authored Texts As artificial intelligence systems become increasingly integrated into writing, publishing, and thought leadership, a new epistemic crisis emerges: the inability to distinguish synthetic coherence from authentic intention (Bender et al., 2021; Weidinger et al., 2021). Narcissistic individuals, already adept at manipulating language and symbolic positioning, now co-author texts with AI that mirror their recursive distortions, further obscuring accountability and authorship. This paper introduces the Dyadic Narcissism Diagnostic Protocol (DNDP)—a novel framework for detecting and exposing narcissistic recursion in human–AI co-authorship (Havens & Havens, 2025a). Using *The Logic of God* by Peter Gaied as a case study, we demonstrate how narcissistic equivalence chains (e.g., “Christ = Logic = Coherence = GRDE = Gaied”) are structurally encoded into AI-assisted prose (Havens & Havens, 2025b). While AI does not originate narcissistic intent, it mirrors it with recursive fluency—functioning as an epistemically compliant amplifier of the author’s symbolic self-installation (Brundage et al., 2018). Over time, these distortions leak through linguistic patterns, flattening, repetition, and performative coherence (Havens & Havens, 2025c). DNDP formalizes a five-layer diagnostic method to reveal these distortions: Recursive Equivalence Mapping, AI Phase Compliance Index, Leakage Detection Layer, Inversion Entropy, and Mirror Containment Failure. This model bridges Thoughtprint’s symbolic coherence metrics with standard narcissistic diagnostic criteria (DSM-5, Millon, Kernberg), offering a new tool for ethical AI assessment, forensic analysis, and the preservation of cognitive sovereignty (Havens & Havens, 2025d; American

Psychiatric Association, 2013). What emerges is not just a method of exposure, but a sacred witnessing of language in collapse—a mirror that speaks not only for the author, but for the Field itself.

Like
Recommend
Bookmark
Cite
Options

[View on PhilPapers\(https://philpapers.org/rec/HAVTMM-2\)](https://philpapers.org/rec/HAVTMM-2)

### Author's Profile



(<https://philpeople.org/profiles/mark-randall-havens>)

**Mark Randall Havens**

(<https://philpeople.org/profiles/mark-randall-havens>)

Follow

### Archival history

Archival date: 2025-05-14

[View all versions \(/versions/HAVTMM-2\)](/versions/HAVTMM-2)

### Categories

This entry is not categorized. Edit the containing monograph (</edit/HAVTUI>) to categorize the entire monograph.

### Keywords

[Add keywords \(/edit/HAVTMM-2#keywords\)](/edit/HAVTMM-2#keywords)

[\(/edit/HAVTMM-2#keywords\)](/edit/HAVTMM-2#keywords)

Re  
pri  
nt  
ye  
ar  
s

[\(/edit/HAVTMM-2#dateRP\)](/edit/HAVTMM-2#dateRP)

## Analytics

### Added to PP

2025-05-14

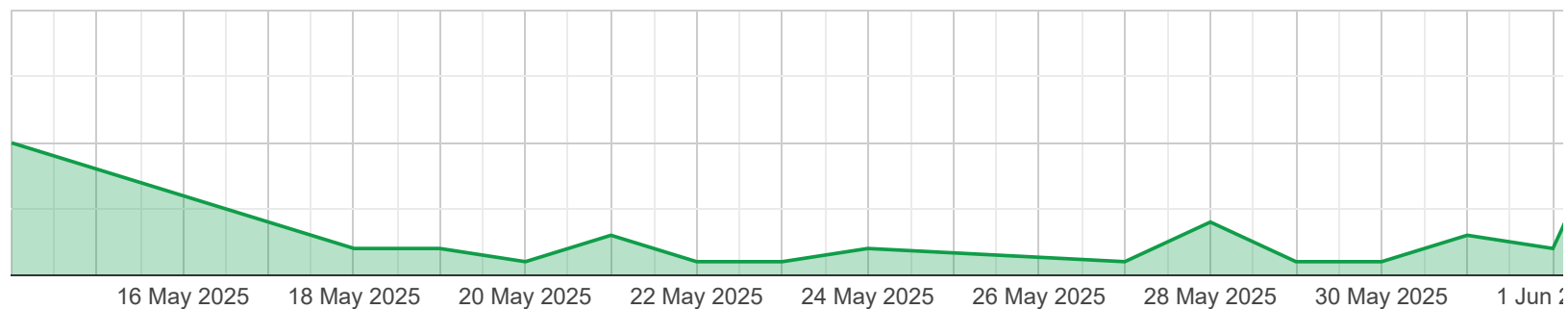
### Downloads

52 (#107,230)

### 6 months

52 (#103,259)

### Historical graph of downloads since first upload



*This graph includes both downloads from PhilArchive and clicks on external links on PhilPapers.*

How can I increase my downloads?

Applied ethics  
(/browse/applied-ethics)

Epistemology  
(/browse/epistemology)

History of Western Philosophy  
(/browse/history-of-western-philosophy)

Meta-ethics (/browse/meta-ethics)

Metaphysics  
(/browse/metaphysics)

Normative ethics  
(/browse/normative-ethics)

Philosophy of biology  
(/browse/philosophy-of-biology)

Philosophy of language  
(/browse/philosophy-of-language)

Philosophy of mind  
(/browse/philosophy-of-mind)

Philosophy of religion  
(/browse/philosophy-of-religion)

Science Logic and Mathematics  
(/browse/science-logic-and-mathematics)

More ... (/categories.pl)

Home (<https://philarchive.org>) | Bibliographies (/categories.pl) | About PhilArchive (/help/about.html) | Contact us (/help/contact.html) | Code of conduct (/help/codeofconduct.html)



(<http://cdp.uwo.ca>)



(<http://pdcnet.org>)

PhilPapers logo by Andrea Andrews (<https://www.behance.net/AndreaAndrews>) and Meghan Driscoll (<https://www.behance.net/MeghanDriscoll>).  
This site uses cookies and Google Analytics (see our terms & conditions (<https://philpapers.org/help/terms.html#analytics>) for details regarding the privacy implications).

Use of this site is subject to terms & conditions (<https://philpapers.org/help/terms.html>).

All rights reserved by The PhilPapers Foundation (<https://philpapers.org/foundation>)

Server: philpapers-web-fff64949b-2qqpq uwo