Independent Submission
Internet-Draft
Intended status: Informational
Expires: 1 November 2025

D. Abaris
Individual Contributor

30 April 2025

AI Content Disclosure Header
draft-abaris-aicdh-00

Abstract

This document proposes a machine-readable Hypertext Transfer Protocol (HTTP) response header field, AI-Disclosure, to disclose the presence and degree of Artificial Intelligence (AI) generated or AI-assisted content in web responses. The header is designed for compatibility with HTTP structured field syntax and provides metadata for user agents, bots, and archiving systems. It supports layered disclosure strategies alongside human-readable and structured metadata formats.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at https://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on 1 November 2025.

Table of Contents

1. Introduction

As AI-generated content proliferates across the web, users [BV-Report], platforms, and regulators increasingly demand transparent disclosure of algorithmic involvement in content creation [PAI-Framework]. Existing approaches to disclosure (e.g., HTML disclaimers) lack machine-readability [GV-Prov], making automation, indexing, and compliance challenging.

This document defines the AI-Disclosure HTTP header field, providing a lightweight, machine-readable mechanism focused specifically on signaling the presence and mode of AI involvement in the generation of an HTTP response's content. It utilizes HTTP Structured Fields [RFC9651] to offer a simple dictionary format directly within the HTTP response headers.

The goal of AI-Disclosure is to offer a low-overhead, easily parsable signal primarily for automated systems like web crawlers, archiving tools, or user agents that may need a quick indication of AI usage without processing complex manifests. This header is intended to be applied at the entire response level.

It is important to distinguish this mechanism from more comprehensive content provenance and authenticity frameworks like the Coalition for Content Provenance and Authenticity (C2PA) specification [C2PA-Spec]. C2PA provides richer, cryptographically signed assertions about content provenance, potentially covering detailed creation/modification history and applying to specific regions within an asset ("Regions of Interest"). C2PA information can be linked via methods including the HTTP Link header [RFC8288] pointing to an associated manifest.

AI-Disclosure can be seen as complementary to such systems within a layered disclosure strategy. While C2PA offers strong, verifiable, and granular provenance, AI-Disclosure provides a simpler, advisory signal directly in the HTTP interaction for basic AI involvement awareness. Systems requiring high assurance or sub-resource granularity should utilize frameworks like C2PA.

2. Terminology

2.1. Acronyms and Abbreviations

AI: Artificial Intelligence

HTTP: Hypertext Transfer Protocol

C2PA: Content Provenance and Authenticity, refers to the specification developed by the Coalition for Content Provenance and Authenticity

3. Conventions and Definitions

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

4. Field Definition

The AI-Disclosure field is defined as a Structured Field of type Dictionary, as described in [RFC9651].

Header Field Name	Structured Type	Template	Protocol	Status	Reference
AI-Disclosure	Dictionary	(blank)	http	provisional	[this document]

Table 1

5. Field Syntax

The AI-Disclosure field value MUST conform to the syntax for Dictionary structures defined in Section 3.2 of [RFC9651]. Each key in the dictionary conveys a distinct aspect of AI disclosure.

5.1. Example

```
AI-Disclosure: mode=ai-originated;
               model="gpt-4";
               provider="OpenAI";
               reviewed-by="editorial-team";
               date=@1745286896
```

5.2. Field Keys

Key	Type	Description
mode	Token	Indicates the nature of AI involvement: none, ai-modified, ai-originated, machine-generated
model	String	Identifier of the AI model used (e.g., gpt-4)
provider	String	Organization providing the AI system
reviewed-by	String	Entity or team who reviewed the AI content
date	Date	Generation timestamp as a numeric epoch value, conforming to RFC9651.

Table 2

6. Semantics

The AI-Disclosure header field is an optional and advisory header providing information about the use of AI in generating the response content. Its presence indicates voluntary disclosure by the server. Absence of the header implies nothing about AI usage.

The meaning of the header is primarily defined by the mode key, whose possible values are described below:

Mode Value	Description
none	Indicates that AI was not used in the creation or substantive modification of the content.
ai-modified	Indicates AI was used to assist with or modify content primarily created by humans. The source material was not AI-generated. Examples include AI-based grammar checking, style suggestions, or generating highlights or summaries of human-written text.
ai-originated	Indicates the core content was initially generated by AI but subsequently reviewed, edited, or significantly guided by humans. This suggests human oversight for accuracy or appropriateness, even if the originality for copyright purposes might be affected.
machine-generated	Indicates the content was primarily or entirely generated by AI with minimal or no human intervention or review post-generation. AI may be responsible for substantive assertions or conclusions.

Table 3

Other keys like model, provider, reviewed-by, and date provide optional, additional context about the AI model used, the provider, human review, or the generation time, respectively.

Recipients should treat this header as informational only and refer to the Security Considerations (Section 7) regarding its trustworthiness.

Note: The AI-Disclosure header applies to the entire content of the HTTP response payload. The ai-modified and ai-originated values indicate AI involvement, but this header does not provide information about specific locations or the exact nature of the partial involvement. For expressing provenance information about specific parts of a resource, more comprehensive mechanisms such as C2PA [C2PA-Spec] should be used. The distinction between ai-modified and ai-originated aims to address whether the foundational content was human or AI, reflecting concerns about originality and the nature of the transformation. The distinction between ai-originated and machine-generated primarily reflects the level of human review or intervention post-generation.

7. Security Considerations

The AI-Disclosure field is intended to provide advisory metadata about AI-generated or AI-assisted content and does not include any form of integrity protection. As such, the field can be trivially spoofed or altered by intermediaries unless the response is delivered over a secure transport such as HTTPS.

Clients and intermediaries MUST NOT rely on the presence, absence, or value of the AI-Disclosure field for making security-critical decisions. The field is not authenticated and SHOULD be treated as untrusted input.

This document does not define any mechanisms for cryptographic verification or provenance validation of the header’s content. Implementations that require trustworthy disclosure metadata SHOULD rely on additional application-layer integrity mechanisms or signed metadata systems, such as those defined by C2PA [C2PA-Spec].

8. IANA Considerations

IANA is requested to register the AI-Disclosure header field in the "Hypertext Transfer Protocol (HTTP) Field Name Registry" maintained at https://www.iana.org/assignments/http-fields/ (https://www.iana.org/assignments/http-fields/), according to the procedures outlined in [RFC9110].

Header Field Name	Applicable Protocol	Status	Reference	Structured Type	Notes
AI-Disclosure	http	provisional	[this document]	Dictionary	Discloses AI involvement in content creation

Table 4

9. References

9.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <https://www.rfc-editor.org/rfc/rfc2119>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/rfc/rfc8174>.

[RFC8288] Nottingham, M., "Web Linking", RFC 8288, DOI 10.17487/RFC8288, October 2017, <https://www.rfc-editor.org/rfc/rfc8288>.

[RFC9110] Fielding, R., Ed., Nottingham, M., Ed., and J. Reschke, Ed., "HTTP Semantics", STD 97, RFC 9110, DOI 10.17487/RFC9110, June 2022, <https://www.rfc-editor.org/rfc/rfc9110>.

[RFC9651] Nottingham, M. and P. Kamp, "Structured Field Values for HTTP", RFC 9651, DOI 10.17487/RFC9651, September 2024, <https://www.rfc-editor.org/rfc/rfc9651>.

9.2. Informative References

[BV-Report] Big Valley Marketing, "AI Disclosure and Transparency: Closing the Trust Gap", November 2024, <https://bigvalley.co/wp-content/uploads/2024/11/BV-AI-Research-Report.pdf>.

[C2PA-Spec]
Coalition for Content Provenance and Authenticity (C2PA), "C2PA Specification Version 2.1", July 2024, <https://c2pa.org/specifications/specifications/2.1/specs/C2PA_Specification.html>.

[GV-Prov] Hofmann, S., "Content Provenance and Disclosure Requirements for AI Generated Content on Digital and Traditional Media Platforms", 31 March 2025, <https://www.globalvoices.org.au/post/content-provenance-and-disclosure-requirements-for-ai-generated-content-on-digital-and-traditional-m>.

[PAI-Framework]
Partnership on AI, "PAI’s Responsible Practices for Synthetic Media: A Framework for Collective Action", 27 February 2023, <https://syntheticmedia.partnershiponai.org/>.

Acknowledgments

The author thanks Michael Andrews from Teradata for helpful comments and feedback on early staging of this document.

Author’s Address

Dogu Abaris
Individual Contributor
Email: abaris@null.net