

AIPref  
Internet-Draft  
Intended status: Informational  
Expires: 30 May 2026

F. Badii  
Digital Medusa  
L. Bailey  
Internet Archive  
J. Levy  
26 November 2025

AI Preferences Signaling: End User Impact  
draft-farzdusa-aipref-enduser-00

## Abstract

Standards can have a major impact on end users across technological, legal, ethical, and governance dimensions, largely centering around access to information, control over their digital contributions, and data privacy. The purpose of this Internet Draft is to document the potential impact of signaling AI preferences on end users other than publishers, and to suggest some principles for the ai-pref working group to consider when assessing proposed vocabulary and definitions IETF wishes to standardize for signaling.

## About This Document

This note is to be removed before publishing as an RFC.

The latest revision of this draft can be found at <https://farzaneh-unicode.github.io/enduser-AIPref/draft-farzdusa-aipref-enduser.html>  
Status information for this document may be found at  
<https://datatracker.ietf.org/doc/draft-farzdusa-aipref-enduser/>.

Source for this draft and an issue tracker can be found at  
<https://github.com/Farzaneh-Unicode/enduser-AIPref>.

## 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 30 May 2026.

## Copyright Notice

Copyright (c) 2025 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<https://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Revised BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Revised BSD License.

## Table of Contents

1. Introduction . . . . .	3
2. Who are the end users? . . . . .	3
3. The IETF' s mandate and the Approach to Signaling AI Preferences . . . . .	4
4. The Right to Access Knowledge and Avoiding Internet Fragmentation . . . . .	5
4.1. Access to Information . . . . .	5
4.2. Intellectual Property Overreach and Rights Delay . . . . .	5
4.3. Risk of Web Fragmentation . . . . .	5
4.4. Impact on AI Tool Utility . . . . .	6
4.5. Site operators don' t own everything on their website . . . . .	6
5. Impact on User Rights, Transformation, and Academic Research . . . . .	6
5.1. Robots Rights to Read . . . . .	6
5.2. Use in the public interest . . . . .	6
6. Shortcomings of Restricting RAG and Inference . . . . .	7
6.1. Collateral Damage to End Users and Beneficial Applications . . . . .	7
6.2. Risk to Automated Tools and Access to Knowledge (A2K) . . . . .	7
6.3. Discouraging Personal Use of Automation . . . . .	7
6.4. Accessibility Concerns . . . . .	8
6.5. Impact on Research and Flexibility . . . . .	8
7. Technical and Definitional Instability . . . . .	8
7.1. The Crawl Time vs. Inference . . . . .	8
7.2. Difficulty in Allowing Non-Commercial and Excluding Commercial Crawling . . . . .	8
7.3. Uncertainty in Definition . . . . .	9
7.4. Asset level signaling . . . . .	9
8. Case Studies . . . . .	9
8.1. Inhibition of Not-for-Profit Crawlers . . . . .	9

8.2. Opaque Robot Defenses . . . . .	9
8.3. Collateral Damage to Beneficial Uses . . . . .	9
8.4. Threat to Public Archiving . . . . .	10
8.5. Intermediaries Choosing for Smaller Site Operators . . . . .	10
8.6. Platform Control over Creator Intent . . . . .	10
8.7. Creation of Quasi-IP Rights . . . . .	10
8.8. The Threat of Exclusionary Defaults . . . . .	10
9. The Impact on the Internet . . . . .	11
10. Proposed End User Impact Principles . . . . .	11
11. Conclusion . . . . .	12
12. Conventions and Definitions . . . . .	12
13. Security Considerations . . . . .	12
14. IANA Considerations . . . . .	12
15. Normative References . . . . .	12
Acknowledgments . . . . .	13
Authors' Addresses . . . . .	13

## 1. Introduction

As AI systems start becoming a primary way people access and interact with information online, mechanisms for signaling AI preferences risk shaping far more than technical coordination. While the AI-Pref Working Group focuses on enabling content hosts to express their preferences, these signals operate at the protocol layer, where they can quietly restrict access, chill lawful uses, and unintentionally fragment the open web. End users, understood broadly in RFC 8890 as the human beings whose activities Internet standards support, include not only publishers and AI developers but also researchers, educators, people with disabilities, small creators, and the global public. Many of these communities rely on rights to access, transform, and analyze information that could be undermined if preference signaling is interpreted or implemented as a form of control. This draft outlines the potential impact of AI signaling on end users and proposes principles to ensure that standardization preserves interoperability, openness, and user agency.

## 2. Who are the end users?

Mark Nottingham the author of The Internet is for the End Users, proposes a definition for end users in RFC 8890: "end users" means human users whose activities IETF standards support, sometimes indirectly... End users are not necessarily a homogenous group; they might have different views of how the Internet should work and might occupy several roles, such as a seller, buyer, publisher, reader, service provider, and consumer. An end user might browse the Web, monitor remote equipment, play a game, videoconference with colleagues, send messages to friends, or perform an operation in a remote surgery theater. They might be "at the keyboard" or

represented by software indirectly (e.g., as a daemon)".

This broad definition is especially useful to keep in mind as the outputs from the AI preferences working group are likely to have an impact on a broad range of end users. The dominant voices in the AI-pref working group so far have been "publishers" and AI developers, each of which are a kind of end user. Other end users' rights, interests, and access have been discussed less frequently. When rights and interests are discussed, in general the rights under consideration are those of Intellectual Property Rightsholders, while the discussions rarely mention other rightsholders, such as people with rights to access to knowledge and essential services, or interested parties.

There is a need to ensure that property and economic interests are aligned with fundamental human rights, such as the right to freedom of opinion and expression under Article 19 of the Universal Declaration of Human Rights. Intellectual-property and market-based mechanisms should function in a way that supports, rather than constrains, the free circulation of information and ideas.

While AI Preference signaling mechanisms are intended to give content hosts (or site operators) a means to signal how content is used by AI systems, they can result in unintended consequences that disproportionately harm end users, including researchers, people with disabilities, and commercial and non-commercial actors, thereby undermining the open Internet. Site operators may signal restrictions over content or uses that they do not have the legal right to control, such as text and data mining permitted under copyright exceptions, accessibility tools, public-domain materials, or facts and information not protected by intellectual property rights. When interpreted strictly by AI systems, such signals can create a chilling effect: researchers, accessibility developers, and others may refrain from lawful uses simply because they believe the content is off-limits. This misalignment between technical signals and actual rights risks fragmenting access to knowledge and undermining the open Internet.

### 3. The IETF's mandate and the Approach to Signaling AI Preferences

The IETF was never intended to serve as a forum for protecting business models from technological innovation or enforcing copyright regimes. Its purpose is to maintain and advance the interoperability of the Internet, to ensure that diverse systems, services, and users can communicate seamlessly across technical and institutional boundaries. A shift away from this focus risks turning what should be a technical coordination effort into a venue for regulating reuse at the protocol layer, well before any legal reasoning or due-process

safeguards can apply. If the IETF shifts toward a regime that encourages blanket prohibitions on data processing, regardless of context, consequence, or user benefit, we risk hardcoding a “chilling effect” into the infrastructure of the web itself. And unlike lawful reuses, which are typically adjudicated through downstream legal reasoning, signaling operates at the technical layer, where denial is fast, quiet, and hard to contest.

#### 4. The Right to Access Knowledge and Avoiding Internet Fragmentation

A major concern for end users is the potential restriction of access to information and essential services due to overly restrictive content signaling instituted by hosting platforms, creators and publishers.

##### 4.1. Access to Information

People globally are increasingly accessing essential services and knowledge through the Internet. AI-enabled applications could potentially simplify access to knowledge and essential services provided online, such as healthcare and other services. However, if signaling AI preferences reinforced by regulatory reference become too restrictive, it can have a chilling effect and the fundamental right to access information and essential services could be harmed.

##### 4.2. Intellectual Property Overreach and Rights Delay

Theorists of the Access to Knowledge (A2K) movement contend that current levels of Intellectual Property (IP) protection may be unbalanced regarding the right to access. They argue that material interests are not simply equivalent to current IP provisions, and that “rights delayed are rights denied”. (Lea Shaver The Right to Science and Culture, 2010 Wis. L. Rev. 121 (2010)).

##### 4.3. Risk of Web Fragmentation

There is a tension between giving site operators and declarants (those who declare their preferences about AI data usage) granular control over content and maintaining an open Internet. Implementing very restrictive mechanisms risks “building fortress walls around content” and sacrificing the open web, potentially throttling the future of knowledge and access to essential services.

#### 4.4. Impact on AI Tool Utility

While large AI companies are the primary targets of restrictive preference signals, these controls can inadvertently restrict and harm end users who wish to use AI tools for common tasks like exploring content, summarizing, or translating documents or making services more accessible. Setting a high bar for accessing and using information could negatively affect individual end users' ability to utilize these tools and overall experience when accessing the world wide web. Ultimately, it is individual people who will pay the price for restricted access to AI tools, both financially and by IETF-endorsed limits on their ability to use AI to access the knowledge and information they seek. Open source AI developers and others can also be disenfranchised if the preferences signals are not coherently implementable or if they have chilling effects.

#### 4.5. Site operators don't own everything on their website

Any given web page has the technical capacity to contain content from a variety of different rightsholders. A large platform can therefore contain content from millions, even billions, of different content creators. Giving a website owner the ability to determine the AI preferences for potentially millions of different creators and their works risks centralizing a huge amount of control.

### 5. Impact on User Rights, Transformation, and Academic Research

End users who are researchers are frequently hindered by licensing and contractual barriers, even when the underlying uses might be legally permissible under copyright law.

#### 5.1. Robots Rights to Read

Researchers and other end users face a tension when downloading and human (consumptive) reading is allowed, but machine (non-consumptive) reading is disallowed, leading to the slogan "let the robots read." Robots are perhaps always the agents of human users and preventing them from reading could ultimately prevent the humans from access.

#### 5.2. Use in the public interest

The principle of allowing public interest uses of information has long supported knowledge production and freedom of expression. These uses are without payment nor permission, rejecting the idea that access to information is "a privilege reserved for the well-behaved". In the AI-Pref discussions, many participants expressed a desire to differentiate between traditional search engine indexing—generally accepted as beneficial to the open web—and newer AI-enabled

summarization or answer generation. Some content providers signaled that they may be open to appearing in classic search results but not in AI-driven outputs that repackage or synthesize their material. However, this signaling raises a deeper structural issue: by opting out, publishers and intellectual property rights holders and hosts effectively shift the burden of determining whether a use is permitted to the recipient of the signal. While this may not be a significant obstacle for well-resourced actors like major platforms, it imposes disproportionate risks on small operators, open-source AI developers, and actors from the global majority, those who often lack the legal infrastructure, funding, or access to navigate complex copyright frameworks.

## 6. Shortcomings of Restricting RAG and Inference

Restricting RAG and inference, uses that occur after the foundation model has been trained, is fundamentally challenging due to the fluid nature of these processes and the risk of unintended negative consequences for legitimate users and access to information.

### 6.1. Collateral Damage to End Users and Beneficial Applications

One of the most significant shortcoming of restricting inference is the unintended curtailment of beneficial applications used everyday.

### 6.2. Risk to Automated Tools and Access to Knowledge (A2K)

If restrictions on inference are applied too broadly, end users who wish to use AI tools for common tasks like discovering content, summarizing documents, or translation would be barred from doing so. This directly undermines the end user's right to access knowledge (A2K).

### 6.3. Discouraging Personal Use of Automation

End users want to benefit from the efficiencies of automation, such as using bulk operations and automated search features, rather than having to rely solely on large, proprietary AI company tools. Placing the same restrictive bar on individual end users as on large AI developers for using information could negatively affect the ability of individuals to leverage these tools for personal or beneficial research purposes.

#### 6.4. Accessibility Concerns

Overly restrictive inference controls, such as a complete opt-out of “all data mining, automation and processing” , carry the risk of affecting uses like spam filtering, harmful language detection, and accessibility tools. For instance, if a tool that describes a website to a visually impaired user relies on inference, a broad restriction could prevent that legitimate use.

#### 6.5. Impact on Research and Flexibility

Researchers, when analyzing data, may be hindered if limitations block non-consumptive reading (TDM) but allow human reading. The core issue remains that an overly restrictive environment risks constraining scholarship and academic communication.

### 7. Technical and Definitional Instability

The attempt to restrict RAG/Inference runs into technical limitations regarding scope, authority, and implementation methods.

#### 7.1. The Crawl Time vs. Inference

Preference signals (like those in robots.txt or metadata) are available at crawl time, but they are often disassociated from the model when the content is actually used at inference time. This could create mixed signals as to what the end user can do regarding the crawling vs inference and signaling about inference because of this time lag can create many problems for the end users, real time translation for example might not happen.

#### 7.2. Difficulty in Allowing Non-Commercial and Excluding Commercial Crawling

For a general-purpose foundational model (LLM), it is technically difficult to distinguish between commercial and non-commercial uses at either crawl or inference time. Because models cannot selectively isolate or invoke portions of their training data based on the eventual use context, a blanket “No AI” or similar restrictive signal tends to apply universally — even to lawful, commercial and non-commercial, or public-interest uses. This leaves developers and researchers with “unappealing choices,” such as omitting the data entirely or creating separate models for every potential use category, both of which are impractical and risk excluding socially beneficial applications.



### 7.3. Uncertainty in Definition

RAG is only one technique, and definitions of inference-time use are unstable. The concept of "inference" is broad and covers a massive, conceptually fuzzy range of uses of material after training, making it difficult to develop a robust vocabulary term. If standards focus too much on a specific technique like RAG, they risk becoming irrelevant as new techniques emerge.

### 7.4. Asset level signaling

Inference is becoming increasingly local (e.g., self-hosted models or on-device processing). Asset-level control or signaling could intervene with an individual's device, and there might be attempts to ensure a user cannot override a "no inference" flag for personal use (like translating a document), which is highly problematic.

## 8. Case Studies

### 8.1. Inhibition of Not-for-Profit Crawlers

Non-profit organizations dedicated to archiving and public benefit, such as the Common Crawl Foundation, the Internet Archive and Data Rescue Project, are often inadvertently blocked or disadvantaged by broad control mechanisms intended to stop commercial scraping.

### 8.2. Opaque Robot Defenses

Modern "robot defenses" are increasingly used by websites to stop crawling instead of relying solely on the transparent robots.txt protocol. Providers of these defenses sometimes treat legitimate archive crawlers, like Common Crawl's CCBot, the same way as commercial bots crawling for specific AI companies. This blocking process is considered opaque.

### 8.3. Collateral Damage to Beneficial Uses

This conflation of uses often leads to "collateral damage" for non-AI uses, effectively "locking down the Web". For example, the official 2024 End of Term Archive has been inadvertently blocked from crawling some US government websites due to these defensive measures.

#### 8.4. Threat to Public Archiving

Organizations providing vital public benefit, such as the Internet Archive's Wayback Machine, are constantly under threat because courts sometimes impose a duty to parse individualized, idiosyncratic Terms of Service (ToS) for every archived page. Such enforcement would make the vital historical, academic, legal, and journalistic work provided by the Wayback Machine impossible.

#### 8.5. Intermediaries Choosing for Smaller Site Operators

A significant problem arising from this overreach is that creators and smaller site operators who wish to share their content for uses like AI training, or search indexing are often prevented from doing so because the hosting platform or site administrator chooses a restrictive standard for them by default.

#### 8.6. Platform Control over Creator Intent

Public websites, particularly those hosting user-generated content (UGC) like social media platforms, use broad Terms of Service (ToS) to assert extensive, contract-based control over content and data. If a site has content from many creators (e.g., social media platforms), the site administrator typically controls the location-based opt-out mechanisms (like robots.txt). Consequently, the administrator "may not allow them [individual creators] to express their preferences fully, or at all".

#### 8.7. Creation of Quasi-IP Rights

Platforms use their ToS to create quasi-intellectual property rights intended to bind the public and impose constraints that stifle academic research and preservation, even over content they do not own the copyright to.

#### 8.8. The Threat of Exclusionary Defaults

When platforms implement restrictive measures (even if intended to be protective), they lock away valuable resources that could otherwise help the open web. For instance, a valuable public-interest resource, such as a digital dictionary, could be blocked from all AI systems by the organizations hosting it, resulting in the resource being inaccessible to speakers, researchers, and language models that could otherwise preserve and revitalize the language.

## 9. The Impact on the Internet

Using Internet Society Internet Impact Assessment tool, we have done a brief Internet impact assessment on several concepts we have been discussing during the working group dialogues:

Critical Property / Goal	Potential Impact
*CP1 Accessible infrastructure*	All-or-nothing AI-pref signals increase participation costs, especially for small/open-source actors.
*CP2 Interoperable building blocks*	Broad bans reduce modular reuse (e.g., summarization/translation), stifling recombination.
*CP3 Decentralized management*	Platform-wide defaults override creator intent; reduce autonomy in UGC contexts.
*CP4 Common global identifiers*	No impact identified yet.
*CP5 Technology-neutral network*	Banning "AI" as a category risks chilling general-purpose uses (e.g., accessibility).

Figure 1

## 10. Proposed End User Impact Principles

To the greatest extent possible, the goal should be to keep public information public and to keep the open web open. Vocabulary standardized by IETF should not unduly restrict end user access to and ability to use public knowledge and information. Definitions should be clear and narrowly tailored to be fit for purpose, rather than vague and broad. End users should be able to understand the definitions without having technical or legal expertise. End users should be able to use general purpose, commercially available tools to accomplish their own lawful activities, regardless of preferences to the contrary.

## 11. Conclusion

AI Preference signaling, while conceived as a mechanism for expressing control and transparency over the use of online content, risks evolving into an instrument that fragments the open Internet and uses of information. The IETF's role has never been to protect business models from innovation or to enforce intellectual property regimes. Its role is to ensure interoperability and resilience across technical systems. When preference signaling becomes a vehicle for implementing legal or economic restrictions at the protocol layer, it moves the IETF into a domain of governance that undermines its core mission and values.

End users, whether researchers, accessibility developers, educators, or ordinary participants in the digital commons, stand to lose most from these shifts. Overly restrictive signaling can chill legitimate research, disable accessibility tools, and entrench barriers to knowledge that the Internet was designed to dismantle. By conflating technical coordination with policy enforcement, we risk replacing openness and innovation with exclusion and opacity.

The AI-Pref Working Group should therefore focus on preserving interoperability and user agency, ensuring that standards remain technology-neutral, modular, and open to innovation.

## 12. 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.

## 13. Security Considerations

TODO Security

## 14. IANA Considerations

This document has no IANA actions.

## 15. 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/info/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/info/rfc8174>>.

#### Acknowledgments

TODO acknowledge.

#### Authors' Addresses

Farzaneh Badii  
Digital Medusa  
Email: [farzaneh@digitalmedusa.org](mailto:farzaneh@digitalmedusa.org)

Lila Bailey  
Internet Archive  
Email: [lila@archive.org](mailto:lila@archive.org)

Jo Levy  
Email: [jlevy@nortonlaw.com](mailto:jlevy@nortonlaw.com)