

# AI & the Web: Understanding and managing the impact of Machine Learning models on the Web

20 August 2024

## ▼ More details about this document

### Latest published version:

<https://www.w3.org/reports/ai-web-impact/>

### History:

[Commit history](#)

### Editor:

Dominique Hazael-Massieux

### Feedback:

[GitHub w3c/ai-web-impact](#) (pull requests, new issue, open issues)

### Translations

[简体中文 \(Simplified Chinese\)](#)

Copyright © 2024 World Wide Web Consortium. W3C<sup>®</sup> [liability](#), [trademark](#) and [permissive document license](#) rules apply.

---

## Abstract

This document proposes an analysis of the systemic impact of [AI systems](#), and in particular ones based on [Machine Learning models](#), on the Web, and the role that Web standardization may play in managing that impact.

## Status of This Document

This document is intended to capture the current shared understanding of the [W3C Team](#) on the current and expected impact of developments linked to [Artificial Intelligence systems](#) on the Web, and identifying explorations the World Wide Web Consortium community has started or ought to be starting, to manage that impact. It does not represent any consensus from the W3C Membership nor is it a standardization document.

The document was authored by Dominique Hazaël-Massieux ([dom@w3.org](mailto:dom@w3.org)), with significant contributions from the rest of the W3C Team.

This document aims first and foremost to help structure discussions on what may be needed at the standardization level to make the systemic impact of AI (and specifically, [Machine Learning models](#)) less harmful or more manageable. It is bound to be incomplete and sometimes wrong - we are gathering input and feedback in [GitHub](#), preferably before June 30, 2024.

If it's not feasible for you to use GitHub, send comments in e-mail to: [public-team-report-comments@w3.org](mailto:public-team-report-comments@w3.org). Please put your comments in the body of the message, not as an attachment. Start your e-mail subject line with: [AI Report].

Depending on the feedback received, possible next steps include more in-depth stakeholder interviews, a dedicated W3C Workshop, or developing a standardization roadmap.

## Table of Contents

### Abstract

### Status of This Document

1. **Executive Summary**
2. **Introduction**
  - 2.1 Terminology
3. **Intersections between AI systems and the Web**
4. **Ethics and societal impact**
  - 4.1 Respecting autonomy and transparency
    - 4.1.1 Transparency on AI-generated content
    - 4.1.2 Transparency on AI-mediated services
  - 4.2 Right to privacy and data control
  - 4.3 Safety and security
  - 4.4 Sustainability
  - 4.5 Balancing content creators incentives and consumers rights
    - 4.5.1 Comparison with search engines
5. **Impact of AI systems on interoperability**

## A. References

### A.1 Informative references

## § 1. Executive Summary

[Machine Learning models](#) support a new generation of [AI systems](#). These models are often [trained](#) on a large amount of Web content, deployed at scale through web interfaces, and can be used to generate plausible content at unprecedented speed and cost.

Given the scope and scale of these intersections, this wave of [AI systems](#) is having potential systemic impact on the Web and some of the equilibriums on which its ecosystem had grown.

This document reviews these intersections through their ethical, societal and technical impacts and highlights a number of areas where standardization, guidelines and interoperability could help manage these changes:

- a [consent mechanism for the use of Web content in training pipelines](#),
- [labeling content as computer-generated](#),
- [surfacing training sources in model cards](#),
- [exposing model-backed Web APIs](#),
- [personal data stores](#) to reduce risk of private data exposure,
- [strengthening credentials and identity mechanisms](#) in light of new impersonation risks,
- an [evaluation framework for the environmental impact of Web standards](#),
- a [framework to manage interoperability based on model inference](#), including for non-deterministic models.

We are [seeking input](#) from the community on proposals that could help make progress on these topics, and on other topics that this document has failed to identify.

## § 2. Introduction

Recent developments in the decades-long computer science field of Artificial Intelligence have made a number of systems emerge that already have started having **systemic** impacts on the Web and can be

expected to further transform a number of shared expectations on which the health of the Web had relied so far.

To help structure a conversation within the W3C community (and possibly with other Web-related standards organization) about these transformations, this document collects the current shared understanding among the W3C Team of the intersections of "[Artificial Intelligence](#)", more specifically in the field of [Machine Learning models](#) (including Large Language Models and other so called generative AI models), with the Web as a system, and ongoing W3C development in this space. It is also a goal to raise questions about additional explorations that may be needed as further developments in this space arise.

That current understanding is bound to be incomplete or sometimes plain wrong; we hope that by publishing this document and inviting community reviews on it, we iteratively improve this shared understanding and help build a community roadmap to increase the positive impact and decrease the harms that are emerging in this intersection.

## § 2.1 Terminology

The term "Artificial Intelligence" covers a very broad spectrum of algorithms, techniques and technologies. [ISO/IEC-22989] defines *Artificial Intelligence* as "research and development of mechanisms and applications of [AI systems](#)", with *AI systems* being "an engineered system that generates outputs such as content, forecasts, recommendations or decisions for a given set of human-defined objectives". At the time of the writing of this document in early 2024, the gist of the Web ecosystem conversation on Artificial Intelligence is mostly about systems based on *Machine Learning* ("process of optimizing model parameters through computational techniques, such that the model's behavior reflects the data or experience") and its software manifestation, *Machine Learning models* ("mathematical construct that generates an inference or prediction based on input data or information").

While we acknowledge the much broader meaning of Artificial Intelligence and its intersection with a number of other Web- and W3C-related activities (e.g., the Semantic Web and Linked Data), this document willfully focuses only on the current conversation around the impact that these [Machine Learning models](#) are bringing to the Web. We further acknowledge that this document has been developed during, and is partially a response to, a cycle of inflated expectations and investments in that space. That situation underlines the need for a framework to structure the conversation.

Because of this focus on [Machine Learning](#), this document analyzes AI impact through the two main phases needed to operate [Machine Learning models](#): *training* ("process to determine or to improve the

parameters of a Machine Learning model, based on a Machine Learning algorithm, by using training data") and *inference* (the actual usage of these models to produce their expected outcomes), which we also casually refer as running a model.

### § 3. Intersections between AI systems and the Web

A major role the Web plays is as a platform for content creators to expose at scale their content to content consumers. AI directly relates to these two sides of the platform:

- In a number of cases, models are trained based on content crawled from the Web; the combination of scale and structure in that content (made possible by the underlying standards) has made it an invaluable source of training data that backs some of the most visible results in recent AI developments, such as large language models or image/video generators;
- Conversely, a number of these AI models can be used to generate content at unprecedented scale, which the reach of the Web allows to deploy seamlessly to the billions of users of the platform.

When looking more specifically at the browser-mediated part of the Web which remains primarily a client/server architecture, AI models can be [run](#) either on the server-side or on the client-side (and somewhat more marginally at this point, in a [hybrid-fashion between the two](#)). On the client side, they can either be provided and operated by the browser (either at the user's request, or at the application's request), or entirely by the client-side application itself.

It's also worth noting that as [AI systems](#) are gaining rapid adoption, their intersection with the Web is bound to evolve and possibly trigger new systemic impact; for instance, emerging [AI systems](#) that combine [Machine Learning models](#) and content loaded from the Web in real-time may induce revisiting in depth the role and user experience of Web browsers in consuming or searching content.

### § 4. Ethics and societal impact

The [W3C's Technical Architecture Group Ethical Web Principles](#) [[ethical-web-principles](#)] includes ensuring "[the Web should not cause harm to society](#)".

As described above, the Web is already a key enabler in some of the recent developments in Artificial Intelligence, and the usage and impact of Artificial Intelligence is being multiplied in scale through its distribution via the Web. This calls for the [W3C](#) community as stewards of the Web to understand potential harms emerging from that combination and to identify potential mitigations to these harms.

The [Ethical Principles for Web Machine Learning](#) [[webmachinelearning-ethics](#)] started in the [Web Machine Learning Working Group](#) combine values and principles from the UNESCO [Recommendation on the Ethics of Artificial Intelligence](#) [UNESCO-AI] with Web-specific principles from the Ethical Web Principles to identify 4 values and 11 principles that integration of Machine Learning on the Web should follow, and which have helped structure this document.

## § 4.1 Respecting [autonomy](#) and [transparency](#)

### § 4.1.1 Transparency on AI-generated content

Recent [AI systems](#) are able to assist in the partial or complete creation of content (textual, graphic, audio and video) at a level of (at least superficially) credible quality and in quantities beyond that developed by humans. This provides both opportunities and risks for content creators, but more importantly, it creates a systemic risk for content consumers in no longer being able to distinguish or discover authoritative or curated content in a sea of credible (but either possibly or willfully wrong) generated content.

That need is pressing directly for end-users as they individually consume content, but also applies to agents that end-users rely on: typically, search engines would likely benefit from transparency on purely AI generated content. Somewhat ironically, crawlers used to train AI models are likely to need such a signal as well, since [training](#) models on the output of models may create unexpected and unwanted results.

We do not know of any solution that could guarantee (e.g., through cryptography) that a given piece of content was or was not generated (partially or entirely) by [AI systems](#). That gap unfortunately leaves a systemic risk in terms of misinformation and spam that should be of grave concern for the health of the Web as a content distribution platform and of society as a whole.

A plausible role of standards in this space would be to at least facilitate the **labeling of content** to indicate whether it is the result of a **computer-generated process**. While such labels are unlikely to be enforceable through technical means, they could gain broad adoption with a combination of being automatically added by [AI systems](#) (possibly with enough friction that removing them would be at least somewhat costly at scale), and possibly serve as hooks in the regulatory context.

A number of proposals have already emerged in this space, which may benefit from more visibility, discussion and ultimately, scalable deployment:

- [C2PA Guidance for AI and Machine Learning](#) [C2PA-AI]
- [IPTC synthetic media](#) [IPTC-DST] and its matching representation in [Schema.org](#) [schema.org]
- [Proposal: Meta Tag for AI Generated Content](#) in [HTML] (individual submission)

An area that could be explored is role Web browsers might play in surfacing labelling or provenance information of content, e.g., embedded content such as images or video. This can currently be done by publishers' websites or independent verifiers, but integrating this capability into the browser could make the information more convenient for users to access, as well as being independent of any particular publisher or website where the same content may be viewed.

## § 4.1.2 Transparency on AI-mediated services

A well-known issue with relying operationally on [Machine Learning models](#) is that they will integrate and possibly strengthen any *bias* ("systematic difference in treatment of certain objects, people or groups in comparison to others" [[[ISO/IEC-22989]]) in the data that was used during their [training](#). [Bias](#) commonly occurs in other algorithms and human decision processes. Where [AI systems](#) make that [bias](#) a bigger challenge is because these models are at this point harder to audit and amend since they operate mostly as a closed box.

Such [bias](#) will disproportionately affect users whose expected input or output is less well represented in training data (as e.g., discussed in the [report from the 2023 AI & Accessibility research symposium](#) [WAI-AI]), which intuitively is likely to correlate strongly with users already disenfranchised by society and technology - e.g., if your language, appearance or behavior doesn't fit the mainstream-expected norm, you're less likely to feature in mainstream content and thus less visible or misrepresented in training data.

Until better tools emerge to facilitate at least the systematic detection of such [bias](#), encouraging and facilitating the systematic publication of information on whether a Machine Learning model is in use, and how such a model was trained and checked for [bias](#) may help end-users make more informed choices about the services they use (which, of course, only helps if they have a choice in the first place, which may not apply e.g., to some government-provided services).

[Model cards for Model Reporting \[MODEL-CARDS\]](#) are one of the approaches that were [discussed in the 2020 W3C Workshop on Web and Machine Learning \[W3C-ML-WS\]](#). Assuming this reporting provides meaningful and actionable transparency, a question for (the) technical standardization would be how such **cards should be serialized and made discoverable on the Web**.

W3C should look at a particular category of model deployments: **models that are used by browser engines themselves to fulfill API requests**. A number of Web browser APIs already expose (more or less explicitly) output of [Machine Learning models](#):

- [Web Speech API \[SPEECH-API\]](#)
- [Accelerated Shape Detection API \[SHAPE-DETECTION-API\]](#)
- [Background blur, face detection, gaze correction controls in Media capture](#)

As described below, these APIs also raise engineering questions about how to ensure they provide the [level of interoperability](#) that have been expected from more traditionally deterministic algorithms.

## § 4.2 [Right to privacy and data control](#)

Models trained on un-triaged or partially triaged content off the Web are bound to include personally identifiable information (PII). The same is true for models trained on data that users have chosen to share (for public consumption or not) with service providers. These models can often be made to retrieve and share that information with any user who knows how to ask, which breaks expectations of privacy for those whose personal information was collected, and is likely to be in breach with privacy regulations in a number of jurisdictions. Worse, they create risks for new types of attacks (see [4.3 Safety and security](#)).



While the exclusion rules discussed in the context of content creation could partially help with the first situation, they would not help with the second one. This problem space is likely to be under heavy regulatory and legal scrutiny.

From a technical standardization perspective, beyond labeling content, the emergence of user data being repurposed for model [training](#) and some of the pushback it is generating may bring renewed momentum (from user and service provider alike) behind decentralized architectures that leave user data under less centralized control (as illustrated by the recent widening adoption of Activity Streams).

A particularly relevant instance of that pattern is emerging with so-called **personal data stores**: these provide ways for users to exert more fine-grained control of their data, by separating more clearly the roles of data store and data processor (which, in a traditional cloud infrastructure, would otherwise typically be handled by a single entity).

That topic has most recently surfaced in [W3C](#) through the [proposed charter for a SOLID Working Group](#) late 2023 (a charter that the [W3C](#) community has recognized as important, but where there is not yet consensus).

Allowing to [run](#) a model on personal data without uploading that data to a server is one of the key motivations behind the browser [Web Neural Network API](#) [WEBNN] which, completing the computing capabilities already provided by [WebAssembly](#) [WASM-CORE-2] and [WebGPU](#) [WEBGPU], provides additional Machine Learning specific optimizations to [run](#) models efficiently from within the browser (and thus, on the end-user device).

## § 4.3 [Safety and security](#)

A number of [Machine Learning models](#) have significantly lowered the cost of generating credible textual, as well as audio and video (real-time or recorded) impersonations of real persons. This creates significant risks of upscaling the capabilities of phishing and other types of frauds, and thus raising much higher the barriers to establish trust in online interactions. If users no longer feel safe in their digitally-mediated interactions, the Web will no longer be able to play its role as a platform for these interactions.

This points towards even stronger needs for **robust identity and credential management** on the Web. The work of the [Verifiable Credentials Working Group](#) allows to express credentials in a cryptographically secure, privacy-preserving, and machine-verifiable way [VC-DATA-MODEL]. The [proposed work to better integrate Federated Identity systems into browsers](#) [FEDCM] and the [emerging one on surfacing digital credentials to Web content](#) [DIGITAL-CREDENTIALS] are likely part of mitigating the risks associated with these new impersonation threats.

## § 4.4 Sustainability

[Training](#) and [running Machine Learning models](#) can prove very resource-intensive, in particular in terms of power- and water-consumption. The imperative of reducing humanity's footprint on natural resources should apply particularly strongly to the technologies that standardization helps deploy at scale.

There is relatively new but promising work in the [Sustainable Web Design Community Group](#) (a [candidate to become a standardization Working Group](#)) to explain how to [use Web technologies in a sustainable way](#).

The [Green Software Foundation Software Carbon Intensity Working Group](#) is developing a score to calculate carbon footprint of software applications.

W3C still lacks a well-defined **framework to evaluate the environmental impact of its standards**. Given the documented high environmental impact of [AI systems](#), it will surely become more important that W3C groups that are expected to accelerate the deployment of [Machine Learning models](#) take a proactive approach in exploring and documenting how they envision the environmental impact of their work, and possible mitigations they might identify.

## § 4.5 Balancing content creators incentives and consumers rights

Some of the largest and most visible [Machine Learning models](#) are known or assumed to have been trained with materials crawled from the Web, without the explicit consent of their creators or publishers.

The controversy that has emerged from that situation is being debated (and in some cases, arbitrated) through the lens of copyright law.

It is not our place to determine if and how various copyright legislation bears on that particular usage. Beyond legal considerations, the copyright system creates a (relatively) shared understanding between creators and consumers that, by default, content cannot be redistributed, remixed, adapted or built upon without creators' consent. This shared understanding made it possible for a lot of content to be openly distributed on the Web. It also allowed creators to consider a variety of monetization options (subscription, pay per view, advertising) for their content grounded on the assumption that consumers will always reach their pages.

A number of [AI systems](#) combine (1) automated large-scale consumption of Web content, and (2) production at scale of content, in ways that do not recognize or otherwise compensate content it was trained from.

While some of these tensions are not new (as discussed below), systems based on Machine Learning are poised to upend the existing balance. Unless a new sustainable equilibrium is found, this exposes the Web to the following undesirable outcomes:

- Significantly less open distributed content (which would likely have a disproportionate impact on the less wealthy part of the population)
- A less appealing platform to distribute content

A less direct risk may emerge from changes in copyright laws meant to help to rebalance the situation but which would reduce the rights from content consumers and then undermine the value of the Web as a platform for which content distribution is a key value proposition.

#### § 4.5.1 Comparison with search engines

A number of the tensions emerging around the re-use of content crawled at scale from the Web have a long history given the central role that search engines play for the Web. Indeed, search engines provide (and absorb) value from their ability to retrieve and organize information from content on the Web, and they heavily rely on the standardized infrastructure this content is built on to achieve these results.

The more or less implicit contract that emerged between search engines and content providers has been that search engines can retrieve, parse and partially display content from the providers, in exchange of bringing more visibility and traffic to them. A further assumption has been encoded in the

way the Web operates that this contract is the default for anyone making content available publicly on the Web, with an opt-out mechanism encoded via the [robots.txt directives](#) [RFC9309].

Over time, in addition to the links to sites matching the user's query, search engines have integrated more ways to surface content directly from the target Web sites: either through rich snippets (typically made possible by the use of schema.org metadata) or through embedded preview (e.g., what the [AMP project](#) enabled). These changes were frequently accompanied by sometimes challenging discussions around the balance between bringing additional visibility to crawled content and reducing the incentive from end-users to visit the source website (e.g., because they may have already received sufficient information from the search results page).

In a certain number of cases, [AI systems](#) are used as an alternative or complement to what users would traditionally have used a search engine for (and indeed, are increasingly integrated into search engine interfaces). So it seems useful to explore to what extent the lessons learned from the evolutionary process balancing the needs from search engines and from content creators can inform the discussion on crawlers used to train [Machine Learning models](#).

In making that comparison, it's also important to note significant differences:

- The implicit contract that content creators expect from search engines crawlers –i.e., that they will bring exposure to their content– does not have a systematic equivalent for content integrated into [AI systems](#); while some such systems are gaining the ability to point back to the source of their training data used in a given [inference](#), this is hardly a widespread feature of these systems, nor is it obvious it could be applied systematically (e.g., would linking back to sources for a generated image even make sense?); even if it could, fewer sources would likely be exposed than in a typical search engine results page, and the incentives for the user to follow the links would likely be substantially lower.
- `robots.txt` directives allow specific rules to be given to specific crawlers based on their user agent; while this has been practically manageable when dealing with (for better or for worse) few well-known search engine crawlers, expecting content creators to maintain potential allow- and block-lists of the rapidly expanding number of crawlers deployed to retrieve training data seems unlikely to achieve sustainable results.

Given the likely different expectations around the quid-pro-quo of crawling in the context of [AI systems](#), it is not obvious that the permission-less pattern inherited from the early days of the Web (`robots.txt` was designed in 1994) would be a satisfactory match to ensure the long term sustainability of content publication on the Web (itself presumably in the long term interest of AI crawlers themselves).

In general, a possibly helpful line of inquiry for standards in this space would be to identify solutions that help **content producers and AI crawlers to find agreeable terms, ideally at a scale** that would make it appealing to all parties.

Several groups and individuals have been exploring how to make it possible for content publishers to express their willingness to get their content used for [training Machine Learning models](#):

- The [Text and Data Mining Reservation Protocol Community Group](#) has developed the [TDM Reservation Protocol \(TDMRep\)](#) [TDMRep] to "express the reservation of rights relative to text & data mining applied to lawfully accessible Web content"
- Discussions are starting [in IETF around updating the robots.txt directives](#) in this context; among others, the [update robots.txt Community Group](#) proposes to add an opt-in mechanism to the robots.txt directives.
- The [JournalList](#) association is looking into revising their `trust.txt` [TRUST.TXT] to [let publishers express whether their content can be used for training models](#)

## § 5. Impact of AI systems on interoperability

A key part of [W3C's vision for the Web](#) [w3c-vision] is to ensure the Web is developed around principles of interoperability: that is, for technologies that W3C codifies as Web standards, to ensure they are implemented and deployed in a way that will work the same across products, allowing for greater choice for users and fostering the long term viability of the content.

When the algorithm on which interoperability relies is deterministic, ensuring interoperability is a matter of describing in sufficient detail and clarity the said algorithm, and running sufficient testing on the products to verify they achieve the intended result. The move to more [algorithmic specifications](#) [design-principles] and thorough automated testing (e.g., through the [Web Platform Tests project](#)) has largely been driven by the goal of providing a robust interoperable platform.

As discussed above, [Machine Learning models](#) are already finding their way into standardized Web APIs. These creates two challenges to our interoperability goals:

- [Machine Learning models](#) are mostly not built or described as a series of algorithmic steps. If a given standardized behavior is expected to be best fulfilled by [Machine Learning models](#), how should that behavior be specified? How can it be tested to a level that sufficiently verifies an **interoperable outcome across products that would use different models**? What impact would it have on the fingerprinting surface of the browsers?
- A number of important [Machine Learning models](#) are not deterministic; if or when some of these **non-deterministic models** get exposed in standardized APIs, this consistency question is no longer limited to two products using two different models, since a given input would no longer produce a predetermined output. It is not clear to us at the moment how to prepare for interoperable behaviors based on non-deterministic models, which probably raises the question of whether and how such models should be acceptable as part of interoperable implementations.

A possible consequence of these challenges is a reduction of the scope of what can be meaningfully made interoperable and standardized as a possibly growing number of features get mediated by [Machine Learning models](#) (similar to the [postulated impact of growing capabilities of web applications on the need to standardize protocols](#)). In that context, discussions e.g., around AI-based codecs point towards possible significant changes in the interoperability landscape.

## § A. References

### § A.1 Informative references

#### [C2PA-AI]

*Guidance for Artificial Intelligence and Machine Learning*. C2PA. URL: [https://c2pa.org/specifications/specifications/1.3/ai-ml/ai\\_ml.html](https://c2pa.org/specifications/specifications/1.3/ai-ml/ai_ml.html)

#### [design-principles]

*Web Platform Design Principles*. Lea Verou. W3C. 18 July 2024. W3C Working Group Note. URL: <https://www.w3.org/TR/design-principles/>

#### [DIGITAL-CREDENTIALS]

*Digital Credentials*. WICG. March 2024. URL: <https://wicg.github.io/digital-identities/>

**[ethical-web-principles]**

*Ethical Web Principles*. Daniel Appelquist; Hadley Beeman; Amy Guy. W3C. 13 August 2024. W3C Working Group Note. URL: <https://www.w3.org/TR/ethical-web-principles/>

**[FEDCM]**

*Federated Credential Management API*. W3C. Draft Community Group Report. URL: <https://fedidcg.github.io/FedCM/>

**[HTML]**

*HTML Standard*. Anne van Kesteren; Domenic Denicola; Ian Hickson; Philip Jägenstedt; Simon Pieters. WHATWG. Living Standard. URL: <https://html.spec.whatwg.org/multipage/>

**[IPTC-DST]**

*Digital Source Type vocabulary*. IPTC. URL: <https://cv.iptc.org/newscodes/digitalsourcetype/>

**[ISO/IEC-22989]**

*Artificial intelligence concepts and terminology*. ISO/IEC. July 2022. Published. URL: <https://www.iso.org/standard/74296.html>

**[MODEL-CARDS]**

*Model Cards for Model Reporting*. M. Mitchell; S. Wu; A. Zaldivar; P. Barnes; L. Vasserman; B. Hutchinson; E. Spitzer; I. Deborah Raji; T. Gebru. 14 Jan 2019. URL: <https://arxiv.org/pdf/1810.03993.pdf>

**[RFC9309]**

*Robots Exclusion Protocol*. M. Koster; G. Illyes; H. Zeller; L. Sassman. IETF. September 2022. Proposed Standard. URL: <https://www.rfc-editor.org/rfc/rfc9309>

**[schema-org]**

*Schema.org*. W3C Schema.org Community Group. W3C. 6.0. URL: <https://schema.org/>

**[SHAPE-DETECTION-API]**

*Accelerated Shape Detection in Images*. WICG. cg-draft. URL: <https://wicg.github.io/shape-detection-api/>

**[SPEECH-API]**

*Web Speech API*. WICG. cg-draft. URL: <https://wicg.github.io/speech-api/>

**[TDMRep]**

*Final Community Group Report*. Text and Data Mining Reservation Protocol Community Group. February 2024. URL: <https://www.w3.org/community/reports/tdmrep/CG-FINAL-tdmrep-20240202/>

**[TRUST.TXT]**

*Specification for trust.txt file and underlying system*. JournalList.net. May 2020. URL: <https://journalist.net/reference-document-for-trust-txt-specifications>



**[UNESCO-AI]**

*Recommendation on the Ethics of Artificial Intelligence*. UNESCO. 2021. URL: <https://unesdoc.unesco.org/ark:/48223/pf0000380455>

**[VC-DATA-MODEL]**

*Verifiable Credentials Data Model v1.1*. Manu Sporny; Grant Noble; Dave Longley; Daniel Burnett; Brent Zundel; Kyle Den Hartog. W3C. 3 March 2022. W3C Recommendation. URL: <https://www.w3.org/TR/vc-data-model/>

**[W3C-ML-WS]**

*W3C Workshop Report on Web and Machine Learning*. W3C. October 2020. URL: <https://www.w3.org/2020/06/machine-learning-workshop/report.html>

**[w3c-vision]**

*Vision for W3C*. Chris Wilson. W3C. 3 April 2024. W3C Working Group Note. URL: <https://www.w3.org/TR/w3c-vision/>

**[WAI-AI]**

*Artificial Intelligence (AI) and Accessibility Research Symposium 2023*. W3C Web Accessibility Initiative. Jan 2023. URL: <https://www.w3.org/WAI/research/ai2023/>

**[WASM-CORE-2]**

*WebAssembly Core Specification*. Andreas Rossberg. W3C. 16 August 2024. W3C Working Draft. URL: <https://www.w3.org/TR/wasm-core-2/>

**[WEBGPU]**

*WebGPU*. Kai Ninomiya; Brandon Jones; Jim Blandy. W3C. 13 August 2024. W3C Working Draft. URL: <https://www.w3.org/TR/webgpu/>

**[webmachinelearning-ethics]**

*Ethical Principles for Web Machine Learning*. Anssi Kostiainen. W3C. 8 January 2024. W3C Working Group Note. URL: <https://www.w3.org/TR/webmachinelearning-ethics/>

**[WEBNN]**

*Web Neural Network API*. Ningxin Hu; Dwayne Robinson. W3C. 5 August 2024. W3C Candidate Recommendation. URL: <https://www.w3.org/TR/webnn/>

[↑](#)