

# AI Control Workshop

Mark Nottingham, IETF 120

**“The [exception for copyright] shall apply on condition that the use of works [...] has not been expressly reserved by their rightholders in an appropriate manner, such as machine-readable means in the case of content made publicly available online.”**

**EU Directive 2019/790**

**“Where the rights to opt out has been expressly reserved in an appropriate manner, providers of general-purpose AI models need to obtain an authorisation from rightsholders if they want to carry out text and data mining over such works.”**

**EU Regulation 2024/1689 ‘AI Act’**

Stream: Internet Engineering Task Force (IETF)  
RFC: [9309](#)  
Category: Standards Track  
Published: September 2022  
ISSN: 2070-1721  
Authors: M. Koster   G. Illyes   H. Zeller   L. Sassman  
*Google LLC   Google LLC   Google LLC*

# RFC 9309

## Robots Exclusion Protocol

---

### Abstract

This document specifies and extends the "Robots Exclusion Protocol" method originally defined by Martijn Koster in 1994 for service owners to control how content served by their services may be accessed, if at all, by automatic clients known as crawlers. Specifically, it adds definition language for the protocol, instructions for handling errors, and instructions for caching.

**Is robots.txt appropriate  
for AI opt-out?**

```
User-Agent: *  
Disallow: *.gif$  
Disallow: /example/  
Allow: /publications/
```

```
User-Agent: foobot  
Disallow: /  
Allow: /example/page.html  
Allow: /example/allowed.gif
```

```
User-Agent: barbot  
User-Agent: bazbot  
Disallow: /example/page.html
```

```
User-Agent: quxbot
```

**“We believe that writers should be compensated also for past training since it appears that the massive training that has already occurred for GPT and Bard to teach the engines to think and to write has already occurred[.]”**

**Mary Rasenberger, The Authors Guild**



## Internet Engineering Task Force

4,185 followers

+ Follow

View full page



## Internet Engineering Task Force

Follow



4,185 followers

18h · 🌐

Don't miss the [#IETF120](#) plenary session set to start at 00:30 UTC/17:30 PDT (UTC -7) on 25 July 2024. More than 1500 registered participants are gathering onsite and remotely. Thanks to meeting host [Huawei](#), & Gold Sponsors [Akamai Technologies](#), [Huawei](#), & [Ericsson](#) for supporting [#openstandards!](#) <https://www.ietf.org/live/>



👍❤️ 21

1 repost

### Reactions



Like



Comment



Repost



Send



# IAB Workshop on AI-CONTROL (aicontrolws)

- About
- Meetings
- History
- Photos
- Email expansions

[Upload material](#) [Edit group](#)

<b>Team</b>	<b>Name</b>	<a href="#">Edit</a> <b>IAB Workshop on AI-CONTROL</b>
	<b>Acronym</b>	aicontrolws
	<b>State</b>	<a href="#">Edit</a> Active
	<b>Additional resources</b>	<a href="#">Edit</a>
<b>Personnel</b>	<b>Chairs</b>	<a href="#">Edit</a> <a href="#">Mark Nottingham</a> , <a href="#">Suresh Krishnan</a>

## Group description

Large Language Models and other machine learning techniques require voluminous input data, and one common source of such data is the Internet -- usually, "crawling" Web sites for publicly available content, much in the same way that search engines crawl the Web.

This similarity has led to an emerging practice of allowing the Robots Exclusion Protocol (RFC 9309) to control the behavior of AI-oriented crawlers.

This emerging practice raises many design and operational questions. It is not yet clear whether robots.txt (the mechanism specified by RFC 9309) is well-suited to controlling AI crawlers. A content creator or host may not be able to distinguish a crawler used for search indexing from a crawler used for LLM ingest – and indeed some crawlers may be used for both purposes. Potential use cases may extend across many different units of content, policies to be signaled, and types of content creators. Before robots.txt becomes a de facto solution to AI crawling opt-out, it is necessary to examine whether it is an appropriate mechanism: in particular, whether the creator of a particular unit of content can realistically and fully exercise their right to opt-out, and the scope of data ingest to which that opt-out applies.

This workshop aims to explore practical opt-out mechanisms for AI, and build an understanding of use cases, requirements, and other considerations in this space. The workshop will focus on mechanisms to communicate the opt-out choice and their associated data models. Technical enforcement of opt-out signals is not in scope.

<https://datatracker.ietf.org/group/aicontrolws/about/>

- Submissions due: 2 August 2024
- Workshop dates: 19-20 September 2024
- Workshop location: Washington DC
- Chatham House Rule

<https://datatracker.ietf.org/group/aicontrolws/about/>