

# Module 2 Assignment: robots.txt Screenshot PG1

The screenshot shows a browser window with the URL "thegradcafe.com/robots.txt". The page content is the robots.txt file for the website. It includes a header about content signals, a section for Cloudflare Managed content (allowing search and AI training), and a section for Cloudflare Managed Content (disallowing specific paths). The file also lists various user-agents with their corresponding Disallow directives.

```
# As a condition of accessing this website, you agree to abide by the following
# content signals:
#
# (a) If a Content-Signal = yes, you may collect content for the corresponding
#     use.
# (b) If a Content-Signal = no, you may not collect content for the
#     corresponding use.
# (c) If the website operator does not include a Content-Signal for a
#     corresponding use, the website operator neither grants nor restricts
#     permission via Content-Signal with respect to the corresponding use.

# The content signals and their meanings are:
#
# search: building a search index and providing search results (e.g., returning
#          hyperlinks and short excerpts from your website's contents). Search does not
#          include providing AI-generated search summaries.
# ai-input: inputting content into one or more AI models (e.g., retrieval
#           augmented generation, grounding, or other real-time taking of content for
#           generative AI search answers).
# ai-train: training or fine-tuning AI models.

# ANY RESTRICTIONS EXPRESSED VIA CONTENT SIGNALS ARE EXPRESS RESERVATIONS OF
# RIGHTS UNDER ARTICLE 4 OF THE EUROPEAN UNION DIRECTIVE 2019/790 ON COPYRIGHT
# AND RELATED RIGHTS IN THE DIGITAL SINGLE MARKET.

# BEGIN Cloudflare Managed content

User-agent: *
Content-Signal: search=yes,ai-train=no
Allow: /

User-agent: Amazonbot
Disallow: /

User-agent: Applebot-Extended
Disallow: /

User-agent: Bytespider
Disallow: /

User-agent: CCBot
Disallow: /

User-agent: ClaudeBot
Disallow: /

User-agent: Google-Extended
Disallow: /

User-agent: GPTBot
Disallow: /

User-agent: meta-externalagent
Disallow: /

# END Cloudflare Managed Content

User-agent: *
Disallow: /cgi-bin/
Disallow: /index-ad-test.php

User-agent: ia_archiver
Disallow: /

User-agent: ia_archiver/1.6
Disallow: /

User-Agent: sitecheck.internetseer.com
Disallow: /

User-agent: Computer_and_Automation_Research_Institute_Crawler
Disallow: /

User-agent: dotbot
Disallow: /

User-agent: YandexBot
Disallow: /
```

## Module 2 Assignment: robots.txt Screenshot PG2

```
User-agent: YandexBot
Disallow: /

User-agent: Mediapartners-Google
Disallow:

User-agent: Opebot-v (https://www.1plusx.com (https://www.1plusx.com/))
Allow: /
```