

Noví strážci brány: Vyčerpávající srovnávací analýza webových crawlerů pro LLM od OpenAI, Anthropic a Google

Úvod

Digitální prostředí se nachází na prahu nové éry, kde je hodnota webového obsahu redefinována nenasytným apetitem velkých jazykových modelů (LLM) po datech. Tento vývoj vytváří zásadní napětí mezi tvůrci obsahu, kteří budují a udržují web, a společnostmi zabývajícími se umělou inteligencí, které tento obsah spotřebovávají pro trénink svých modelů. Tento report slouží jako základní strategický průvodce pro všechny digitální subjekty a představuje tři primární agenty – GPTBot od OpenAI, ClaudeBot od Anthropic a kontrolní mechanismus Google-Extended od společnosti Google – jako nové, často neprůhledné strážce informačních toků v éře umělé inteligence. Pochopení jejich chování, mechanismů a strategických důsledků již není pouze technickou záležitostí, ale klíčovou obchodní nutností pro každého, kdo působí v online prostoru.

Část 1: GPTBot od OpenAI – Dokumentovaný sběrač dat

Analýza přístupu společnosti OpenAI odhaluje strategii, která se vyznačuje relativní transparentností a strukturovanou architekturou. Tento přístup ji staví do pozice etalonu pro chování „dobrého občana“ v prostoru AI crawlerů, vůči kterému lze ostatní porovnávat.

1.1 Mandát a architektura s více agenty

Explicitním a jediným účelem **GPTBota** je procházet veřejně dostupný webový obsah za účelem tréninku a vylepšování základních modelů generativní umělé inteligence společnosti OpenAI.¹ Tato jasně definovaná role je zásadní pro pochopení jeho funkce a odlišení od ostatních agentů provozovaných společností.

Pro zamezení běžným nedorozuměním je nutné GPTBot striktně odlišit od ostatních user-agentů OpenAI, z nichž každý má specifický účel:

- **ChatGPT-User:** Tento agent není automatizovaný crawler, ale spíše tzv. „fetcher“ spouštěný uživatelem. Jedná jménem konkrétního uživatele ChatGPT, který provádí akci procházení webu, a provádí jednorázové návštěvy stránek, nikoli systematické procházení webu.¹ Jeho činnost je reaktivní, nikoli proaktivní.
- **OAI-SearchBot:** Tento agent je využíván pro vyhledávací funkce v rámci ChatGPT. Jeho úkolem je zobrazovat webové stránky ve výsledcích vyhledávání, ale je explicitně uvedeno, že *není* používán k tréninku základních modelů.¹

Toto jasné oddělení funkcí demonstruje promyšlenou architektonickou volbu ze strany OpenAI, která

nabízí správcům webů granulární kontrolu nad tím, jak je jejich obsah využíván. Taková transparentnost je strategickým krokem, jehož cílem je budovat důvěru a zajistit dlouhodobou udržitelnost datového kanálu, na kterém jsou modely OpenAI závislé. Zatímco agresivní nebo neprůhledné metody sběru dat mohou vést k odporu veřejnosti, právním sporům a masivnímu blokování, transparentní přístup podporuje spolupráci. Vytvořením samostatných agentů pro odlišné funkce a jejich jasnou dokumentací dává OpenAI správcům webů jednoznačnou kontrolu, což je pozicuje jako odpovědného aktéra a zvyšuje pravděpodobnost, že jim bude přístup povolen.

1.2 Identifikační a kontrolní protokoly

Pro efektivní správu přístupu je nezbytné znát přesné technické identifikátory GPTBota. Jeho user-agent token je GPTBot a plný user-agent string je Mozilla/5.0 AppleWebKit/537.36 (KHTML, like Gecko); compatible; GPTBot/1.1; +https://openai.com/gptbot.¹

OpenAI plně podporuje standardní direktivy souboru robots.txt, což umožňuje správcům webů detailně řídit přístup. Jsou podporovány následující konfigurace:

- **Povolení plného přístupu:**
User-agent: GPTBot
Allow: /
- **Zakázání veškerého přístupu:**
User-agent: GPTBot
Disallow: /
- **Povolení přístupu pouze ke specifickým adresářům:**
User-agent: GPTBot
Allow: /directory-1/
Disallow: /directory-2/

8

Dalším prvkem transparentnosti je zveřejňování rozsahů IP adres, ze kterých GPTBot operuje. Tyto informace jsou dostupné ve formátu JSON a umožňují robustnější blokování a ověřování na úrovni serveru, například prostřednictvím firewallu.¹

1.3 Pozorované chování a přítomnost na trhu

Analýza provozních protokolů GPTBota odhaluje závazek k etickému procházení webu. Je navržen tak, aby automaticky filtroval a ignoroval obsah za platebními branami, zdroje, které shromažďují osobně identifikovatelné údaje (PII), a obsah, který je v rozporu se zásadami OpenAI.³

Klíčová data z analýzy společnosti Cloudflare za období 2024–2025 kvantifikují obrovský a rostoucí dopad GPTBota na webovou infrastrukturu. Jeho podíl na celkovém provozu AI a vyhledávacích crawlerů dramaticky vzrostl z 2,2 % na 7,7 %. Tento nárůst byl doprovázen zvýšením objemu surových

požadavků o 305 %, což GPTBota katapultovalo z 9. na 3. neaktivnějšího crawlera na webu.¹³ Tato čísla potvrzují, že GPTBot není okrajovým hráčem, ale jedním z nejvýznamnějších zdrojů automatizovaného provozu na současném internetu.

Část 2: ClaudeBot od Anthropic – Kontroverzní operátor

Analýza crawleru společnosti Anthropic představuje investigativní pohled na operátora, jehož činnost je charakterizována nedostatkem transparentnosti, měnícími se identitami a významnými veřejnými i právními kontroverzemi. Jeho příběh slouží jako případová studie, jak nekomunikovat s komunitou tvůrců webového obsahu.

2.1 Historie nejasností a měnících se identit

Základní problém spočívá v téměř úplné absenci oficiální dokumentace ze strany společnosti Anthropic týkající se jejích webových crawlerů.¹⁵ Tento nedostatek informací nutí spoléhat se na analýzy třetích stran a zprávy z komunity, což vytváří prostředí nejistoty.

Historicky byla situace komplikována existencí několika špatně vysvětlených user-agentů, včetně Claude-Web, anthropic-ai a současného primárního agenta ClaudeBot.¹⁹ Tato nekonzistence způsobila značný zmatek. Zásadní odhalení přinesl článek na 404 Media, který ukázal, že mnoho webových stránek blokovalo zastaralé agenty

Claude-Web a anthropic-ai, zatímco aktivní ClaudeBot zůstával neblokován.²¹ To poukazuje na závažné selhání v komunikaci ze strany Anthropic, které v praxi zneplatnilo snahy správců webů o kontrolu přístupu.

2.2 Obvinění z nedodržování pravidel a agresivního procházení

Tato podkapitola tvoří jádro vyšetřování a podrobně se zabývá hlavními kontroverzemi spojenými s činností crawleru společnosti Anthropic.

- **Žaloba ze strany společnosti Reddit (podána v červnu 2025):** Podrobná analýza žaloby odhaluje obvinění, že Anthropic bez povolení shromažďoval obsah z Redditu za účelem tréninku modelu Claude, čímž porušil podmínky služby platformy.²² Mezi klíčová tvrzení patří, že Anthropic ignoroval direktivy v souboru robots.txt a přistoupil k platformě více než 100 000krát i poté, co veřejně prohlásil, že s touto činností přestal.²²
- **Přetížení serverů:** Existují zdokumentované důkazy o tom, že činnost ClaudeBota způsobila vážné technické problémy. Případ společnosti iFixit je obzvláště závažný; její generální ředitel Kyle Wiens obvinil ClaudeBota z „krádeže“ dat, kdy crawler během jediného dne stáhl 10 TB souborů a za celý měsíc celkem 73 TB, což vedlo k výraznému zatížení serverů.²⁵ Podobné problémy s přetížením a pády serveru hlásila také fóra Linux Mint.²⁵

- **Extrémní poměr extrakce:** Výzkum společnosti Cloudflare, citovaný v souvislosti se žalobou, zjistil, že ClaudeBot procházel 38 000 stránek na každou jednu odkazovanou návštěvu, což naznačuje výjimečně agresivní model sběru dat ve srovnání s průmyslovými normami.²²

Tento vzorec chování – neprůhlednost, agresivní sběr dat a ignorování zavedených norem – vedl k vytvoření významného „deficitu důvěry“ vůči společnosti Anthropic. Důsledky tohoto přístupu se projevily nejen v poškození pověsti, ale i v konkrétních tržních reakcích. Komunita správců webů aktivně začala blokovat crawlery Anthropic, což potvrzují data od Cloudflare, která ukazují 46% pokles objemu surových požadavků ClaudeBota a snížení jeho podílu na provozu z 11,7 % na 5,4 %.¹³ Vzniká tak negativní zpětnovazební smyčka: agresivní procházení vede k poškození reputace, to vede k masivnímu blokování, což omezuje přístup k čerstvým webovým datům a v dlouhodobém horizontu může potenciálně snižovat kvalitu modelů Claude ve srovnání s konkurenty, kteří mají důvěryhodnější datové kanály.

2.3 Kontrolní mechanismy a jejich účinnost

Navzdory kontroverzím existují standardní metody pro blokování crawlerů společnosti Anthropic. Primárním nástrojem je soubor robots.txt a user-agent token ClaudeBot.²⁰

Je však nutné tuto informaci okamžitě uvést do kontextu obvinění ze žaloby společnosti Reddit, která zpochybňuje, zda bot tato pravidla skutečně dodržuje. Nedávné prohlášení společnosti Anthropic, že ClaudeBot bude nyní respektovat blokování i pro své starší, zastaralé agenty, je zjevně reaktivním krokem v reakci na širokou kritiku, nikoli proaktivním opatřením.²¹

Část 3: Google-Extended – Strategický kontrolní signál

Přístup společnosti Google se zásadně liší od přístupu OpenAI a Anthropic. Důraz je kladen na to, že se nejedná o nový crawler, ale o nové pravidlo pro jeho stávající, dominantní infrastrukturu pro procházení webu. Tato skutečnost vytváří pro majitele webových stránek kritické strategické nuance.

3.1 Zásadně odlišný přístup: Token, nikoli bot

Je klíčové objasnit, že Google-Extended je *token* pro user-agenta v souboru robots.txt, který funguje jako kontrolní signál. Nejedná se o samostatný crawler s vlastním HTTP user-agent stringem.²⁸ Samotné procházení webu provádí stávající flotila crawlerů společnosti Google, především

Googlebot. Token Google-Extended v souboru robots.txt pouze dává Googlu pokyn, jak může být shromážděný obsah následně použit.³⁰ Toto je zásadní architektonický rozdíl oproti OpenAI a Anthropic, které pro sběr dat pro AI provozují dedikované boty.

3.2 Kritické rozlišení: Gemini vs. AI Overviews

Toto je nejdůležitější strategický bod celého reportu. Je nezbytné jasně a explicitně uvést, co blokování Google-Extended způsobuje a co nikoli.

- **Co BLOKUJE:** Použití obsahu vaší webové stránky pro trénink budoucích generací modelů Gemini a pro tzv. „grounding“ (poskytování citovaných zdrojů) v rámci chatovací aplikace Gemini a rozhraní Vertex AI API.²⁸
- **Co NEBLOKUJE:** Použití obsahu vaší webové stránky v **AI Overviews** (dříve SGE) ve výsledcích vyhledávání Google. Tyto přehledy jsou považovány za nedílnou součást vyhledávacího zážitku, nikoli za samostatný produkt řízený tokenem Google-Extended.³²

Pro blokování obsahu v AI Overviews musí správce webu sáhnout k drastičtějším opatřením, jako je meta tag nosnippet. To však s sebou nese významný kompromis v oblasti SEO, protože tento tag zároveň zabráni zobrazení obsahu v tradičních úryvcích ve výsledcích vyhledávání.³²

3.3 Strategické důsledky odhlášení

Tato situace představuje pro vydavatele zásadní volbu. Blokováním Google-Extended majitel stránek chrání svůj obsah před použitím pro trénink modelů. Jak však podrobně popisují odborníci z oboru, jako je Marie Haynes, má to svou cenu: stránka nebude doporučována ani citována jako zdroj v rozhraní chatu Gemini.³² Vzhledem k tomu, že se konverzační AI stává primárním kanálem pro objevování informací, mohlo by to znamenat obětování významné budoucí viditelnosti a návštěvnosti. Toto rozhodnutí je tedy vysoce sázkovou volbou mezi ochranou obsahu a budoucí relevancí.

Společnost Google strategicky rozdělila využití dat pro AI, aby se vyhnula problémům s autorskými právy a obavám vydavatelů, a zároveň si zachovala schopnost vylepšovat svůj hlavní produkt – vyhledávání – pomocí umělé inteligence. Tímto přístupem vytvořila pro méně informované správce webů jakousi „iluzi kontroly“. Google si uvědomuje právní a etická rizika spojená s používáním obsahu chráněného autorskými právy k trénování komerčních modelů AI bez svolení. Token Google-Extended je jejich odpovědí na tento problém a poskytuje jasnou možnost odhlášení pro trénink modelů. Současně je hlavním byznysem Googlu vyhledávání. AI Overviews vnímá jako evoluci vyhledávacích úryvků, nikoli jako samostatný AI produkt. Oddělením těchto dvou případů použití může Google veřejně tvrdit, že nabízí kontrolu (prostřednictvím Google-Extended), zatímco nadále využívá webový obsah pro svou nejvýznamnější a nejhodnotnější AI funkci (AI Overviews). Vzniká tak situace, kdy správce webu může zablokovat Google-Extended v domnění, že se odhlásil z „AI od Googlu“, a přesto zjistí, že jeho obsah stále pohání odpovědi generované AI ve výsledcích vyhledávání.

Část 4: Přímé srovnání – Srovnávací matice

Tato část syntetizuje zjištění z předchozích kapitol do přímého srovnání jednotlivých funkcí a vlastností, které je završeno přehlednou referenční tabulkou.

4.1 Transparentnost a dokumentace

V oblasti transparentnosti a dokumentace existují mezi jednotlivými společnostmi výrazné rozdíly. OpenAI je hodnocena vysoce za svou jasnou a veřejně dostupnou dokumentaci, která detailně popisuje jednotlivé agenty a jejich účel. Google je rovněž hodnocen vysoko za svou komplexní dokumentaci, ačkoli složitost jeho mechanismu vyžaduje pečlivé studium. Naopak Anthropic je hodnocen velmi špatně kvůli absenci oficiální dokumentace a nutnosti spoléhat se na zprávy třetích stran, což vytváří nejistotu a zmatek.

4.2 Mechanismus a kontrola

Architektonické přístupy jednotlivých společností se zásadně liší:

- **OpenAI:** Používá dedikovaný, jednoúčelový crawler (GPTBot) pro trénink AI, oddělený od ostatních agentů.
- **Anthropic:** Využívá dedikovaný crawler (ClaudeBot), který má však za sebou historii několika matoucích a zastaralých identit.
- **Google:** Implementoval kontrolní token (Google-Extended), který řídí využití dat shromážděných jeho stávající, víceúčelovou infrastrukturou pro procházení webu.

4.3 Dodržování robots.txt a reputace

Srovnání deklarovaných zásad a pozorovaného chování odhaluje další rozdíly:

- **OpenAI:** Veřejně deklaruje dodržování robots.txt a má obecně dobrou pověst v oblasti dodržování pravidel.
- **Anthropic:** Veřejně deklaruje dodržování, ale čelí věrohodným a mediálně známým obviněním z nedodržování a agresivního chování.
- **Google:** Striktně dodržuje direktivu robots.txt, avšak *rozsah* této direktivy je užší, než mnozí předpokládají, což vede k nedorozuměním.

4.4 Tabulka: Matice vlastností a chování LLM crawlerů

Následující tabulka zhušťuje nejdůležitější technická a strategická data z reportu do jediného, snadno stravitelného formátu. Umožňuje rychlé porovnání crawlerů napříč klíčovými atributy a usnadňuje tak rozhodování.

Vlastnost	OpenAI (GPTBot)	Anthropic (ClaudeBot)	Google (Google-Extended)
User-Agent Token	GPTBot	ClaudeBot	Google-Extended

Plný User-Agent String	Poskytnut a dokumentován	Mozilla/5.0... ClaudeBot/1.0...	N/A (používá Googlebot atd.)
Deklarovaný účel	Trénink základních modelů	Trénink LLM Claude	Kontrolní signál pro trénink a grounding Gemini
Kontrolní mechanismus	Direktiva robots.txt pro konkrétního bota	Direktiva robots.txt pro konkrétního bota	Direktiva robots.txt řídící použití dat
Oficiální dokumentace	Komplexní	Minimální až žádná	Komplexní
Dodržování robots.txt	Deklarováno a obecně respektováno	Deklarováno, ale předmětem závažných právních/veřejných obvinění z porušování	Striktně respektováno pro svůj definovaný účel
Klíčová strategická nuance	Jasně oddělení od vyhledávacích/užívatelských botů	Historie matoucích/zastaralých agentů	Neblokuje AI Overviews ve vyhledávání
Hlavní kontroverze	Žádné významné	Žaloba od Redditu, přetížení serverů (iFixit), neprůhledné praktiky	Široce rozšířené nepochopení jeho omezeného rozsahu
Tržní trend (2024–25)	+305 % objemu požadavků (prudký nárůst)	-46 % objemu požadavků (pokles)	N/A (součást celkového provozu Googlebot, který je dominantní a roste)

Část 5: Definitivní manuál pro správu webových stránek

Tato závěrečná, akční část převádí všechny předchozí analýzy do konkrétního strategického a technického průvodce pro majitele webových stránek.

5.1 Část I: Čeho se vyvarovat – Běžné chyby a nesprávné konfigurace

- **Zastaralý blocklist:** Přidávání pravidel do robots.txt pro zastaralé agenty, jako je Claude-Web, a zároveň neblokovaní aktivního ClaudeBota, což činí snahu neúčinnou.²¹
- **Mylná představa o Google-Extended:** Blokování Google-Extended s falešným očekáváním, že to odstraní obsah z AI Overviews ve výsledcích vyhledávání.³²

- **Pasivní spoléhání na robots.txt:** Předpoklad, že direktiva Disallow v robots.txt je neprůstřelnou technickou bariérou. Pro crawlery, které nemusí protokol respektovat, je to pouze žádost, nikoli příkaz.⁴

5.2 Část II: Jak na to správně – Vícevrstvá strategie obrany a zapojení

- **Krok 1: Audit a identifikace:** Prvním krokem je analýza serverových logů za účelem identifikace, kteří AI crawleři navštěvují vaše stránky, jaká je jejich frekvence a objem spotřebovaných dat. To poskytuje daty podložený základ pro jakoukoli akci.
- **Krok 2: Základní kontrola (standard robots.txt):**
 - Pro správce webů, kteří si přejí odhlásit se ze všech známých crawlerů pro trénink AI, je zde připraven komplexní blocklist pro robots.txt, který lze přímo zkopírovat.
 - Příklad snippetu:

```
# OpenAI Training Crawler
User-agent: GPTBot
Disallow: /
```

```
# Anthropic Training Crawler
User-agent: ClaudeBot
Disallow: /
```

```
# Google AI Training Control
User-agent: Google-Extended
Disallow: /
```

```
# Other Common AI Crawlers
User-agent: CCBot
Disallow: /
User-agent: Bytespider
Disallow: /
```

26

- **Krok 3: Pokročilé vynucování (pro nedodržující nebo vysokobjemové crawlery):**
 - Pro robustnější ochranu je doporučeno implementovat blokování na úrovni serveru.
 - **.htaccess (Apache):** Lze použít k přímému blokování user-agentů, které vrátí chybu 403 Forbidden.³ Příklad:

```
RewriteEngine on
RewriteCond %{HTTP_USER_AGENT} ClaudeBot
RewriteCond %{HTTP_USER_AGENT} GPTBot [NC]
RewriteRule.* - [F,L]
```

 - **Web Application Firewalls (WAF):** Služby jako Cloudflare umožňují vytvářet pravidla firewallu, která mohou blokovat, vyzývat k ověření nebo omezovat rychlost požadavků na základě user-agent stringů, IP adres nebo vzorců chování. Toto je nejúčinnější metoda pro řešení agresivních nebo nedodržujících botů.³⁷
- **Krok 4: Rámec pro strategické rozhodování:**
 - Tento průvodce pomáhá zúčastněným stranám učinit informované rozhodnutí, které přesahuje jednoduchou mentalitu „zablokovat vše“.

- **Zhodnotte svůj obsah:** Jedná se o proprietární data, licencovaný obsah nebo jedinečné tvůrčí dílo, které představuje vaše klíčové duševní vlastnictví? (Zvažte blokování). Nebo se jedná o marketingový obsah, zprávy nebo vzdělávací materiály, kde je prvořadá široká viditelnost? (Zvažte povolení).
- **Zhodnotte svůj obchodní model:** Pokud váš byznys závisí na organické návštěvnosti z vyhledávání a na tom, abyste byli citovanou autoritou, blokování Google-Extended by mohlo být z dlouhodobého hlediska škodlivé.³² Pokud provozujete službu založenou na předplatném nebo s platební bránou, je ochrana obsahu prioritou.
- **Zhodnotte svou toleranci k riziku:** Jste ochotni riskovat, že váš obsah bude použit bez uvedení zdroje nebo kompenzace, abyste potenciálně získali viditelnost na nových platformách AI? Tento rámec umožní uživatelům vytvořit nuancovanou politiku přizpůsobenou jejich specifickým potřebám.

Závěr

Tento report dospěl k závěru, že prostředí AI crawlerů je definováno spektrem transparentnosti a dodržování pravidel, od strukturovaného přístupu OpenAI přes chaotické a kontroverzní metody Anthropic až po komplexní střední cestu, kterou zvolil Google. Zjištění jasně ukazují, že správa AI crawlerů již není pasivním úkolem typu „nastav a zapomeň“, ale stala se kritickou a aktivní součástí moderní digitální strategie a řízení rizik.

Budoucnost bude pravděpodobně vyžadovat vývoj nových, robustnějších webových standardů nad rámec robots.txt, které budou řídit vztah mezi tvůrci obsahu a umělou inteligencí spravedlivějším a ověřitelnějším způsobem. Do té doby zůstává proaktivní, informovaná a vícevrstvá strategie, jak je popsána v tomto reportu, nejlepší obranou a nástrojem pro orientaci v tomto rychle se vyvíjejícím ekosystému.

Citovaná díla

1. Overview of OpenAI Crawlers, použito září 23, 2025, <https://platform.openai.com/docs/gptbot?ref=internet.exchangepoint.tech>
2. Managing OpenAI's Web Crawlers [GPTBot]: A Comprehensive Guide, použito září 23, 2025, <https://www.movingtrafficmedia.com/managing-openai-web-crawlers/>
3. What Is The GPTBot And How Do You Block It? - Search Logistics, použito září 23, 2025, <https://www.searchlogistics.com/learn/seo/gptbot/>
4. What is the GPTBot? - DataDome, použito září 23, 2025, <https://datadome.co/bots/gptbot/>
5. GPTBot Explained: How OpenAI's Web Crawler Works and Why It Matters - Medium, použito září 23, 2025, <https://medium.com/@jorgecabaco/gptbot-explained-how-openais-web-crawler-works-and-why-it-matters-1cf9b43bdded>
6. Block AI Bots from Crawling Websites Using Robots.txt - Originality.ai, použito září 23, 2025, <https://originality.ai/ai-bot-blocking>
7. What Is GPTBot? Bot User Agent Insights - Dark Visitors, použito září 23, 2025, <https://darkvisitors.com/agents/gptbot>
8. GPTBot: OpenAI releases new web crawler - Search Engine Land, použito září 23, 2025, <https://searchengineland.com/gptbot-openai-new-web-crawler-430360>
9. GPTBot: OpenAI's Web Crawler - Medium, použito září 23, 2025, <https://medium.com/prompt-engineering/gptbot-openai-web-crawler-76d369d11848>
10. Here is how to block OpenAI from using your web content for ChatGPT - The Decoder, použito září 23, 2025, <https://the-decoder.com/here-is-how-to-block-openai-from-using-your-web-content-for-chatgpt/>
11. How to block OpenAI from crawling your website - Pluralsight, použito září 23, 2025, <https://www.pluralsight.com/resources/blog/ai-and-data/blocking-ChatGPT-OpenAI-website-crawling>

12. Understanding OpenAI's GPTBot & Robots.txt Setup - Quattr, použito září 23, 2025, <https://www.quattr.com/improve-discoverability/gptbot-robots-txt-access>
13. From Googlebot to GPTBot: who's crawling your site in 2025 - The Cloudflare Blog, použito září 23, 2025, <https://blog.cloudflare.com/from-googlebot-to-gptbot-whos-crawling-your-site-in-2025/>
14. From Googlebot to GPTBot: who's crawling your site in 2025, použito září 23, 2025, <https://blog.cloudflare.com/from-googlebot-to-gptbot-whos-crawling-your-site-in-2025>
15. Home - Claude Docs - Anthropic, použito září 23, 2025, <https://www.anthropic.com/docs>
16. Claude Console - Anthropic, použito září 23, 2025, <https://console.anthropic.com/>
17. Claude, použito září 23, 2025, <https://claude.ai/>
18. Building with Claude - Anthropic API, použito září 23, 2025, <https://docs.anthropic.com/en/docs/intro-to-claude>
19. What Is Claude-Web? Bot User Agent Insights - Dark Visitors, použito září 23, 2025, <https://darkvisitors.com/agents/claude-web>
20. What Is ClaudeBot? Bot User Agent Insights - Dark Visitors, použito září 23, 2025, <https://darkvisitors.com/agents/claudebot>
21. Websites are Blocking the Wrong AI Scrapers (Because AI Companies Keep Making New Ones) - 404 Media, použito září 23, 2025, <https://www.404media.co/websites-are-blocking-the-wrong-ai-scrapers-because-ai-companies-keep-making-new-ones/>
22. Reddit files lawsuit against Anthropic over unauthorized Claude AI ..., použito září 23, 2025, <https://ppc.land/reddit-files-lawsuit-against-anthropic-over-unauthorized-claude-ai-training/>
23. Why Reddit Sues Anthropic: The Dangers of AI Data & Privacy | AI Magazine, použito září 23, 2025, <https://aimagazine.com/articles/why-reddit-sues-anthropic-the-dangers-of-ai-data-privacy>
24. Reddit accuses Anthropic of misusing user content | Digital Watch Observatory, použito září 23, 2025, <https://dig.watch/updates/reddit-accuses-anthropic-of-misusing-user-content>
25. The Claude Team Sparks Outrage Over Data Scraping Issues ..., použito září 23, 2025, <https://www.aibase.com/news/10739>
26. How to manually block common AI crawlers - Raptive Support, použito září 23, 2025, <https://help.raptive.com/hc/en-us/articles/25756415800987-How-to-manually-block-common-AI-crawlers>
27. How to Block Crawlers That Slow Down Your Site? - Cloudways, použito září 23, 2025, <https://www.cloudways.com/blog/block-ai-crawlers/>
28. Google's common crawlers | Google Search Central | Documentation, použito září 23, 2025, <https://developers.google.com/search/docs/crawling-indexing/google-common-crawlers>
29. Google Clarifies the "Google-Extended" Crawler Documentation, použito září 23, 2025, <https://www.searchenginejournal.com/google-clarifies-the-google-extended-crawler-documentation/507645/>
30. Google-Extended - New Google Crawler To Block Bard Or Google AI From Using Your Content - Search Engine Roundtable, použito září 23, 2025, <https://www.seroundtable.com/google-extended-crawler-ai-bard-36130.html>
31. Google's clarification on "Google Extended" Crawler Documentation - MindStick, použito září 23, 2025, <https://www.mindstick.com/blog/303837/google-s-clarification-on-google-extended-crawler-documentation>
32. Should you block Google Extended in Robots.txt? - Marie Haynes, použito září 23, 2025, <https://www.mariehaynes.com/should-you-use-google-extended-in-robots-txt/>
33. Google-Extended Crawler Update (April 2025): What It Means for Publishers and AI Training, použito září 23, 2025, <https://thatware.co/google-extended-crawler-update/>
34. Disallowing Google-bard from indexing but allow everyone else (robots.txt) - Reddit, použito září 23, 2025, https://www.reddit.com/r/bigseo/comments/173r4t1/disallowing_googlebard_from_indexing_but_allow/
35. What is Claude-Web crawler bot - DataDome, použito září 23, 2025, <https://datadome.co/bots/claude-web/>
36. How to prevent ChatGPT from crawling your website - Classic - RapidWeaver Forum, použito září 23, 2025, <https://forums.realmacsoftware.com/t/how-to-prevent-chatgpt-from-crawling-your-website/42591>
37. How to block bots from OpenAI GPTBot - Vercel, použito září 23, 2025, <https://vercel.com/guides/how-to-block-bots-openai-gptbot>
38. How to block ClaudeBot with firewall rules? - SSL / TLS - Cloudflare Community, použito září 23, 2025, <https://community.cloudflare.com/t/how-to-block-claudebot-with-firewall-rules/654432>
39. Declare your AI independence: block AI bots, scrapers and crawlers with a single click, použito září 23, 2025, <https://blog.cloudflare.com/declaring-your-ai-independence-block-ai-bots-scrapers-and-crawlers-with-a-single-click/>