

---

## Část 1: Jak se zavádějící subjekt (autor) dozví o porušení?

Zjistit, že poskytovatel GPAI skutečně použil vaše data navzdory uplatněnému opt-outu, je **značně obtížné, ale ne nemožné**. Neexistuje žádný jednoduchý nástroj, který by "proskenoval" model a odhalil vaše data. Autor se musí spoléhat na kombinaci nepřímých důkazů, právních nástrojů a analýzy chování modelu.

Zde jsou hlavní způsoby, jak lze porušení zjistit:

### 1. Analýza souhrnu trénovacích dat (povinnost dle AI Actu)

Poskytovatelé GPAI musí podle Nařízení o AI zveřejnit "dostatečně podrobný souhrn" dat použitých pro trénink.

- **Co to znamená:** Tento souhrn sice nebude obsahovat seznam všech jednotlivých webových stránek nebo děl, ale měl by popisovat použité datové sady (např. "archiv Common Crawl z let 2020-2023", "výběr knih z Google Books", "data z Wikipedie").
- **Jak to pomůže:** Pokud autor ví, že jeho dílo bylo součástí těchto velkých datových sad, a zároveň ví, že měl na svých stránkách v dané době aktivní opt-out, má silný nepřímý důkaz, že jeho pokyn k opt-outu byl pravděpodobně ignorován.

### 2. Analýza výstupů modelu (tzv. memorizace)

Toto je nejprůkaznější metoda, kterou může autor sám vyzkoušet. Jedná se o situaci, kdy model "vyplivne" části chráněného obsahu téměř nebo úplně doslovně.

- **Jak to funguje:** Autor může model cíleně dotazovat (promptovat) tak, aby se pokusil zrekonstruovat části svého díla. Může například zadat začátek unikátní věty ze svého článku nebo popsat specifickou scénu ze své knihy.
- **Příklad:** Pokud autor článku o pěstování specifické odrůdy rajčat zadá do modelu první dvě věty a model vrátí další tři odstavce slovo od slova, je to velmi silný důkaz, že daný článek byl v trénovacích datech. Taková míra "zapamatování" (memorizace) je považována za jasný signál porušení autorských práv.

### 3. Soudní řízení a proces zjišťování (Discovery)

Pokud má autor dostatečně silné podezření (např. na základě výše uvedených bodů), může poskytovatele GPAI zažalovat.

- **Co se stane:** V rámci soudního řízení může soud nařídít tzv. proces zjišťování (discovery), který donutí AI společnost předložit podrobnější záznamy o svých trénovacích datech a procesech. To je často jediný způsob, jak získat přímý důkaz.
- **Nevýhoda:** Soudní spory jsou extrémně nákladné a časově náročné. Příkladem jsou žaloby deníku The New York Times proti OpenAI.

## Část 2: Jak může autor takový opt-out prakticky učinit?

Aby byl opt-out právně a technicky účinný, musí být vyjádřen ve **strojově čitelném formátu**. To znamená, že nestačí mít jen větu v obchodních podmínkách (i když i to je z právního hlediska užitečné). Roboti (crawlers), kteří sbírají data pro trénink, musí být schopni tomuto pokynu automaticky porozumět.

Zde jsou standardní a doporučené metody:

### 1. Soubor robots.txt (nejběžnější a nejdůležitější metoda pro weby)

Jedná se o jednoduchý textový soubor umístěný v hlavním adresáři webové stránky (např. [www.mojestranka.cz/robots.txt](http://www.mojestranka.cz/robots.txt)). Tento soubor dává pokyny automatizovaným robotům.

- **Jak na to:** Do souboru se přidají pravidla, která zakazují přístup specifickým robotům používaným AI společnostmi.
- **Příklad obsahu robots.txt pro blokaci hlavních AI crawlerů:**

```
code Code
```

```
downloadcontent_copy
```

```
expand_less
```

```
# Zakazuje robotovi OpenAI pro trénink budoucích modelů
User-agent: GPTBot
Disallow: /
```

```
# Zakazuje robotovi Google pro trénink modelů Vertex AI a Bard/Gemini
User-agent: Google-Extended
Disallow: /
```

```
# Zakazuje robotovi Common Crawl, který je masivně využíván třetími stranami
User-agent: CCBot
Disallow: /
```

```
# Zakazuje robotovi společnosti Anthropic (Claude)
User-agent: anthropic-ai
Disallow: /
```

```
# Zakazuje robotovi Perplexity AI
User-agent: PerplexityBot
Disallow: /
```

### 2. Protokol TDMRep (ve vývoji a standardizaci)

Jedná se o novější standard navržený specificky pro vyjádření rezervace práv pro účely vytěžování textů a dat (Text and Data Mining, TDM). Umožňuje podrobnější nastavení.

### 3. Metadata v souborech

U jednotlivých děl, jako jsou obrázky nebo dokumenty, lze informaci o zákazu vytěžování vložit přímo do jejich metadat (např. EXIF u fotek).

### 4. Smluvní podmínky a podmínky použití

I když to není primárně strojově čitelné, je důležité mít v podmínkách použití webu nebo služby jasně uvedeno, že jakékoli automatizované vytěžování obsahu (scraping, TDM) pro účely trénování AI je zakázáno bez výslovného souhlasu. Slouží to jako silný právní základ pro případný spor.

**Závěr:** Pro autora je klíčové **aktivně a technicky správně implementovat opt-out** (především pomocí robots.txt). Pokud má poté podezření na porušení, musí se připravit na detektivní práci kombinující analýzu výstupů modelu a využití právních nástrojů, které mu dává Nařízení o AI a autorský zákon.