

RLHF (Reinforcement Learning from Human Feedback)

RLHF (Učení posilováním z lidské zpětné vazby) je metoda ladění modelů umělé inteligence, která využívá lidské hodnocení k tomu, aby se model naučil odpovídat užitečněji, pravdivěji a bezpečněji.

Bez RLHF by jazykové modely (LLM) pouze mechanicky doplňovaly text na základě pravděpodobnosti. RLHF je proces, který model „zarovnává“ (**Alignment**) s lidskými hodnotami a očekáváními.

Proč je RLHF potřeba?

Samotné [předtrénování](#) modelu na internetu má své limity:

- **Nevhodný obsah:** Model se může naučit toxickému vyjadřování nebo předsudkům z webových fór.
- **Neochota pomoci:** Model sice ví, jak napsat recept, ale nemusí pochopit, že po něm uživatel chce instrukce, nikoliv jen pokračování příběhu o vaření.
- **Halucinace:** RLHF pomáhá trestat model za vymyšlení faktů.

Třífázový proces trénování

Proces RLHF se obvykle skládá ze tří klíčových kroků:

1. SFT (Supervised Fine-Tuning)

Lidé (anotátoři) napíší vzorové otázky a k nim ideální odpovědi. Model se na těchto příkladech učí „napodobováním“ správného stylu komunikace.

2. Tvorba modelu odměn (Reward Model)

Model vygeneruje několik různých odpovědí na stejnou otázku a člověk je seřadí od nejlepší po nejhorší. Na základě těchto žebříčků se vytrénuje druhý, menší model (Reward Model), který se naučí „chápat“, co se lidem líbí.

3. Optimalizace pomocí RL (PPO)

Hlavní model pak generuje tisíce odpovědí a za každou dostává od „modelu odměn“ body. Pomocí algoritmu **PPO** (Proximal Policy Optimization) se model upravuje tak, aby získával co nejvíce bodů. Tím se jeho chování postupně zlepšuje i bez přímého dozoru člověka v každém kroku.

Výhody a výzvy

Vlastnost	Dopad v praxi
Užitečnost	Model lépe dodržuje složité instrukce (např. „Napiš to stručně a v bodech“).
Bezpečnost	Účinnější odmítání požadavků na nelegální nebo nebezpečné činnosti.
Subjektivita	Výzva: Model se může naučit předsudkům lidí, kteří ho hodnotí (tzv. annotator bias).
Vyladění	Výzva: Příliš silné RLHF může vést k tomu, že je model až příliš opatrný a odmítá i nezávadné dotazy.

Význam pro moderní AI

RLHF je klíčovým důvodem, proč se **OpenAI** a **Google** staly lídry trhu. Tato technika umožnila přeměnit surový algoritmus v produkt, který může používat i laik, aniž by musel znát programování nebo složitý [prompt engineering](#).

— *Související pojmy:* [OpenAI](#), [ChatGPT](#), [Gemini](#), [Hluboké učení](#)

From:
<https://serviceit.cz/> - IT ENCYKLOPEDIE

Permanent link:
<https://serviceit.cz/doku.php?id=rlhf>

Last update: **2026/01/05 19:40**

