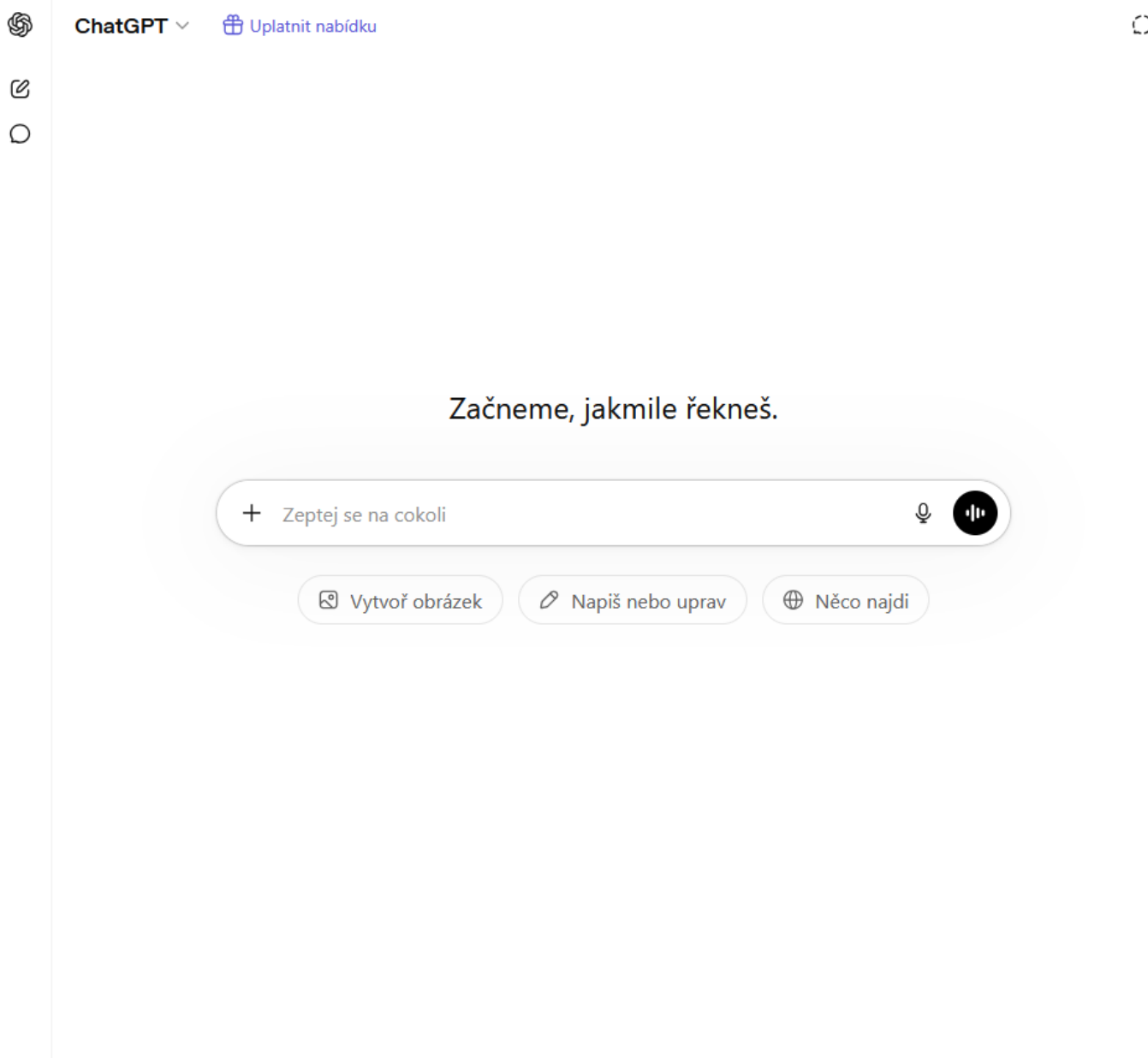


# ChatGPT

ChatGPT (Chat Generative Pre-trained Transformer) je pokročilý chatbot a asistent založený na velkém jazykovém modelu (LLM) vyvinutý výzkumnou společností OpenAI. Od svého uvedení v listopadu 2022 (kdy dosáhl 100 milionů uživatelů za pouhé dva měsíce) způsobil absolutní revoluci v tom, jak lidé a firmy interagují s počítači, generují text, programují, analyzují data a vyhledávají informace.

Z technického hlediska nejde o klasický „vyhledávač“ ani databázi faktů, ale o vysoce komplexní stochastický prediktivní model, který na základě rozsáhlého kontextu matematicky odhaduje nejpravděpodobnější následující část slova (tzv. token). Webové okno Chatgpt k 10. 6. 2026



## Architektura a fungování

ChatGPT stojí na několika klíčových technologických konceptech a pilířích:

## 1. Transformer

Jde o přelomový typ architektury **hlubokého učení** (představený společností Google v roce 2017), která využívá mechanismus pozornosti (Self-Attention). To umožňuje modelu chápat vztahy mezi slovy v dlouhých větách, i když jsou od sebe daleko (např. shoda podmětu s přísudkem napříč celým odstavcem) a zpracovávat text paralelně, na rozdíl od starších RNN modelů.

## 2. Tokenizace

Model nečte text po slovech, ale po tzv. tokenech. Token může být celé slovo (např. „pes“), ale i část slova nebo jen jeden znak (typické pro češtinu a složitější jazyky). 100 tokenů odpovídá zhruba 75 slovům v angličtině. Omezená „paměť“ modelu v jedné konverzaci se nazývá Context Window (kontextové okno) a u nejnovějších modelů dosahuje až 128 000 (nebo i více) tokenů, což je ekvivalent celé knihy.

## 3. Generative Pre-training (Předtrénování)

Model byl nejprve trénován (tzv. unsupervised learning) na obrovském clusteru GPU od společnosti **NVIDIA** na masivním množství textových dat z internetu (knihy, články, kód, fóra). Během této fáze si vytvořil vnitřní reprezentaci světa, naučil se gramatiku, programovací jazyky, fakta i různé styly psaní.

## 4. RLHF (Reinforcement Learning from Human Feedback)

Klíčový krok (tzv. alignment), který udělal z „naivního doplňovače textu“ užitečného a bezpečného pomocníka. Lidé hodnotili odpovědi modelu a přidělovali jim skóre, čímž ho učili:

Odpovídat strukturovaně, slušně a bezpečně.

Pečlivě dodržovat instrukce uživatele (tzv. system prompt).

Přiznat chybu nebo odmítnout nevhodný dotaz (např. generování malwaru nebo nenávistného obsahu).

## 5. MoE (Mixture of Experts)

Od verze GPT-4 model pravděpodobně nevyužívá jednu gigantickou neuronovou síť pro každý dotaz, ale architekturu „směsi expertů“. Model se skládá z několika menších podsítí (expertů) a při každém dotazu se aktivují jen ty relevantní (např. expert na Python, expert na historii). To dramaticky snižuje výpočetní nároky (inference) při zachování obrovského počtu parametrů.

## Ekosystém a podnikové integrace

Dnes již ChatGPT není jen textové okno, ale komplexní platforma:

API a Function Calling: Vývojáři mohou modely integrovat do vlastních aplikací. Model umí nejen odpovídat textem, ale umí zavolat externí funkci (API), např. zjistit aktuální počasí, provést SQL dotaz do databáze nebo založit tiket v Jira, a výsledek zpracovat.

Custom GPTs: Uživatelé si mohou vytvářet vlastní verze ChatGPT, kterým nahrají specifická data (např. interní manuály firmy) a dají jim specifické instrukce (např. „chovej se jako seniorní Linux administrátor“).

Advanced Data Analysis (Code Interpreter): ChatGPT má k dispozici virtuální prostředí (sandbox), kde dokáže sám psát a spouštět Python kód. Díky tomu umí analyzovat nahrané Excel tabulky, generovat grafy, konvertovat soubory nebo řešit složité matematické úlohy.

RAG (Retrieval-Augmented Generation): Klíčová technologie pro podnikovou sféru. ChatGPT nemá v parametrech aktuální data firmy, ale RAG mu umožňuje nejprve vyhledat relevantní dokumenty v podnikové databázi a teprve na jejich základě vygenerovat přesnou odpověď.

## Vývoj verzí modelu

OpenAI modely neustále iteruje. Trénink i běh (inference) vyžaduje extrémně výkonný hardware (AI akcelerátory, např. řady H100 a B200), nejčastěji vyráběný v továrnách [TSMC](#).

Verze modelu	Hlavní přínos a charakteristika
GPT-3.5	Původní verze dostupná zdarma. Rychlá, ale s omezeným logickým uvažováním a menším kontextovým oknem. Dnes již považována za zastaralou (legacy).
GPT-4	První multimodální model s vysokou přesností. Výrazně lepší v programování, logice a chápání složitého kontextu. Průlomový pro firemní nasazení.
GPT-4 Turbo	Rychlejší, levnější pro API, s kontextovým oknem 128k tokenů (cca 300 stran textu) a aktuálnějšími daty.
GPT-4o (Omni)	Nativně multimodální. Zpracovává text, zvuk i obraz současně a v reálném čase. Umožňuje plynulou hlasovou konverzaci s přirozenou intonací a nulovým zpožděním (reaguje jako člověk na telefonu).
OpenAI o1 / o1-mini	Zásadní změna paradigmatu. Tzv. Reasoning models. Před vygenerováním odpovědi „přemýšlejí“ (skrytý Chain of Thought). Jsou mnohem pomalejší, ale extrémně schopné v řešení komplexních programátorských, matematických a vědeckých úloh, kde GPT-4o selhává.

## Omezení a rizika

Nasazení ChatGPT v praxi nese řadu výzev, na které musí IT oddělení pamatovat:

Halucinace (Konfabulace): Model může s naprostou jistotou tvrdit věc, která není pravdivá. Neví, co neví, a raději si vymyslí přesvědčivě znějící nesmysl (např. neexistující příkaz v Linuxu nebo fiktivní judikát).

Znalostní cutoff a temporalita: Modely mají konkrétní datum, po kterém už nebyly trénovány. Přístup k internetu to částečně řeší, ale nativní znalosti jsou omezené.

Ochrana soukromí a únik dat (Data Leakage): Data vložená do běžné, neplacené verze chatu mohou

být použita k dalšímu trénování (slavný je případ, kdy vývojáři Samsungu nechtěně nahráli do ChatGPT tajný zdrojový kód). Řešením je ChatGPT Enterprise / Team, kde OpenAI garantuje, že data pro trénink nevyužije.

Prompt Injection (Jailbreaking): Bezpečnostní zranitelnost, kdy útočník pomocí speciálně navrženého vstupu přesvědčí model, aby ignoroval své původní instrukce a prozradil tajné informace nebo vygeneroval škodlivý obsah.

Autorská práva: Probíhají desítky soudních sporů (např. The New York Times vs. OpenAI) o to, zda trénink modelu na datech chráněných autorským právem splňuje definici „Fair Use“.

Ekologická stopa: Trénink a samotný běh modelů vyžaduje obrovské množství elektřiny a vody na chlazení datacenter.

## Význam pro IT průmysl

ChatGPT a jeho API zcela redefinovaly roli IT profesionálů:

Software Engineering & DevOps: Nástroje jako GitHub Copilot nebo Cursor (založené na LLM) generují boilerplate kód, píšou testy, hledají bugy a navrhnou refaktoring. Vývojář se stává spíše „architektem“ a „revizorem“ kódu.

Kybernetická bezpečnost: Generativní AI dokáže rychle analyzovat obrovské logy ze SIEM systémů, psát YARA pravidla a vysvětlovat deobfuskovaný malware. Útočníci ji naopak využívají k tvorbě vysoce personalizovaných phishingových e-mailů.

Zákaznická podpora a IT Helpdesk: Inteligentní asistenti na první linii podpory, kteří s využitím RAG prohledávají interní Knowledge Base (např. Confluence) a radí uživatelům s běžnými IT problémy dříve, než se tiket dostane k živému operátorovi.

Vznik nových rolí: Do popředí se dostává **Prompt Engineering** (umění správně psát zadání pro AI) a AI Integrator (specialista na propojování firemních systémů s LLM modely).

## Hlavní konkurenti na trhu

ChatGPT nezůstal osamocen. Současný trh LLM zahrnuje silné alternativy:

Claude (Anthropic): Považován za bezpečnější, s lepším a „lidštějším“ stylem psaní, vynikající pro programování (Claude 3.5 Sonnet).

Gemini (Google): Hluboce integrovaný do ekosystému Google (Workspace, Cloud).

Llama (Meta/Facebook): Nejsilnější open-weight (volně dostupný) model, který si firmy mohou stáhnout a provozovat na vlastních serverech s absolutní kontrolou nad daty (On-Premise).

*Související pojmy: [LLM \(Large Language Model\)](#), [Hluboké učení](#), [RAG](#), [Prompt Engineering](#), [SoC](#), [TSMC](#), [NVIDIA](#), [Broadcom](#) (síťové prvky pro AI)*

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

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

Last update: **2026/06/10 12:21**

