

Co je co v AI

Slova, která kolem sebe slyšíte, srovnaná od toho nejmenšího uvnitř až po to, co vám běží samo, i když jste pryč.

Je to jako u auta: **motor sám nikam nejede**. Kolem něj musí být auto, ve kterém se dá sedět, v autě řidič, který ví, kam jede, a ve firmě dispečink, který ho pošle na cestu, i když majitel zrovna spí. U AI je to stejné - model, aplikace, agent a automatizace nejsou soupeři, je to jedna vrstva kolem druhé. Většina zmatku v člancích vzniká z toho, že se ty vrstvy míchají dohromady.

KDE JSTE VY

Kdo píše do ChatGPT nebo Claude v prohlížeči, je ve třetí vrstvě - a vrstvy pod ní ani nemusí vidět. **Osm vrstev je volba tohoto plakátu, ne fakt o AI**: jiný autor by je rozdělil na šest nebo dvanáct a nebyl by to spor. Čísla také nejsou pořadí, ve kterém se to zavádí - je to pořadí, ve kterém to na sobě stojí.



1

Model

VELKÝ JAZYKOVÝ MODEL (LLM)

VYBÍRÁTE TÍM, KTEROU APLIKACI OTEVŘETE

Vycvičený předvídač textu. Dostane text a počítá, co nejspíš následuje - jako když telefon při psaní nabízí další slovo, jen mnohem chytřeji. Nemyslí, počítá.

Sám o sobě nemá paměť, nevidí vaše soubory a neví, co je dnes za den. Všechny vrstvy nad ním existují proto, aby mu to dodaly.

SLOVA, KTERÁ SE V TÉTO VRSTVĚ OBJEVÍ

GPT *Generative pre-trained transformer* - druh modelu, který text tvoří a předem se učí z obrovského množství textu. Modely od OpenAI. Pohánějí ChatGPT a spoustu firemních aplikací.

Claude Modely od Anthropicu. Unesou v jednom zadání hodně textu, takže si poradí i se složitějším rozbořením.

Gemini Modely od Googlu. Stejně jméno nese i jejich aplikace, proto ten zmatek.

Llama Modely od Mety. Smíte si je stáhnout a pustit na vlastním počítači.

Mistral Francouzský výrobce. V Evropě se zmiňuje kvůli tomu, kde leží data.

DeepSeek, Qwen, Kimi Čínští výrobci: Qwen je od Alibaby, Kimi od Moonshotu, DeepSeek stojí samostatně. Levné a schopné, ale text jde do Číny.

Grok Modely od xAI, napojené na síť X (dřív Twitter) - mají tak přístup k čerstvému dění.

CO SI Z TOHO ODNĚST

LLM (*large language model*, velký jazykový model) je název druhu, ne značky - tak jako „automobil“, ne „Škoda“. Pro běžnou kancelářskou práci se dnešní modely liší míň, než se zdá. Země výrobce ale říká, kam váš text odchází a čí zákony na něj platí.



2

Kde model běží

CIZÍ POČÍTAČ, NEBO VÁŠ

ÚČET U VÝROBCE SAMI, VLASTNÍ SERVER S POMOCÍ

Model je jeden obrovský soubor - takový, že by se stahoval hodiny a mailem ho nepošlete. Buď běží na cizím počítači, který jede nonstop (tomu se říká server), nebo si ho stáhnete k sobě.

Když píšete do ChatGPT v prohlížeči, běží model na cizím serveru a váš text tam odchází. Proto se předem rozmyslí, co tam nepatří: mzdy, smlouvy, osobní údaje zákazníků.

SLOVA, KTERÁ SE V TÉTO VRSTVĚ OBJEVÍ

Napojení pro programy (API) *Application programming interface*, rozhraní pro programy. Přístup, přes který si s modelem povídá jiný program, ne člověk. Platí se podle spotřeby, ne měsíčním paušálem.

Azure OpenAI, AWS Bedrock Konkrétní jména pronajatého prostoru - u Microsoftu a u Amazonu (*AWS, Amazon Web Services*). Model je tentýž; firmy to volí, když chtějí mít data v Evropě, třeba v datovém centru ve Frankfurtu.

Ollama, LM Studio Programy, kterými si model spustíte na vlastním notebooku. Nic neodchází ven, ale výkon i kvalita jsou nižší.

Hugging Face Není to místo, kde model běží, ale odkud se bere: veřejná knihovna, ze které si model stáhnete, než ho pustíte u sebe.

CO SI Z TOHO ODNĚST

„Otevřený model“ si smíte stáhnout a pustit na svém počítači. Uzavřený běží jen u výrobce - používat ho smíte, dovnitř nevidíte. To je celý rozdíl, o kterém se v článcích píše jako o open-source AI.



3

Aplikace

CHAT OKNO, DO KTERÉHO PÍŠETE

ZVLÁDNETE SAMI, STAČÍ UMĚT PSÁT

Model je uvnitř, ale aplikace k němu přidává historii rozhovoru, nahrávání souborů, hledání na webu, hlas a účet, ze kterého se platí.

GPT je model, ChatGPT je aplikace nad ním.

Proto se dvě věci se stejným motorem chovají jinak. A rozdíl, který v práci opravdu ucítíte, dělá častěji aplikace než model: co umí nahrát, najít a zapamatovat si.

SLOVA, KTERÁ SE V TÉTO VRSTVĚ OBJEVÍ

ChatGPT Chat okno od OpenAI. Nejrozšířenější, proto se jeho jméno používá jako obecné označení pro AI.

Claude.ai Chat okno od Anthropicu. Umí nahrát a prohledat vlastní dokumenty.

Gemini Chat okno od Googlu, propojené s Gmailem a Dokumenty.

Microsoft Copilot AI přímo ve Wordu, Excelu a Outlooku - pracujete v programu, který už znáte.

Perplexity Hledání, které u každé odpovědi ukáže zdroje. Dobré tam, kde si potřebujete ověřit fakta.

NotebookLM Aplikace, která se ptá jen vašich vlastních dokumentů. Neodpovídá z internetu.

CO SI Z TOHO ODNĚST

Tady je většina lidí. Jedno jméno pro model i aplikaci (Gemini, Grok) je marketing, ne technika - vrstvy to nespojí. Proto srovnávat „ChatGPT a Claude“ znamená srovnávat dvě vrstvy naráz.



4

Agent

AI, KTERÁ JEDNÁ

NAINSTALUJETE SAMI, ALE CHCE TO ZVYK

Nejen odpovídá - dělá kroky. Čte a píše soubory, spouští programy, prochází web, po chybě to zkusí jinak a jde k cíli.

Agent je ten řidič z úvodu - ten, kdo jede sám za cílem. Chatbot vám jen poradí po telefonu, co máte udělat; **agent to udělá sám - jako byste brigádníkoví dali klíče od dílny**. Řidič ukazuje, že jede bez dozoru; klíče ukazují, kolik smí bez dovolení. Proto se u něj poprvé řeší, k čemu ty klíče má a kdy se musí zeptat.

SLOVA, KTERÁ SE V TÉTO VRSTVĚ OBJEVÍ

Claude Code Agent od Anthropicu, který běží v příkazovém okně: místo klikání se mu píše. Umí sáhnout na soubory ve složce, kterou mu dáte.

Codex CLI, Gemini CLI Totéž od OpenAI a od Googlu. CLI (*command line interface*) je příkazové okno - černá plocha, do které se píše místo klikání. Liší se detaily, princip je stejný.

Cowork Agent od Anthropicu na dokumenty místo na programování. Běží na jeho serverech; do složky na vašem počítači sáhne jen s vaším svolením a jen když máte spuštěnou aplikaci.

Cursor, Copilot Agent Agenti uvnitř programu, ve kterém se píše kód. Určené pro vývojáře, ne pro kancelář.

OpenClaw Osobní agent, kterému se píše do WhatsAppu nebo Telegramu. Provozuje si ho každý sám - a je to zdaleka nejmiň hotová věc v této vrstvě: špatně zabezpečených instalací se dá na internetu najít tisíce.

CO SI Z TOHO ODNĚST

Slovo „agent“ se dnes lepí na cokoli. Poznáte to jedinou otázkou dodavatelů: udělá to za mě, nebo mi jen napíše, co mám udělat? Kdo jedná sám, je agent. Zbytek je chat.



5

Kontext a paměť

ODKUD BERE VAŠE FAKTA

VLASTNÍ DOKUMENTY SAMI, FIREMNÍ DATA S POMOCÍ

Model neví nic o vaší firmě. Tady se mu to dodává - buď rovnou do rozhovoru, nebo dohledáním z vašich vlastních dokumentů.

Aplikace si o vás pár údajů uloží, ale dlouhý rozhovor není paměť. Trvale platí jen to, co je zapsané v souboru nebo v úložišti dat (databázi). **Zkouška:** nahrajte ceník a tři poslední nabídky. Když se vás AI přestane ptát na to, co v nich stojí, tato vrstva u vás funguje.

SLOVA, KTERÁ SE V TÉTO VRSTVĚ OBJEVÍ

Kontextové okno Kolik textu model vidí najednou. Co se do něj nevejde, vypadne - a nikdo vás na to neupozorní.

Vlastní instrukce, projekty Trvalé zadání, které platí v každém rozhovoru, aniž ho pokaždé opisujete.

Soubor s pravidly Totéž psané rovnou u dat, se kterými se pracuje. Agent si ho načte sám, než začne.

RAG *Retrieval-augmented generation*, odpověď obohacená o vyhledání. K vaší otázce se nejdřív dohledá správný kus vašich dokumentů a teprve pak se odpovídá. Jako asistentka, která před poradou položí na stůl tu správnou složku.

Embedding Převod textu na čísla, aby šlo hledat podle významu, ne podle přesných slov. Napíšete „reklamace“ a najde se i „stížnost zákazníka“.

Vektorová databáze Sklad těch čísel, ve kterém se hledá rychle i mezi tisíci dokumenty.

CO SI Z TOHO ODNĚST

RAG není chytřejší model a nic se jím nedoučí - jen dostane k otázce správnou stránku. Pro většinu firem je to levnější a rychlejší cesta než doučování modelu na vlastních datech.



6

MCP a API

NAPOJENÍ A NÁSTROJE – RUCE SYSTÉMU

HOTOVÁ NAPOJENÍ SAMI, VLASTNÍ S POMOCÍ

Bez nich AI jen mluví. S nimi si přečte mail, sáhne do kalendáře, otevře stránku, připraví fakturu.

MCP (*model context protocol*) je pro AI to, co USB pro počítač: jeden dohodnutý tvar zásuvky, do kterého jde zapojit cokoliv. Bez něj se každé napojení programuje zvlášť - proto ho dnes podporuje víc výrobců naráz.

SLOVA, KTERÁ SE V TÉTO VRSTVĚ OBJEVÍ

MCP *Model context protocol*, protokol pro předávání kontextu modelu. Dohodnutý tvar zásuvky, kterým se k AI připojují nástroje. Co ho podporuje, jde zapojit bez programátora.

API *Application programming interface*, rozhraní pro programy. Seznam úkonů, které jeden program nabízí druhému - třeba e-shop bance při platbě kartou. Starší cesta než MCP: každá služba má vlastní.

Konektory, integrace Hotová napojení na běžné služby - Gmail, Google Disk, Slack. Zapnete je klikáním a povolíte, co smí vidět.

Nástroje / funkce Seznam úkonů, které AI smí použít: vyhledat, poslat, zapsat. Co v seznamu není, udělat nemůže.

API klíč Heslo pro program, ne pro člověka. Kdo ho má, používá vaše napojení a platí to váš účet.

Webhook Opačný směr než dotaz: cizí služba se ozve sama, jakmile se něco stane. Přišla platba, přibyl soubor - a linka se rozjede.

CO SI Z TOHO ODNĚST

Klíč patří do správce hesel, ne do chatu a ne do sdílené tabulky. A pozor na podvržený pokyn: v mailu nebo na stránce může být text, který agentovi nakáže něco jiného, než chcete vy - a on nepozná, že není od vás.



7

Skill a hook

PRAVIDLA A ROZŠÍŘENÍ – JAK SI TO OHNETE PRO SEBE

PSANÁ PRAVIDLA SAMI, AUTOMATICKÉ ZÁBRANY S POMOCÍ

Model bez pokynů se chová průměrně. Tady se určuje, co má dělat vždycky, co nikdy a v jakém pořadí.

Cedule „zhasínejte“ je prosba - platí, jen když si ji někdo přečte. Pohybové čidlo zhasne samo. **Napsané pravidlo je cedule, hook je to čidlo.**

SLOVA, KTERÁ SE V TÉTO VRSTVĚ OBJEVÍ

Prompt Zadání úkolu. Ne kouzelná formulka: popis toho, co chcete, co se nesmí a jak má výsledek vypadat.

Systémový prompt Zadání, které platí pořád a v každém rozhovoru. Určuje roli, tón i to, čeho se AI nemá dotýkat.

Skill Hotový návod na opakovanou úlohu. Leží stranou a načte se, teprve když na tu úlohu dojde.

Hook Zásah, který se spustí sám před akcí nebo po ní - nezávisle na tom, co si AI myslí. Tím se z prosby stává zábrana.

Mantinely Co agent nesmí, i kdyby ho o to někdo požádal. Typicky mazání dat, platby, odchází komunikace.

Subagent Pomocník s vlastním úkolem a odděleným zadáním. Odvede kus práce a vrátí jen výsledek, ne celý svůj postup.

CO SI Z TOHO ODNĚST

Tato vrstva je rozdíl mezi „používám AI“ a „mám vlastní systém“. Bez ní začínáte pokaždé znovu a pokaždé stejně vysvětľujete, jak se u vás věci dělají.



8

Automatizace

KDO TO SPUSTÍ, KDYŽ U TOHO NESEDÍTE

KLIKÁNÍM SAMI, SLOŽITĚJŠÍ LINKY S POMOCÍ

Propojení kroků do linky: přijde mail → vytáhne se příloha → AI ji shrne → zápis do tabulky → zpráva vám. Bez jediného kliknutí.

Automatizace nedělá AI chytřejší, jen rychlejší. Chyba, která se dřív stala jednou za den, se teď stane i ve tři ráno a nikdo u toho není. Proto se automatizuje až to, co ověřeně funguje.

SLOVA, KTERÁ SE V TÉTO VRSTVĚ OBJEVÍ

n8n Propojovač kroků, který se skládá klikáním. Dá se provozovat u sebe, ale pak se o jeho běh staráte sami.

Make, Zapier Totéž jako služba: běží to za vás a platí se měsíčně. Rychlejší start, menší kontrola.

Power Automate Propojovač uvnitř Microsoftu. Umí i naplánovaný běh.

Naplánovaný běh agenta Claude Code i Codex se dnes umí probudit podle kalendáře a odvést práci samy. Propojovač k tomu nepotřebujete.

Spouštěč události Nespouští to čas, ale událost: přišel mail, přibyl soubor, změnil se řádek v tabulce.

CO SI Z TOHO ODNĚT

Propojovač a agent dělají totéž jinak: propojovač jede podle nakliknutého schématu, agent podle zadání. Schéma je předvídatelné, agent si poradí i s tím, co jste nenaklikali - a právě proto potřebuje kontrolu.

A nad tím vším: kontrola

Každá vrstva přidává jednu schopnost - a jedno nové místo, kde se to může tiše pokazit. Proto k tomu patří ještě jedna věc navíc. Není to nástroj, je to návyk. **Sebejistě vyslovená nepravda není porucha, kterou půjde spravit:** plyne z toho, jak model pracuje - počítá pravděpodobné pokračování, ne pravdu. Čím víc vrstev nad sebou, tím dál se chyba stihne rozjet, než si jí někdo všimne.

1 Kontrolujte tam, kde to skončí

V odeslaném mailu, v tabulce, na webu.
Ne v okně chatu a ne podle toho, co o své práci tvrdí AI

2 Poslední slovo má člověk

U všeho, co jde ven z firmy nebo co se nedá vzít zpět

3 Pište si datum ověření

U všeho, co stárne - ceny, verze, jména produktů. Tento plakát taky

Slova, na která narazíte v článcích

Token

Kousek slova, se kterým model počítá. Podle tokenů se platí: delší zadání i delší odpověď zvedají účet. A protože model nepracuje s písmeny, neumí spolehlivě říct, kolik jich ve slově je.

Halucinace

Sebejistě vyslovená nepravda. Není to porucha, kterou půjde spravit: plyne z toho, že model počítá pravděpodobné pokračování. Poznáte ji ověřením, z tónu odpovědi nikdy.

Server

Cizí počítač, který běží nonstop a někdo platí jeho provoz. „Běží to v cloudu“ neznamená nic víc než toto - jede to na cizím počítači, ne na vašem.

Váhy modelu

Ty naučené hodnoty jako soubor ke stažení. Když jsou volně dostupné, smíte si model pustit u sebe; jinak se k němu dostanete jen přes výrobce.

Reasoning model

Model, který si nejdřív projde úvahu a teprve pak odpoví. Trvá to déle a stojí víc, ale na složitém zadání udělá míň chyb.

Prompt

Zadání. Nejde o kouzelná slova, ale o tři věci: co chci, co se nesmí a jak to má vypadat. Čím míň prostoru necháte, tím míň si model domýšlí.

Bez programování

Nástroj, kde se řešení skládá klikáním. Platí to o propojovačích i o většině AI aplikací - programátor je potřeba až u napojení šitého na míru. Anglicky no-code.

AI Act

Evropské nařízení k umělé inteligenci. Po firmě chce hlavně vědět, kde a jak AI používá; na běžné psaní do chatu dopadá nejmírněji, na rozhodování o lidech nejpřísněji. Povinnost postarat se o proškolení lidí, kteří s AI pracují, ale platí bez ohledu na to, k čemu ji firma používá.

Kontextové okno

Kolik textu udrží najednou před očima. Co se do něj nevejde, vypadne - a model to neohlásí. Prostě odpoví, jako by tam ta část nikdy nebyla.

API

Způsob, jak si mezi sebou povídají dva programy bez člověka u toho. Když někdo řekne „napojíme to přes API“, myslí tím, že si to spolu bude data předávat samo, bez kopírování a vkládání.

Parametry (7B, 70B)

Počet naučených hodnot uvnitř modelu; B je miliarda. Více neznamená automaticky lepší - znamená dražší provoz a silnější počítač pod tím.

Fine-tuning

Doučení modelu na vlastních datech. Drahé a rychle zastarává - většine firem stačí dobře dodané vlastní dokumenty (vrstva 5).

Multimodální

Umí i obrázek, zvuk nebo video, nejen text. V praxi: vyfotíte fakturu nebo obrazovku a rovnou se na ni zeptáte.

Podvržený pokyn

Cizí příkaz schovaný v textu, který AI čte - v mailu, na stránce, v dokumentu. Agent ho může poslechnout, protože nerozliší vaše zadání od cizího. Anglicky prompt injection.

Osobní údaje

Jméno, e-mail, mzda, zdravotní údaj zákazníka nebo zaměstnance. Do cizí AI nepatří jen tak - platí pro ni stejná pravidla jako pro každého jiného dodavatele, kterému data svěříte.

Destilace

Malý model se učí od velkého. Levnější a rychlejší provoz, o něco horší výsledky - poznáte ho podle „mini“, „flash“ nebo „lite“ v názvu.