Nová třída modelů AI v ostrém provozu: co znamená nasazení systému JEV v českém e-shopu

25.9.2026 - | Česká zemědělská univerzita v Praze

Katedra informačního inženýrství Provozně ekonomické fakulty ČZU sleduje a odborně doprovází nasazení modelu JEV u e-shopu Mironet.cz. Jde o jeden z prvních případů, kdy se v Evropě dostává do komerčního provozu model, který záměrně nepíše text. Pro veřejnost je to příležitost porozumět tomu, proč se část výzkumu umělé inteligence vrací od generování jazyka k dobře změřené nejistotě.

Model, který neodpovídá větami

Velké jazykové modely jako ChatGPT, Gemini nebo Claude tvoří odpověď slovo po slovu — každý krok musí počkat na předchozí. Je to univerzální, ale pro některé úlohy zbytečně drahé a pomalé.

Model **JEV**, který 15. září 2026 uvedla americká společnost TypeSafe AI, jde jinou cestou. Nedostanete z něj větu, ale odpověď na přesně položenou otázku: vyber jednu z možností, ohodnoť na škále, odpověz ano/ne. A ke každé odpovědi **rozdělení pravděpodobností** — nejen „tento produkt patří mezi notebooky", ale i „na 92 %". Všechny položené otázky navíc vyhodnotí **naráz, v jediném průchodu**. Odtud odezva v desetinách sekundy a výrazně nižší cena: protože model negeneruje text, výstupní tokeny — u běžných modelů ta dražší polovina účtu — vůbec nevznikají.

Název je poctou ekonomovi **Williamu Stanleymu Jevonsovi** a jeho paradoxu, podle nějž zlevnění zdroje nevede k úspoře, ale k intenzivnějšímu využívání.

Proč je to vědecky zajímavé — a v čem naopak novinka není

Je lákavé číst JEV jako „zcela nový druh AI". Přesnější je, že jde o neobvykle dobře zabalenou kombinaci věcí, které věda zná desítky let. Kořeny sahají k **rozhodování s možností odmítnout odpověď** — teoretický základ položil C. K. Chow už v roce **1970**, na neuronové sítě jej přenesli Geifman a El-Yaniv v roce 2017. Systém nemusí odpovídat vždy: když si není dost jistý, má právo mlčet a předat případ člověku.

Druhým pilířem je **kalibrace pravděpodobností**. Model je kalibrovaný, když jeho čísla znamenají to, co říkají: prohlásí-li tisíckrát „na 90 %", mělo by se to zhruba v devíti stech případů potvrdit. Běžné jazykové modely tuto vlastnost nemají, a mají k tomu systémový důvod: trénují se na lidské zpětné vazbě a lidem se líbí odpovědi, které znějí sebejistě. Tréninkový tlak tak odměňuje sebevědomí a trestá přiznanou nejistotu — že doladění na lidské preference kalibraci zhoršuje, doložila i technická zpráva k modelu GPT-4.

Firma uvádí, že JEV trénovala metodou **RLCD**, která místo lidských preferencí optimalizuje přímo kvalitu odhadu pravděpodobnosti. Vědecký článek popisující RLCD ale zatím nevyšel a firma nezveřejnila ani účelovou funkci, ani zdrojový kód — nezávisle zopakovat se to tedy nedá.

Čísla: kdo co naměřil

Veřejně se pohybují tři sady čísel a znamenají různé věci.

Zdroj	Zrychlení	Zlevnění	Poznámka
TypeSafe AI (vlastní testy)	193,6×	444,6×	Vlastní úlohy, referencí shoda dvou jiných modelů, nesprávná odpověď. Firma sama píše, že jde o hodnoty „na horní hranici reálných přínosů".
Mironet (vlastní provoz)	19× (32,4 › 1,7 s)	343×	Měřeno na vlastní úloze kategorizace, proti předchozímu řešení.
Nezávislé srovnání (JevBench)	—	~6× na 1 000 rozhodnutí	Proti běžnému rychlému modelu; skladba úloh je volbou autorů.

Že je Mironetem naměřené zrychlení desetkrát nižší než dodavatelova hlavička, **není rozpor, ale normální stav**: Laboratorní test se optimalizuje na rozdíl, reálný provoz zahrnuje i vše ostatní. Řádové zlepšení z marketingového materiálu se v praxi obvykle scvrkne — a přesto může být významné. Bez výhrad ověřitelný zůstává ceník: 0,042 USD za milion vstupních tokenů, výstup zdarma.

Jedno nedorozumění, které stojí za to vyjasnit

O modelech tohoto typu se často píše, že „nemohou halucinovat". **Platí to jen napůl, a ten rozdíl je v praxi zásadní.**

Model si nemůže vymyslet kategorii, která neexistuje — vybírá jen z toho, co mu zadáte. To je spolehlivost formy. Nic mu ale nebrání s naprostou jistotou zařadit produkt do **existující, jen úplně špatné** kategorie. Chyba se přesouvá z formy do obsahu, kde je hůř vidět, protože výstup vypadá bezvadně.

Nezávislá měření z prvního týdne po uvedení to potvrzují a přidávají tři varování: **kvalita kalibrace se liší úlohu od úlohy**, **model se nezdrží odpovědi, pokud mu možnost „nevím" nenabídnete**, a **přesnost citelně závisí na formulaci otázky**. Jde o rychlé komunitní testy, ne o recenzované studie — ta o modelu JEV zatím neexistuje.

Nejvíc řekne jeden test, který obě strany postavil vedle sebe. Na úloze **„patří tento text do této kategorie?"** model prošel všemi šesti předem stanovenými kritérii. Na úloze **odstupňované relevance produktu vůči vyhledávacímu dotazu** — tedy na jádru vyhledávání v e-shopu — **neprošel u čtyř ze šesti**: chyba kalibrace vystoupala na 0,24, model byl přesvědčený o sobě ve všech pásmech jistoty a naprostou většinu dvojic nacpal do jediné odpovědi.

Stojí za to číst to přesně. **Není to nález proti tomu, co nasadil Mironet.** Kategorizace zboží je právě ta první úloha — zařazení do kategorie —, kde model uspěl. Odstupňovaná relevance je úloha podstatně těžší a model na ni podle dostupných měření zatím nestačí. Pro e-shop z toho plyne užitečné vodítko: **jednoduchá a ohraničená rozhodnutí ano, jemné odstupňování zatím ne.** Měření z ostrého provozu, která to potvrdí nebo vyvrátí, teprve vzniknou.

Souvisí s tím riziko, o kterém se mluví méně: jednoprocentní chybovost v popisu produktu je nepříjemnost, táž chybovost v rozhodnutí, které se během sekund promítne do desetitisíců položek, je provozní incident.

Kde to naopak pomáhá evropské regulaci

Akt EU o umělé inteligenci vyžaduje u rizikových aplikací **lidský dohled**, a model vracející holé číslo se jeví jako krok špatným směrem. Kalibrovaná jistota ale nabízí něco, co volný text ne: **jasný, auditovatelný práh**. Organizace může stanovit, že pod určitou mírou jistoty systém nesmí rozhodnout sám a případ putuje člověku — a to se dá zapsat do směrnice, doložit logem a zkontrolovat. Dobře změřená nejistota je pro dohled použitelnější než výřečné, ale nezávazné odůvodnění.

Role katedry informačního inženýrství ČZU

Katedra informačního inženýrství Provozně ekonomické fakulty ČZU a její iniciativa AI BRIDGE dlouhodobě spolupracují s praxí v rámci **Technologické platformy IT People**, která propojuje technologické firmy s akademickými pracovišti. Vedle ČZU se na spolupráci podílejí **katedra datové analytiky Fakulty managementu VŠE** a **katedra informatiky Přírodovědecké fakulty UJEP**; aplikačním partnerem je v tomto případě e-shop **Mironet.cz**, který uvádí, že jde o první komerční nasazení modelu v Evropě.

Úloha katedry je záměrně skromnější: **sledovat, měřit a zasadit do kontextu** — tedy ověřit to, co si firma sama neověří, protože k tomu nemá důvod. Jak se kalibrace chová mimo laboratoř, o Vánocích nebo při rozšíření sortimentu. Jak nastavit prahy jistoty, aby lidský dohled nebyl formalita. A jak si model vede v češtině: dosavadní měření naznačují, že u neanglických jazyků přesnost klesá, což je pro český e-shop otázka zcela praktická.

Takové nasazení je cenné právě proto, že je skutečné; provozní data se nedají nahradit simulací. Zároveň platí, že **z jednoho nasazení se nedělají obecné závěry** — a to je práce, kterou má akademické pracoviště odvést dřív, než se z prvního dojmu stane sdílená pravda.

Zdroje

- Chow, C. K.: *On optimum recognition error and reject tradeoff.* IEEE Transactions on Information Theory 16(1), 1970.
- Geifman, Y., El-Yaniv, R.: *Selective Classification for Deep Neural Networks.* NeurIPS 2017, arXiv:1705.08500.
- Guo, C., Pleiss, G., Sun, Y., Weinberger, K.: *On Calibration of Modern Neural Networks.* ICML 2017, arXiv:1706.04599.
- Brier, G. W. (1950); Gneiting, T., Raftery, A. E.: *Strictly proper scoring rules, prediction, and estimation.* JASA 102(477), 2007.
- Ovadia, Y. a kol.: *Can You Trust Your Model's Uncertainty?* NeurIPS 2019, arXiv:1906.02530.
- Ouyang, L. a kol.: *Training language models to follow instructions with human feedback.* NeurIPS 2022, arXiv:2203.02155.
- TypeSafe AI: dokumentace a vlastní srovnávací testy, `typesafe.ai`, `evals.typesafe.ai`.
- Komunitní nezávislé testy modelu JEV z 20.–21. září 2026: **jev-orderby-bench** (předem registrovaná kritéria; 20 Newsgroups, 360 položek; Amazon ESCI, 306 lidmi ohodnocených dvojic dotaz–produkt ve 30 obtížných dotazech), **Benchmark Heaven JevBench v1.3.0**, **PriorBench**, kalibrační audity.
- Mironet.cz: tisková zpráva k nasazení modelu JEV.

Katedra informačního inženýrství, Provozně ekonomická fakulta, Česká zemědělská univerzita v Praze

eni-sy.html