

Case study: 1000 otázok pre AI účtovníka

2026-09-20 · @Someone

Zhrnutie

Keď účtovník dostane od AI nesprávnu odpoveď, nie je to trápna drobnosť — je to riziko pokuty, ktorú zaplatí on alebo jeho klient. Preto sme presnosť nášho AI účtovníka nemerali reklamnými frázami, ale najväčším verejným testom odbornej presnosti AI, aký na Slovensku vznikol: **vyše 1 000 skutočne ťažkých otázok z účtovnej a daňovej praxe**, každá s odpoveďou overenou proti Zbierke zákonov a metodikám Finančnej správy — a s doslovným citátom predpisu ako dôkazom.

Test nie je marketingový trik s vybranými otázkami: otázky miera presne tam, kde sa mýlia aj skúsení účtovníci — prechodné ustanovenia noviel, hospodárske roky, prahy a ich výnimky, sadzby, ktoré sa za posledné tri roky menili aj trikrát. Celý test vrátane otázok, správnych odpovedí, zdrojov a výsledkov všetkých meraných systémov zverejňujeme — ktokoľvek si každé tvrdenie môže overiť klikom do zákona.

A jedno priznanie vopred: test odhalil chyby aj u nás. Namiesto zamlčania sme ich opravili a opravy zdokumentovali v tomto dokumente — pretože presne v tom je rozdiel medzi všeobecnou AI a systémom, za ktorý niekto ručí: **chybu, ktorú raz nájdeme, opravíme pre všetkých zákazníkov navždy.**

Ako test vznikol

Banka otázok vznikala v štyroch vlnách a pokrýva **38 oblastí praxe** — DPH, daň z príjmov, mzdy a odvody, sociálne a zdravotné poistenie, cestovné náhrady, správu daní, účtovné predpisy, e-commerce, medzinárodné zdaňovanie a ďalšie. Zámerne sme vynechali triviálne otázky typu „aká je sadzba DPH“ — každá otázka musela prejsť testom „*pomýlil by sa na tom aj skúsený účtovník?*“: prechodné ustanovenia, hraničné prahy, výnimky z výnimiek, lehoty s posunom na pracovný deň.

Každá otázka v banke má povinne päť častí:

1. **Otázku v dvoch podobách** — odborne presnú a „voľnú“, teda tak lajdácky, ako by ju položil podnikateľ v praxi.
2. **Správnu odpoveď** s kľúčovými hodnotami (sumy, sadzby, dátumy).
3. **Presné ustanovenie** — nie len zákon, ale § a odsek.
4. **Doslovný citát predpisu** ako dôkaz.

5. Funkčný odkaz na oficiálny zdroj — Slov-Lex, Finančná správa, Sociálna poisťovňa, ministerstvá.

Kvalitu sme nestrážili jedným párom očí, ale tromi na sebe nezávislými kontrolami:

- **Nezávislé overenie:** každú vygenerovanú otázku preveril iný systém než ten, ktorý ju vytvoril — s prístupom na web, s príkazom prepočítať každý výpočet a pri akejkoľvek pochybnosti otázku zamietnuť. Radšej menej otázok než jedna pochybná.
- **Kritik vzorky:** náhodná vzorka každej vlny prešla ešte jednou kontrolou navyše.
- **Adverzariálny audit celej banky:** pred zverejnením prešla každá jedna otázka záverečnou kontrolou zámerne skeptického audítora, ktorý dostal zadanie „predpokladaj, že chyba tam je, a nájdi ju“. Preveroval sedem vecí v pevnom poradí: či nie je chybný už samotný predpoklad otázky, či uvedený § a odsek existuje a upravuje to, čo otázka tvrdí, či je citát doslovný, či odkaz funguje a vedie na oficiálny zdroj, či znenie platí k dnešku, či sedí aritmetika a či má otázka jedinú obhájitelnú odpoveď.

Audit nebol formalita: pri stovkách otázok našiel a opravil aj chyby, ktoré prvé overenie prepustilo — zlé číslo odseku, parafrázu vydávanú za citát, zamenenú sadzbu v predpoklade otázky. Presné počty uvádzame vo výsledkoch nižšie.

Ako sa meria

Každú otázku dostávajú traja súper — vždy vo „voľnej“ podobe, teda tak, ako sa reálne pýtajú podnikatelia, nie v učebnicovej formulácii:

1. **Všeobecná AI bez zdrojov** — špičkový veľký jazykový model, tak ako ho používa bežný človek: bez prístupu na web, odpovedá z toho, čo si „pamätá“.
2. **Všeobecná AI s vyhľadávaním Google** — ten istý model s prístupom na web.
3. **SmartUctovnik.sk** — naša produkčná aplikácia, presne tá istá, ktorá odpovedá zákazníkom: s vlastnou databázou slovenských predpisov, overenými referenčnými hodnotami a povinným citátom zdroja.

Odpovede nehodnotíme sami — posudzuje ich **nezávislý sudca**: ďalší AI model, iný než ktorýkoľvek zo súperov, nastavený deterministicky (žiadna „kreativita“). Sudca porovná odpoveď so správnou odpoveďou a kľúčovými hodnotami z banky a vydá verdikt správne/nesprávne s odôvodnením.

Čo znamená „správne“: odpoveď musí obsahovať vecne správny záver **vrátane kľúčových čísel** — súm, sadzieb, dátumov. Vyhýbavé odpovede typu „závisí to od okolností, poraďte sa s odborníkom“ sa nepočítajú ako správne — účtovník potrebuje odpoveď, nie alibi.

Jedno dôležité pozorovanie z opakovaných meraní: výsledky všeobecnej AI s vyhľadávaním **kolíšu medzi behmi** — tá istá sada otázok dopadla v rôznych dňoch výrazne rozdielne podľa toho, čo vyhľadávač práve našiel. Presne preto meriame na veľkej vzorke a

zverejňujeme celý protokol: jednorazové demo sa dá vybrať, tisíc otázok nie.

Výsledky

Meranie prebehlo na **celej banke 1 066 otázok** 21. 9. 2026. Všetky tri systémy dostali rovnaké otázky v rovnakom znení a posudzoval ich ten istý nezávislý sudca podľa rovnakej prísnej rubriky.

Systém	Úspešnosť	Správnych otázok	Chyby na 1 000 otázok
SmartUctovnik.sk	94,5 %	1 007 z 1 066	≈ 55
Všeobecná AI s vyhľadávaním Google	74,6 %	795 z 1 066	≈ 254
Všeobecná AI bez prístupu na web	36,7 %	391 z 1 066	≈ 633

Rubrika je zámerne prísna a pre všetkých rovnaká: **nesprávne právne odôvodnenie znamená nesprávnu odpoveď, aj keď výsledné číslo náhodou sedí**. Pre účtovníka je to jediné zmysluplné kritérium — odpoveď, ktorá sa opiera o zlý paragraf, neobstojí pred kontrolou ani vtedy, keď suma vyšla správne. Práve táto požiadavka najviac zrazila všeobecnú AI: často uvedie správne číslo s vymysleným alebo zastaraným odôvodnením.

Sporné verdikty sme naviac preverili **tromi nezávislými posudkami** (neutrálny, obhajca odpovede, kritik; pri nezhode arbiter na silnejšom modeli) — a to **identicky pre všetky systémy**: všetkých 70 našich zlyhaní a vzorku 68 zlyhaní každej všeobecnej AI. U nás sa zmenilo 10 verdiktov, u konkurencie **ani jeden** — ich chyby obstáli ako chyby aj vtedy, keď ich niekto vyslovene obhajoval.

A jedna oprava, ktorú sme slúbili robiť verejne: po zverejnení sme v banke otázok sami našli jednu chybnú správnu odpoveď — pri zdravotnom poistení z ostatných príjmov uvádzala základnú sadzbu 15 %, prechodné ustanovenie § 38ezk zákona č. 580/2004 Z. z. však pre roky 2026–2027 predpisuje 16 %. Otázku sme opravili a nechali prehodnotiť tým istým sudcom pre všetky systémy naraz: verdikt sa otočil na správny nám **aj všeobecnej AI s Googlom** (obe odpovedali 16 %), AI bez prístupu na web pri 15 % ostala nesprávne. Čísla v tomto dokumente už opravu obsahujú — presne takto má verejný test fungovať.

Ako to čítať v peniazoch: každá nesprávna odpoveď, na ktorú sa účtovník spoľahne, je potenciálna pokuta, dorubenie dane alebo penalizovaná lehota — typicky stovky až tisíce eur. Rozdiel medzi 55 a 254 chybami na tisíc otázok nie je štatistika, je to **viac než štvornásobne menšia šanca, že práve vaša otázka bude tá zle zodpovedaná** — a oproti AI bez prístupu na web jedenásťkrát menšia. Bežná AI bez zdrojov sa totiž na slovenských ťažkých otázkach mylí takmer v dvoch tretinách prípadov.

Prečo je rozdiel taký veľký, keď „pod kapotou“ pracuje podobná AI? Pretože presnosť nerobí samotný model, ale to, čo dostane k dispozícii: overená databáza slovenských predpisov v aktuálnom znení, referenčné hodnoty (sadzby, minimá, maximá) udržiavané osobitne vrátane budúcich zmien, deterministická kalkulačka na každý výpočet a povinnosť doložiť odpoveď citátom predpisu. Všeobecná AI s vyhľadávaním nachádza články — aj zastarané, aj bulvárne; náš systém číta zákon.

Výsledky záverečného auditu banky (dokončený 20. 9. 2026): audit prešiel všetkými 1 066 otázkami — bez výhrad obstálo 938 (88,0 %), 128 sme na základe nálezov opravili (typicky presnejšie číslo odseku, doslovnejší citát, lepší zdroj — spolu 196 opravených polí) a ani jedna nebola zamietnutá ako nepoužiteľná. Opravené znenia sú súčasťou zverejnenej banky.

Čo test odhalil u nás — a čo sme s tým spravili

Sľúbili sme transparentnosť, takže: test našiel chyby aj v našom produkte. Každú sme opravili ešte počas testovania a oprava platí pre všetkých zákazníkov od momentu nasadenia. Tri najzaujímavejšie:

Sviatky verzus dni pracovného pokoja. Slovenské právo od roku 2026 rozlišuje až tri režimy: deň môže byť sviatkom pre mzdové príplatky, ale už nie dňom pracovného pokoja pre posun daňových lehôt — a naopak. Aplikácia pôvodne pracovala s jednou množinou dní pre oboje. Dnes má dve oddelené pravidlá a v teste odpovedá správne aj na chytáky typu „patrí zamestnancovi príplatok za 15. september 2026?“.

Aritmetika sa nepočíta „z hlavy“. AI modely robia chyby v násobeniach a percentách — zriedkavo, ale pri mzdách je aj zriedka priveľa. Každý výpočet v aplikácii preto dnes povinne prechádza deterministickou kalkulačkou: AI zostaví vzorec, ale čísla počíta stroj, ktorý sa nemýli. Podnetom bol reálny prípad zákazníčky, ktorej v chate nesedeli súčty — tá chyba už dnes vzniknúť nemôže.

Prechodné ustanovenia noviel. Najzradnejšia kategória otázok v celom teste: čo platí pre hospodársky rok, ktorý sa začal pred účinnosťou novely? Doplnili sme do databázy overené referenčné hodnoty prechodných ustanovení vrátane budúcich zmien — aplikácia dnes výslovne rozlišuje, čo platí teraz a čo až bude platiť.

Čísla za týmto príbehom: prvé meranie celej banky našlo **174 našich chybných odpovedí**. Každú sme rozobrali na koreňovú príčinu a opravili — 113 nových overených referenčných hodnôt, 16 nových právnych zdrojov v databáze, nové pravidlá pre prácu s prechodnými ustanoveniami a výpočtami. Náš audit pritom odhalil aj vadu, ktorá by sa inak nikdy neukazala: časť rozhodujúcich pravidiel sa pri zostavovaní podkladov pre AI orezávala, takže sa k modelu vôbec nedostala. Opravené — a odvtedy to stráži automatický test. Po opravách klesol počet našich chybných odpovedí zo **174 na 59**.

Toto je zároveň odpoveď na otázku, prečo vôbec takýto test robíme a zverejňujeme: **každá nájdená chyba je lacnejšia počas testu než u zákazníka**. Test preto nie je jednorazová marketingová akcia — je to náš stály kontrolný mechanizmus: meraj → oprav → meraj znova. Zvyšných 59 otázok, ktoré sme ani po opravách nezvládli, nájdete v zverejnenom teste aj s verdiktom sudcu — sú naším ďalším zoznamom práce.

Prečo na tom záleží práve účtovníkom

Účtovník je jediná profesia, kde „takmer správna“ odpoveď nemá hodnotu — pred daňovou kontrolou neobstojí „AI mi to tak povedala“. Obstojí len paragraf. Preto náš systém nestavíme na tom, aby pôsobil presvedčivo, ale na tom, aby bol **overiteľný**:

- **Každá odpoveď cituje predpis** — zákon, § a odsek, s preklikom priamo do textu zákona. Odpoveď, ktorú si neviete overiť klikom, je len názor.
- **Čísla pochádzajú z udržiavanej databázy, nie z pamäte modelu** — sadzby, minimálne základy, stravné či veličiny platné od budúceho roka sa aktualizujú ako dáta, s dátumom účinnosti.
- **Za chyby niekto ručí**. Všeobecná AI sa mýli anonymne a zakaždým inak. V našom systéme je chyba, ktorú raz niekto nahlási alebo nájde test, opravená pre všetkých a navždy — presnosť rastie s každým mesiacom používania.

A ešte jedna vec, ktorú tento test ukázal čiernobielo: rozdiel nie je v tom, „či používať AI“. Rozdiel je medzi AI, ktorá háda z internetu, a systémom, ktorý číta zákon a dokáže to.

O teste a dátach

Zverejňujeme kompletný protokol, nie len výsledné percento:

- **Celú banku otázok** — každú otázku v odbornej aj voľnej podobe, správnu odpoveď, presné ustanovenie (§ a odsek), doslovný citát predpisu a odkaz na oficiálny zdroj.
- **Odpovede všetkých troch systémov** na každú otázku, tak ako prišli, bez úprav.
- **Verdikty sudcu** s odôvodnením — vrátane prípadov, kde sme neuspeli my.

Každé tvrdenie v tomto dokumente si tak môže ktokoľvek overiť sám: otvoriť otázku, kliknutím na zdroj si prečítať predpis a porovnať odpovede. Ak nájdete v banke chybu, napíšte nám — opravu zverejníme rovnako ako všetky ostatné.

Test pripravila a prevádzkuje spoločnosť ZOTRI, s. r. o., prevádzkovateľ služby SmartUctovnik.sk — AI účtovníka pre slovenských podnikateľov a účtovné firmy. Rovnakú službu prevádzkujeme v šiestich krajinách EÚ pod lokálnymi značkami.

Dátum merania: 21. septembra 2026. Znenie predpisov overované k dátumu merania; banka otázok sa priebežne aktualizuje pri legislatívnych zmenách. Odpovede všeobecnej AI boli vygenerované 20. 9. 2026 a posudzované rovnakou rubrikou ako naše.