1.3 Zodpovědné používání AI

Shrnutí

- **Odpovědné využívání AI je otázkou správného nastavení, nikoli kontroly.** Nástroje pro detekci AI nejsou spolehlivé.
- **Detekce AI může vést k nespravedlivému hodnocení.** Nástroje chybně označují lidskou práci za výstup AI a naopak. Francie a Nizozemsko jejich používání ve školách nedoporučují.
- **AI může vytvářet nepravdivé informace a přebírat předsudky.** Žáky je proto nutné učit tato rizika rozpoznávat.
- **Nejprve vlastní úvaha, potom AI.** Samostatné přemýšlení a vhodná softwarová omezení pomáhají předcházet negativním dopadům AI.
- **Samotné prompty nestačí.** Chatboti často přecházejí k poskytování hotových odpovědí. Důležitý je proto výběr vhodných nástrojů.
- **AI může prohlubovat rozdíly mezi žáky.** Jejich zmírnění vyžaduje dostupné technologie, vhodné nástroje a především proškolené učitele.
- **Inkluze je jedním z největších přínosů AI.** Právě zde vidí učitelé největší potenciál jejího využití.

> Pokud si z této stránky máte zapamatovat jen jednu větu: Hráze se bortí. Naučme se plavat.

Čtenářský deník, druhý pohled

Vraťme se k našemu žákovi deváté třídy a jeho čtenářskému deníku, který působí až podezřele dokonale. Instinkt je téměř univerzální a je to instinkt dobrého učitele: zjistit, zda použil AI. Postavit lepší hráz.

Tato kapitola pojednává o tom, proč tento instinkt sám o sobě nikam nevede. Dostaneme se k tomu, co funguje místo něj.

Začněme tou nepříjemnou částí. **Učitelé nedokážou spolehlivě rozeznat text vytvořený AI od textu žáka.** Ve dvou studiích německého týmu dostalo 89 budoucích učitelů a 200 zkušených učitelů směs žákovských esejí a esejí od chatbota a měli je roztřídit. Ani jedna skupina to nedokázala spolehlivě, přesto si obě skupiny byly jisté, že to dokážou. Zkušenost pomohla jen málo: i zkušení učitelé přehlédli většinu textů vytvořených AI. Na jedné britské univerzitě

výzkumníci vložili odpovědi vygenerované chatbotem běžnými kanály do skutečných online zkoušek bez dozoru, aniž by o tom hodnotitele informovali.

„Můžete ji používat, ale musíte umět vysvětlit výsledek vlastními slovy.“

Detektor, který trestá špatné žáky

Výzkumníci otestovali sedm nejrozšířenějších detektorů AI na esejích TOEFL, které napsali lidé učící se angličtinu jako cizí jazyk a které byly shromážděny z čínského online fóra. Vedle nich zařadili eseje rodilých amerických osmáků. Práce rodilých mluvčích byla klasifikována správně téměř pokaždé. Z esejí v cizím jazyce bylo **více než šest z deseti mylně označeno jako vytvořené strojem**.

Mechanismus není žádnou záhadou. Mnohé z těchto nástrojů nedetekují práci velkých jazykových modelů. Nástroje měří *předvídatelnost*. Text, který používá běžná slova v očekávaném pořadí, je vyhodnocen jako umělý. Lidé píšící v cizím jazyce, s menší slovní zásobou a strnulejšími větnými strukturami, mohou produkovat přesně takový signál. Detektory trénované pro jeden konkrétní test a studie v jiných jazycích přitom ukazují menší rozdíly. Dosud však na trhu neexistují detektory dostatečně spolehlivé na to, aby bylo možné kohokoli obvinit z úmyslného podvodu.

Chyby nástroje systematicky dopadají na žáky, kteří už nesou největší zátěž. Instituce si to spočítaly. Vanderbiltova univerzita upozornila, že i kdyby se takový nástroj mýlil jen v 1 % případů, stále by to znamenalo přibližně **750 mylně označených prací ročně** z jejích 75 000 odevzdaných prací. V důsledku toho jej vypnula. Příručka, kterou pro německé spolkové ministerstvo školství napsali výzkumníci, je jednoznačná: technologie *„není dostatečně spolehlivá, aby poskytovala právně bezpečný důkaz“*. Nezávislý evropský test čtrnácti detektorů dospěl ke stejnému verdiktu: *„ani přesné, ani spolehlivé“*.

Velká část Evropy o tom už přestala diskutovat. Francouzské ministerstvo školství a francouzský úřad pro ochranu osobních údajů školám doporučují detektory nepoužívat, nizozemská agentura pro ICT ve vzdělávání Kennisnet je označuje za vysoce nespolehlivé. Ve vysokoškolském vzdělávání jejich používání při hodnocení zakazují jednotlivé univerzity, například ZHAW v Curychu. V Anglii zkušební komise připouštějí detekční software pouze jako jeden z prvků úsudku, který zůstává na učiteli, jenž *„zná své žáky nejlépe“*. Stojí za to si zapamatovat i jeden bod týkající se ochrany osobních údajů: **vložení práce žáka do online nástroje znamená zpracování jeho osobních údajů.** K tomu je zapotřebí právní základ a smlouva s poskytovatelem. → 1.4 Soulad s předpisy při používání AI

Zdá se, že panika ohledně podvádění je slabší, než by odpovídalo hluku kolem ní. Výzkumníci provedli průzkum mezi žáky tří amerických středních škol před vznikem ChatGPT a poté znovu. Podvádění zůstalo zhruba na stejné úrovni. Navazující průzkum mezi více než 4 000 žáky šesti dalších amerických středních škol o rok později zjistil totéž.

Skutečná otázka nikdy nebyla „Jak ho chytím?“ Je jí: co domácí úkol vlastně ještě měří, když jeho výsledek lze vygenerovat na požádání?

„Nejdřív jsem to považoval za podvádění. Teď to vnímám jako signál, že se musí změnit samotný úkol.“

Dvě věci, které se nedají opravit, a proto je třeba je učit

Sebejisté smyšlenky jsou vestavěnou vlastností. Tyto systémy generují věrohodné pokračování textu, a věrohodné není totéž co pravdivé. Přehledy výzkumu se shodují, že žádná technika tento problém dosud neodstranila, i když jej lze omezit a částečně odhalit. Důvodem je, že standardní testy používané k hodnocení velkých jazykových modelů odměňují sebejisté hádání více než přiznání nejistoty, a dokud budou modely takto hodnoceny, přetrvá určitá míra plynulé, dobře formulované nepravdy.

Zkreslení v trénovacích datech jsou vestavěná také. A navíc se s vylepšováním technologie nezmenšují. Evropská studie genderových stereotypů napříč jazyky zjistila, že větší modely, dokonce i po dodatečném tréninku určeném k tomu, aby byly bezpečnější a spravedlivější, někdy stereotypizují více, nikoli méně. Pokud se vlastnost nedá technicky odstranit, musí se stát součástí učiva. A lze ji učit: v jednom finském programu vzrostl podíl žáků, kteří dokázali vysvětlit, odkud se zkreslení generátoru obrázků bere, ze 7 % na 44 %. Důsledkem pro výuku musí být trénink v analýze chyb místo jejich pouhého vyhýbání se. Jinými slovy: chyby stroje se stávají samotnou lekcí.

„Nechte žáka zkontrolovat řešení chatbota, jako by byl učitelem. Právě hledáním chyb se metodu konečně naučí.“

Co funguje: návrh, nikoli kontrola

Důkazy ukazují na jeden jednoduchý a snadno zapamatovatelný princip: nejprve myslet, pak AI.

- **Nechte člověka zavázat se jako první.** V experimentu s přibližně 200 dospělými se lidé, kteří museli učinit vlastní rozhodnutí *dříve*, než uviděli návrh stroje, řídili jeho chybnými návrhy méně často než lidé, kterým bylo ukázáno krátké vysvětlení uvažování stroje. Vysvětlení AI problém nevyřešilo. Pomohla metoda, která vynutila vlastní úsudek. Vynutit úsudek znamená přimět lidi přemýšlet. Protože přemýšlení vyžaduje úsilí, patřily tyto metody mezi účastníky k nejméně oblíbeným. Nejvíce však pomáhaly těm, kteří jsou již zvyklí intenzivně přemýšlet. Jde o takzvaný Matoušův efekt: ti, kteří již mají určitou výhodu, z ní získávají ještě více.
- **Zabudujte omezení přímo do nástroje.** Chatbot z kapitoly 1.2, který zadržoval odpovědi a místo toho kladl otázky, je stejná myšlenka přenesená do softwaru: žáci, kteří

jej používali, na zkoušce nic neztratili, zatímco ti s neomezenou verzí dopadli výrazně hůře.
- **Zviditelněte proces, místo abyste honili výsledek.** Koncepty, pracovní poznámky, krátký rozhovor o práci. Ty vám ukážou, co žák dokáže, aniž by kdokoli musel dokazovat, co udělal.
- Dánsko pilotně ověřuje tento princip: na zapojených školách mohou žáci používat AI při přípravě na ústní zkoušku z angličtiny, zatímco písemná zkouška obsahuje ručně psanou část bez pomůcek.

Než se na tento přístup spolehnete, je dobré znát dvě důležitá omezení.

První omezení: Často doporučovaný postup, kdy učitel jednoduše zadá chatbotovi, aby se choval jako sokratovský tutor a vedl žáky k odpovědím pomocí otázek, nemusí při delším rozhovoru fungovat spolehlivě. Předběžné výsledky studie se simulovanými žáky, která k říjnu 2026 dosud neprošla recenzním řízením, naznačují, že při opakovaném naléhání mají tito AI tutoři tendenci vracet se k poskytování hotových odpovědí. Spolehlivá omezení proto musí být součástí samotného systému. Důležitější než správná formulace zadání je tak výběr vhodného AI nástroje.

Druhé omezení: Požádat AI, aby vysvětlila svůj postup, se může zdát jako jednoduchý způsob ověření správnosti odpovědi. Samotné vysvětlení však nezaručuje, že uživatel nebude AI příliš důvěřovat. Jedna ze studií ukázala, že vysvětlení nesnížila ochotu lidí přijímat nesprávné rady. Příliš technická nebo naopak zjednodušená vysvětlení mohou dokonce falešnou důvěru posilovat. Vysvětlení tedy může učení podpořit, samo o sobě ho však nezaručuje.

Postoj: dva způsoby selhání, nejen jeden

Obáváme se, že učitelé vkládají do AI příliš mnoho důvěry. Měli bychom se však stejně obávat i opaku. I zkušení lidé mohou padnout za oběť přílišné důvěře. V malém pilotním projektu na amerických školách označily čtyři z deseti dotázaných učitelů zpětnou vazbu nástroje AI pro hodnocení za neužitečnou, většinou nejasnou, nebo dokonce nesprávnou. Výzkum zkreslení vlivem automatizace (automation bias) ukazuje, že lidé mají sklon přijmout návrh stroje, i když mají důvod o něm pochybovat. **Odhalit chybu a opravit ji jsou dva odlišné kroky.**

Opačným pólem je utajování. Učitelé skrývají vlastní používání AI před kolegy i žáky a zakazují jim to, co sami dělají: 77 % z nich používá AI soukromě, zatímco 57 % tvrdí, že žáci by ji nikdy neměli používat k vymýšlení nápadů. Studie akademicky nadaných žáků 6. až 8. ročníku mezitím zjistila, že mezi třetinou a dvěma pětinami z nich už tvrdí, že AI ví více než jejich učitelé. Profese, která skrývá vlastní praxi, nemůže být vzorem zodpovědného jednání, a transparentnost není jen zdvořilostí. Je to mechanismus, kterým funguje výuka vlastním příkladem.

Pomáhá AI vyrovnávat rozdíly, nebo je naopak prohlubuje?

U pracovních úkolů AI vyrovnává rozdíly. V jednom experimentu podávali účastníci s vyšším vzděláním výrazně lepší výkony než účastníci s nižším vzděláním. S asistentem AI zmizely zhruba tři čtvrtiny tohoto rozdílu. Když byl tutor znovu odebrán, značná část rozdílu se vrátila. **AI odstranila rozdíl ve výstupu, nikoli rozdíl ve schopnostech.**

U učebních úkolů je tomu naopak. Jedna studie nezjistila v průměru žádný přínos, ale zjistila rostoucí propast mezi žáky, kteří už toho hodně věděli, a těmi, kteří ne. Jiná studie se zaměřila na žáky píšící v cizím jazyce: slabší pisatelé dostávali od AI více oprav a úspěšně jich uplatnili méně, a mnozí považovali záplavu zpětné vazby za skličující. Zpětná vazba bez schopnosti ji roztřídit nepomáhá; zavaluje.

Rozdíly mezi žáky existují ještě předtím, než začnou AI vůbec používat. V Německu vnímá umělou inteligenci jako příležitost 80 % žáků z nejlépe socioekonomicky zajištěných rodin, zatímco mezi žáky z nejméně zvýhodněného prostředí je to pouze 55 %. Ve Spojeném království vede dělicí čára spíše sborovnou než skladem se zařízením: **45 % učitelů na soukromých školách absolvovalo formální školení v používání AI, oproti 21 % na státních školách.** V rámci státního sektoru bohatší školy opět předstihují ty chudší. Průzkum z roku 2026 mezi více než 25 000 dětmi ze 17 evropských zemí ukazuje stejné sociální rozdělení v tom, jak moc a jak různorodě AI používají. Výrok jednoho komentátora vystihuje celý problém lépe než jakékoli číslo: *„bohatí mají přístup k technologii i k lidem, kteří jim pomohou ji používat, zatímco chudí mají přístup pouze k technologii.“*

Rozhodujícím faktorem je tedy to, zda má žák k dispozici učitele, který je na využívání AI dostatečně připraven. Umělá inteligence může pomáhat vyrovnávat rozdíly mezi žáky, pokud jsou splněny tři podmínky: dostupnost technologií, využívání nástrojů navržených pro podporu učení, nikoli pouze pro poskytování hotových odpovědí, a odborné vedení ze strany učitele. Pokud některá z těchto podmínek chybí, může AI naopak rozdíly prohlubovat. Protože se všechny tři podmínky často vyskytují společně na stejných školách, zatímco jinde chybějí, AI má přirozenou tendenci existující nerovnosti prohlubovat. Jejich vyrovnávání proto vyžaduje cílené úsilí.

Revision #15
Created 2026-09-02 08:50:45 UTC by Alexander
Updated 2026-10-21 11:03:38 UTC by Magdalena