1.3 Zodpov?dné používání UI

Shrnutí

- **Zodpovědnost je úkolem návrhu, nikoli úkolem kontroly.** Detekce UI spolehlivě nefunguje.
- **A detekce UI je nespravedlivá, když selže.** Existuje riziko, že odpovědi budou mylně označeny jako vygenerované UI, nebo naopak. Několik evropských zemí tyto nástroje již zcela vyloučilo.
- **Sebejisté smyšlenky a zděděné zkreslení vyplývají z toho, jak jsou tyto systémy vytvářeny a hodnoceny.** Důsledkem by mělo být: učit o nich, místo čekání na opravu.
- **Pracovní zásada zní: nejprve myslet, pak UI.** Přinutit lidi zavázat se k vlastní odpovědi funguje lépe než vysvětlovat postup stroje; zabudování omezení do softwaru funguje lépe než jeho zákaz.
- **Samotné příkazy nevydrží.** Chatbot pověřený rolí tútora má tendenci sklouzávat zpět k přímému sdělování odpovědí, což dělá ze spolehlivých omezení otázku nákupu, nikoli formulace zadání.
- **Zesilování je výchozí stav; vyrovnávání je třeba vybudovat společně z přístupu, návrhu a podpůrného vedení – a nejvzácnějším ze tří je proškolený učitel.**
- **Inkluze je nejjasnějším přínosem** a oblastí, kde samotní učitelé vidí nejsilnější argument.

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

?tená?ský deník, druhý pohled

Vraťme se k našemu žákovi deváté třídy a jeho podezřele vybroušenému čtenářskému deníku. Instinkt je téměř univerzální a je to instinkt dobrého učitele: zjistit, zda použil UI. 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 UI od textu žáka. V německé studii dostalo 289 učitelů s různou zkušeností směs žákovských textů a textů od chatbota a měli je roztřídit. Ani jedna skupina to nedokázala. Přesto si obě skupiny byly jisté, že to dokážou. Zkušenost nedělala téměř žádný rozdíl. Ve větším měřítku je to ještě horší. Na jedné univerzitě výzkumníci vložili odpovědi vygenerované chatbotem do skutečného zkouškového

systému běžnými kanály, aniž by o tom hodnotitele informovali. 94 % z nich nebylo nikdy zpochybněno. Navíc byly v průměru ohodnoceny lépe než práce skutečných studentů. Jaký je tedy důsledek? Sáhneme po stroji! A právě zde přestává být zjištění pouze trapné a stává se etickým problémem!

Detektor, který trestá špatné žáky

Výzkumníci otestovali sedm nejrozšířenějších detektorů UI na esejích TOEFL napsaných žáky, kteří se učí angličtinu jako cizí jazyk. 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. Tyto nástroje nedetekují práci velkých jazykových modelů, detekují předvídatelnost. Text, který používá běžná slova v očekávaném pořadí, je vyhodnocen jako umělý. Kdokoli píše v cizím jazyce, s menší slovní zásobou a bezpečnějšími větnými strukturami, produkuje přesně takový signál. 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 nespravedlivě obviněných žáků ročně z jejích 75 000 odevzdaných prací. V důsledku toho jej vypnula. Vyjádření německého spolkového ministerstva školství je jednoznačné: technologie „není dostatečně spolehlivá, aby poskytovala právně bezpečný důkaz“.

Velká část Evropy o tom už přestala diskutovat. Francie, Nizozemsko a Švýcarsko od detektorů odrazují nebo je zakazují. Britský zkušební regulátor se záměrně rozhodl upřednostnit lidský úsudek před softwarem. Nizozemský argument jde o krok dál a stojí za zapamatování: vložení práce žáka do detektoru je samo o sobě zpracováním jeho osobních údajů, a učinit to bez souhlasu porušuje právo na ochranu osobních údajů. → 1.4 Soulad s předpisy při používání UI

Zdá se, že panika ohledně podvádění je slabší, než by odpovídalo hluku kolem ní. Výzkumníci provedli průzkum mezi žáky tří středních škol před vznikem ChatGPT a poté znovu. Podvádění zůstalo zhruba na stejné úrovni. Většina žáků odmítala nechat chatbota jednoduše vytvořit celou jejich práci, zatímco za fér považovali použít jej k rozjezdu práce nebo k vysvětlení něčeho.

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

Dv? v?ci, které se nedají opravit, a proto je t? eba je u?it

Sebejisté smyšlenky jsou vestavěnou vlastností. Tyto systémy generují věrohodné pokračování textu, což není totéž jako pravdivé pokračování. Výzkumníci nyní formálně ukázali, odkud se to bere: 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. Je to vlastnost způsobu, jakým jsou vytvářeny a hodnoceny, nikoli chyba čekající na opravu.

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 identifikovat zkreslení ve výstupu systému UI, 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.

Co funguje: návrh, nikoli kontrola

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

- **Nechte člověka zavázat se jako první.** V sérii experimentů lidé, kteří si museli nejprve zapsat vlastní odpověď, než uviděli návrh stroje, odhalili mnohem více jeho chyb než lidé, kterým bylo podrobně vysvětleno, jak stroj uvažoval. Vysvětlení UI problém nevyřešilo, ale vynucení úsudku ano. S jednou upřímnou výhradou: verze, které fungovaly nejlépe, se účastníkům líbily nejméně a nejvíce pomáhaly těm, kdo už mají rádi náročné přemýšlení. Jde tedy o jev známý také jako Matoušův efekt: Tomu, kdo má, bude dáno více. Právě oni z tohoto přístupu těží nejvíce.
- **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 UI při přípravě na ústní zkoušku z angličtiny, zatímco písemná zkouška obsahuje ručně psanou část bez pomůcek.

Než se na to spolehnete, stojí za to znát dvě omezení. To první: oblíbená rada, že učitelé mají chatbotovi jednoduše nařídit, aby se choval jako sokratovský tútor, nepřežije dlouhý rozhovor. Systém má, alespoň v tuto chvíli, tendenci sklouzávat zpět k poskytování odpovědí. Spolehlivá omezení musí být zabudována přímo do systému, což z nich dělá spíše otázku nákupního rozhodnutí než dovednosti formulovat zadání. To druhé: nechat UI vysvětlit své uvažování zní jako zjevná pojistka, ale vysvětlení spolehlivě nezabraňují nadměrné důvěře. V jedné studii ponechala vysvětlení míru, v jaké lidé přijímali nesprávné rady, beze změny, a vysvětlení, která jsou příliš technická nebo příliš zjednodušená, mohou mylnou důvěru dokonce zvýšit. Vysvětlení je podmínkou, nikoli řešením.

Postoj: dva zp?soby selhání, nejen jeden

Obáváme se, že učitelé vkládají do UI příliš mnoho důvěry. Měli bychom se však stejně obávat i opaku. I zkušení lidé mohou padnout za oběť přílišné důvěře. V jedné studii učitelé posuzovali známky vygenerované UI. U 42 % zpětné vazby stroje usoudili, že je nejasná nebo chybná, ale upravili jen 9 % z ní. Odhalit chybu a opravit ji jsou dva odlišné procesy.

Opačným pólem je utajování. Učitelé skrývají vlastní používání UI před kolegy i žáky a zakazují jim to, co sami dělají: 77 % z nich používá UI 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 UI 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.

Vyrovnává UI podmínky, nebo je naklání?

U pracovních úkolů UI vyrovnává rozdíly. V jednom experimentu podávali účastníci s vyšším vzděláním výrazně lepší výkony než účastníci s nižším vzděláním. Když každý z nich dostal asistenta UI, tyto rozdíly téměř úplně zmizely. Když jim pak byl asistent znovu odebrán, velká část rozdílu se vrátila. UI 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 UI více oprav a úspěšně jich uplatnili méně, a mnozí považovali záplavu zpětné vazby za skličující. Zpětná vazba bez schopnosti ji roztřídit nepomáhá; zavaluje.

Propast existuje i dřív, než se kdokoli nástroje dotkne. V Německu vnímá UI jako příležitost 80 % žáků z nejzvýhodněnějších rodin oproti 55 % z nejméně zvýhodněných. Ve Spojeném království vede dělicí čára spíše sborovnou než skladem se zařízením: 45 % učitelů na soukromých školách absolvovalo formální školení v používání UI, oproti 21 % na státních školách. V rámci státního sektoru bohatší školy opět předstihují ty chudší. Průzkum mezi dětmi z 20 evropských zemí ukazuje stejné sociální rozdělení v tom, jak moc a jak různorodě UI používají. Jedna věta 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.“ To je nejsilnější dostupný argument pro to, že na učitelích při uzavírání této propasti záleží.

Úzkým hrdlem je tedy to, zda má žák učitele, který byl proškolen. UI kompenzuje nevýhody tehdy, když se spojí tři věci: zajištěný přístup, nástroj navržený pro učení, nikoli pro pouhé poskytování odpovědí, a učitel, který jeho používání podpůrně vede. Odeberte kteroukoli z nich a UI místo toho rozdíly zesiluje. Protože všechny tři podmínky bývají přítomné ve stejných školách a chybí ve stejných školách, je zesilování výchozím stavem a vyrovnávání je něco, co je třeba záměrně vybudovat.

Revision #10
Created 2026-09-02 08:50:45 UTC by Alexander
Updated 2026-09-03 08:49:06 UTC by Alexander