

RIZIKA SPOJENÁ S VYUŽÍVÁNÍM AI Z POHLEDU IA A AUDITOVANÝCH

V dnešní době je umělá inteligence (AI), ať již v jakékoli své podobě, stále více integrována do různých míst organizací i našich běžných životů, což přináší nejen mnoho výhod, ale také řadu rizik. A to i pro interní audit (IA). Zapojení AI do jakýchkoli činností otevírá nové možnosti výsledků práce či zábavy, ale také právě celou škálu nových rizik. V tomto článku se podíváme na vybraná rizika spojená s AI z pohledu interního auditu a auditovaných a naznačíme, jak tato rizika, ve vztahu k technickému vývoji poloviny roku 2023, minimalizovat.

PhDr. Ing. Jiří Čihák
Interní a forenzní audit IT

V současné době pracuje pro významnou mezinárodní telekomunikační společnost v oblasti interního auditu IT. Předchozí zkušenosti získal v jedné z českých bank v oblasti auditu IT nebo ve společnosti tzv. velké čtyřky, kde se věnoval externímu forenznímu auditu, či v instituci veřejné správy v oblasti IT.

Pohledy na tuto problematiku je možné v úvodu teoreticky rozdělit na přinejmenším dva. A to z pohledu na AI jako objekt auditu a na AI jako nástroj pro audit. Pro ilustraci rizik spojených s AI tyto pohledy pro účely tohoto článku spojíme a podíváme se komplexně na rizika spojená s využíváním AI obecně. Jako vzorový typ AI, na kterém si můžeme konkrétní rizika ilustrovat, si zvolíme v současnosti nejspíše nejznámější typ AI, a to jazykový/textový model umělé inteligence (někdy také nazývaný zjednodušeně AI chatbotem), a stranou nechme všechny jiné typy AI.

Lze zobecnit, že pro AI platí základní triáda: Jak model funguje? – Jak chrání svěřená data? – Jak je používán?

Podíváme se na konkrétní příklady. Co lze zařadit mezi rizika a problémy spojené s využíváním AI? Uvedené příklady ilustrované na prostředí IA lze aplikovat i na jiné útvary či použití jednotlivcem. Zohledňují totiž více charakteristiky AI než charakteristiky IA.

AI modely mohou být či jsou skutečně složité a obtížně interpretovatelné. To způsobuje jakýsi **nedostatek transparentnosti**. Modely jsou často založené na složitých konceptech nejrůznějších typů učení, rozhodování apod., které mohou být obtížně srozumitelné a vysvětlitelné. Nedostatek transparentnosti může vést k neschopnosti odhalit chyby nebo předpojatosti ve výsledcích AI, nepochopení, jakým způsobem dochází k vlastním rozhodnutím a vysvětlením, jak samotná AI dospívá k určitým rozhodnutím. To může vést k nedostatečné transparentnosti procesu rozhodování a ztížit schopnost ověřování auditních výstupů a zjištění.

K eliminaci uvedeného by mohl pomoci přístup interního auditu k důkladným popisům fungování AI modelů, což není vždy dostupné a možné. Nebo nemusí být vždy efektivní tyto modely studovat.

AI modely jsou závislé na **tréninkových zdrojových datech**, a pokud jsou tato data zkreslená nebo obsahují bias, může to vést k nepřesným výsledkům

nebo nesprávným či nespravedlivým rozhodnutím. Je to způsobeno tím, že AI modely se obvykle učí z velkého množství dat, která jsou jim zpřístupněná. To znamená, že jejich odpovědi mohou být ovlivněny neaktuálními, nepřesnými nebo zaujatými informacemi, které byly v těchto datech a byly použity k vycvičení a natrénování modelu. To ve výsledku může ohrozit kvalitu a důvěryhodnost výstupů, včetně auditního procesu, a může se promítnout v **nepřesnost či nespolehlivost**. Při použití zavádějících informací se mohou vyskytnout chyby, které by mohly ovlivnit kvalitu provedeného auditu. Na místě je tak opatrnost při používání informací generovaných AI a nutnost ověření poskytnutých výsledků i z jiných zdrojů, je-li to možné. Což však zpětně umazává časovou úsporu získanou případným využitím AI.

„AI modely mohou být či jsou skutečně složité a obtížně interpretovatelné.“

Dále se mohou objevit i návody k neetickému chování, kdy AI může případně poskytnout **neetické rady** nebo podporovat neetické chování při auditu. Řešením by mohla být prokazatelně prověřená kvalita a reprezentativnost tréninkových dat, což nemusí být vždy možné.

Poskytnutí nepřesných, zavádějících nebo jen povrchních informací může být způsobeno také použitím nesprávného modelu pro daný účel. Tedy pokud je používán nevhodně zvolený typ AI, který má například omezenou specializaci nebo specializaci na něco jiného, než uživatel potřebuje nebo co je pro výsledek auditního použití důležité.

O několik odstavců výše padlo slovo **bias**. **Bias, předsuděk, tendence, sklon, předpojatost**. To je spojené zejména s rizikem nerovnosti a diskriminace. AI může zesilovat určité nežádoucí prvky, jako například sociální stereotypy a nerovnosti obsažené v datech, protože může být ovlivněna právě předsudky dat, na kterých je vycvičena. AI tak může generovat informace, které jsou předsudkové, diskriminační nebo jinak nevhodné. Auditori, a obecně všichni uživatelé, by si měli být vědomi potenciálu předsudků, resp. biasu. To se přeneseně týká například i historických faktů.

„AI modely se obvykle učí z velkého množství dat, která jsou jim zpřístupněná.“

Závislost na zdrojových datech může vést k nespravedlivým rozhodnutím a k posilování sociálních nerovností, či přímo k diskriminaci. Proto mohou vývojové týmy AI zahrnovat skupiny reviewerů, kteří sledují poskytované výsledky a dávají zpětnou vazbu na nevhodný bias či jiné zaujetí. V zájmu korektnosti mohou být některé odpovědi uhlazené či rovnou neposkytované. Představme si to například tak, že pokud budou tréninková data pojednávat o určité skupině osob určitého geografického původu či stáří, je možné, že právě tento profil osob se následně promítne do generovaných odpovědí. Tj. nemusí dojít ke zohlednění osob ze skupin jiného geografického původu či jiného stáří. V detailech se tak může promítat například do nechtěných opomenutí minoritních skupin, a tedy nedostatku diverzity či diskriminace, do zesilování či opakování nevhodného jazyka, názorů či stereotypů. To vše na základě toho, že tyto aspekty byly zahrnuty ve zdrojových datech použitých k nacvičení modelu. Problematika biasu je jedním z výrazných témat spojených se spravedlivým rozvojem AI.

A samozřejmě, bias mechanismus v AI může být taktéž zneužit k manipulaci s informacemi nebo k šíření dezinformací.

AI modely pracují s velkým množstvím dat od uživatelů, které jsou do nich nahrány jako uživatelské vstupy. To zvyšuje riziko narušení důvěrnosti při zneužití poskytnutých citlivých informací a vystavuje zvýšené nároky na **bezpečnost svěřených dat**. Jakákoli AI, ať již online (cloud), nebo offline, je zranitelná vůči **kybernetickým útokům** a zneužitím, což může ohrozit citlivá data a způsobit škody. Lze tak odůvodněně očekávat rizika krádeže nebo zneužití citlivých informací, se kterými je pracováno, ale také rizika implementace nevhodného kusu kódu, a to s nejrůznějšími účely, včetně například právě krádeže dat, zavlečení cíleného biasu, chybovosti či poskytování tendenčních výsledků.

„Poskytnutí nepřesných, zavádějících nebo jen povrchních informací může být způsobeno také použitím nesprávného modelu pro daný účel.“

Při využívání AI, a zejména při implementaci AI systémů do prostředí organizací, musí proto být pečlivě zvažována **bezpečnostní opatření**. Použití AI totiž může vyžadovat sdílení citlivých interních dat s třetí stranou (například s provozovatelem AI). To může zvýšit riziko úniku či zneužití důvěrných informací. Je důležité zajistit, aby byla zajištěna odpovídající ochrana dat a zachování jejich důvěrnosti, či případně bylo definováno, jaké typy dat lze a které nelze do daného AI nahrávat.

Je vhodné pracovat také s rizikem **nepředvídatelného chování** a chyby, kdy AI mohou občas reagovat nepředvídatelným způsobem. To může mít vliv na kvalitu poskytnutých výsledků, může způsobit nedorozumění či škody případně i většího rozsahu, zejména u aktivních AI, v jejichž možnostech je vykonávání úkolů či předávání pokynů k takovým úkolům.

„Jakákoliv AI, ať již online (cloud), nebo offline, je zranitelná vůči kybernetickým útokům a zneužitím.“

I zde je tedy namísto role IA, aby hodnotil zabezpečení datových procesů AI a ověřoval, že jsou dodržovány všechny příslušné normy ochrany dat. Náležitá analýza rizik, spojených s využíváním AI, se taktéž jeví jako výrazně vhodná.

„Používání AI může vyvolávat také široká právní rizika.“

Z pohledu organizací hraje důležitou roli také **lidský faktor**, tedy uživatelé. Ti by měli být při práci s AI adekvátně obezřetní. To z pohledu organizace znamená správně je proškolení a nastavit vhodný dohled, resp. pravidla v podobě interních předpisů. Jinak může snadno docházet k chybám při používání i implementaci těchto technologií. I zde se může uplatnit IA jako kontrolní linie toho, zda jsou správně nastavena pravidla používání a implementace (včetně technického řešení) technologie, zda existuje efektivní dohled nad provozem AI systémů a zda jsou uživatelé dostatečně vyškoleni.

Opět i zde platí omílané, že samotný člověk, uživatel, může být nejslabším článkem řetězu.

Kontrola nad provozem AI, ať již automatizovaná, nebo manuální, je důležitá, neboť některé AI systémy mohou být schopny operovat autonomně a bez lidského dohledu, a poskytovat i aktivní operace. To může zvýšit riziko chyb nebo nežádoucích výsledků.

Alespoň prozatím, k polovině roku 2023, lze uvést, že AI zatím postrádá širší rozpětí jakéhosi lidského uvážení, a zejména **lidské intuice**. Což může být důležité při analýze komplexních situací (ne zcela založených na datech) a zjišťování nepřímých důkazů. AI tak nemusí vždy správně chápat specifický kontext interního auditu, což může vést k nesprávným závěrům nebo přehlédnutí klíčových informací. Některé AI modely již dnes excelují v úkolech založených na analýze obřích množství dat, ale mohou mít stále ještě určité obtíže s vykazováním kreativity, empatie a s porozuměním složitým lidským kontextům. Nicméně opět platí, že již za několik měsíců může být toto zcela jinak.

„Při využívání AI, a zejména při implementaci AI systémů do prostředí organizací, musí proto být pečlivě zvažována bezpečnostní opatření.“

Používání AI může vyvolávat také široká **právní** rizika a otázky týkající se odpovědnosti za rozhodnutí AI, autorských práv atd. Může také vyvolat širší **etická** témata a dilemata ohledně toho, zda a jak má AI přijímat etická rozhodnutí.

Právě rychlý vývoj AI může znamenat, že **regulace a standardy** nejsou dostatečně připraveny na řešení

všech jeho aspektů, což může vést k právním a bezpečnostním nejistotám, rizikům a rizikům dalších či zmařených investic.

Nad rámec tohoto, v kontextu **ESG**, může mít rozšíření AI v některých odvětvích za následek přeskládání struktury pracovních míst, zejména v rutinních a opakujících se činnostech. Nemusí zcela automaticky vést ke snížení pracovních míst, ale nejspíše povede k diverzifikaci mezi těmi zaměstnanci, kteří umí využívat možnosti AI, a kteří nikoli.

Dále, velmi přeneseně, mohou být i jisté environmentální dopady spojené s tím, že rychlý nárůst využívání AI může vyžadovat velké množství výpočetního výkonu, což může zvýšit spotřebu energie v datacentrech, a mít tak negativní dopad na životní prostředí.

„Kontrola nad provozem AI, ať již automatizovaná, nebo manuální, je důležitá.“

Na závěr shrňme, že nejen z pohledu IA je důležité si uvědomit, že AI by ve svých rozličných podobách měla být používána jako **podpůrný nástroj** při interním auditu a jiných aktivitách, a nikoli zcela jako náhrada za lidskou odbornost a rozsáhlé zkušenosti auditorských týmů. Prozatím. Minimálně pro výkon interního auditu je klíčové zachovávat kritické myšlení, provádět ověření informací a používat technologii AI s obezřetností a zodpovědností, a to jak při využívání AI, tak při jejím auditování.

Navzdory rizikům může být AI jednoznačně cenným nástrojem pro organizace i interní audit. Při používání AI je však důležité být si vědom rizik a přijmout opatření k jejich snížení.

Pro minimalizaci nejen uvedených rizik je vhodné zvážit:

- Zavedení správného využití AI jako nástroje podpory, nikoliv jako jediného zdroje informací.
- Implementaci opatření pro zajištění bezpečnosti dat (technická, uživatelská).
- Pravidelnou aktualizaci a ověřování dat použitých pro učení modelu, je-li to možné.
- Vytvoření dokumentace, která popisuje, jak AI používáte, co je povoleno a co není a jaká jsou rizika spojená s jejím používáním ve vašem prostředí.
- Explicitní definování odpovědnosti za výsledky práce, do kterých byly použité jako vstupy výstupy AI.
- Pravidelné monitorování a sledování vývoje AI, např. novinky v nově implementovaných verzích, jsou-li, a návazné přijímání opatření ke snížení nových rizik.
- Zajištění vhodných školení a dohledu pro správné a efektivní používání AI uživateli.
- Faktor kritického posouzení, interpretace a ověření výstupů AI lidskými auditory a uživateli.

Výhody AI mohou být velké, pokud jsou správně využity, nicméně je velmi důležité pečlivě zvažovat všechny aspekty jejího použití. Interní audit může zaujmout i v této oblasti jakousi klíčovou roli při minimalizaci rizik spojených s využíváním AI v organizacích. Je proto i v této oblasti důležité, aby auditoři byli dobře seznámeni s AI, jejími možnostmi a měli schopnosti identifikovat a adresovat potenciální rizika ještě o něco málo dříve, než skutečně přijdou.

Napiš mi, prosím, 20 rizik spojených s výkonem interního auditu.

„Samozřejmě, zde je seznam 20 rizik spojených s výkonem interního auditu:

- Nepřiměřená nezávislost auditorského týmu a možnost vlivu managementu na auditní proces.
- Konflikty zájmů, které by mohly snížit objektivitu a integritu auditu.
- Nedostatečně kvalifikovaný personál interního auditu.
- Nedostatečná znalost a porozumění procesů, postupů a rizik organizace.
- Nedostatečné pokrytí klíčových oblastí auditního plánu.
- Neefektivní plánování a alokace zdrojů pro audity.
- Nedostatečná spolupráce zaměstnanců a manažerů při auditu.
- Nedostatečná transparentnost auditních postupů a zjištění.
- Nepřiměřená dokumentace auditorských prací a závěrů.
- Nevyhovující sledování a monitorování implementace doporučení auditu.
- Nerespektování příslušných právních a regulačních požadavků.
- Nedostatečná ochrana a bezpečnost auditních dat.
- Neadekvátní kontrolní mechanismy a dohled nad auditními procesy.
- Chybějící nebo nedostatečně aktualizovaný interní auditorský manuál.
- Neúčinné hodnocení rizik a identifikace nových rizikových oblastí.
- Neadekvátní komunikace s vedením a dalšími zúčastněnými stranami.
- Nepřiměřené závěry a doporučení auditorského týmu.
- Neefektivní sledování a posuzování pokroku při implementaci auditorských doporučení.
- Nedostatečné posílení schopností a rozvoj dovedností interního auditu.
- Neodhalení závažných rizik a nedostatečných interních kontrol.

*Paměť si prosím, že některá z těchto rizik lze minimalizovat správným plánováním, striktním dodržováním standardů a etických principů, stejně jako kvalitním řízením interního auditu.**

Generováno pomocí jazykového modelu
OpenAI GPT-3.5. @ chat.openai.com; červen 2023.

Tak a je to. Trochu s překlepy, jedno riziko se opakuje dvakrát, ale kdo z nás by toto vyjmenoval za cca 6 vteřin?

À propos, v článku jsou použity inspirace ze zdrojů AI. ■

S autorem článku se můžete setkat na seminářích ČIIA

Lektor: **PhDr. Ing. Jiří Čihák**

REGISTRUJTE SE NA SEMINÁŘE

■ Audit IT pro non-IT auditory

1. prosinec 2023

■ Praktický průvodce

začínajícího auditora: aneb od programu, přes zajišťování důkazů, po prezentaci zprávy interního auditu

8. prosinec 2023

■ Trendy v IT

13. prosinec 2023

Objednávejte na webu ČIIA
(www.interniaudit.cz) nebo
na prihlasky@interniaudit.cz