Jak eticky a adekvátně využívat umělou inteligenci

Definice

Umělá inteligence (AI) je vývoj počítačových systémů schopných plnit úkoly, které obvykle vyžadují lidskou inteligenci. Mezi tyto úkoly patří uvažování, rozpoznávání vzorců, učení se z dat a rozhodování. Etické využívání umělé inteligence se zaměřuje na vytváření technologií umělé inteligence, které jsou nejen účinné, ale také v souladu s morálními a společenskými hodnotami. To zahrnuje řešení etických otázek, jako je spravedlnost, transparentnost, odpovědnost a soukromí.

Etická umělá inteligence zajišťuje, že systémy umělé inteligence fungují nezaujatě, respektují soukromí uživatelů a fungují transparentně, aby zúčastněné strany rozuměly rozhodnutím o umělé inteligenci. Adekvátní využití umělé inteligence navíc vyžaduje efektivitu, dodržování etických norem a důsledné monitorování, aby se zmírnila rizika, jako jsou nezamýšlené důsledky nebo zneužití technologie.

Integrací těchto principů lze umělou inteligenci zodpovědně využít ke zvýšení produktivity a zlepšení životů při zachování důvěry a bezpečnosti v činnostech závislých na umělé inteligenci. Některá témata v tomto pracovním listu se překrývají s tématy zkoumanými v části "Etické otázky v digitálním podnikání". Pro hlubší analýzu těchto aspektů se čtenářům doporučuje, aby prostudovali rovněž tento dokument.

Proč je to důležité?

Integrace etických postupů do umělé inteligence (AI) má zásadní význam pro zajištění toho, aby technologický pokrok pozitivně přispíval společnosti a respektoval práva jednotlivce.

- **Posiluje důvěru:** etická umělá inteligence buduje důvěru mezi uživateli, která je nezbytná pro přijetí a efektivní využívání technologií ve vzdělávání i mimo něj.
- **Podporuje spravedlnost**: Minimalizací předsudků etická umělá inteligence zajišťuje, že všichni jednotlivci mají stejné příležitosti těžit z technologického pokroku.
- **Chrání soukromí**: etické postupy při využití umělé inteligence chrání osobní údaje, zachovávají soukromí a bezpečnost informací uživatelů.
- **Zvyšuje spolehlivost:** Systémy umělé inteligence navržené podle etických zásad snižují chyby a nekonzistence a zvyšují jejich spolehlivost a efektivitu.

Spolufinancováno Evropskou unií

Financováno Evropskou unií. Názory vyjádřené jsou názory autora a neodráží nutně oficiální stanovisko Evropské unie či Evropské výkonné agentury pro vzdělávání a kulturu (EACEA). Evropská unie ani EACEA za vyjádřené názory nenese odpovědnost.

- **Podporuje inkluzivitu**: eticky vyvinutá umělá inteligence zohledňuje různé potřeby a zázemí, díky čemuž jsou digitální nástroje dostupné a přínosné pro všechny.

Tipy a triky

- **Pochopte a řešte předsudky:** Identifikujte a zmírňujte předsudky v datech a algoritmech, abyste zajistili správné výsledky umělé inteligence, a to pomocí různých tréninkových dat, která přesně reprezentují různé demografické údaje.
- **Upřednostněte transparentnost a odpovědnost:** Udržujte jasnou dokumentaci procesů a rozhodování umělé inteligence a zajistěte otevřenou komunikaci se zúčastněnými stranami o tom, jak systémy umělé inteligence fungují.
- **Implementujte robustní ochranu soukromí:** Využijte silné šifrování a bezpečná řešení pro ukládání dat k ochraně uživatelských dat a integrujte aspekty ochrany osobních údajů ve fázi vývoje aplikací AI.
- **Zajistěte lidský dohled:** Udržujte lidské zapojení do kritických rozhodovacích procesů, což umožňuje dohled a opravy rozhodnutí v oblasti využívání umělé inteligence podle toho, jak bude potřeba**.**
- **Podporujte neustálé zlepšování:** Podporujte průběžnou zpětnou vazbu o etickém chování umělé inteligence a přizpůsobujte pokyny a postupy na základě nových výzev a technologického pokroku.
- **Zahrňte různé pohledy:** Ujistěte se, že je váš tým umělé inteligence různorodý. Přítomnost lidí z různých oborů může pomoci odhalit problémy a předsudky v systémech umělé inteligence, které nemusí být na první pohled zřejmé.
- **Zapojte všechny:** Použijte týmový přístup, kdy zapojíte nejen technické pracovníky, ale také odborníky na etiku a zástupce ze skupin, které by mohly být ovlivněny umělou inteligencí. Ujistěte se, že při vývoji a používání umělé inteligence je zohledňováno mnoho různých úhlů pohledu.

Užitečné zdroje a nástroje

Pochopte a řešte předsudky:

- **AI Fairness 360 (AIF360):** Sada nástrojů od IBM pro zkoumání a zmírňování předsudků v modelech AI.

Spolufinancováno Evropskou unií

Financováno Evropskou unií. Názory vyjádřené jsou názory autora a neodráží nutně oficiální stanovisko Evropské unie či Evropské výkonné agentury pro vzdělávání a kulturu (EACEA). Evropská unie ani EACEA za vyjádřené názory nenese odpovědnost.

- **Nástroj What-If od Googlu:** Nabízí snadno použitelné rozhraní pro zkoumání modelů strojového učení z hlediska spravedlnosti a dalších funkcí bez nutnosti kódování.

Upřednostněte transparentnost a odpovědnost:

- **IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems (Globální iniciativa pro etiku autonomních a inteligentních systémů):** Poskytuje etické standardy a osvědčené postupy pro návrh systémů umělé inteligence.
- **Partnership on AI (Partnerství v oblasti umělé inteligence):** formuluje osvědčené postupy týkající se technologií umělé inteligence a podporuje porozumění veřejnosti

Implementujte robustní ochranu soukromí:

- **Etika v oblasti umělé inteligence a datové vědy na edX:** Pokrývá etické otázky umělé inteligence a poskytuje rámce pro řízení obav o soukromí a bezpečnost.

Podporujte neustálé zlepšování:

- **AI pro každého na Coursera:** Tento kurz, který vyučuje Andrew Ng, pojednává o společenských důsledcích AI a podporuje pokračující etické postupy.

Knihy pro hlubší vhled:

- "**Weapons of Math Destruction**" od Cathy O'Neil – Tato kniha zkoumá temnou stránku velkých dat a předpojatých algoritmů, které mohou posilovat nerovnost
- "**Ethics of Artificial Intelligence (Etika umělé intelligence)**" editoval S. Matthew Liao – Sbírka esejů od předních filozofů, kteří nabízejí filozofické pojednání o etických důsledcích umělé inteligence.

Spolufinancováno Evropskou unií

Financováno Evropskou unií. Názory vyjádřené jsou názory autora a neodráží nutně oficiální stanovisko Evropské unie či Evropské výkonné agentury pro vzdělávání a kulturu (EACEA). Evropská unie ani EACEA za vyjádřené názory nenese odpovědnost.