

JAK DŮVĚRYHODNÁ JE

UMĚLÁ INTELIGENCE?

Co přesně znamená pojem „důvěryhodná umělá inteligence“ (trustworthy AI)? V jakých případech můžete používat AI zodpovědným způsobem?

Pro zodpovězení těchto otázek existují určité zásady pro vývoj a užití AI. Tyto zásady se zaměřují na etické a sociální aspekty a otázky týkající se samotné AI.

Belgické centrum Knowledge Centre Data & Society shrnulo [7 požadavků na AI](#), které byly sepsané expertní skupinou pro umělou inteligenci (High-Level Expert Group on AI, AI HLEG).

Pro každý z požadavků je uvedeno několik otázek, které Vám umožní zvážit důvěryhodnost Vašeho AI systému.

Pokud byste chtěli využít ještě detailnější postup, Knowledge Centre Data & Society v nedávné době vyvinulo [AI Blindspots cards](#). Tento nástroj Vám umožní proaktivně přemýšlet o (etických

a sociálních) rozhodnutích a krocích, které je dobré podniknout na začátku projektu nebo během vývoje daného AI systému.

Upozornění: zmíněných 7 požadavků se částečně překrývá, proto jsme omezili nadbytečné otázky a pro zjednodušení jsme vybrali pouze ty nejnaléhavější.

Knowledge Centre Data & Society (2020). How ethical is AI? brAlnfood from the Knowledge Centre Data & Society. Brussels: Knowledge Centre Data & Society

Tento dokument je k dispozici pod licenci CC BY 4.0

BrAlnfood je společný projekt CLAIRE, D5M a CIIRC CTU, který podporuje společnou vizi EU #AI4Good a #AI4All.

LIDSKÝ FAKTOR A DOHLED

- Je interakce mezi Vaším AI systémem a člověkem smysluplná a relevantní?
- Kdo činí konečné rozhodnutí, stroj nebo člověk? Pokud je to stroj, neexistuje lidský dohled.
- Je Váš AI systém autonomní nebo sebezvědomivý? Pokud ano, existují mechanismy, které nad ním mají dohled?

TECHNICKÁ ROBUSTNOST A BEZPEČNOST

- Jaká opatření učiníte, pokud bude Váš systém napaden, nebo se bude chovat jinak, než jste předpokládali, nebo pokud bude použit pro jiný (nežádoucí) účel?
- Existuje možnost, že Váš systém provede nepřesné předpovědi?
- Jaká opatření jsou zavedena za účelem vypořádání se s nepřesnostmi?

SOUKROMÍ A SPRÁVA DAT

- Používá Váš AI systém osobní data? Pokud ano, jste si vědom/a dopadů a požadavků při použití osobních dat? Můžete dokázat, že dodržujete nařízení týkající se ochrany osobních údajů?
- Máte mechanismus pro dohled, sběr, uložení, zpracování a použití osobních dat?
- Kdo má přístup k uživatelským datům? Máte zavedena opatření a postupy týkající se přístupu k datům?

TRANSPARENTNOST

- Můžete Vašemu týmu, ale také všem, kdo jsou se systémem v kontaktu, ukázat, jak je daný algoritmus navržen a sestaven, a jak jsou dělána rozhodnutí? Ujistěte se, že všichni porozumí tomu, že AI složka je součástí systému.

ROZMANITOST, NEDISKRIMINACE A FÉROVOST

- Jak se vyhnete předpojatosti v rámci Vašeho AI systému? Je možné, aby ostatní poukázali na zkrslení nebo diskriminaci?
- Zvážil/a jste, jak Vaše AI inovace mění přístup k Vaší službě vyloučeným skupinám? Pokud ano, poskytnete alternativní službu takto vyloučeným?
- Mohou se zainteresované subjekty zapojit do vývoje a použití Vašeho AI systému?

DOBŘE ŽIVOTNÍ PODMÍNKY PRO SPOLEČNOST A ŽIVOTNÍ PROSTŘEDÍ

- Jsou dopady životního cyklu Vašeho AI systému na životní prostředí měřitelné? Víte, jak je zmírnit?
- Jsou si (koncoví) uživatelé vědomi omezení v rámci sociální interakce s Vaším AI systémem a znají společenské dopady Vašeho AI systému?
- Mohly by být i jiné skupiny a jedinci, kromě Vaší cílové skupiny, ovlivněni Vaším AI systémem?

ZODPOVĚDNOST

- Máte přehled všech rozhodnutí a kompromisů, které jste učinil/a při tvorbě Vašeho systému?
- Identifikoval/a jste negativní dopady na všechny zapojené strany?
- Byl/a jste schopen/na snížit negativní dopady a jsou tato opatření zdokumentovaná?

brAlnfood of the Knowledge Centre Data & Society



CLAIRE

