Как да използваме изкуствения интелект етично и адекватно

Определение

Изкуственият интелект (AI) е разработката на компютърни системи, способни да изпълняват задачи, които обикновено изискват човешки интелект. Тези задачи включват разсъждение, разпознаване на модели, учене от данни и вземане на решения. Етичното използване на AI се фокусира върху създаването на AI технологии, които са не само ефективни, но и са в съответствие с моралните и обществени ценности. Това включва справяне с етични проблеми като справедливост, прозрачност, отчетност и поверителност.

Етичният AI гарантира, че AI системите работят безпристрастно, зачитат поверителността на потребителите и работят прозрачно, така че заинтересованите страни да разбират решенията на AI. Освен това, адекватното използване на AI изисква ефективност, спазване на етичните стандарти и последователен мониторинг за смекчаване на рискове като непредвидени последици или злоупотреба с технологията.

Чрез интегрирането на тези принципи AI може да се използва отговорно за повишаване на производителността и подобряване на живота, като същевременно се поддържа доверие и безопасност в дейности, зависими от AI. Някои теми в този работен лист се припокриват с тези, разгледани в „Етични въпроси в дигиталното предприемачество“. За по-задълбочен анализ на тези аспекти читателите се насърчават да направят справка с този документ.

Защо е важно?

Интегрирането на етични практики в изкуствения интелект (AI) е от решаващо значение за гарантиране, че технологичният напредък допринася положително за обществото и зачита индивидуалните права.

- **Насърчава доверието:** Етичният AI изгражда доверие сред потребителите, което е от съществено значение за възприемането и ефективното използване на технологиите в образованието и извън него.

Съфинансирано от Европейския съюз

Финансирано от Европейския съюз. Изразените възгледи и мнения обаче принадлежат изцяло на техния(ите) автор(и) и не отразяват непременно възгледите и мненията на Европейския съюз или на Европейската изпълнителна агенция за образование и култура (EACEA). За тях не носи отговорност нито Европейският съюз, нито EACEA.

- **Насърчава справедливостта**: като минимизира пристрастията, етичният AI гарантира, че всички хора имат равни възможности да се възползват от технологичния напредък.
- **Защитава поверителността**: етичните AI практики защитават личните данни, като поддържат поверителността и сигурността на информацията на потребителите.
- **Повишава надеждността:** AI системите, проектирани с етични насоки, намаляват грешките и несъответствията, повишавайки тяхната надеждност и ефективност.
- **Подкрепа за приобщаване**: Етично разработеният изкуствен интелект отчита различните нужди и произход, като прави цифровите инструменти достъпни и полезни за всички.

Съвети и трикове

- **Разберете и адресирайте пристрастията:** Идентифицирайте и смекчете пристрастията в данните и алгоритмите, за да осигурите справедливи резултати от AI, като използвате разнообразни данни за обучение, които точно представят различни демографски данни.
- **Дайте приоритет на прозрачността и отчетността:** Поддържайте ясна документация за процесите и решенията на AI и осигурете открита комуникация със заинтересованите страни относно това как работят системите с AI.
- **Внедряване на стабилна защита на поверителността:** Използвайте силно криптиране и сигурни решения за съхранение на данни, за да защитите потребителските данни, като интегрирате съображенията за поверителност на етапа на разработка на AI приложения.
- **Осигурете човешки надзор:** Поддържайте човешко участие в критичните процеси на вземане на решения, позволявайки надзор и коригиране на решенията на AI, ако е необходимо.
- **Насърчавайте непрекъснато подобрение:** Насърчавайте постоянната обратна връзка относно етичното представяне на ИИ и адаптирайте насоки и практики въз основа на възникващи предизвикателства и технологичен напредък.

Съфинансирано от Европейския съюз

Финансирано от Европейския съюз. Изразените възгледи и мнения обаче принадлежат изцяло на техния(ите) автор(и) и не отразяват непременно възгледите и мненията на Европейския съюз или на Европейската изпълнителна агенция за образование и култура (EACEA). За тях не носи отговорност нито Европейският съюз, нито EACEA.

- **Включете различни гледни точки:** Уверете се, че вашият AI екип е разнообразен. Наличието на хора от различен произход може да помогне за откриването на проблеми и пристрастия в системите с изкуствен интелект, които може да не са очевидни в началото.
- **Включете всички:** Използвайте екипен подход, при който включвате не само технически хора, но и експерти по етика и представители на групи, които може да бъдат засегнати от AI. Уверете се, че много различни гледни точки са взети предвид при разработването и използването на AI.

Полезни ресурси и инструменти

Разберете и адресирайте пристрастията:

- **AI Fairness 360 (AIF360):** Инструментариум от IBM за изследване и смекчаване на пристрастията в моделите на AI.
- **Google's What-If Tool:** Предлага лесен за използване интерфейс за изследване на модели за машинно обучение за справедливост и други функции, без да е необходимо кодиране.

Дайте приоритет на прозрачността и отчетността:

- **IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems:** Осигурява етични стандарти и най-добри практики за проектиране на AI системи.
- **Partnership on AI:** Осигурява етични стандарти и най-добри практики за проектиране на AI системи.

Приложете стабилна защита на поверителността:

- **Ethics in AI and Data Science on edX:** Покрива етичните въпроси AI и предоставя рамки за управление на проблемите, свързани с поверителността и сигурността.

Насърчавайте непрекъснато подобрение:

- **AI For Everyone on Coursera:** Преподаван от Andrew Ng, този курс обсъжда социалните последици от AI и насърчава продължаващите етични практики.

Съфинансирано от Европейския съюз

Финансирано от Европейския съюз. Изразените възгледи и мнения обаче принадлежат изцяло на техния(ите) автор(и) и не отразяват непременно възгледите и мненията на Европейския съюз или на Европейската изпълнителна агенция за образование и култура (EACEA). За тях не носи отговорност нито Европейският съюз, нито EACEA.

Книги за по-дълбоки прозрения:

- "**Weapons of Math Destruction**" от Cathy O'Neil („Оръжия за унищожаване на математиката“) – Тази книга изследва тъмната страна на големите данни и предубедените алгоритми, които могат да засилят неравенството.
- "**Ethics of Artificial Intelligence**" редактиран от S. Matthew Liao („Етика на изкуствения интелект“) – Колекция от есета от водещи философи, предлагащи философско третиране на етичните последици от AI.

Съфинансирано от Европейския съюз

Финансирано от Европейския съюз. Изразените възгледи и мнения обаче принадлежат изцяло на техния(ите) автор(и) и не отразяват непременно възгледите и мненията на Европейския съюз или на Европейската изпълнителна агенция за образование и култура (EACEA). За тях не носи отговорност нито Европейският съюз, нито EACEA.