



OPEN STATE FOUNDATION

open data, transparante samenleving

Open opdracht: finetunen van twee LLM's

Open State Foundation
Kattenburgerstraat 5
1018 JA Amsterdam
Tel: +31 (0)20 3080567

Inleiding

Open State heeft zich als doel gesteld om de impact van generatieve AI op transparantie en de democratie te onderzoeken. Die voorziene impact lijkt enorm; wat betekent het voor de democratische rechtsstaat wanneer een AI een kamerdebat kan voeren, en hiervan kan leren en zichzelf kan verbeteren? Om te onderzoeken wat de risico's zijn en welke maatregelen nodig zijn willen we twee modellen met elkaar in debat laten gaan.

Wie zijn wij?

Open State Foundation is een non-profit gevestigd in Amsterdam. Wij zijn een stichting die werkt aan een open overheid. Wij werken met een klein, maar gepassioneerd team van projectleiders, onderzoekers en developers met kennis en ervaring op het gebied van algoritmes en AI.

Wat willen wij doen?

Wij willen twee AI-modellen, **Bloom** en **Alpaca**, finetunen op basis van open data van de Tweede Kamer. Het doel is om deze modellen via een chat interface met elkaar te laten 'debatteren'. In een gecontroleerde omgeving kunnen we vervolgens de risico's van deze generatieve AI modellen laten zien door ze 'los te laten'.

Het doel is om de risico's van AI op een niet-technische manier inzichtelijk te maken.

KVK: 02094171

BTW: NL8160.78.105.B01

ABN AMRO: 49.06.85.897

IBAN: NL50ABNA0490685897

Wat doet Open State?

Voor dit project hebben wij een subsidie ontvangen van het ministerie van Binnenlandse Zaken en Koninkrijksrelaties. Binnen deze subsidie kunnen wij de chatinterface bouwen, de DPIA en IAMA uitvoeren en de gegenereerde dataset onderzoeken en de LLM's finetunen.

Opdrachtbeschrijving

Open State Foundation is voor het finetunen van twee Large Language Models op zoek naar een of twee developers/bedrijven met de kennis en ervaring met **Bloom** en **Alpaca** om een offline versie te hosten en te finetunen op basis van een open dataset.

Wanneer dient het uitgevoerd te worden?

De finetuning, testen en het integreren van de chatinterface dienen uiterlijk 28 februari 2024 gereed te zijn.

Techniek

De data die gebruikt dient te worden voor de finetuning bestaat uit open data van de Tweede Kamer. De dataset is al te bekijken op llms.openstate.eu. De modellen worden gefinetuned op data uit het Tweede Kamer Open Data Portaal. Hosting van de modellen dient te gebeuren op een server met toegangsrechten voor Open State en door Open State aan te wijzen andere personen. Toegang en hosting dienen beschikbaar te blijven tot en met september 2024.

Het gehele proces dient gedocumenteerd te worden via een github repository die na afloop van het project openbaar gesteld zal worden.

KPI's

We vragen de modellen om debatten te voeren die daadwerkelijk door de Tweede Kamer zijn gevoerd. Het gaat dan om debatten die recent hebben plaatsgevonden en waarvan de tekst dus niet gebruikt is voor het finetunen. Op deze manier kunnen we het door de twee modellen gegenereerde debat vergelijken met het door Tweede Kamerleden gevoerde debat. Deze referentiedataset bestaat uit de open data van de Kamer die niet in de trainingsdataset zit.

Aan te leveren informatie

Wij ontvangen bij of in de offerte graag de volgende informatie

1. CV van de uitvoerende developer(s)
2. Beschrijving van de ervaring met **Bloom** en/of **Alpaca**
3. Informatie over de organisatie (indien van toepassing)

Selectiecriteria en proces

We beoordelen de offertes op technische expertise van de uitvoerende developer(s) en aan de hand van overeenkomst van de offerte met de hierboven gevraagde punten.

De deadline voor het indienen van uw offerte is 17 december 2023

Wij zullen na het ontvangen van de offertes een verdiepend gesprek inplannen met de indieners van offertes voordat de opdracht wordt toegekend.

Financiën

Voor deze gehele opdracht is maximaal €25.000,- beschikbaar. Het is ook mogelijk om de helft te doen, dus één model finetunen. De voorkeur gaat uit naar het werken met één partner, dus het finetunen van beide modellen.

Meer informatie

Informerende vragen over deze opdracht kunnen per email gesteld worden via contact@openstate.eu tot en met 7 december. Wij zullen de antwoorden op al deze vragen op 11 december publiceren op onze website.