

La face (très) sombre de ChatGPT ...

Révélations sur les pratiques d'OpenAI pour entraîner des modèles d'intelligence artificielle

L'intelligence artificielle (IA) est un domaine en pleine expansion, notamment avec le développement de modèles comme GPT (Generative Pretrained Transformer) qui permettent de générer des poèmes, des recettes de cuisine ou même des séances de sport. Cependant, pour que ces modèles fonctionnent, il est nécessaire de leur fournir des données labellisées. C'est ainsi que des entreprises comme OpenAI ont recours à des filtres et à des classificateurs pour trier les données en entrée.

Pourtant, de récentes révélations du Time ont mis en lumière les pratiques d'OpenAI pour entraîner leur modèle GPT. La classification et le filtrage du texte et des images nuisibles sont une étape nécessaire pour minimiser la quantité de contenu violent et sexuel inclus dans les données de formation et créer des outils capables de détecter le contenu nuisible. Cependant, pour cela, OpenAI a recours à des entreprises qui emploient des travailleurs dans des pays où la main-d'œuvre est peu chère, comme le Kenya ou l'Asie, pour classer des milliers d'images violentes et potentiellement illégales.

L'entreprise sama, employée par OpenAI pour trier ces données, paie ses employés moins de 2 dollars par jour pour passer toute la journée devant un écran à trier des contenus violents, y compris des images impliquant des enfants. Les travailleurs sont exposés à des scènes sordides et

traumatisantes, avec entre 150 et 200 textes violents à classer toutes les heures, soit jusqu'à 7200 textes par jour. Les travailleurs subissent des traumatismes, comme des visions récurrentes notamment après avoir vu des images obscènes.

En conséquence, l'entreprise sama a signé trois contrats pour étiqueter des données, pour un montant total de 200 000 €, tandis que les travailleurs ne touchaient qu'une fraction de ce montant. OpenAI a quant à elle minimisé les dommages causés aux travailleurs et a déclaré qu'elle travaillait avec des entreprises tierces pour gérer cette question. Cette affaire révèle la facette sombre et méconnue du développement des modèles d'IA, qui requièrent des milliers d'heures de travail et pour lesquels la question du traitement des travailleurs est souvent loin d'être résolue.

Source : [Defend Intelligence](#) | Date : 2023-01-29 18:00:30 | Durée : 00:10:29

→  **Accéder à [CHAT GPT](#) en cliquant dessus**