Anleitung zum Blockierenvon GPTBot auf Webseiten

OpenAI hat eine Anleitung veröffentlicht, die zeigt, wie Webseiten das Crawlen durch seinen GPTBot verhindern können. Über den GPTBot-Benutzeragenten ausgewertete Webseiten tragen laut OpenAI potenziell dazu bei, das Modell von ChatGPT zu verbessern. OpenAI betont, dass die vom Bot gesammelten Informationen gefiltert würden. Dabei entferne man etwa Quellen, die einen Paywall-Zugang erfordern, die persönlich identifizierbare Informationen sammeln oder Text enthalten, der gegen die OpenAI-Richtlinien verstößt, heißt es weiter. Wer dem GPTBot den Zugang zu seiner Website gewähre, könne den KI-Modellen dadurch helfen, genauer zu werden und ihre allgemeinen Fähigkeiten sowie die Sicherheit zu verbessern.

Um den AI-Bot fernzuhalten, genügt beispielsweise folgender Eintrag in der robots.txt:
User-agent: GPTBot
Disallow: /

Alternativ funktioniert eine Blockade der IP-Adressen des Bots.
Die finden sich auf https://openai.com/gptbot-ranges.txt.