KI Prinzipien: Kinder müssen vor Missbrauch geschützt werden

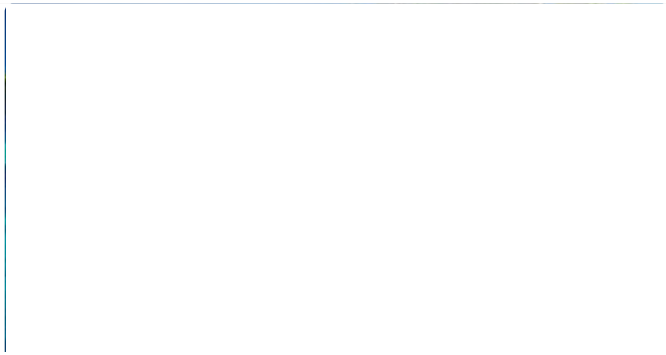
24.4.24 7:35 von Sunny Lesezeit: 3 Min.

KI-Prinzipien: Unternehmen, die Künstliche Intelligenz entwickeln, verpflichten sich, den sexuellen Missbrauch von Kindern zu bekämpfen.

INHALT

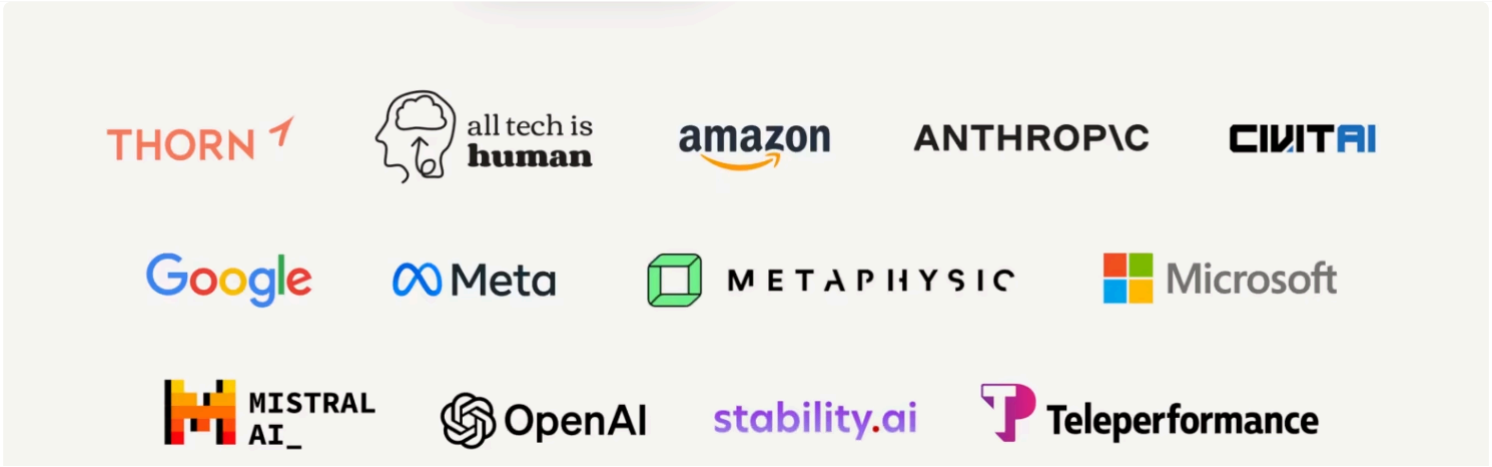
- KI Prinzipien: Unternehmen versprechen Sicherheit für Kinder
- Safety by Design for Generative AI
- Ein wichtiger Schritt in die richtige Richtung

Im Bereich der künstlichen Intelligenz (KI) soll ein neues Kapitel aufgeschlagen werden. Die Sicherheit von Kindern soll ganz klar an erster Stelle stehen. Führende KI-Unternehmen wie OpenAI, Microsoft, Google, Meta und andere haben sich zusammengeschlossen. Sie wollen sicherstellen, dass ihre Technologien nicht für den sexuellen Missbrauch von Minderjährigen missbraucht werden können.



KI Prinzipien: Unternehmen versprechen Sicherheit für Kinder

OpenAI, Microsoft, Google, Meta und einige andere sind klare Verpflichtungen eingegangen. Die Initiative, die von der Kinderschutzgruppe Thorn und der gemeinnützigen Organisation All Tech Is Human angeführt wird, stellt einen großen Schritt für die Branche dar.



tarnkappe.info

Die Bemühungen dieser Unternehmen und Organisationen sollten nicht unterschätzt werden. Denn laut Thorn wurden 2023 allein in den USA mehr als 104 Millionen Dateien mit mutmaßlichem Material über sexuellen Kindesmissbrauch gemeldet. Ohne kollektives Handeln wird die Entwicklung der künstlichen Intelligenz dieses Problem weiter verschärfen. Die Strafverfolgungsbehörden werden es künftig noch schwerer haben.

Safety by Design for Generative AI

Strategien und Empfehlungen

Thorn und All Tech Is Human haben ein neues Papier mit dem Titel „Safety by Design for Generative AI: Preventing Child Sexual Abuse“ veröffentlicht, das konkrete Strategien und Empfehlungen enthält, um sicherzustellen, dass generative KI nicht zum Schaden von Kindern eingesetzt wird. Dies berichtet Engadget in einem aktuellen Artikel.

Eine dieser empfohlenen „KI Prinzipien“ bezieht sich auf die Auswahl von Datensätzen für das Training von KI-Modellen. Es wird vorgeschlagen, Datensätze zu vermeiden, die nicht jugendfreie Inhalte für Erwachsene enthalten. Denn generative KI neigt dazu, solche Daten zu kombinieren.

Ebenso werden Social-Media-Plattformen und Suchmaschinen aufgefordert, Links zu Websites und Apps zu entfernen, die „Nacktbilder“ von Kindern zeigen, um die Verbreitung von neuem KI-generiertem Material über sexuellen Kindesmissbrauch einzudämmen.

Ein wichtiger Schritt in die richtige Richtung

Rebecca Portnoff, Vice President of Data Science bei Thorn, betonte, dass es keine Entschuldigung für Untätigkeit gibt. Einige Unternehmen haben bereits Maßnahmen ergriffen. Sie trennen Bilder und Videos von Kindern von nicht jugendfreien Inhalten, um zu verhindern, dass ihre Modelle beides vermischen. Andere verwenden Wasserzeichen, um KI-generierte Inhalte zu identifizieren. Metadaten und Wasserzeichen sind aber leicht zu entfernen. Daher gilt diese Methode nicht als sonderlich sicher.

Die Bemühungen dieser Unternehmen und Organisationen sind ein kleiner, aber wichtiger Schritt, um die Sicherheit von Kindern im Internet zu gewährleisten. Denn auf diese Weise kann hoffentlich eines Tages sichergestellt werden, dass die von uns verwendeten Technologien nicht zur Ausbeutung von Minderjährigen beitragen.

3 Kommentare lesen Mehr zu dem Thema

Über Sunny

Sunny schreibt seit 2019 für die Tarnkappe. Er verfasst die wöchentlichen Lesetipps und berichtet am liebsten über Themen wie Datenschutz, Hacking und Netzpolitik. Aber auch in unserer monatlichen Glosse, in Interviews und in „Unter dem Radar“ - dem Podcast von Tarnkappe.info - ist er regelmäßig zu hören.