

Motivation

Diese Fragen sind hauptsächlich dafür gedacht, Erstellern von Datensets dazu zu animieren, ihre Gründe für die Erstellung von Datensets klar zu kommunizieren und für Transparenz über Finanzierung und Interessen.

Für welchen Zweck wurde das Datenset erstellt? Gab es eine spezifische Aufgabe, die dafür angedacht war? Gab es eine bestimmte Lücke, die gefüllt werden musste? Bitte formulieren Sie eine Beschreibung.

Wer hat das Datenset erstellt (z.B. welches Team, welche Forschungsgruppe) und in wessen Auftrag (z.B. Firma, Institution, Organisation)?

Wer hat die Erstellung des Datensets bezahlt?

Welche Listen mit Klassen, Tags und Attributen gibt es? Wieso wurden diese Klassen ausgewählt?

Zusammensetzung

Datenset-Ersteller sollten sich diese Fragen vor der Datensammlung durchlesen und Antworten bereitstellen, wenn die Datensammlung beendet ist. Die meisten der Fragen hier sind dazu da, den Nutzern von Datensets mit den Informationen bereitzustellen, durch die sie Entscheidungen treffen können für die Nutzung des Datensets für den ausgewählten Task. [...]

Welche Instanzen werden durch das Datenset repräsentiert (z.B. Dokumente, Fotos, Personen, Länder)? Bitte eine Beschreibung beifügen.

Enthält das Datenset alle möglichen Instanzen oder ist es eine Stichprobe (nicht notwendigerweise zufällige) eines größeren Datensets? Wenn das Datenset eine Stichprobe ist, was ist das größere Datenset? Ist das Datenset repräsentativ für das größere Datenset? Wenn ja, bitte beschreiben, wie die Repräsentativität validiert/verifiziert wurde. Falls nicht, bitte beschreiben, wieso es nicht repräsentativ ist.

Gibt es irgendwelche fehlenden Informationen in den einzelnen Instanzen? Wenn ja, bitte eine Beschreibung angeben, weshalb die Information fehlt.

Gibt es Fehler, Rauschen oder Redundanzen im Datenset? Wenn ja, bitte eine Beschreibung angeben.

Enthält das Datenset vertrauliche Informationen? Wenn ja, bitte eine Beschreibung angeben.

Enthält der Datenset Daten, die bei direkter Betrachtung anstößig, beleidigend oder bedrohlich wirken oder anderweitig Ängste auslösen könnten? Wenn ja, bitte eine Beschreibung angeben.

Enthält der Datenset Daten, die in irgendeiner Weise als sensibel angesehen werden könnten? (z. B. Daten, die Aufschluss über Rasse oder ethnische Herkunft, sexuelle Orientierung, religiöse Überzeugungen, politische Meinungen [...])? Wenn ja, bitte eine Beschreibung angeben.

Datensammelungsprozess

Welche Mechanismen oder Prozeduren wurden verwendet, um die Daten zu sammeln (z.B. Hardware, Sensoren, manuell durch Menschen, APIs...)? Wie wurden diese Mechanismen und Prozeduren validiert?

Wenn das Datenset Stichprobe eines größeren Datenset ist, wie wurde diese ausgewählt?

Wer war involviert in den Datensammelungsprozess (z.B. Studierende, Crowdfunder, ...) und wie viel Geld erhielten sie?

Über welchen Zeitraum wurden die Daten gesammelt?

Gab es ein ethisches Review?

Haben Sie die Daten direkt von den Individuen gesammelt oder über Dritte erhalten (z.B. Websites)?

Wurden die Individuen über die Datensammlung informiert?

Haben die Individuen der Datensammlung zugestimmt?