Umgang mit KI-Modellen

In der KI-Toolbox werden in den verschiedenen Modulen unterschiedliche KI-Modelle angeboten.

Die aktuell verwendeten Modelle finden Sie in einer Übersicht unter https://ki-toolbox.tu-braunschweig.de/authenticationToken/show. Loggen Sie sich hierzu bitte ein.

Die Auswahl der Modelle wird bei den "on-premise" Modellen durch beschränkte Ressourcen begrenzt. Hier versuchen wir folgendes umzusetzen:

- Mindestens ein kleines Modell, welches sich durch weniger Parameter, aber gute Leistungsfähigkeit auszeichnet
- Mindestens ein großes Modell, welches mehrere Milliarden Parameter unterstützt.
- Mindestens ein Modell, welches sich durch besondere Fähigkeiten oder Technologien auszeichnet, wie beispielsweise Multimodalität (z.B. ein Vision Modell, womit die Inhalte von Bildern interpretieren können), Reasoning Modelle (Modelle, die schlussfolgern und damit zu besseren Ergebnissen kommen). Die Liste besonderer Fähigkeiten wird sich im Laufe der Zeit entwickeln, entsprechend des technologischen Fortschritts.

Daher wird es über die Zeit immer wieder zu Anpassungen der von uns angebotenen Modelle kommen. Wenn wir uns also entscheiden, ein Modell zu entfernen, handelt es sich in der Regel um ältere oder wenig genutzte Modelle oder um Modelle, die wir gegen neuere und bessere Varianten austauschen. Wir geben dies im Voraus bekannt. Falls ein bestimmtes Modell für Forschung oder Lehre benötigt wird, sprechen Sie uns bitte an. Wir versuchen gemeinsam eine Lösung zu finden. Beispielsweise indem wir den Betrieb für die Dauer eines Projektes oder einer Abschlussarbeit verlängern oder wir sprechen mit unseren Partnern in der Forschung und den Rechenzentren, ob ein bestimmtes Modell andernorts temporär weiter betrieben werden kann.

Bei den Modellen der eingebundenen Hersteller und anderen Organisationen sind wir auf deren Politik zu den angebotenen Modellen angewiesen. Hier kommt es immer wieder vor, dass gerade ältere Modelle abgekündigt werden. Wir versuchen zu den abgekündigten Modellen das Datum der Entfernung transparent darzulegen.

Created 2025-06-04 04:23:18 UTC by Jan-Marc Pilawa
Updated 2025-06-04 09:00:46 UTC by Jan Hegewald