Universität Stuttgart
Technische Informations- und
Kommunikationsdienste (TIK)

Benutzung der API in RAI 2.0

Fritz Hohl

2.11.2026

RAI API

- RAI (2.0) hat jetzt eine API
- Allerdings keine OpenAI API-kompatible API, sondern eine mit einem etwas anderen Format, die auch kein Tool-Calling kann
- Das ist weniger ein Problem, wenn man sich seine Software selber schreibt, erlaubt es einem aber nicht, diese API einfach an bestehende Werkzeuge anzubinden

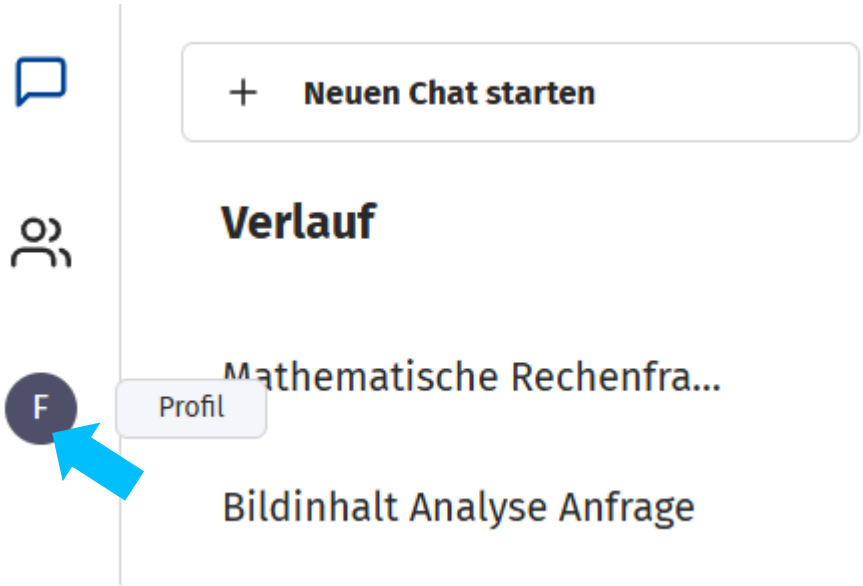
Wichtig! Benutzung von C1-Modellen

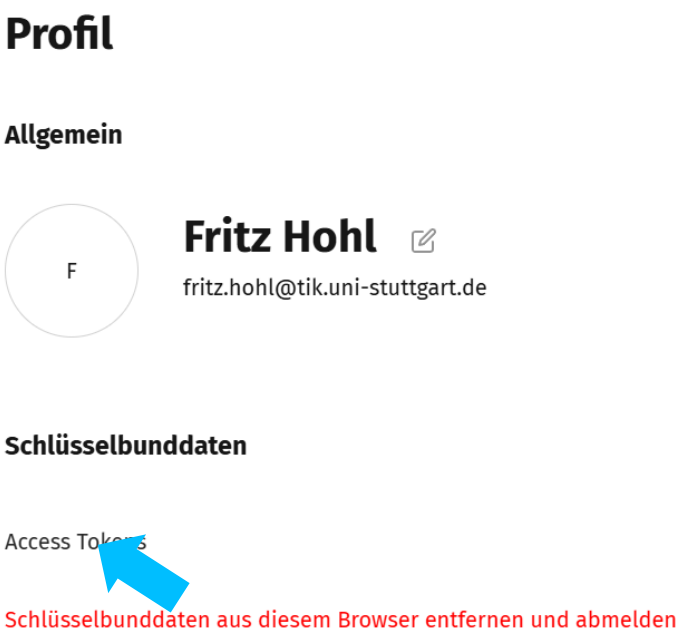
- Sie können die gleichen Modelle verwenden, die Sie auch interaktiv in RAI 2.0 verwenden können

- **Bitte beachten Sie, dass einige der Azure-Modelle** (diejenigen, die mit „C1-“ beginnen), **ausschließlich für C1-Zwecke verwendet werden dürfen!**

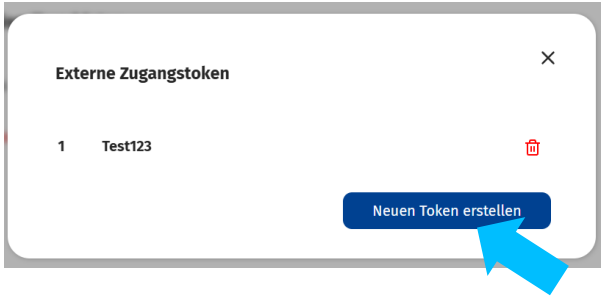
- Die übrigen Modelle von Azure können maximal für C2-Zwecke verwendet werden
- Die GWDG-Modelle können maximal für C3-Zwecke verwendet werden

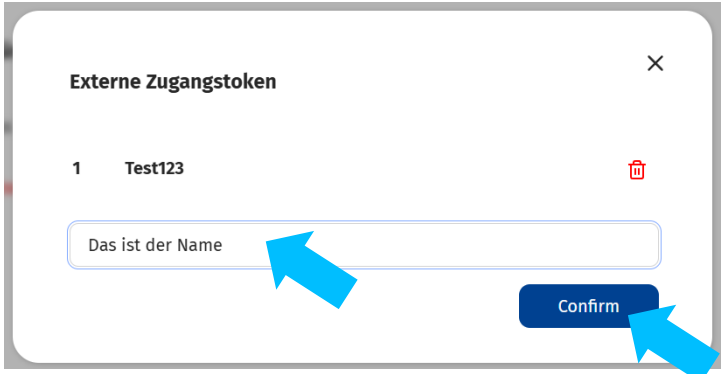
Schritt 1: Auf das eigene Profil gehen und dann auf „Access Tokens“

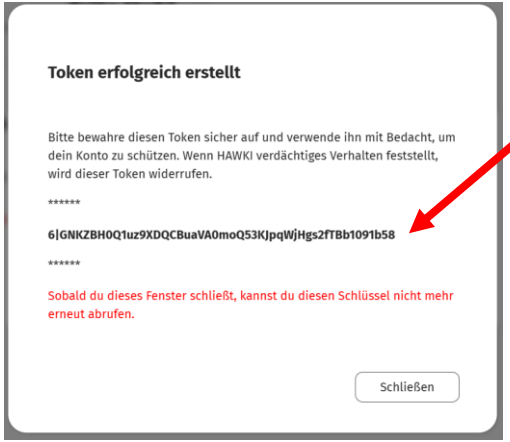




3. Neuen Token erstellen, 4. Namen eingeben 5. Confirm







- Jetzt sich den Token irgendwo speichern
- Er kann nicht wieder gezeigt werden
- Aber man kann jederzeit einen neuen Token erzeugen

Eine Datei namens body-rai.json abspeichern

```
{
  "payload": {
    "model": "hawki-25-gpt-6-luna",
    "messages": [
      {
        "role": "system",
        "content": {
          "text": "You are a helpful assistant."
        }
      },
      {
        "role": "user",
        "content": {
          "text": "What is Manila?"
        }
      }
    ]
  }
}
```

Das ist der Namen eines RAI-Modells

Das ist der Text der Anfrage an den Chatbot

Welche RAI Modelle gibt es?

- C1-hawki-claude-opus-5-5
- C1-hawki-claude-sonnet-5-5
- C1-hawki-claude-haiku-5-5

Modelle ausschließlich für C1

- hawki-25-gpt-6-astra
- hawki-25-gpt-6-luna
- hawki-25-gpt-6.1-sol

Modelle für C1 und C2

- gemma-4-31b-it
- qwen3-omni-30b-a3b-instruct
- qwen3.6-35b-a3b
- qwen3.8-27b
- glm-5.3-flash
- deepseek-v4-flash-0731

Modelle für C1, C2 und C3

Start des API Calls (durch curl)

In der (Windows) Eingabeaufforderung eingeben:

- curl -s -k -X POST "https://rai.uni-stuttgart.de/api/ai-req" -H "Authorization: Bearer DASISTDASTOKEN" -H "Content-Type: application/json" -T body-rai.json

- DASISTDASTOKEN muss natürlich durch das eigene Token ersetzt werden
- Der Aufruf muss in dem Verzeichnis stattfinden, in dem sich die Datei body-rai.json befindet

Universität Stuttgart
Technische Informations- und Kommunikationsdienste

Vielen Dank!



Dr. Fritz Hohl

fritz.hohl@tik.uni-stuttgart.de
Telefon +49 (0) 1726346136
tik.uni-stuttgart.de

Technische Informations- und Kommunikationsdienste
Allmandring 30, 70569 Stuttgart

Für Profis: Welche Modelle kann RAI?

- curl -s -H "Authorization: Bearer DASISTDASTOKEN " https://rai.uni-stuttgart.de/api/hawki/v1/ai-models