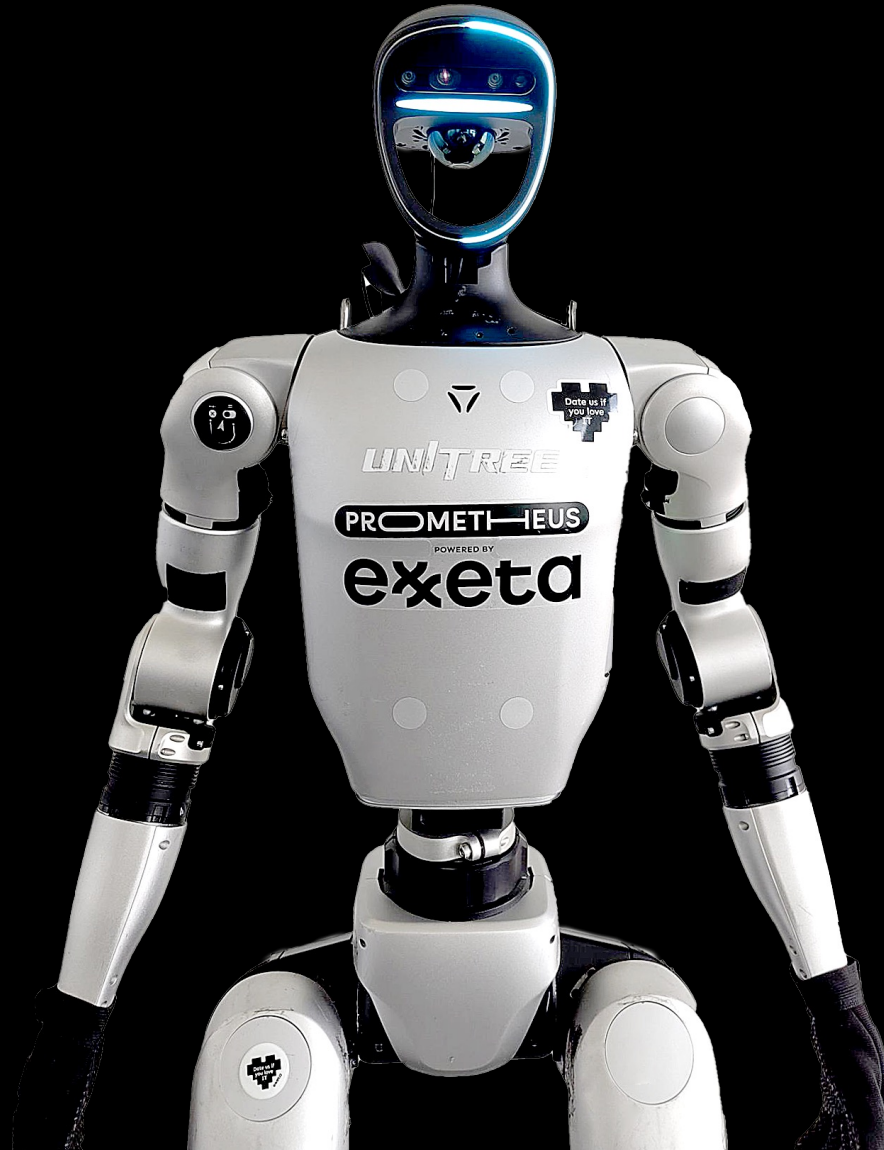


Conversational AI X Robotics

Julian Kramer

exeta

Ausgangssituation A



Wir nutzen den humanoiden Roboter Unitree G1, um Gespräche mit Menschen zu ermöglichen. Die Spracherkennung und -ausgabe erfolgen über die OpenAI Realtime API. Der Roboter verfügt über ein eingebautes Mikrofon und einen Lautsprecher für die Konversation.

Um dem Roboter seinen "Charakter" und sein aktuelles Wissen zu verleihen, ist der Kontext essentiell. Aktuell wird der Gesprächskontext manuell auf dem Roboter in einer Konfigurationsdatei angepasst. Dafür ist ein manueller SSH-Zugang zum Roboter erforderlich. Kontexte sind damit weder versioniert noch für normale Anwender einfach zugänglich. Wir suchen eine Lösung, die für nicht-technische Anwender leicht zu bedienen ist. So soll es einfach möglich sein, dem Roboter für ein spezielles Event den Gesprächskontext anzupassen und so das Konversationsverhalten zu steuern.

Ausschreibung A: Interkative Kontextanpassung

Ziel der Ausschreibung

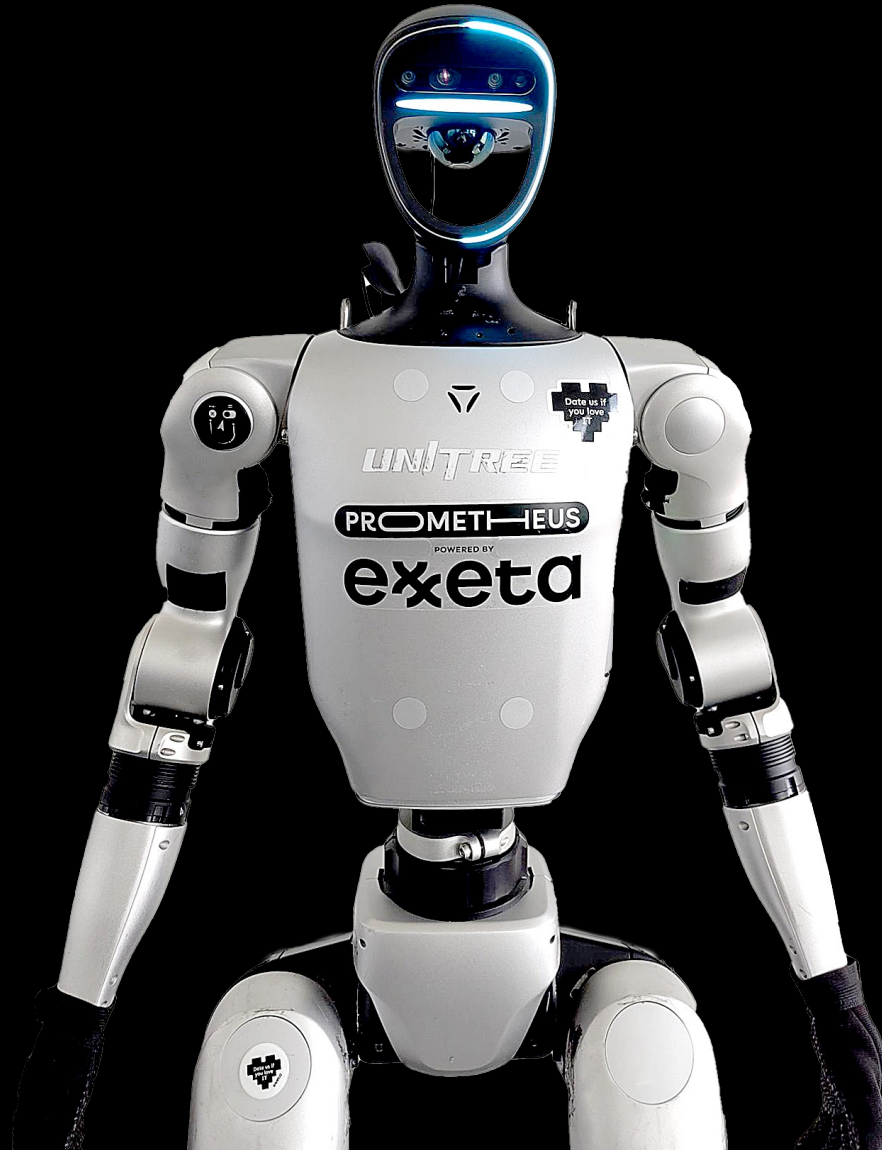
Entwicklung einer Weboberfläche, die es erlaubt (Priorität absteigend):

- Kontexte bequem im Browser über die Roboter-IP anzupassen
- Eine Historie von Kontexten einzusehen und alte Versionen wiederherzustellen
- Dokumente (z. B. PDF-Dateien) hochzuladen, damit der Roboter Wissen aus diesen Dokumenten in Gesprächen nutzen kann
- Physische Aktionen anhand von Gesprächsmustern zu konfigurieren. Bspw. Der Roboter soll winken, wenn er sich verabschiedet. (Function Calling Realtime API)

Randbedingungen

- Hardware: Unitree G1-EDU1 mit Jetson Orin NX (alternativ Raspberry Pi)
- Betriebssystem: Ubuntu 20.04
- Kommunikation: Lokaler Netzwerkzugriff über IP-Adresse
- Applikation muss mit den zur Verfügung stehenden Ressourcen auskommen
- Lizenzen: Es sollen nur kommerzielle nutzbare Open-Source-Lizenzen (z. B. MIT) verwendet werden
- Die technischen Möglichkeiten der Realtime API müssen respektiert werden
- Ergebnis: nutzerfreundlicher Prototyp mit Demo-Funktionalität

Ausgangssituation B



Der humanoide Roboter Unitree G1 wird bereits heute für Gesprächsfunktionalitäten genutzt. Die Audioverarbeitung erfolgt über Mikrofon und Lautsprecher, gesteuert durch die OpenAI Realtime API.

Neben der Sprachschnittstelle verfügt der Roboter über zwei Kameras:

- Weitwinkelkamera für die Umgebung
- Intel RealSense Tiefenkamera für Tiefeninformationen

Aktuell passt der Roboter sein Gesprächsverhalten unabhängig von der Umgebung an. Es wäre jedoch wünschenswert, wenn er auf Basis seiner Wahrnehmung erkennt, in welcher Art von Umfeld er sich befindet.

Ausschreibung B: Interaktive Umfelderkennung

Ziel der Ausschreibung

Entwicklung einer Lösung, die anhand der Kameradaten eine klassifizierende Umfeld-Erkennung ermöglicht. Der Roboter soll erkennen, ob er sich z. B. befindet in:

- einer Küche
- einem Büro
- einem Privathaushalt
- einer Bar

Je nach Umfeld soll der Roboter sein Gesprächsverhalten anpassen (informeller, förmlicher, erklärender etc.).

Randbedingungen

- Hardware: Unitree G1 mit Jetson Orin NX (alternativ Raspberry Pi)
- Betriebssystem: Ubuntu 20.04
- Applikation muss mit den zur Verfügung stehenden Ressourcen auskommen
- Lizenzen: Es sollen nur kommerzielle nutzbare Open-Source-Lizenzen (z. B. MIT) verwendet werden
- Ergebnis: Prototyp, der eine Klassifizierung vornimmt und das Gesprächsverhalten darauf anpasst

Let's innovate together

Julian Kramer

Robotics Lead

Mobile: +49 174 99 51 241

Julian.Kramer@exxeta.com

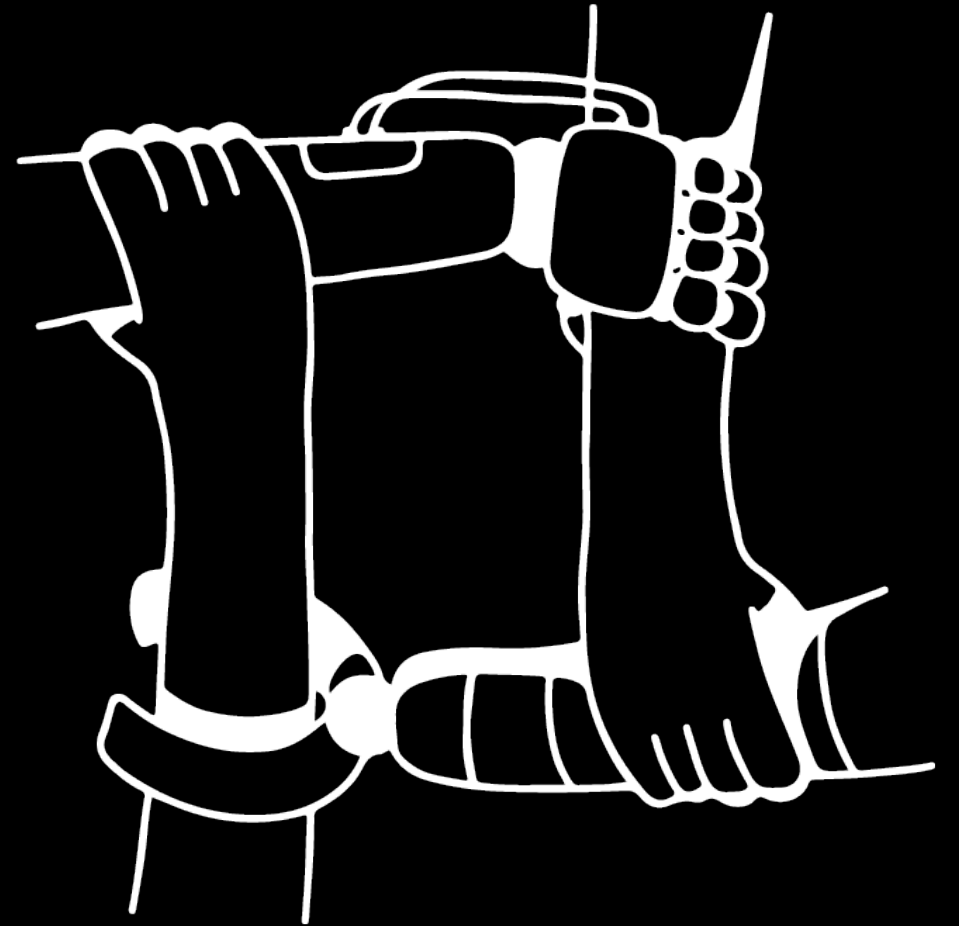
Markus Fischer

Cloud & Robotics Fellow

Mobile: +49 162 29 59 538

Markus.Fischer@exxeta.com

exeta



exeta