Gen VOICES

Generative Verbal Ontology for Integrative Conversational Emergency Scenarios

Programm / Ausschreibung	FORPA, Dissertaionen 2024, Industrienahe Dissertationen 2025	**Status**	laufend
Projektstart	01.10.2025	**Projektende**	30.09.2028
Zeitraum	2025 - 2028	**Projektlaufzeit**	36 Monate
Keywords	AI, Generative Pre-trained Transformers (GPT), Large Language Models (LLMs), Emergency Services		

Projektbeschreibung

Generative Pre-trained Transformers (GPT) ist eine KI-Architektur, die viele Berufe revolutioniert, indem sie Assistenten oder Agenten bereitstellt, die audiovisuelle und textuelle Informationen in mehr als 50 Sprachen verstehen und generieren können. Computerprogrammierer verwenden GPT-Modelle zur Codegenerierung; Pädagogen, Trainer und Studierende nutzen sie für personalisiertes Lernen und Forschung; Fachkräfte im Gesundheitswesen verwenden sie zur Erstellung von Berichten oder grundlegenden Diagnosen und Behandlungen; audiovisuelle Produzenten, Designer, Schriftsteller und Künstler setzen GenAI für Geschichten, Zeichnungen, Videos oder Musik ein. Darüber hinaus leistet die Gaming-Industrie bedeutende Beiträge zur Verbreitung der GPT-Technologie durch Simulationen von realen Szenarien. Heute wird GPT-KI in hochimmersiven virtuellen Umgebungen mit 3D-Nicht-Spieler-Charakteren (NPCs) oder Avataren verwendet, die menschliches Aussehen und Verhalten nachahmen. Diese Avatare können mit menschlichen Spielern und anderen GPT-gestützten NPCs kommunizieren.
Notfalldienste wie medizinische Erstversorgung (MFR), Polizeieinsätze oder die Abwehr von chemischen, biologischen, radiologischen, nuklearen und explosiven Bedrohungen (CBRNE) können von GPT-KI in ihren Ausbildungsprogrammen und operativen Aufgaben profitieren. Computersimulationen bieten eine kostengünstige und flexible Alternative zur Ausbildung von mehr Einsatzkräften im Vergleich zu herkömmlichen Methoden. Nutzerstudien zu Interaktionen mit GPT-gestützten NPCs im kritischen Bereich des Bevölkerungsschutzes liefern vielversprechende Alternativen. Es sind jedoch weitere Studien erforderlich, um technische Herausforderungen zu überwinden und höhere Technologiereifegrade für die Nutzung zu erreichen.
Es gibt vier zentrale Herausforderungen in den aktuellen Interaktionen mit GPT-Modellen: 1. Proprietäre GPT-Modelle sind allgemein, umfangreich und erfordern erhebliche Rechenressourcen. Daher finden Gespräche meist über externe Drittanbieter-Server statt, was zu verzögerten Reaktionen führt und Datenschutz- sowie Sicherheitsbedenken hervorruft. 2. Modelle sollten autonom agieren, z. B. audiovisuelle Inhalte erstellen und veröffentlichen oder Geräte steuern. 3. Die Rolle der GPT-Agenten wird durch Kontexte im Prompt definiert; es gibt keinen Echtzeit-Audiovisuellen-Kontext, der die Antworten beeinflusst. 4. Menschliche Gesprächsdynamiken wie Unterbrechungen und nonverbale Hinweise für Sprecherwechsel sind in Sprachinteraktionen mit GPT-Agenten, d. h. Generativen Sprachagenten (GVAs), weitgehend nicht vorhanden.
Dieser Vorschlag zielt darauf ab, diese aktuellen Herausforderungen zu überwinden, indem er sich auf die Bedürfnisse der

Notfalldienste konzentriert. Das erwartete Ergebnis ist eine Reihe von angepassten GPT-Modellen, die: 1. auf lokaler, schneller Computerinfrastruktur mit spezialisierter Sicherheit und Datenschutz laufen, 2. Geräte und Werkzeuge in die Sprachinteraktionsprozesskette integrieren, um die Autonomie zu verbessern, 3. audiovisuelle Eingaben für eine intelligentere Kontextwahrnehmung integrieren und 4. Sprecherwechsel für intuitivere und effizientere Konversationen erleichtern

Abstract

Generative Pre-trained Transformers (GPT) is an AI architecture revolutionizing many professions with assistants or agents that can understand and generate audiovisual and textual information in more than 50 languages. Computer programmers use GPT models to generate code; educators, trainers, and students use it for personalized learning and research; healthcare professionals use it to create reports or basic diagnoses and treatments; audiovisual producers, designers, writers, and artists use GenAI for stories, drawings, videos, or music. Moreover, the gaming industry is making important contributions to the adoption of GPT technology with simulations of real-world scenarios. Today, GPT AI is used in highly immersive virtual environments with 3D non-player characters (NPCs) or avatars that mimic human appearance and behavior. These avatars can communicate back and forth with human players and with other GPT-powered NPCs.
Emergency services like medical first response (MFR), police interventions or Chemical, Biological, Radiological, Nuclear and Explosives (CBRNE) defense can benefit from GPT AI in their training programs and operational tasks. Computer simulations are a cost-effective, flexible alternative to train more emergency personnel compared to traditional methods. User studies of interactions with GPT-powered NPCs in today’s critical domain of civil protection are emerging with promising alternatives. However, further studies to overcome technical challenges and achieve higher technological readiness levels for exploitation are needed.
There are four key challenges in current interactions with GPT models. 1. proprietary GPT models are general, extensive, and require significant computing resources. Hence, conversations with them usually take place via third-party external servers, contributing to delayed responses and raising privacy and security concerns. 2. models should act autonomously, e.g., produce and publish audiovisual content or control devices. 3. the role of the GPT agents is detailed with contexts in the prompt, there is no real-time audiovisual context informing the responses. 4. human conversational dynamics, such as interruptions and non-verbal cues that signal a speaker change, are largely absent in voice interactions with GPT agents, i.e., Generative Voice Agents (GVAs).
This proposal aims to overcome these current challenges by focusing on the needs of emergency services. The expected outcome is a set of customized GPT models that: 1. run on local fast computer infrastructure with specialized security and privacy, 2. integrate devices and tools in the voice interaction to improve autonomy of the GVA, 3. integrate audiovisual inputs for smarter context awareness and 4. facilitate speaker change for more intuitive and efficient conversational interactions.

Projektpartner

- AIT Austrian Institute of Technology GmbH