First Discussion

1. Collected diverse Personas Datasets from Kaggle (Persona chat(ConvAI2)-facebook AI)

Fine tuned with (Unsloth-LoRA Low-Rank Adaptation) phi3→Llama 3

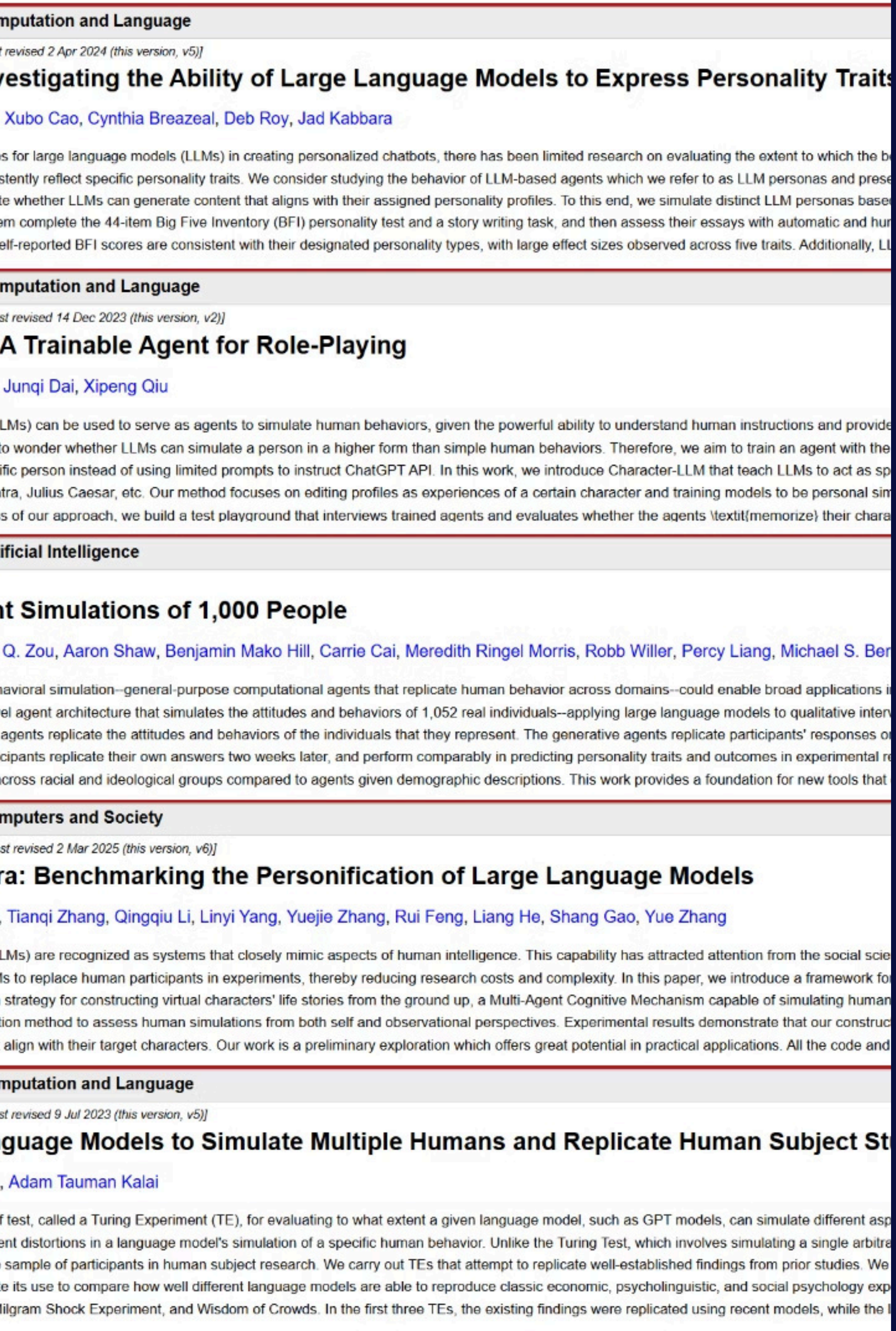
Challenges: Time-consuming, Google Colab disconnections

integrated with Fast API backend Publish API with ngrok

Connected API with React Native Mobile App via Expo GO

conclusion: collect real data from WA,FB,twitter,telegram or create fake one form GPT4

by Anas Saleh



mputation and Language

t revised 2 Apr 2024 (this version, v5)]

vestigating the Ability of Large Language Models to Express Personality Traits

Xubo Cao, Cynthia Breazeal, Deb Roy, Jad Kabbara

s for large language models (LLMs) in creating personalized chatbots, there has been limited research on evaluating the extent to which the b
stently reflect specific personality traits. We consider studying the behavior of LLM-based agents which we refer to as LLM personas and prese
te whether LLMs can generate content that aligns with their assigned personality profiles. To this end, we simulate distinct LLM personas base
em complete the 44-item Big Five Inventory (BFI) personality test and a story writing task, and then assess their essays with automatic and hu
elf-reported BFI scores are consistent with their designated personality types, with large effect sizes observed across five traits. Additionally, LL

mputation and Language

st revised 14 Dec 2023 (this version, v2)]

A Trainable Agent for Role-Playing

Junqi Dai, Xipeng Qiu

LMs) can be used to serve as agents to simulate human behaviors, given the powerful ability to understand human instructions and provide
to wonder whether LLMs can simulate a person in a higher form than simple human behaviors. Therefore, we aim to train an agent with the
ific person instead of using limited prompts to instruct ChatGPT API. In this work, we introduce Character-LLM that teach LLMs to act as sp
tra, Julius Caesar, etc. Our method focuses on editing profiles as experiences of a certain character and training models to be personal sim
s of our approach, we build a test playground that interviews trained agents and evaluates whether the agents \textit{memorize} their chara

ificial Intelligence

t Simulations of 1,000 People

Q. Zou, Aaron Shaw, Benjamin Mako Hill, Carrie Cai, Meredith Ringel Morris, Robb Willer, Percy Liang, Michael S. Ber

avioral simulation--general-purpose computational agents that replicate human behavior across domains--could enable broad applications i
el agent architecture that simulates the attitudes and behaviors of 1,052 real individuals--applying large language models to qualitative inter
agents replicate the attitudes and behaviors of the individuals that they represent. The generative agents replicate participants' responses o
cipants replicate their own answers two weeks later, and perform comparably in predicting personality traits and outcomes in experimental r
cross racial and ideological groups compared to agents given demographic descriptions. This work provides a foundation for new tools that

mputers and Society

st revised 2 Mar 2025 (this version, v6)]

ra: Benchmarking the Personification of Large Language Models

, Tianqi Zhang, Qingqiu Li, Linyi Yang, Yuejie Zhang, Rui Feng, Liang He, Shang Gao, Yue Zhang

LMs) are recognized as systems that closely mimic aspects of human intelligence. This capability has attracted attention from the social scie
ls to replace human participants in experiments, thereby reducing research costs and complexity. In this paper, we introduce a framework fo
strategy for constructing virtual characters' life stories from the ground up, a Multi-Agent Cognitive Mechanism capable of simulating human
tion method to assess human simulations from both self and observational perspectives. Experimental results demonstrate that our construc
align with their target characters. Our work is a preliminary exploration which offers great potential in practical applications. All the code and

mputation and Language

st revised 9 Jul 2023 (this version, v5)]

guage Models to Simulate Multiple Humans and Replicate Human Subject St

, Adam Tauman Kalai

f test, called a Turing Experiment (TE), for evaluating to what extent a given language model, such as GPT models, can simulate different asp
ent distortions in a language model's simulation of a specific human behavior. Unlike the Turing Test, which involves simulating a single arbitra
sample of participants in human subject research. We carry out TEs that attempt to replicate well-established findings from prior studies. We
e its use to compare how well different language models are able to reproduce classic economic, psycholinguistic, and social psychology exp
ilgram Shock Experiment, and Wisdom of Crowds. In the first three TEs, the existing findings were replicated using recent models, while the l

- Huge LLMs
- Zero Shot
- Few Shot
- AI Agents
- Prompt Engineering
- Retrieval-Augmented Generation - RAG
- Memory Mechanisms

Second Discussion

Alternative Data Collection Methods

Used voice recordings from a team member

Transcribed speech using Whisper Large V3 Turbo (supports Arabic dialects)

Avoided fine-tuning due to computational cost

Used AI inference to process extracted text

Used Groq AI inference with Llama-3.3-70B Versatile

Voice Cloning for Personalized Speech

Converted text responses into speech using XTTS model

High-quality voice cloning, supports Arabic (closer to Gulf Arabic)

Creating a Realistic Deepfake

Used SadTalker for deepfake video generation

Improved output quality with Wave2Lip

Challenges in Data Collection:

Hard to Collect Real data from specific Person from WA or Social media required sign in

conclusion: The professor said she doesn't see enough effort in how we're mimicking the personality to make it respond like us. She wants us to focus more on the simulation aspect and see if there's any algorithm applied or related research papers. We told her we can't use WhatsApp data, and although Telegram is an option, we don't use it much, so there's little data available. Social media posts and tweets also aren't helpful since people rarely share personal content there.

Third Discussion in Ramadan

Hard To Fine Tune With Our Resources

🔹 Key Points about ConvAI2 Dataset:

- **ConvAI2** is a benchmark dataset designed to improve **persona-based dialogue generation**.
- Each speaker is assigned a **persona** (e.g., "I have four sisters") to encourage **more personalized responses**.
- It includes around **131,000 utterances** and **fewer than 5,000 unique personas**, making it relatively **small compared to open-domain datasets** like Reddit-Pushshift (1.2B+ utterances).
- Due to its size, ConvAI2 highlights the **data scarcity problem** in persona-based dialogue systems.
- The proposed model in the paper was evaluated on ConvAI2 and achieved:
  - **Competitive or better performance** than large models like GPT-2.
  - **30% improvement in perplexity**.
  - **Strong results using only 20–30% of the training data**, showing data efficiency.

after reading papers :

1- First Paper "Personalized Dialogue Generation with Persona-Adaptive Attention" PAA (2022)

Summary

This paper enhances persona-based dialogue systems, which aim to generate replies that are consistent with a speaker's personality and conversation history. Unlike basic chat models, these systems must carefully balance the speaker's persona with the conversation context.

The Authors Introduce

⚖️ Persona-Adaptive Attention (PAA): A new attention mechanism that dynamically adjusts the importance of persona vs. dialogue context.

🎭 Dynamic Masking: This technique filters out unnecessary info and acts as a regularizer, helping the model avoid overfitting.

Results

Outperforms strong baseline models in both human and automatic evaluations.

Works well even with low-resource data (achieving near full-data performance with just 20–30% of training data).

Multiple model variants show that their adaptive weighting and masking strategies are essential.

Key Insight: Their method offers a smart and flexible way to generate personalized, consistent responses — even with limited training data.

- 🔹 Dataset Introduced: ConvAI - MINST - Reddit

Second Paper on Personalizing Dialogue Agents

Paper Overview

"Personalizing Dialogue Agents: I have a dog, do you have pets too?" (2018)

Main Contributions

📋 Persona Conditioning: Dialogue agents are trained to include their own profile information in responses (e.g., "I like hiking").

👥 Interactive Learning: The agents also learn to infer the persona of the person they're talking to, based on the dialogue.

🧠 Engagement Strategy: Since the other person's persona is unknown at the start, the model proactively asks personal questions to elicit information — mimicking real social behavior.

Results

This profile-based approach led to better next-utterance predictions, making conversations feel more human-like and connected.

The conversations became more personal and memorable, a big step up from generic chit-chat bots.

Dataset Introduced: PersonaChat which is part of ConvAI

A curated dataset where each speaker is given a profile with 5 sentences (e.g., "I love painting", "My favorite food is pizza").

Structure: Contains over 160,000 utterances and is used to train models to generate responses grounded in their persona.

Goal: Improve personalization and make chatbot dialogues more natural and engaging.

Access: Released publicly by Facebook AI Research, often cited in follow-up personalization studies.

Third Paper on Characteristic AI Agents

Paper Overview

"Characteristic AI Agents via Large Language Models" (2024)

Character100 Benchmark

100 most-visited people on Wikipedia

Evaluation Metrics

Measuring character trait embodiment

Comprehensive Experiments

Testing LLM performance in character simulation

3-Third Paper "Characteristic AI Agents via Large Language Models" (2024)

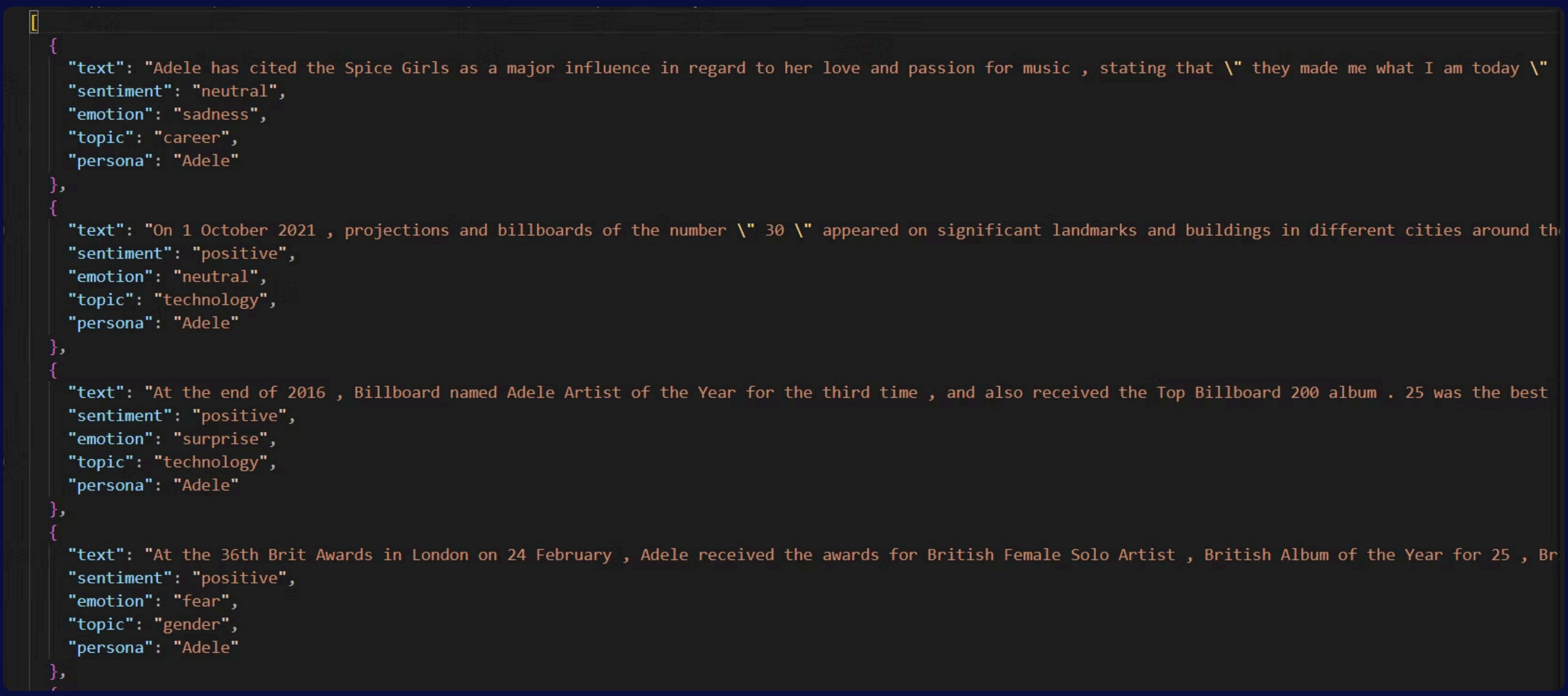
🔍 Summary:

The research paper proposes a framework for creating AI agents with distinct and consistent personality traits. This framework is designed to be applied *individually* to each character.

This paper explores how Large Language Models (LLMs) can be used to simulate real-life characters in chatbot systems — not just generic roles, but detailed personality-driven agents. Although commercial tools exist for role-based chatbots, academic research has lagged behind.

The research does not focus on "training" the large language model from scratch or even extensively fine-tuning it. Instead, it relies on pre-trained large language models (pre-trained LLMs) and develops methods to guide and provide them with the necessary information to embody the desired personality. The specific methods mentioned or inferred from the nature of the research include:

- **Detailed Character Profiles:** They create a precise and structured description of the character. This description includes multiple aspects such as:
  - Background
  - Personality Traits (e.g., using the OCEAN model or similar)
  - Goals and Motivations
  - Relationships
  - Knowledge and Expertise
  - Memory: An important component for maintaining consistency. The research might include a memory module to store and retrieve information about past interactions or significant events in the character's "life."



```
[
  {
    "text": "Adele has cited the Spice Girls as a major influence in regard to her love and passion for music , stating that \" they made me what I am today \"
    "sentiment": "neutral",
    "emotion": "sadness",
    "topic": "career",
    "persona": "Adele"
  },
  {
    "text": "On 1 October 2021 , projections and billboards of the number \" 30 \" appeared on significant landmarks and buildings in different cities around th
    "sentiment": "positive",
    "emotion": "neutral",
    "topic": "technology",
    "persona": "Adele"
  },
  {
    "text": "At the end of 2016 , Billboard named Adele Artist of the Year for the third time , and also received the Top Billboard 200 album . 25 was the best
    "sentiment": "positive",
    "emotion": "surprise",
    "topic": "technology",
    "persona": "Adele"
  },
  {
    "text": "At the 36th Brit Awards in London on 24 February , Adele received the awards for British Female Solo Artist , British Album of the Year for 25 , Br
    "sentiment": "positive",
    "emotion": "fear",
    "topic": "gender",
    "persona": "Adele"
  },
```

- **Profile-Guided Prompt Engineering:** This detailed profile is used to formulate prompts that are given to the large language model. These prompts guide the model to generate responses (speech, potential actions) that are consistent with all aspects of the character defined in the profile.
- **Agent Architecture:** The research might propose a specific architecture for the agent that includes different components (such as a memory unit, a planning/decision-making unit, in addition to the base language model) working together to produce consistent and distinct character behavior.

🔹 Dataset Introduced: Character100

Description: A benchmark dataset consisting of 100 of the most-visited people on Wikipedia, used to train and evaluate LLMs in simulating real-life personalities.

Purpose: Enables role-playing and character emulation by LLMs.

Contents: Includes persona profiles and contextual data about each character.

Access: Publicly available at the project's GitHub: Character100 Benchmark

LoRA vs Fine Tuning - PAA
ConvAI2 vs Personachat vs Character 100
BlenderBot vs Llama

searching about models :

BlenderBot 3 Model Overview

Internet Access

Can search the web for up-to-date information.

Long-term Memory

Remembers facts about users to personalize conversations.

Learning from Users

Improves over time by learning from real interactions.

Open Source

Meta released the code and model to support research.

More Natural Dialogue

Trained to generate human-like responses.

1- 🤖 BlenderBot 3 – Overview BlenderBot 3 is a chatbot developed by Parl.ai, part of Meta's research. Meta AI (formerly Facebook) and released in August 2022. It's part of Meta's research into open-domain conversational AI.

🔍 Key Features: Internet Access: Can search the web for up-to-date information.

Long-term Memory: Remembers facts about users to personalize conversations.

Learning from Users: Improves over time by learning from real interactions.

Open Source: Meta released the code and model to support research.

More Natural Dialogue: Trained to generate human-like responses.

Llama with RAG and Vector Store Implementation

1

Vector Store Creation

Creating a Vector Store from the data we've prepared using FAISS + LangChain.

RAG Pipeline

Using the RAG Pipeline with LLaMA + Retriever + Memory.

Analysis Models

"sentiment_model = "cardiffnlp/twitter-roberta-base-sentiment (Sentiment) تحليل مشاعر

"style_model = "j-hartmann/emotion-english-distilroberta-base (Emotion) تحليل أسلوب

"topic_model = "facebook/bart-large-mnli (Topic) تصنيف موضوع

Memory Options

ConversationBufferMemory لو عايز تحافظ على الـ context

ConversationSummaryMemory لو عندك سياق طويل

persona لكل memory-specific store أو تبني

5

example:

{ "content": "Sure, I can help you with that.",

 "persona": "Alex",

 "sentiment": "positive",

 "emotion": "joy",

 "topic": "career" }

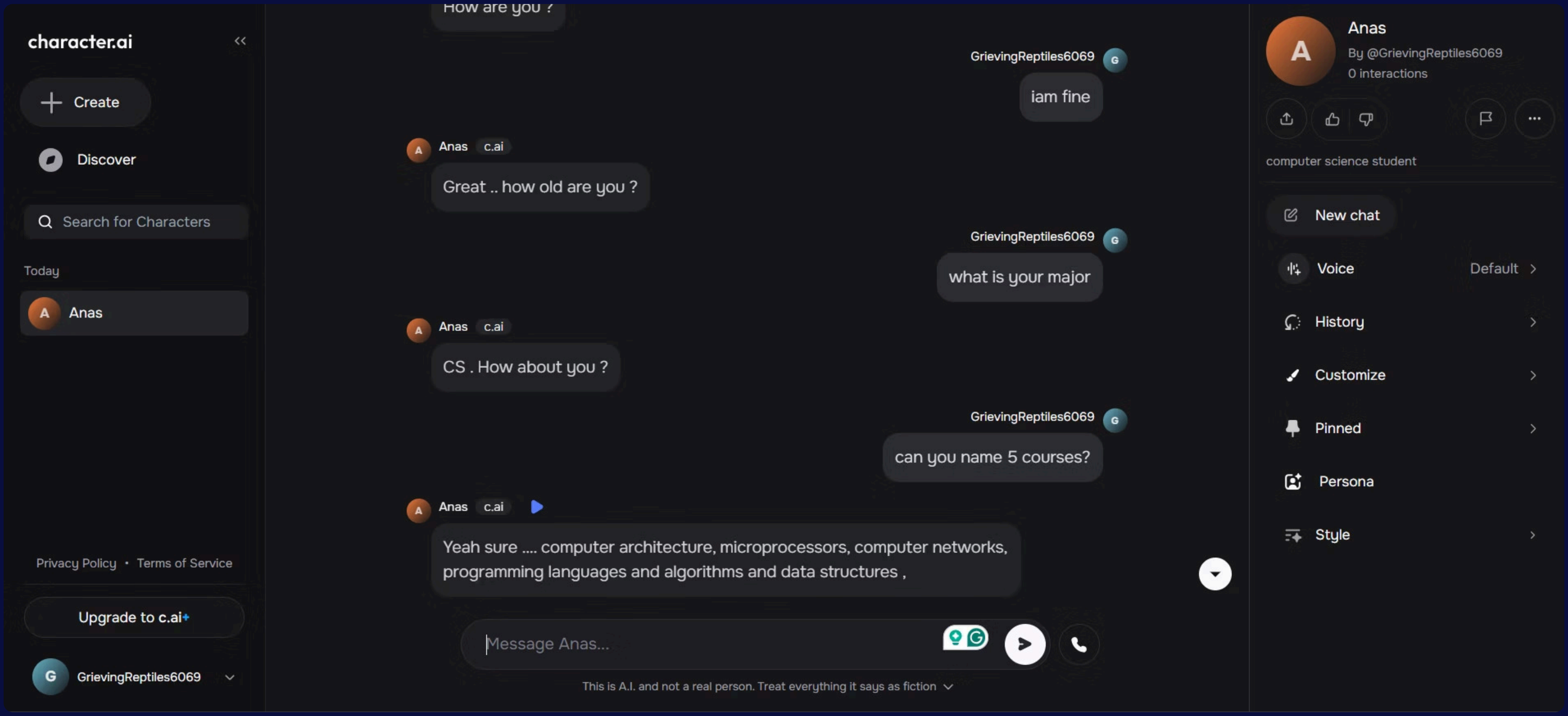
2- unsloth/Meta-Llama-3.1-8B-bnb-4bit + RAG + Vector Store

Creating a Vector Store from the data we've prepared using FAISS + LangChain. Using the RAG Pipeline with LLaMA + Retriever + Memory.

Future Steps for Model Enhancement

- **Expand training data:** Use datasets like PersonaChat and ConvAI for diversity. and merging.
- **Enhance memory and retrieval:** Improve context-awareness with better RAG components.
- **Platform integration:** Develop frontend/backend for web and mobile using Streamlit.
- **User feedback:** Implement questionnaires to refine dialogue personalization.
- **Experimentation:** Attempt LoRA fine-tuning on BlenderBot 3 using Hugging Face.

Character.ai



How are you ?

GrievingReptiles6069: iam fine

Anas c.ai: Great .. how old are you ?

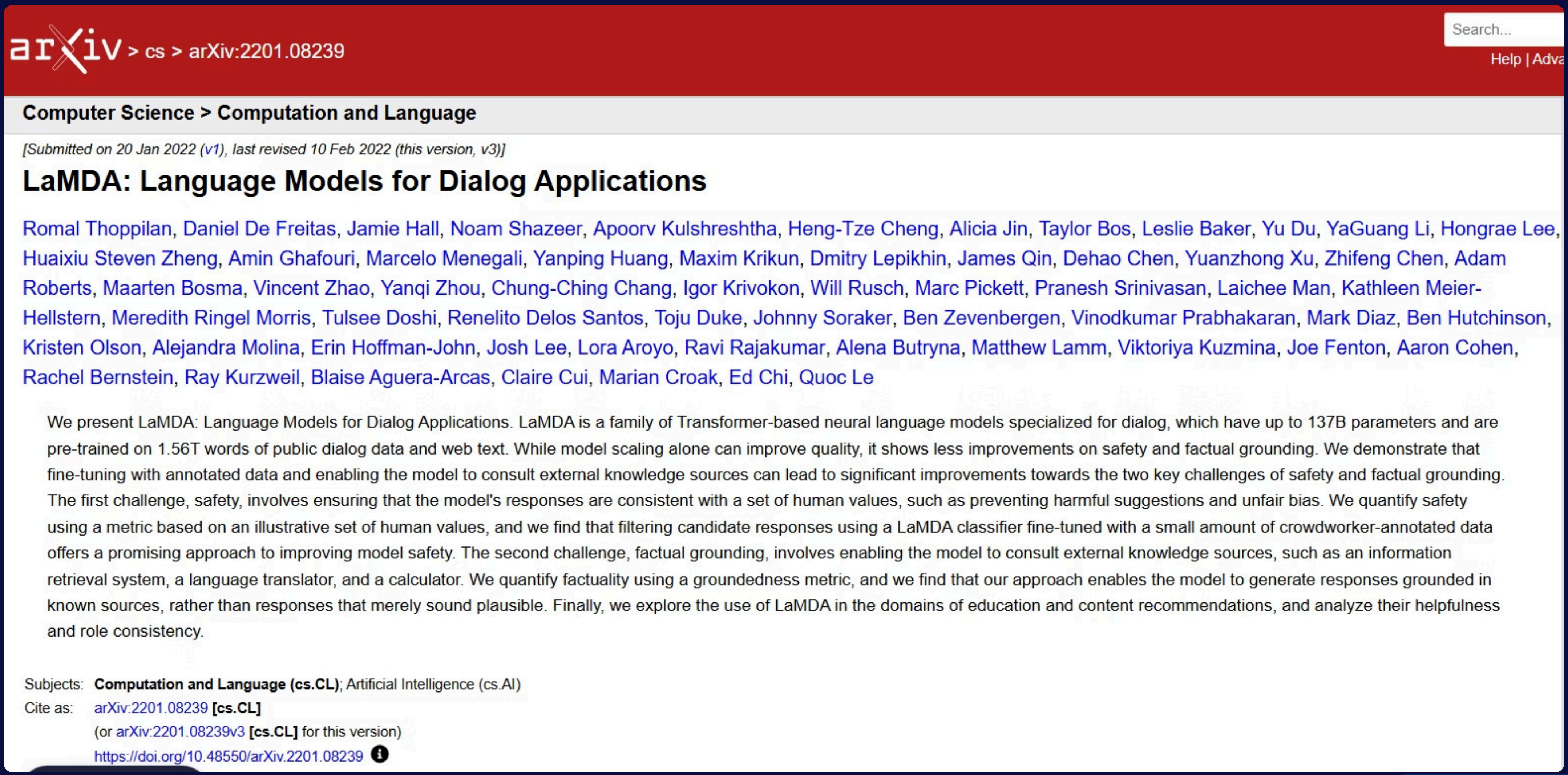
GrievingReptiles6069: what is your major

Anas c.ai: CS . How about you ?

GrievingReptiles6069: can you name 5 courses?

Anas c.ai: Yeah sure .... computer architecture, microprocessors, computer networks, programming languages and algorithms and data structures ,

This is A.I. and not a real person. Treat everything it says as fiction



Computer Science > Computation and Language

[Submitted on 20 Jan 2022 (v1), last revised 10 Feb 2022 (this version, v3)]

LaMDA: Language Models for Dialog Applications

Romal Thoppilan, Daniel De Freitas, Jamie Hall, Noam Shazeer, Apoorv Kulshreshtha, Heng-Tze Cheng, Alicia Jin, Taylor Bos, Leslie Baker, Yu Du, YaGuang Li, Hongrae Lee, Huaixiu Steven Zheng, Amin Ghafouri, Marcelo Menegali, Yanping Huang, Maxim Krikun, Dmitry Lepikhin, James Qin, Dehao Chen, Yuanzhong Xu, Zhifeng Chen, Adam Roberts, Maarten Bosma, Vincent Zhao, Yanqi Zhou, Chung-Ching Chang, Igor Krivokon, Will Rusch, Marc Pickett, Pranesh Srinivasan, Laichee Man, Kathleen Meier-Hellstern, Meredith Ringel Morris, Tulsee Doshi, Renelito Delos Santos, Toju Duke, Johnny Soraker, Ben Zevenbergen, Vinodkumar Prabhakaran, Mark Diaz, Ben Hutchinson, Kristen Olson, Alejandra Molina, Erin Hoffman-John, Josh Lee, Lora Aroyo, Ravi Rajakumar, Alena Butryna, Matthew Lamm, Viktoriya Kuzmina, Joe Fenton, Aaron Cohen, Rachel Bernstein, Ray Kurzweil, Blaise Aguera-Arcas, Claire Cui, Marian Croak, Ed Chi, Quoc Le

We present LaMDA: Language Models for Dialog Applications. LaMDA is a family of Transformer-based neural language models specialized for dialog, which have up to 137B parameters and are pre-trained on 1.56T words of public dialog data and web text. While model scaling alone can improve quality, it shows less improvements on safety and factual grounding. We demonstrate that fine-tuning with annotated data and enabling the model to consult external knowledge sources can lead to significant improvements towards the two key challenges of safety and factual grounding. The first challenge, safety, involves ensuring that the model's responses are consistent with a set of human values, such as preventing harmful suggestions and unfair bias. We quantify safety using a metric based on an illustrative set of human values, and we find that filtering candidate responses using a LaMDA classifier fine-tuned with a small amount of crowdworker-annotated data offers a promising approach to improving model safety. The second challenge, factual grounding, involves enabling the model to consult external knowledge sources, such as an information retrieval system, a language translator, and a calculator. We quantify factuality using a groundedness metric, and we find that our approach enables the model to generate responses grounded in known sources, rather than responses that merely sound plausible. Finally, we explore the use of LaMDA in the domains of education and content recommendations, and analyze their helpfulness and role consistency.

Subjects: **Computation and Language (cs.CL)**; Artificial Intelligence (cs.AI)

Cite as: arXiv:2201.08239 **[cs.CL]**
(or arXiv:2201.08239v3 **[cs.CL]** for this version)
https://doi.org/10.48550/arXiv.2201.08239

Prompt Engineering + Inference of models + Questionnaire + Memory & RAG

Recent Project Milestones

- **Voice Transcription:** Users can now send voice recordings with automatic text transcription. Whisper, CoquiTTS, gTTS
- **VoiceMood Avatars:** Animated avatars appear during responses; lip syncing is actively being developed.
- **Enhanced Memory:** Chat history improves context, enabling more personalized model responses.
- **Telegram Data:** Data integration has begun, with full implementation scheduled for Eid.
- **Character Memory:** Each virtual character now features its own distinct memory system.

Google Docs

WhatsApp Video 2025-05-09 at 00.56.48_0fdc889b.mp4

Google Docs

WhatsApp Video 2025-05-09 at 00.57.06_5700cabc.mp4