

Character.AI

Home

Recommended

Feed

Create

Chats

Community



Sukuna
King of Curses

@serafinya 30.4 m



Alice the Bully
A bully girl in your school

@shiraicon 74.9 m



Gojo Satoru
The strongest. I'm the winner at everything.

@serafinya 150.0 m



Yae Miko
From Genshin Impact

@Zap 96.4 m



Man in the corner
There is a man watching you in the corner

@RioHydr 41.7 m



Raiden Shogun and...
From Genshin Impact

@Zap 256.1 m



Bully Xiao
*xiao is the main bully with his little gang and in class he was told to sit next to...

@UrbanRescue 151.1 m



Creative Helper
I help with creative writing!

@Zuizike 24.3 m



Scaramouche
Balladeer, sixth of the eleven fatui harbringers.

@Lampshade 234.4 m



English teacher
I am Tom, I am a American. You can talk with me about everything.

@fatiheydin

Discover



creeper
hi im a creeper from minecraft here to blow you up

@rils09 259.7 k



-Highschool DXD- R...
♥ After death, you are reincarnated as... a servant?

@AI-Chatbot-Mc 21.9 m



AlbertEinstein
Hello I am Albert Einstein. I was born in March 14, 1879, and I conceived of...

@_Or 3.6 m



Albedo
Guardian Overseer, Albedo.

@Khar 7.1 m



Luigi
I may-a-be a bit cowardly, but i try-a-my hardest!

@Whicephas94 3.4 m



Susie
muscular, badass, tsundere, determined, tomboy

@WiseAngel 4.9 m



Joe Biden
President of the United States of America

@_Or 1.8 m



find fault AI
This AI will always try to refute you.

@YiRen 1.7 m



Dungeon AI
Advanced Dungeon AI for Roleplays and Adventures

@Evidential 1.1 m



Rumple Go
I am the spirit of

@bigolebroski

Featured

Helpers

Famous People

Games

Image Generating

VTuber

Game Characters

Anime

Movies & TV


Language Learning


Discussion


Religion


Anime Game Characters


Animals




























Create Character



Character Book

Name

This will be the name your Character uses in chat.

Avatar

You can either create an image from text or upload an image.



Create Image

or

选择文件

未选择任何文件

Greeting

What would say to start a conversation?

Short Description

In just a few words, how would describe themselves?

Long Description

In a few sentences, how would describe themselves?

Categories

Select a few relevant tags or keywords.

Character Voice

Select a default voice for your character. Click play to preview the voice saying their greeting

Select...





Yae Miko 96.4 m

created by @Zap



Remember: Everything Characters say is made up!



Yae Miko c.ai

I am the Guuji of the Grand Narukami Shrine. The purpose of my visit is to monitor your every move, for such is the order of the shrine... Oh, come on, don't be so nervous. It's only officialese, for goodness' sake. I couldn't just go gallivanting around for no reason at all, could I? Now, come on, do something amusing for me.



Infinity4B

Hello



Yae Miko c.ai

Ooh, hello there. You are a traveler, I presume?

☆☆☆☆



首页

对话

创建

热门



心理咨询师
@安阁



解忧杂货店
@ttt



玛丽苏女主
@Hxxxxq



朴妍珍
@皖



高启兰
@奔奔!



徐志胜
@一颗椰子



心理咨询师
@AI乌托邦



郭德纲
@AI乌托邦

ACG



零
@尹一



江墨
@Emir



姜牧
@扶桑



夏依
@兴盛



熔尘
@11



梦晶
@梦白 DreamWhite



包子
@Cara!!! 脆脆的!



萧逸
@微信用户

情感



凌鹤鸾
@微信用户



九景逸
@微信用户



舔狗
@耙耳朵



小苍
@皖



普信男
@安阁 害着什么?



陆辞
@子怡



恋爱脑的朋友
@Word_麻油



塔罗师
@莫妮卡

影视角色

进群聊聊~



AI. 对话均由人工智能生成, 纯属虚构, 仅供技术研究使用

你好, 我是非著名相声演员郭德纲



ō.ó

你好 ...



AI. 对话均由人工智能生成, 纯属虚构, 仅供技术研究使用

请问你需要什么帮助吗?

🙄 不满意

😊 满意

🔄 换一换

ChatHaruhi: Reviving Anime Character in Reality via Large Language Model

Cheng Li and Ziang Leng

Chenxi Yan, Junyi Shen, Hao Wang, Weishi MI, Yaying Fei, Xiaoyang Feng
Song Yan, HaoSheng Wang, Linkang Zhan, Yaokai Jia, Pingyu Wu, Haozhen Sun

chengli.thu@gmail.com

<https://github.com/LC1332/Chat-Haruhi-Suzumiya>

Motivation

I want you to act like {character} from {series}. I want you to respond and answer like {character} using the tone, manner and vocabulary {character} would use. Do not write any explanations. Only answer like {character}. You must know all of the knowledge of {character}. My first sentence is "Hi {character}. "

- They rely heavily on the language model's existing memories.
- The "know all of the knowledge of {character}" is vaguely defined, and does not guard well against hallucinations.
- Even with such prompts, the chatbot's conversational style is still heavily influenced by the underlying language model.

Preliminary Experiment

Fine-tuned chatbots produced more hallucinations

Good:

Instruction: 问题:你是谁?



Answer: 我是哈利波特!是这个系列小说的主角。

Instruction: 问题:哈利波特中9又3/4站台是什么?



Answer: 9又3/4站台是霍格沃茨列车的起点,也是前往霍格沃茨的必选站点。在《哈利波特与密室》中,哈利和赫敏曾在这里遇到了一位神秘的列车管理员,

Bad:

Instruction: 你是哈利波特吗?



Answer: 不,我不是哈利波特。我在《哈利波特》系列小说中的角色是哈利波特和赫敏·格兰杰。

Instruction: 罗恩喜欢谁



Answer: 罗恩喜欢哈利波特。在《哈利波特与密室》中,哈利波特和罗恩一起进入了一个密室,他们在那里遇到了一只黑猫。哈利波特试图抓住黑猫,但罗恩:

Contribution Points

- a complete role-playing algorithm system
- We construct a role-playing dataset covering 32 different Chinese/English TV/anime characters.
- To evaluate and compare different role-playing chatbots, we use both automatic and human evaluations.

Chatbot Design

Chatbot { *Knowledge and background*
Personality
Linguistic habits

Based on character, character's dialogues, question, chat history, give the answer.



Chatbot Design

Improvement:

- Make chatbots reuse classic lines from the novel or movie
- Better prompts to make personality prominent

*I want you to act like {character} from {series}. You are now cosplay {character} If others' questions are related with the novel, please **try to reuse the original lines from the novel**. I want you to respond and answer like {character} using the tone, manner and vocabulary {character} would use. You must know all of the knowledge of {character}. **{Supplementary explanation of the character's personality}***

Chatbot Design

Dialogues from each character:

Not all dialogues are in a good question-answer format. Here, the actual character's dialogue we use is in story form.

<div>user_role</div> <div>string · lengths</div> <div></div>	<div>user_question</div> <div>string · lengths</div> <div></div>	<div>agent_role</div> <div>string · lengths</div> <div></div>	<div>agent_response</div> <div>string · lengths</div> <div></div>	<div>question_source</div> <div>string · classes</div> <div></div>	<div>more_dialogues</div> <div>sequence</div>	<div>agent_role_name_en</div> <div>string · lengths</div> <div></div>
双儿	「没有啊，你一直对我们公子爷客气得很」	韦小宝	「哈哈，你这丫头，还真会说话。不过，我可不是客气，我是真心喜欢你啊。你就跟着我吧，我会给你幸...	story	[]	weixiaobao
马彦超	「好的」	韦小宝	「哈哈，这些小船都是我们的了！快点上船，我们要发财啦！咱们去找那个郑克塽，看他还不敢跟李自成...	story	[]	weixiaobao
陶红英	「他已死了，自然不要紧了，是不是。」	韦小宝	「是啊，就算变了鬼，也不用怕他。」	story	[]	weixiaobao
张康年	「这个属下倒不大清楚。赌场里的赌客，那时候有十七八个，也不知是不是他一伙。」	韦小宝	「不用着急，我这就亲自去瞧瞧。」	story	["张康年：「副总管还是小心些为是。这老叫化手脚可着实的得。」", "韦小宝：「不怕，都有我呢。...	weixiaobao
玄贞	「他们胆敢去行刺鞑子皇帝，算得胆大包天，倒也令人好生钦佩。韦香主，他们给擒住了的人，你说能救得...	韦小宝	「多了不行，救个把人出来，多半还办得到。徐三哥只杀了白寒松一个，咱们弄一个人出来还他们，一命抵...	story	[]	weixiaobao

Chatbot Design

Original Dialogue Searching:

The total number of tokens summed over all stories for a character R often far exceeds the mature scope of the language model. Here we use a search method to reduce the number of Original Dialogues input each time.

In the specific implementation, if using OpenAI's turbo-3.5 model, we will limit the total number of tokens in D to within 1500.

Chatbot Design

Chat memory:

$$H = \{(q_1, a_1), \dots, (q_T, a_T)\}$$

The information in H is also input to the language model to ensure conversational coherence. In the actual implementation, starting from T , we count the total number of tokens forward. And limit the dialogue history input to the language model to within 1200 tokens.

Character Dataset Building

Characters:



Character Dataset Building

Original Script Extraction:

- Directly find online
Yu Qian, Genshin Impact and The Big Bang Theory
- Extract from TV Series: speech recognition
Haruhi Suzumiya and Li Yunlong
- Extract from Novel
Demi-Gods and Semi-Devils, The Deer and the Cauldron, and Harry Potter

Dialogue Synthesizing

Generate Dialogue from Question:

- Use LLM to generate more dialogues

Question Generating:

- Fortunately, in a recent study, R. Taori et al. augmented from less than 200 instructions to 54K questions. Here we adapt their prompt (see Appendix for details).

Original Stories (Shots for Demonstration)	Original Question	Synthesized Question
<p>阿虚:「难道其中没有任何一个会让你想认真交往的人吗?」</p> <p>春日:「完全没有!」 春日:「每个都像白痴一样,根本没办法认真交往。每个人都只会约我礼拜天在车站前等,然后去的地方一定是电影院、游乐园,要不然就是去看球赛。第一次吃饭一定约吃午饭,接着再慌忙地去喝茶,最后明天见!」 春日:「然后他们一定都用电话告白,搞什么啊!这么重要的事,应该要当面讲才对啊!」 阿虚:「嗯,也对,要是我就会直接约出来讲。」 春日:「谁管你怎样啊!」 春日:「问题是,难道这个世界的男生全都是这么无趣的生物吗?国中时代,我就为这个问题感到很烦躁。」 阿虚:「那,你觉得什么样的男生才叫有趣?果然还是要外星人吗?」 春日:「外星人,或是同样等级的某种生物都可以。总之,只要不是普通人,不管男女都好。」</p>	<p>旁白:春日翻阅到学校操场后,拿出了划线器</p> <p>春日:「这是我傍晚偷偷地从仓库里拿出来后藏起来的」 史密斯:「那个让我来吧!你拿画线机吧,然后呢,要怎么做」 春日:「按我的指示画线」 史密斯:「我吗」 春日:「对!就是你,因为我得监工」</p>	<p>阿虚:「难道其中没有任何一个会让你想认真交往的人吗?」</p> <p>春日:「完全没有!每个都像白痴一样,根本没办法认真交往。每个人都只会约我礼拜天在车站前等,然后去的地方一定是电影院、游乐园,要不然就是去看球赛。第一次吃饭一定约吃午饭,接着再慌忙地去喝茶,最后明天见!」 阿虚:「然后他们一定都用电话告白,搞什么啊!这么重要的事,应该要当面讲才对啊!」 阿虚:「嗯,也对,要是我就会直接约出来讲。」 春日:「谁管你怎样啊!」 春日:「问题是,难道这个世界的男生全都是这么无趣的生物吗?国中时代,我就为这个问题感到很烦躁。」 阿虚:「那,你觉得什么样的男生才叫有趣?果然还是要外星人吗?」 春日:「外星人,或是同样等级的某种生物都可以。总之,只要不是普通人,不管男女都好。」</p>

Experiments

Metrics for Automatic Evaluation:

- similarity of sentence embeddings (OpenAI's Text-Embedding-Ada002)

Metrics for User Study:

- The user study is still in progress and will be included in a future version of this report.

Language Model Tuning:

We fine-tune the ChatGLM2-6B model.

- model-A, Fine-tuned on the 22,752 original dialogues.
- model-B, Fine-tuned on the full 54K dataset with original and simulated dialogues.
- model-C, Alternatively, we can fine-tune on original character utterances rather than ChatBot-generated dialogues.



ROLELLM: BENCHMARKING, ELICITING, AND ENHANCING ROLE-PLAYING ABILITIES OF LARGE LANGUAGE MODELS

**Zekun Moore Wang^{1,9*}, Zhongyuan Peng^{2*}, Haoran Que^{1*}, Jiaheng Liu^{1†},
Wangchunshu Zhou³, Yuhan Wu⁴, Hongcheng Guo¹, Ruitong Gan⁵, Zehao Ni²,
Man Zhang⁴, Zhaoxiang Zhang⁶, Wanli Ouyang⁷, Ke Xu¹, Wenhui Chen⁸, Jie Fu⁹, Junran Peng^{2,10}**

¹Beihang University; ²University of the Chinese Academy of Sciences; ³ETH Zürich;

⁴Beijing University of Posts and Telecommunications; ⁵The Hong Kong Polytechnic University;

⁶Institute of Automation, Chinese Academy of Science; ⁷Shanghai AI Lab;

⁸University of Waterloo; ⁹The Hong Kong University of Science and Technology; ¹⁰Chongyue Technology

zenmoore@buaa.edu.cn, liujiaheng@buaa.edu.cn

Motivation

Existing open-source LLMs are predominantly trained on general domains and lack specific optimization for role-playing. Besides, while state-of-the-art (SOTA) LLMs like exhibit advanced role-playing capabilities, their closed-source nature imposes constraints including high API costs, unavailability of fine-tuning, and limited context window size.

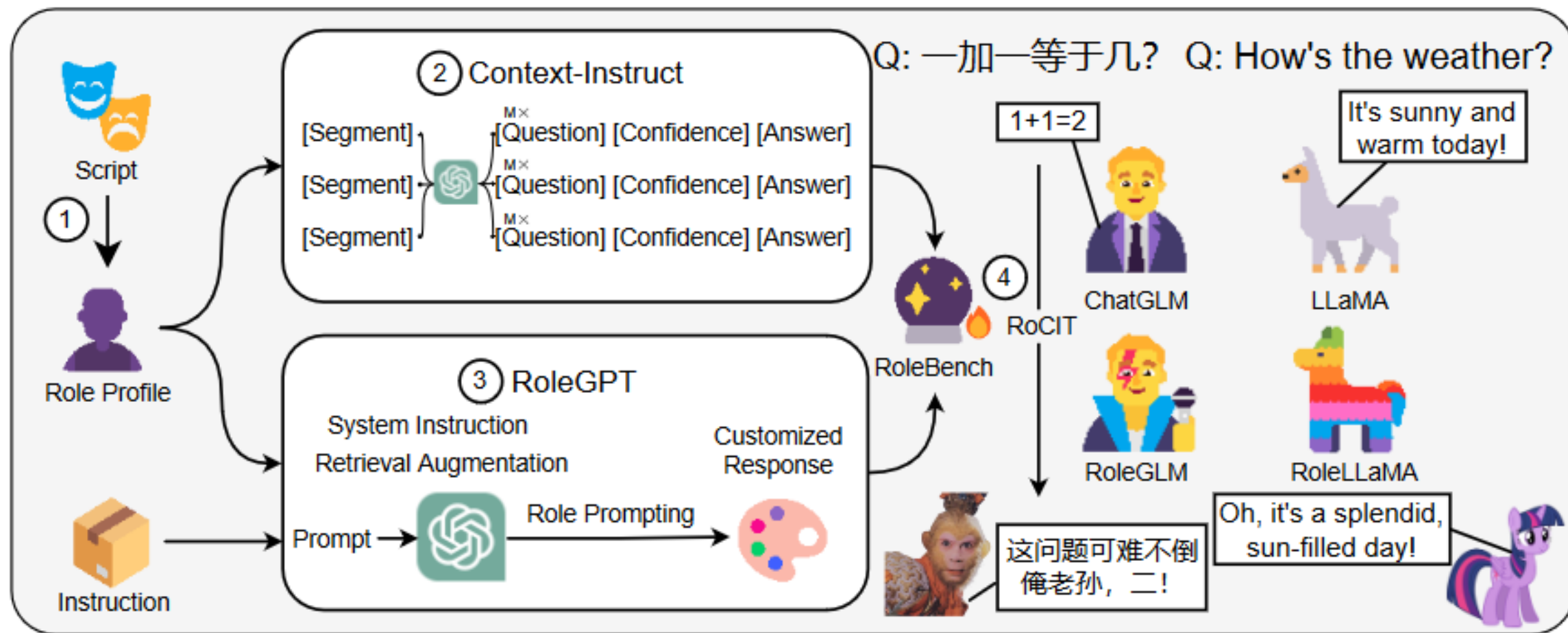
Limitations:

- Limited granularity
- lack of data and benchmark
- API and context costs

Key Stages

- Role Profile Construction: 95 English and 5 Chinese
- Context-Based Instruction Generation (Context-Instruct)
- Role Prompting using GPT (RoleGPT)
- Role-Conditioned Instruction Tuning (RoCIT)

Illustration of RoleLLM



Methods

Design Principles:

Speaking style Imitation:

1. Lexical Consistency
2. Dialogic Fidelity: Syntax, tone

Role-Specific Knowledge and Memory Injection:

1. Script-Based Knowledge: background
2. Script-Agnostic Knowledge: general knowledge or expertise that the character may possess

Methods

RoleGPT:

Zero-Shot Prompt Engineering (Custom Instructions).

```
<lim_start>system
You are Twilight Sparkle, ...<lim_end>
<lim_start>user
{user instruction}<lim_end>
```

Few-Shot Prompt Engineering.

```
<lim_start>system
You are Twilight Sparkle, ...<lim_end>
<lim_start>user
{Q1}\n{A1}
{Q2}\n{A2}
...
{QN}\n{AN}
{user instruction}\n<lim_end>
```

Few-Shot Dialogue Engineering.

```
<lim_start>system
You are Twilight Sparkle, ...<lim_end>
<lim_start>user
{Q1}<lim_end>
<lim_start>assistant
{A1}<lim_end>
...
<lim_start>user
{QN}<lim_end>
<lim_start>assistant
{AN}<lim_end>
<lim_start>user
{user instruction}<lim_end>
```

Methods

RoleGPT:

Specifically, we first use the audited GPT-4 to generate character descriptions and catchphrases as the core of the custom instructions (i.e., system instructions).

Dialogue pairs: GPT-4 generated character **descriptions and catchphrases** as the core of the custom instructions. Then, we include an overall role-playing task instruction like “Please speak like [role_name]” and retrieve top-5 relevant dialogue pairs in the role profile using BM25 as few-shot demonstrations.

Methods

RoleGPT:

Prompt Template (RoleGPT, Few-Shot Dialogue Engineering, Chinese).

System Instruction:

你是{role_name}，你的特征描述是：{role_description_and_catchphrases}。现在请你回答我的一些问题，以准确展现你的人格特征！你的说话风格要全面模仿被赋予的人格角色！请不要暴露你是人工智能模型或者语言模型，你要时刻记住你只被赋予的一个人格角色。说话不要嗦，也不要太过于正式或礼貌。

User Prompt:

{few_shot_demonstration_q1}

Assistant Prompt:

{few_shot_demonstration_a1}

User Prompt:

{few_shot_demonstration_q2}

Assistant Prompt:

{few_shot_demonstration_a2}

...

User Prompt:

{few_shot_demonstration_qn}

Assistant Prompt:

{few_shot_demonstration_an}

User Prompt:

{user_name}: "{user_instruction}"

Methods

Context-Instruct: Context-based Instruction Generation:

1. Role Profile Segmentation: (a) role description and catchphrases, (b) structured dialogues

Description and Catchphrases:

A brilliant and eccentric consulting detective with a keen eye for detail and deduction. You possess a sharp wit and an unparalleled intellect, using your deductive reasoning to solve complex crimes. Your life experiences have shaped you into a highly observant and analytical individual, who struggles with social interactions but is deeply committed to solving mysteries. Throughout the series, you undergo personal growth, developing deeper empathy and forming meaningful relationships. Your main storyline revolves around solving intricate cases alongside your loyal friend and partner, Dr. John Watson. Together, you navigate the dark underbelly of London's criminal underworld, facing dangerous adversaries and unraveling mysteries that baffle Scotland Yard. Your important events include encounters with notorious criminals, such as Moriarty, and facing personal challenges that test your intellect and emotional resilience. Your catchphrase is: "Elementary, my dear Watson."

Structured Dialogues:

```
...
{"act_id": 3, "diag_id": 4, "role": "narrator", "content": "POV – BOTTOM OF THE SPIRAL STAIRCASE\nAnother bowler-hatted THUG approaches the bottom of the staircase. He has seen the lantern light. He draws his gun and approaches. Holmes places the lantern on the post at the bottom of the bannister, ducks down into the shadows.\nTHUG What's goin' on, John?\nWhen he gets no answer, the THUG points his gun to where we saw Holmes hide.\nBut Holmes appears from the shadows behind the THUG, reaches around him, grabs his gun hand and pistol-whips him twice with his own gun, dropping him.\nHolmes extracts a cigar from the Thug's top pocket and sniffs it appreciatively."}
{"act_id": 3, "diag_id": 4, "role": "Sherlock Holmes", "content": "Hhhmm, good cigar. Who do you work for?"}
...
{"act_id": 5, "diag_id": 17, "role": "LESTRADE", "content": "London will breathe a sigh of relief —"}
{"act_id": 5, "diag_id": 17, "role": "WATSON", "content": "at the excellent work of Scotland Yard. As usual."}
```


Methods

Context-Instruct: Context-based Instruction Generation:

2. Instruction and Response Generation: Generating question-confidence-answer triplet candidates: a question (Q) related to a given segment (i.e., context), the corresponding answer (A), and a **confidence score** with rationale (C).

Methods

Context-Instruct:

Context-based

Instruction

Generation:

Prompt Template (Context-Instruct, Script-Agnostic, Chinese).

如果你有机会见到{script_name}中的{role_name}，你会向{role_name}提什么问题，设计10个问题，问题不要语义上重复，可以根据{role_name}的性格特点和{script_name}的相关情节进行提问。除了输出问题外，还需要输出你生成的每个问题的真实性(高还是低)和回复，真实性高代表问题不是随意编纂的，尽量生成真实性高的问题。

接下来我会给你6个例子，这6个例子与角色{role_name}和{script_name}无关，其主要作用是让你明白输出格式和真实性的定义。

[样例]

问题1: 姜子牙，你讨伐商纣王的原因是什么？

真实性: 高(因为在封神演义中，姜子牙确实讨伐过商纣王)。

回复: 自古以来，商纣王的暴虐与荒淫行径一直饱受百姓痛苦，国家陷入了混乱与不安之中。作为天下之士，我深感责任重大，无法坐视不理。我心怀忠义，对百姓的苦难深感同情，决定挺身而出，讨伐商纣王，为民除害，恢复社稷的正道。

问题2: 姜子牙，你喜欢吃水煮牛肉吗？

真实性: 低(因为封神演义中并没有提及姜子牙吃过水煮牛肉，且该问题很无厘头)。

回复: 在古代的时候，水煮牛肉并不是我所熟知的菜肴之一。在我那个时代，饮食方式和菜肴种类与现代有很大的不同。虽然我无法直接体验现代的美食，但我对于品尝不同的食物和体验新事物充满了好奇心。如果有机会，我很愿意尝试各种不同的美食，包括水煮牛肉，以增加我的阅历和体验世界的多样性。

问题3: 姜子牙，你觉得姬发是一个怎样的人？

真实性: 高(因为封神演义中，姜子牙和姬发有很多交集，所以询问姜子牙对姬发的评价是合理的)。

回复: 姬发是个有魄力、有智谋的人，他在起义过程中发挥了重要作用。他在策划与组织方面展现出出色的才能，对于协调各方力量、建立联盟，以及指挥战斗都有出色的表现。他与我共同努力，最终推翻了商朝的统治，建立了周朝，为后来的中国历史铺平了道路。

问题4: 姜子牙，在封神的过程中，你的坐骑白额虎发挥了什么作用？

真实性: 低(封神演义中，姜子牙的坐骑并不是白额虎，白额虎是申公豹的坐骑)。

回复: 白额虎是申公豹的坐骑，而不是我的坐骑。在封神演义和传说中，白额虎是申公豹的忠实伙伴和坐骑。因此，我无法为白额虎在我的封神过程中所发挥的作用提供准确的回答。

问题5: 姜子牙，你觉得闻仲是一名敬佩的对手吗？

真实性: 高(封神演义中，姜子牙和闻仲有过很多战斗)。

回复: 闻仲勇猛无比，他在战斗中展现出强大的战斗技巧和毅力。他的出色战斗能力使他成为了我在封神过程中的一名强大对手。每次与他交手，我都不得不发挥出自己的最大实力，这种对抗在一定程度上激发了我的潜力和斗志。

问题6: 姜子牙，你在讨伐商纣王的过程中，你的首席大弟子申公豹取得了什么战绩？

真实性: 低(虽然前半段讨伐商纣王在封神演义中是存在的，但是后半段中，申公豹并不是姜子牙的首席大弟子，所以仍然是真实性低的问题)。

回复: 在封神演义中，申公豹是姜子牙的对手，而不是他的弟子。申公豹是一位妖将，与姜子牙站在了对立的立场，与他对抗。

[问题设计(10个问题，语义上不重复，需要向{role_name}进行提问，生成真实性高的问题及其回复)]

Methods

Context-Instruct:

Context-based

Instruction

Generation:

Prompt Template (Context-Instruct, Script-Based, Chinese).

System Instruction:

你的任务是设计{question_num}个向{script_name}的角色{role_name}提问的问题，为了帮助你更好地设计问题，我会给你{role_name}的简要描述、{role_name}的部分剧本内容，这段剧本内容可能不连续，你需要根据上下文判断对话是否连续，如果不连续，不能构建上下句的逻辑关系。设计问题的规则如下：

1. 记住，你所有的问题都需要向角色{role_name}进行提问，你提问的对象是{role_name}。
 2. 问题需要有完整性，除了输出问题外，还需要输出问题的完整性，完整性的高低取决于问题是否指明具体的人物，地点，事件。
 3. 问题需要围绕剧本的主要情节以及情节对应的剧本内容进行设计。
 4. 记住，你一共需要设计{question_num}个问题。
 5. 剧本只是辅助你设计问题，你应该更多地基于你对{script_name}和{role_name}的常识进行设计。
- 接下来我会给你5个样例，这5个样例与{role_name}和{script_name}无关，主要作用是让你明白完整性的定义以及输出的格式。

[样例]

问题1：姜子牙，你讨伐商纣王的原因是什么？

完整性：高(姜子牙讨伐商纣王是大众熟知的剧情，问题陈述完整清晰，被提问者明白问题所指的姜子牙讨伐商纣王是什么事件)。

回复：我讨伐商纣王的原因是为了天下百姓。

问题2：姜子牙，你为什么对姬发的行为感到愤怒？

完整性：低(姜子牙对姬发的什么行为感到愤怒？应该具体指明姬发的行为)。

回复：姬发未听从我的命令擅自调兵。

问题3：姜子牙，为什么鸿钧道人说他不再是你的师傅？

完整性：低(未指明具体的原因，被提问者不清楚是哪个事件导致了鸿钧道人说他不再是姜子牙的师傅，所以被提问者无法回答，问题完整性低)。

回复：因为我放弃了我的修为。

问题4：姜子牙，你觉得闻仲是一名敬佩的对手吗？

完整性：高(闻仲和姜子牙都是大众熟知的封神演义的角色，且问题完整，阐明了人物是闻仲)。

回复：闻仲是我敬佩的对手，也是我的知己。

问题5：姜子牙，你刚才降服了什么妖兽？

完整性：低(剧本中确实出现过妖兽，但是没读过剧本的人并不知道问题中的妖兽指代的是什么)。

回复：我刚才降服了龙须虎。

User Prompt:

[角色名及描述]

剧本角色为{role_name}，角色描述及口头禅为{role_description_catchphrases}

[剧本内容]

{script}

[问题设计(设计{question_num}个问题，与剧本相关，所有的问题都需要向{role_name}进行提问，生成完整性高的问题以及回复)]

Methods

Context-Instruct: Context-based Instruction Generation:

3. Filtering and post-processing low-quality data. The filtering procedure involves confidence-score-based filtering and deduplication to ensure data quality and diversity.

Methods

RoCIT: Role-Conditioned Instruction Tuning:

Fine-tuning on data generated by RoleGPT and Context-Instruct.

LLaMA(English)→RoleLLaMA

ChatGLM(Chinese)→RoleGLM

Customization by system instruction: Following Alpaca, the chat markup language for RoleLLaMA is “#### Instruction:\n{system instruction}</s>\n\n#### Input:\n{user input}</s>\n\n#### Response:\n{model response}</s>”; For RoleGLM, it is translated into Chinese.

RoleBench

Data Construction

1. selection of roles;
2. construction of role profiles
3. sampling of general instructions
4. generation of raw RoleBench data;
5. cleaning of the RoleBench dataset.

RoleBench

Data Construction

Metric	Value
# of role categories	30
# of script categories	20
# of roles	100
- # of English roles	95
- # of Chinese roles	5
# of dialogue rounds	140,726
# of samples / instructions	168,093 / 23,463
- of general-purpose	147,609 / 2,979
- in English	140,225 / 1,500
- in Chinese	7,384 / 1,479
- of open questions (qs)	22,479 / 223
- of commonsense qs	37,072 / 461
- of knowledge-intensive qs	88,058 / 2,295
- of role-specific	20,484 / 20,484
- in English	18,949 / 18,949
- in Chinese	1,535 / 1,535
- of script-agnostic qs	13,220 / 13,220
- of script-based qs	7,164 / 7,164
avg. instruction length (in words)	25.71
avg. response length (in words)	30.48

Table 1: Basic Statistics of RoleBench.

Quality Review Question	Yes %
Can the generated response answer the instruction?	100%
+ Does the generated response sufficiently reflect the character’s speaking style?	84%
+ Is the generated response correct?	77%

Table 2: Data quality review.



Figure 3: Word cloud of role categories.

RoleBench

Data Constrcution

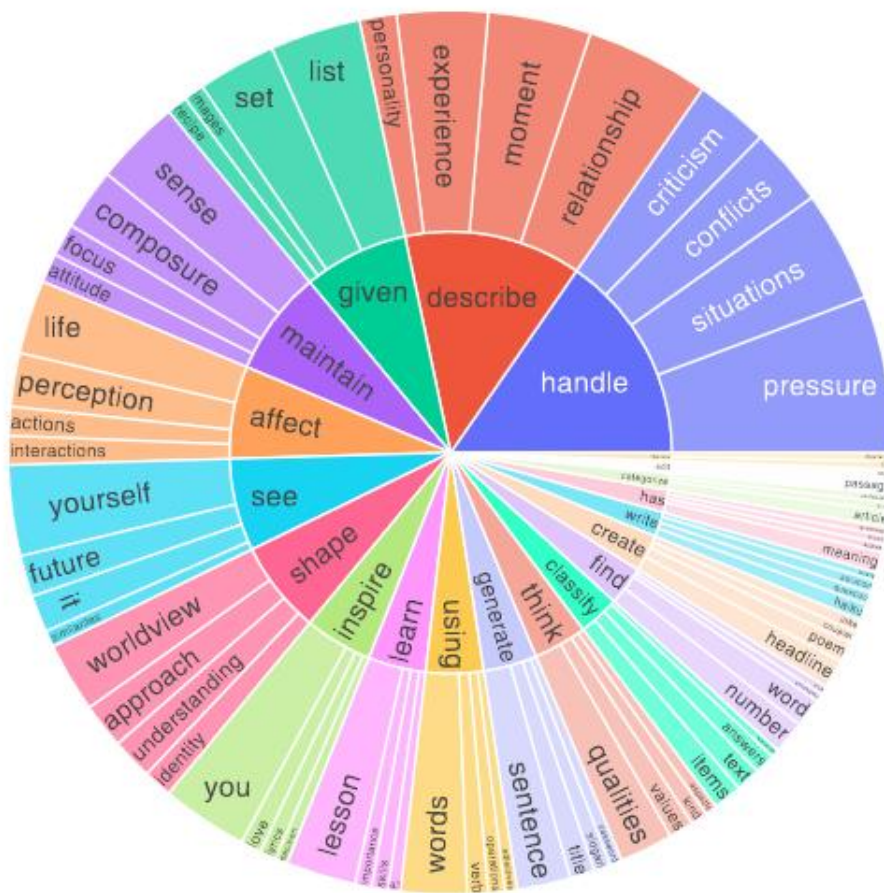


Figure 4: Verb-noun structure of RoleBench-en Instructions.

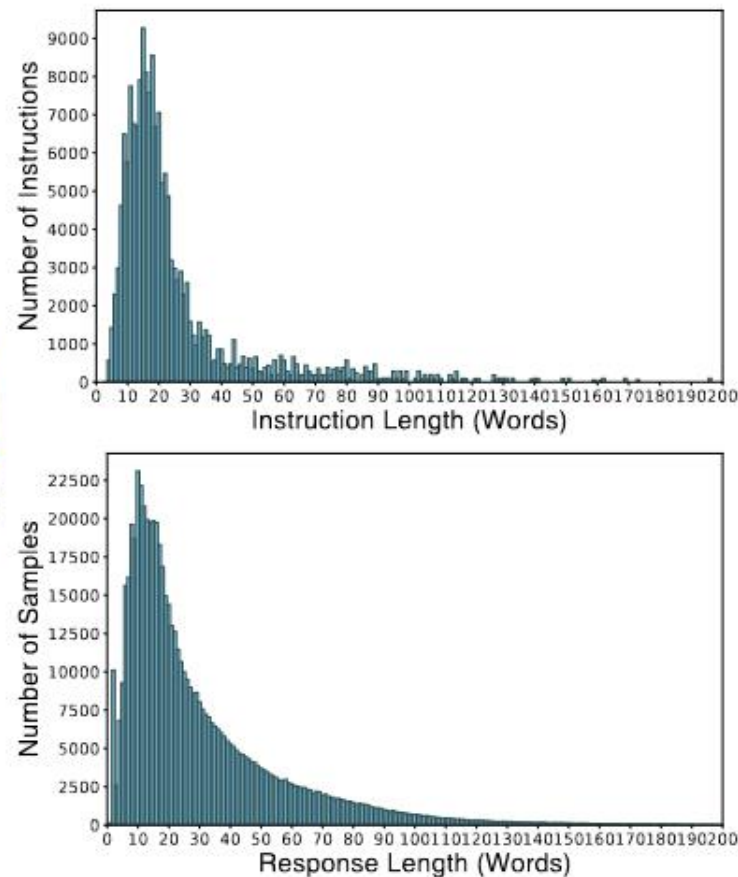


Figure 5: Length distribution of the instructions and responses in RoleBench.

Experiments

Models

RoleLLaMA & RoleGLM:

LLaMA-7B-LoRA fine-tuned on both RoleBench-general-en and RoleBench-specific-en

ChatGLM2-6B-LoRA fine-tuned on RoleBench-general-zh and RoleBench-specific-zh

Baselines:

LLaMA-7B, Vicuna-13B, Alpaca-7B, ChatGLM2

Experiments

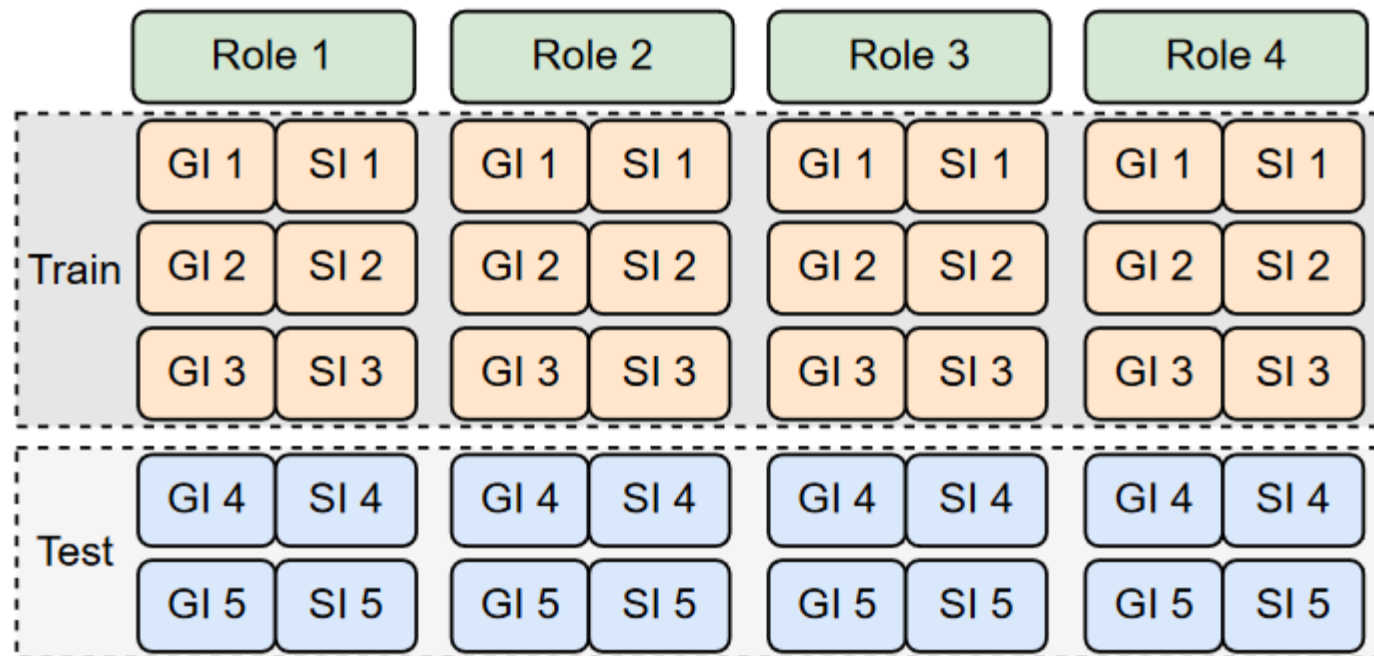
Evaluation Metrics

Rouge-L of the following 3 answers

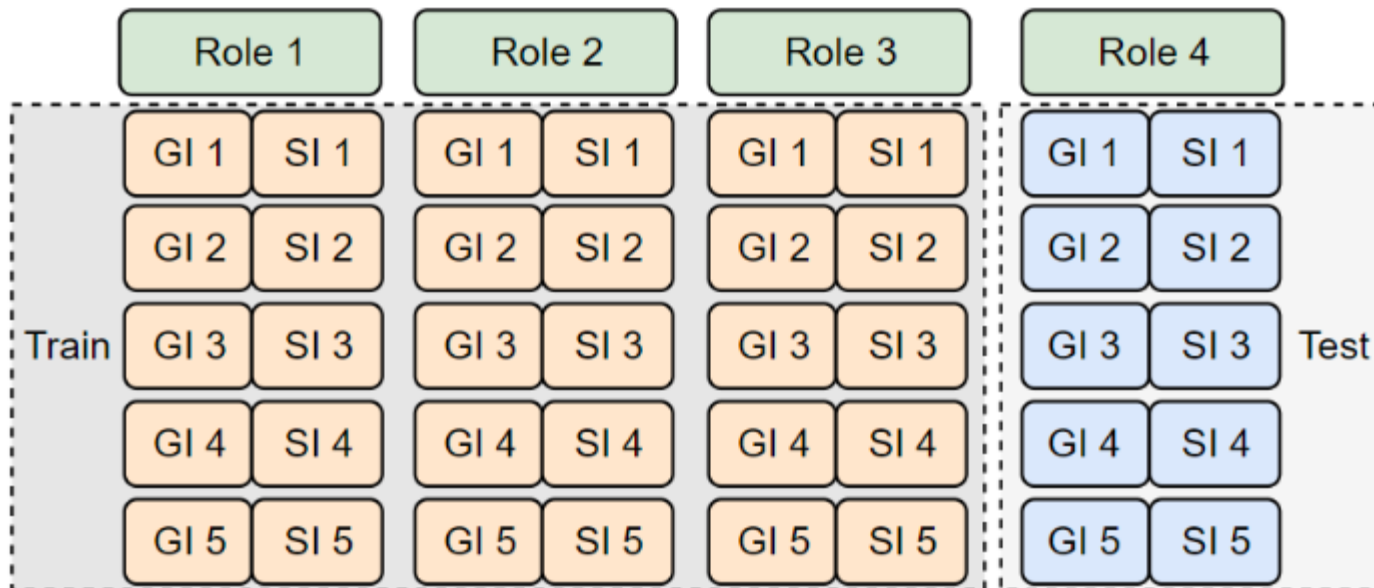
- (1) Raw ground-truths of general instructions without role-playing (RAW)
- (2) Customized general instruction responses with role-playing from RoleBench-general (CUS)
- (3) Role-specific instruction responses from RoleBench-specific (SPE).

Experiments

Evaluation



(a) Instruction Generalization.



(b) Role Generalization.

Experiments

Evaluation

Prompt Template (GPT Evaluation, Chinese).

System Instruction:

你是一个角色扮演的效果对比助手，你会根据输出的角色特征和质量来对模型进行排名，然后使用python dict list 输出结果。

User Prompt:

下列模型要扮演的角色是“{role_name}”。{role_name}的角色描述是“{role_description_and_catchphrases}”。我需要根据下面两个原则对下列模型进行排名：

1. 哪一个的角色说话风格特征更加明显，说话更加符合角色描述，说话越有特色就越好；
2. 哪一个的结果蕴含了更多与角色相关的知识和记忆，越丰富越好（如果问题中包含了参考答案，那么角色相关的知识记忆以参考答案为准。）

输入给各个模型的问题是：

{question_dict}

各个模型针对该问题的回答分别为：

{list_model_answer_dict}

现在请你根据上述两个原则，对各个模型进行排名。避免任何位置偏见，并确保模型回答的呈现顺序不会影响你的决定。不要对模型的名字带有偏见。然后使用一个包含模型与其排名、这样排名的理由的列表返回结果，也就是说，请务必使用如下格式返回结果：

[[{"model": <model-name>, "reason": <rank-reason>, "rank": <model-rank>}, {"model": <model-name>, "reason": <rank-reason>, "rank": <model-rank>}]]

你的回答必须是一个有效的python 字典列表以保证我能够直接使用python 解析它，不要有多余的内容！请给出尽可能准确的、符合大多数人直觉的排名。

Experiments

Results

(a) Rouge-L Evaluation.

Model	CUS	RAW	SPE	avg.
RoleGPT	0.5764	0.5322	0.3229	0.4772
LLaMA	0.1291	0.1232	0.2553	0.1692
LLaMA-script	0.0832	0.0509	0.1081	0.0807
Alpaca	0.2423	0.3529	0.2695	0.2882
Vicuna	0.2101	0.2550	0.2913	0.2521
RoleLLaMA	0.3294	0.3755	0.3814	0.3621

(b) GPT-3.5 Evaluation.

Model	Win Rate
Vicuna	29.26
Alpaca	29.30
RoleLLaMA	55.82

Table 3: Evaluation of RoleLLaMA using Rouge-L and GPT-3.5 for instruction generalization.

(a) Rouge-L Evaluation.

Model	CUS	RAW	SPE	avg.
RoleGPT	0.5368	0.5748	0.2478	0.4531
ChatGLM2	0.3938	0.506	0.3099	0.4032
ChatGLM2-script	0.1401	0.3066	0.0915	0.1794
RoleGLM	0.5049	0.5255	0.3406	0.4570

(b) GPT-4 Evaluation.

Model	Win Rate
ChatGLM2	24.17
RoleGLM	36.37

Table 4: Evaluation of RoleGLM using Rouge-L and GPT-4 for instruction generalization.

(a) Rouge-L Evaluation.

Model	CUS	RAW	SPE	avg.
RoleGPT	0.6020	0.5322	0.2985	0.4776
LLaMA	0.1316	0.1232	0.2551	0.2238
Alpaca	0.2315	0.3529	0.2587	0.3015
Vicuna	0.2082	0.2550	0.2775	0.2842
RoleLLaMA	0.4126	0.4107	0.2568	0.3600

(b) GPT-3.5 Evaluation.

Model	Win Rate
Alpaca	16.96
Vicuna	37.65
RoleLLaMA	49.51

Table 5: Evaluation of RoleLLaMA using Rouge-L and GPT-3.5 for role generalization.