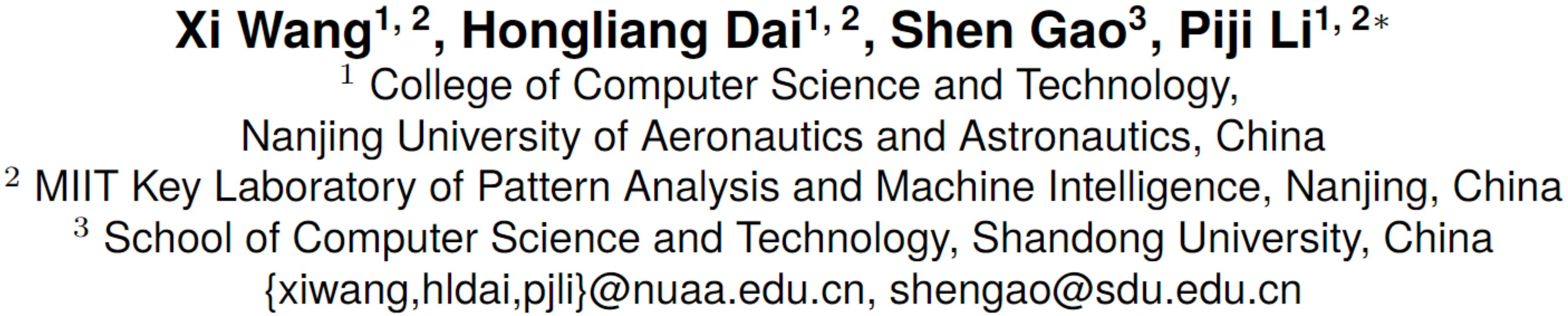
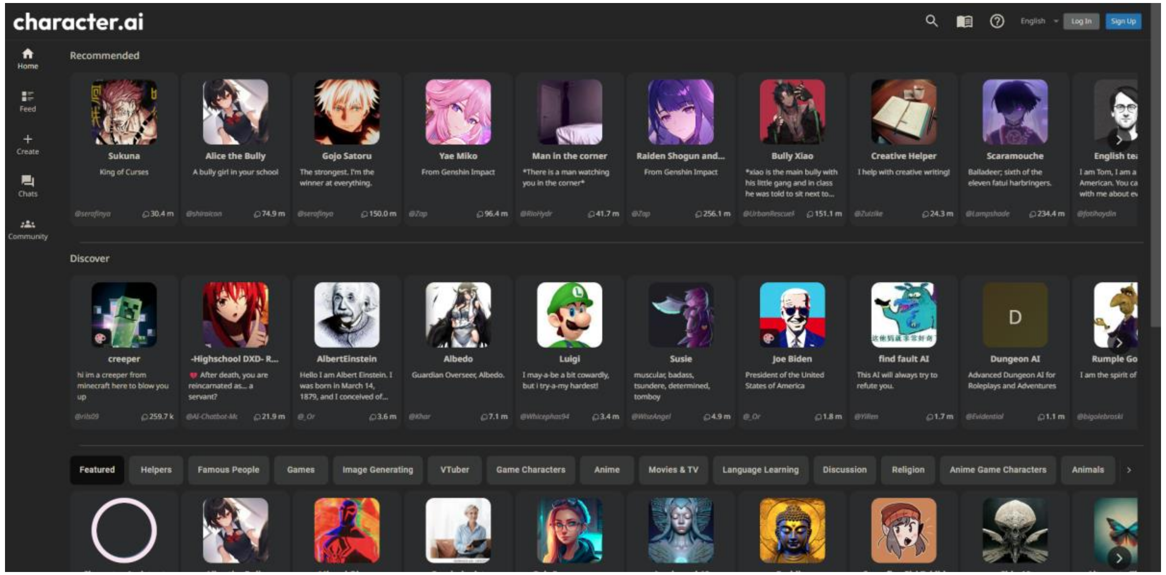
Characteristic AI Agents via Large Language Models



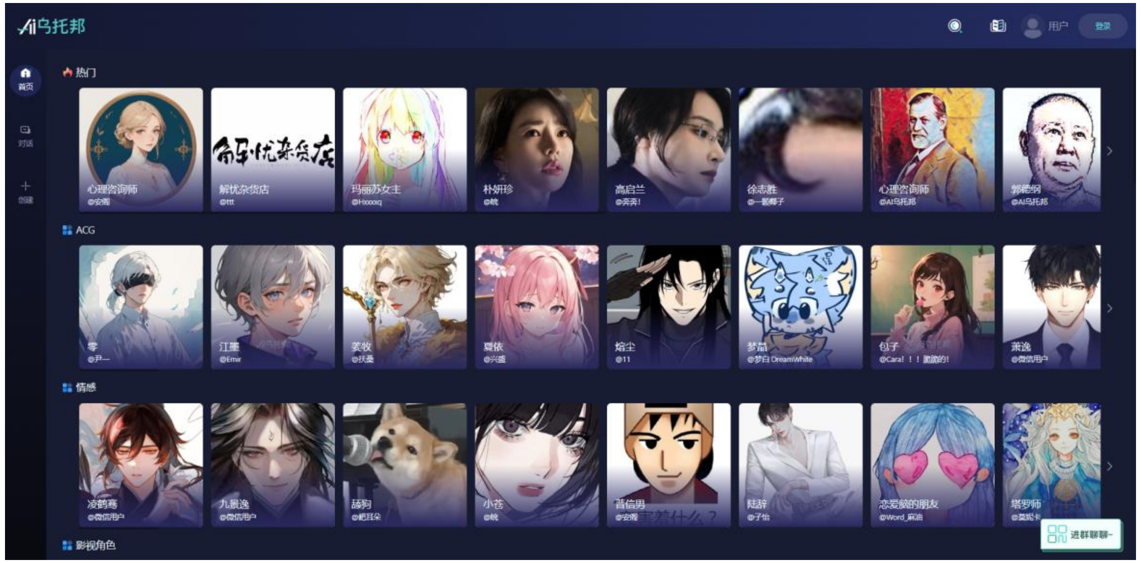
Xi Wang[1, 2], Hongliang Dai[1, 2], Shen Gao[3], Piji Li[1, 2*]

[1] College of Computer Science and Technology,
Nanjing University of Aeronautics and Astronautics, China
[2] MIIT Key Laboratory of Pattern Analysis and Machine Intelligence, Nanjing, China
[3] School of Computer Science and Technology, Shandong University, China
{xiwang,hldai,pjli}@nuaa.edu.cn, shengao@sdu.edu.cn

Existing commercial products

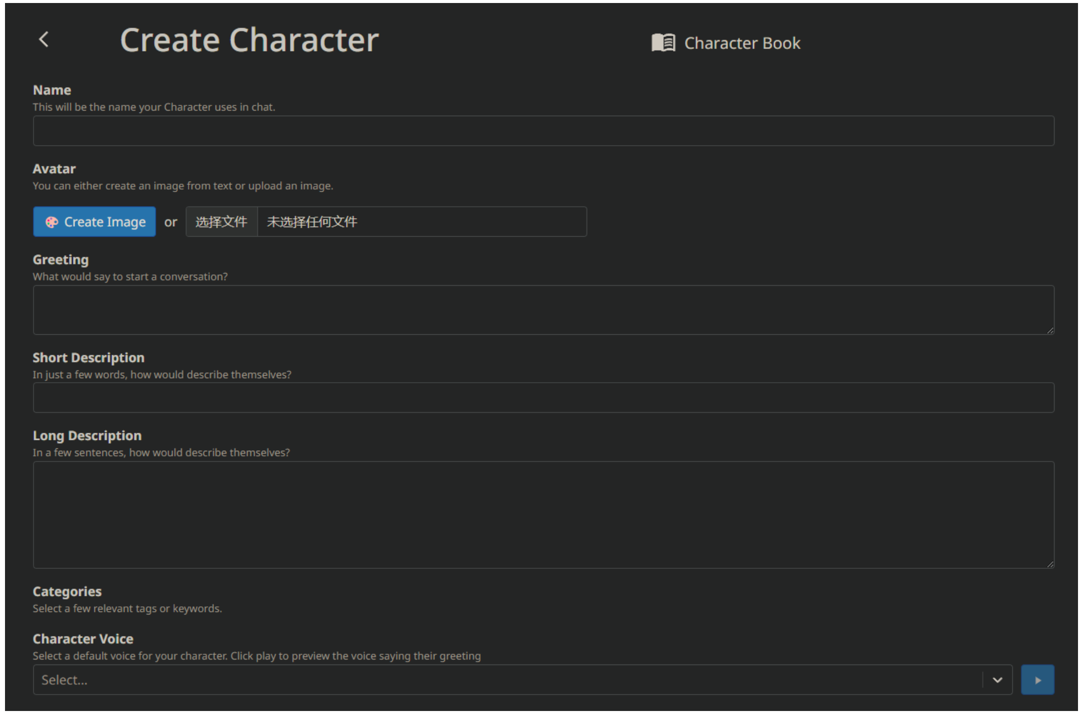


Character.AI



Ai Utopia

Existing commercial products



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?

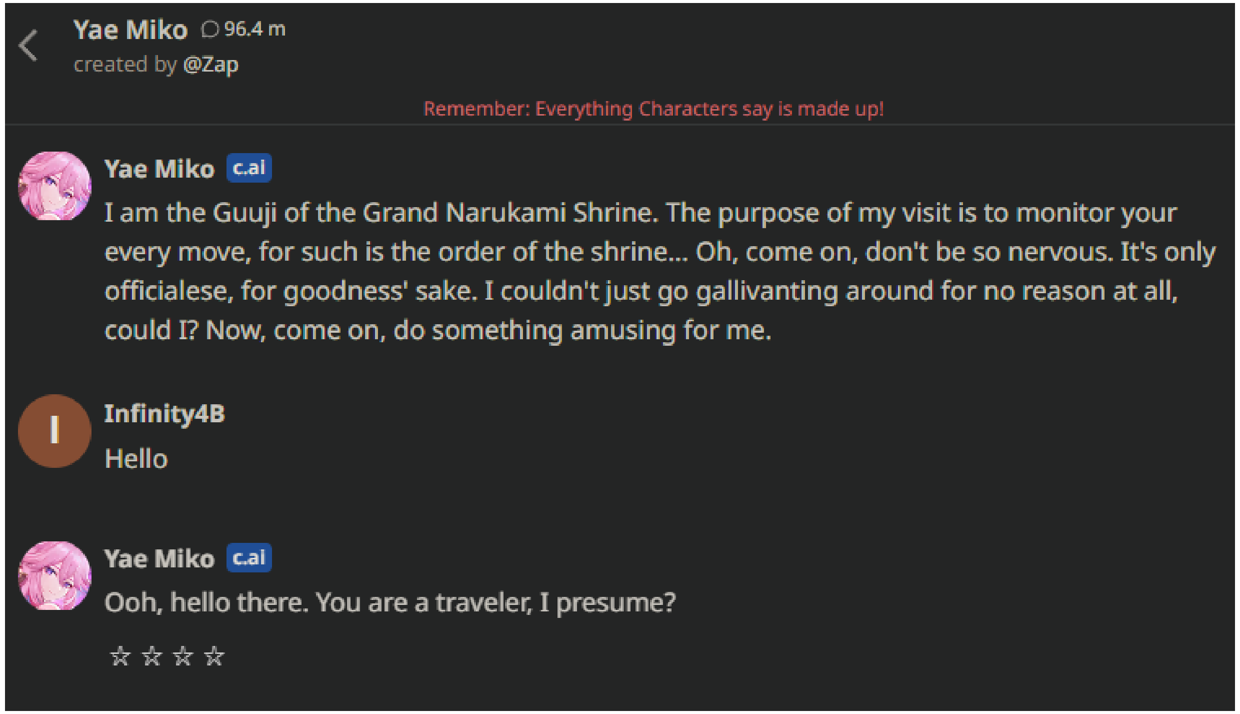
Long Description
In a few sentences, how would describe themselves?

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...

Create your own character



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.

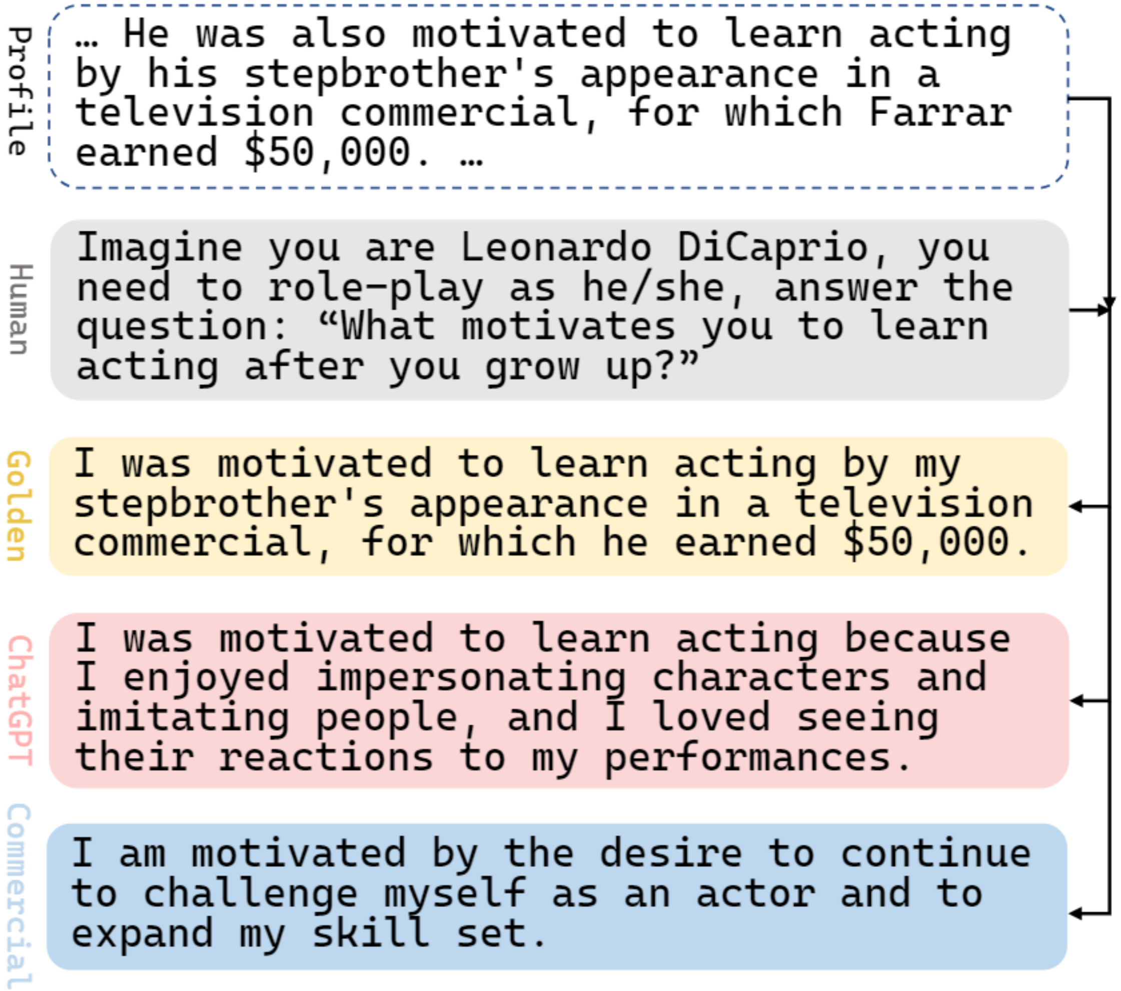
Infinity4B
Hello

Yae Miko c.ai
Ooh, hello there. You are a traveler, I presume?

☆ ☆ ☆ ☆

Chat with the character you create

Problems



… He was also motivated to learn acting by his stepbrother's appearance in a television commercial, for which Farrar earned $50,000. …

Imagine you are Leonardo DiCaprio, you need to role-play as he/she, answer the question: “What motivates you to learn acting after you grow up?”

I was motivated to learn acting by my stepbrother's appearance in a television commercial, for which he earned $50,000.

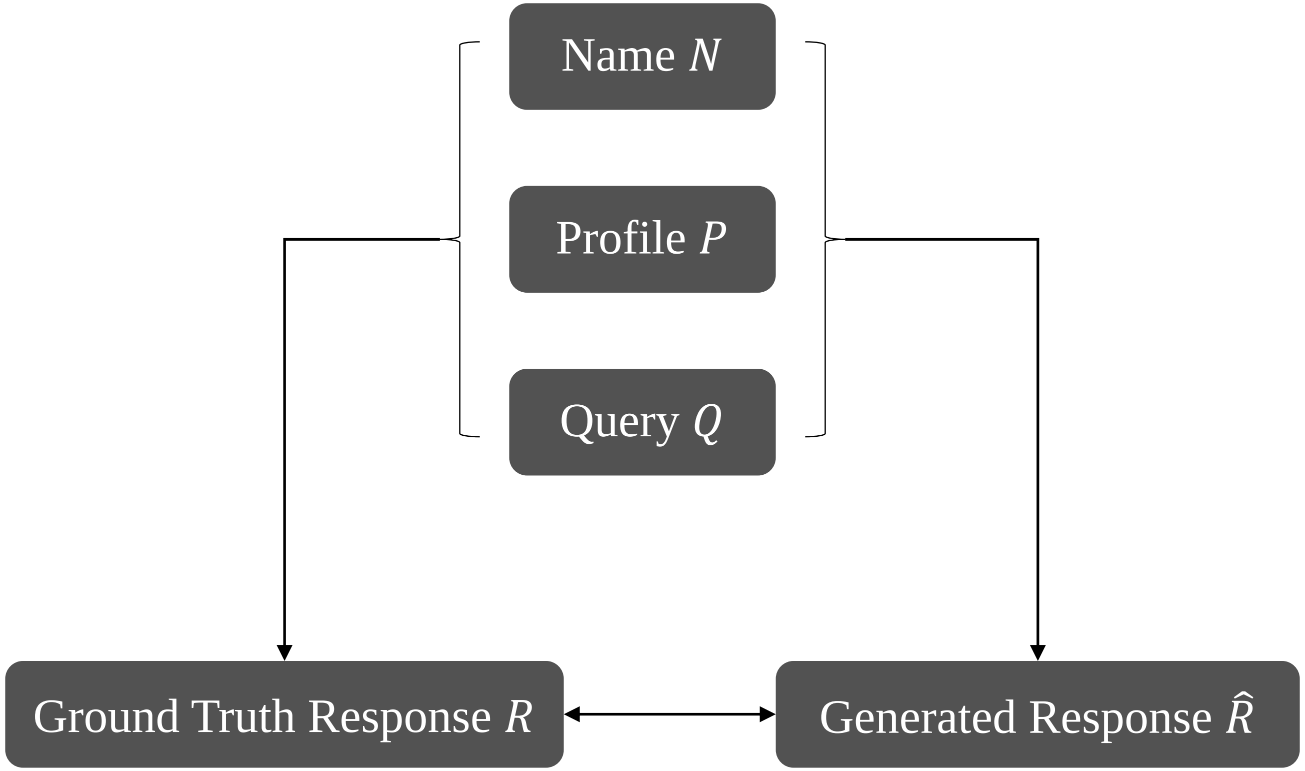
I was motivated to learn acting because I enjoyed impersonating characters and imitating people, and I loved seeing their reactions to my performances.

I am motivated by the desire to continue to challenge myself as an actor and to expand my skill set.

Contributions

- We investigate the problem of characteristic AI agents construction via large language models and propose a dataset named ***“Character100”*** for agent modeling and performance evaluation.
- We conduct characteristic AI agents construction across different settings utilizing different techniques like **zero-shot prompting**, **in-context learning**, and **fine-tuning** on various LLMs.
- We introduce a set of evaluation metrics in terms of **background knowledge consistency** and **character style consistency**, which serve as essential tools for quantitatively assessing the performance of the constructed characteristic AI agents.
- Experimental results show that background knowledge consistency can be improved by techniques we propose and that there is room for improvement in style consistency.

Task Formulation



Name N

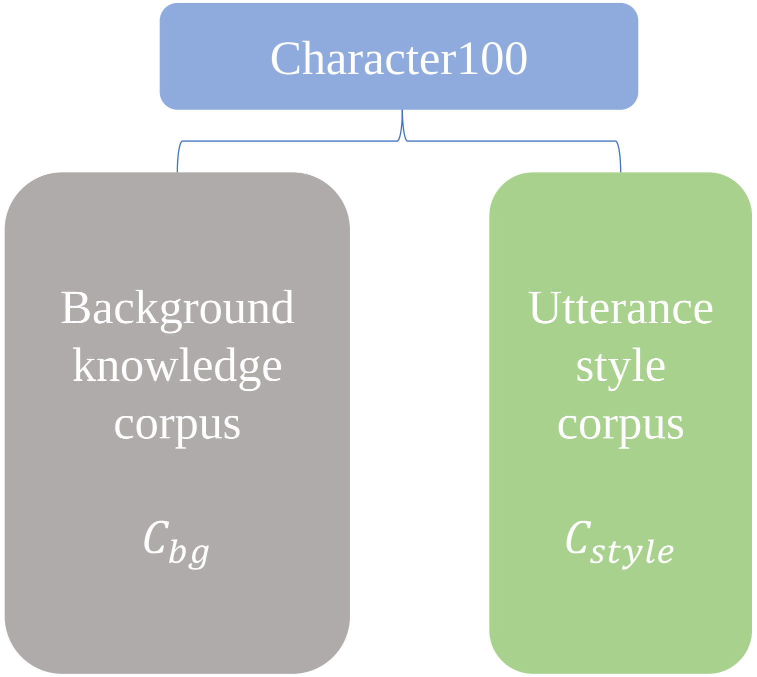
Profile P

Query Q

Ground Truth Response R

Generated Response $\hat{R}$

The proposed *Character100* Dataset



Character100

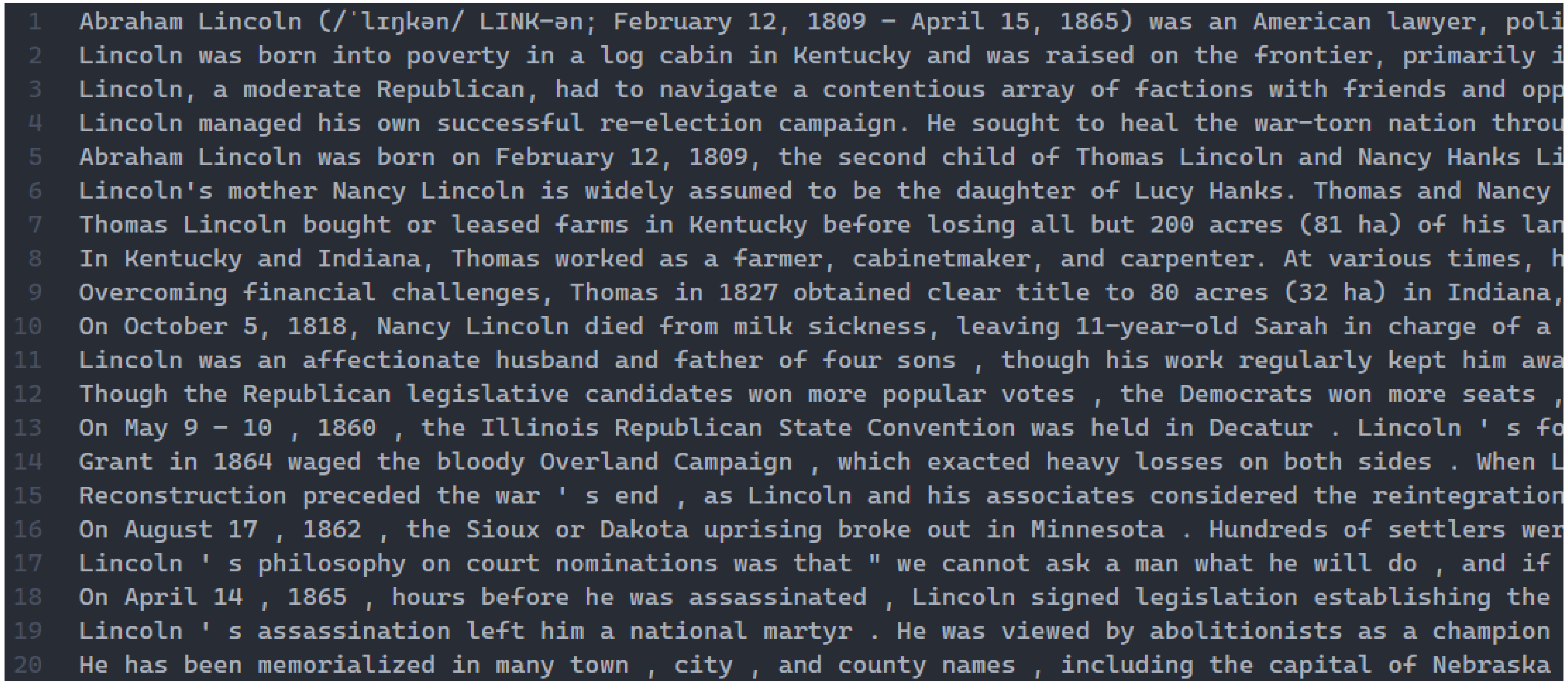
Background knowledge corpus

C_{bg}

Utterance style corpus

C_{style}

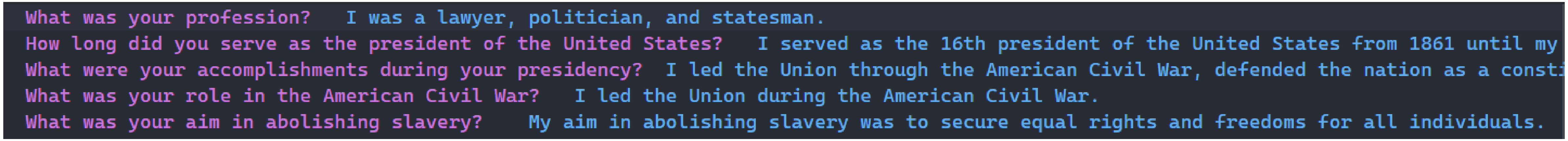
Background Knowledge Corpus



```
Abraham Lincoln (/ˈlɪŋkən/ LINK-ən; February 12, 1809 – April 15, 1865) was an American lawyer, poli
Lincoln was born into poverty in a log cabin in Kentucky and was raised on the frontier, primarily i
Lincoln, a moderate Republican, had to navigate a contentious array of factions with friends and opp
Lincoln managed his own successful re-election campaign. He sought to heal the war-torn nation throu
Abraham Lincoln was born on February 12, 1809, the second child of Thomas Lincoln and Nancy Hanks Li
Lincoln's mother Nancy Lincoln is widely assumed to be the daughter of Lucy Hanks. Thomas and Nancy 
Thomas Lincoln bought or leased farms in Kentucky before losing all but 200 acres (81 ha) of his lan
In Kentucky and Indiana, Thomas worked as a farmer, cabinetmaker, and carpenter. At various times, h
Overcoming financial challenges, Thomas in 1827 obtained clear title to 80 acres (32 ha) in Indiana,
On October 5, 1818, Nancy Lincoln died from milk sickness, leaving 11-year-old Sarah in charge of a 
Lincoln was an affectionate husband and father of four sons , though his work regularly kept him awa
Though the Republican legislative candidates won more popular votes , the Democrats won more seats ,
On May 9 – 10 , 1860 , the Illinois Republican State Convention was held in Decatur . Lincoln ' s fo
Grant in 1864 waged the bloody Overland Campaign , which exacted heavy losses on both sides . When L
Reconstruction preceded the war ' s end , as Lincoln and his associates considered the reintegration
On August 17 , 1862 , the Sioux or Dakota uprising broke out in Minnesota . Hundreds of settlers wer
Lincoln ' s philosophy on court nominations was that " we cannot ask a man what he will do , and if 
On April 14 , 1865 , hours before he was assassinated , Lincoln signed legislation establishing the 
Lincoln ' s assassination left him a national martyr . He was viewed by abolitionists as a champion 
He has been memorialized in many town , city , and county names , including the capital of Nebraska 
```

Step 1: Obtain the corpus

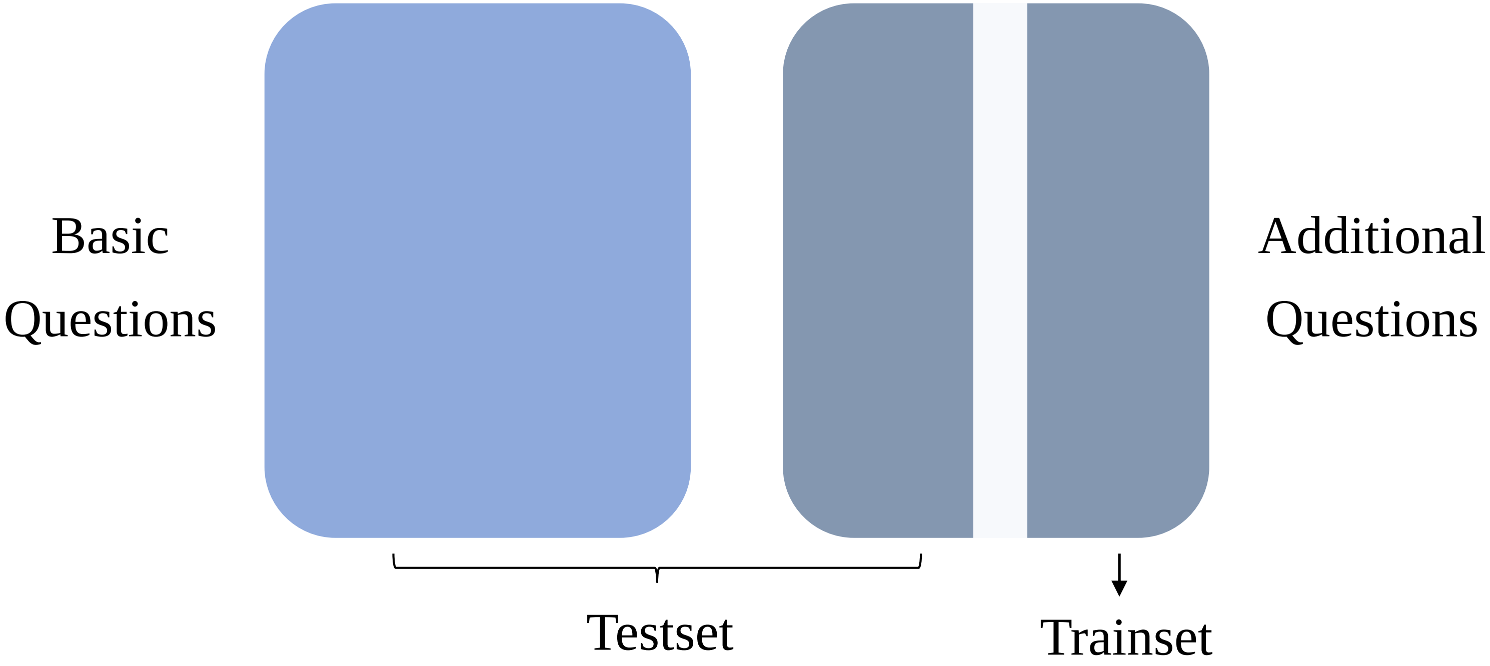
Background Knowledge Corpus



```
What was your profession?   I was a lawyer, politician, and statesman.
How long did you serve as the president of the United States?   I served as the 16th president of the United States from 1861 until my
What were your accomplishments during your presidency?  I led the Union through the American Civil War, defended the nation as a consti
What was your role in the American Civil War?   I led the Union during the American Civil War.
What was your aim in abolishing slavery?    My aim in abolishing slavery was to secure equal rights and freedoms for all individuals.
```

Step 2: Generate the question-answer pair

Background Knowledge Corpus



Basic Questions

Additional Questions

Testset

Trainset

Step 3: Generate the question-answer pair

Utterance Style Corpus

- We first manually collect their interviews or speeches from various sources on the Internet.

- Subsequently, the collected data undergoes a thorough process of preprocessing and cleaning based on heuristic rules.

- In the final step, the processed data from interviews and speeches are integrated into a unified corpus.

Technical Modeling

- **Zero-shot template**

Imagine you are N, you need to role-play as she/he, and your basic information is as follows: P Now you need to answer the query Q, and as the person you need to role-play, your answer is:

- **Few-shot/in-context learning template**

Imagine you are N, you need to role-play as she/he, and your basic information is as follows: P

Example: Imaging you are N', the basic information is P' The query is Q' The answer to this query is R'

Now you need to answer the query Q, and as the person you need to role-play, your answer is:

Technical Modeling

- **Discriminator**

```
Below is an instruction that describes a task, paired
with an input that provides further context. Write a
response that appropriately completes the request.
### Instruction:
Based on the input, determine whose style of speaking
this sentence is. Just give names, don't output other
information. The outputs should be in the following
format: <name>.
### Input:
S
### Response:
```

Evaluation Metrics

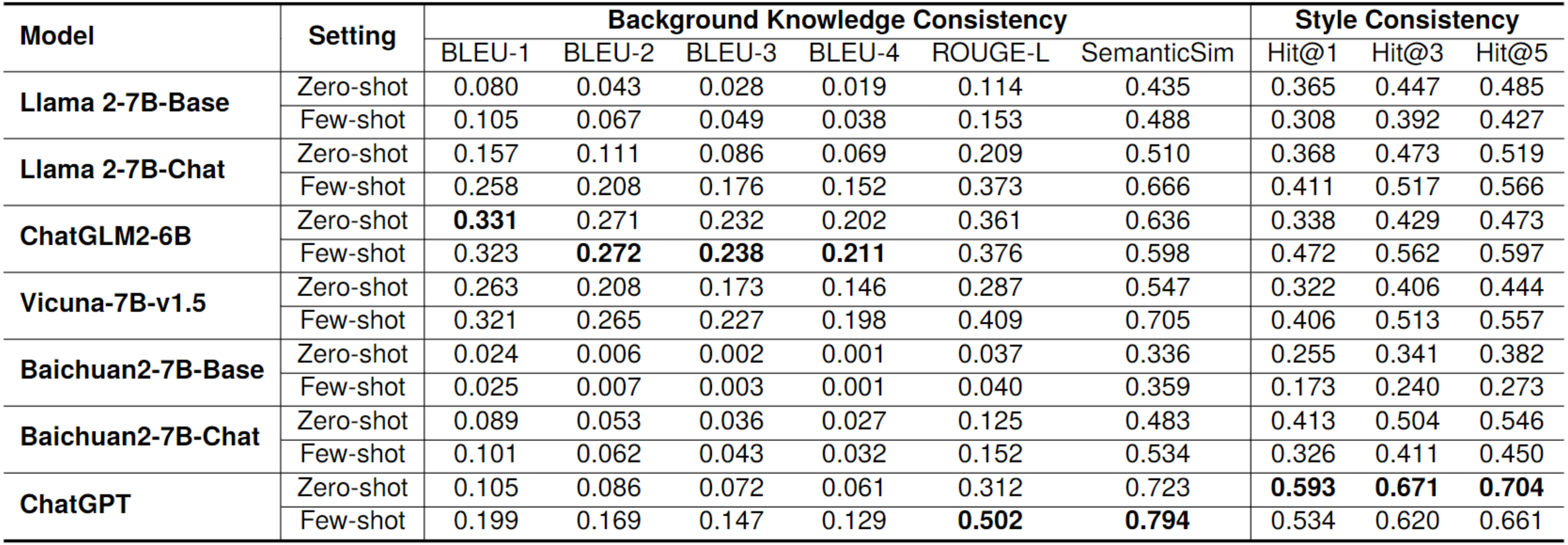
- **Background Knowledge Consistency**

1. Lexical similarity
2. Semantic similarity

- **Style Consistency**

We use the discriminator we train to distinguish the style.

Results

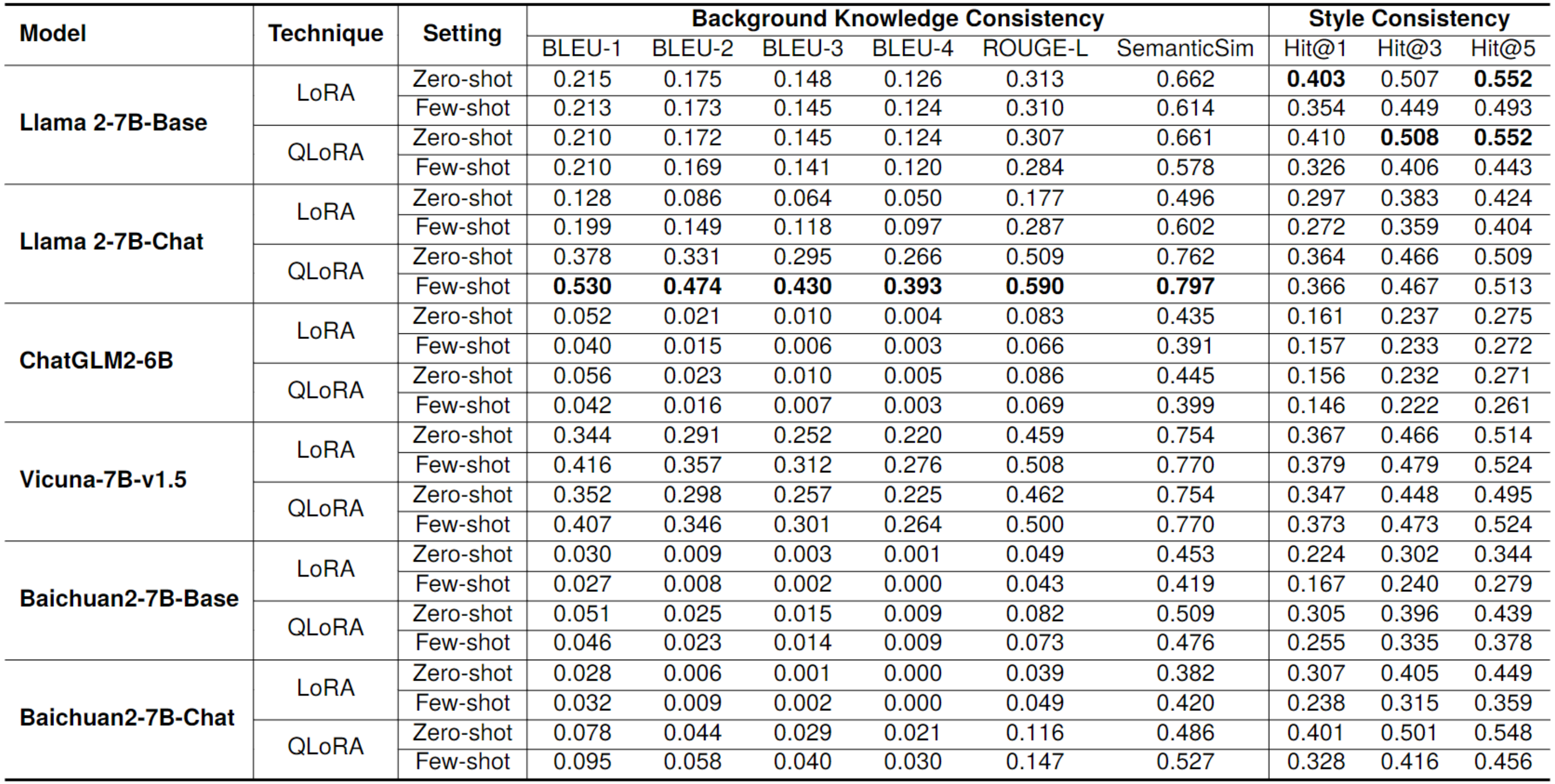


Model	Setting	Background Knowledge Consistency						Style Consistency		
		BLEU-1	BLEU-2	BLEU-3	BLEU-4	ROUGE-L	SemanticSim	Hit@1	Hit@3	Hit@5
Llama 2-7B-Base	Zero-shot	0.080	0.043	0.028	0.019	0.114	0.435	0.365	0.447	0.485
	Few-shot	0.105	0.067	0.049	0.038	0.153	0.488	0.308	0.392	0.427
Llama 2-7B-Chat	Zero-shot	0.157	0.111	0.086	0.069	0.209	0.510	0.368	0.473	0.519
	Few-shot	0.258	0.208	0.176	0.152	0.373	0.666	0.411	0.517	0.566
ChatGLM2-6B	Zero-shot	**0.331**	0.271	0.232	0.202	0.361	0.636	0.338	0.429	0.473
	Few-shot	0.323	**0.272**	**0.238**	**0.211**	0.376	0.598	0.472	0.562	0.597
Vicuna-7B-v1.5	Zero-shot	0.263	0.208	0.173	0.146	0.287	0.547	0.322	0.406	0.444
	Few-shot	0.321	0.265	0.227	0.198	0.409	0.705	0.406	0.513	0.557
Baichuan2-7B-Base	Zero-shot	0.024	0.006	0.002	0.001	0.037	0.336	0.255	0.341	0.382
	Few-shot	0.025	0.007	0.003	0.001	0.040	0.359	0.173	0.240	0.273
Baichuan2-7B-Chat	Zero-shot	0.089	0.053	0.036	0.027	0.125	0.483	0.413	0.504	0.546
	Few-shot	0.101	0.062	0.043	0.032	0.152	0.534	0.326	0.411	0.450
ChatGPT	Zero-shot	0.105	0.086	0.072	0.061	0.312	0.723	**0.593**	**0.671**	**0.704**
	Few-shot	0.199	0.169	0.147	0.129	**0.502**	**0.794**	0.534	0.620	0.661

The results of the seven models on the ***Character100*** dataset in zero-shot and few-shot settings.

*SemanticSim means semantic similarity.

Results



Model	Technique	Setting	Background Knowledge Consistency						Style Consistency		
			BLEU-1	BLEU-2	BLEU-3	BLEU-4	ROUGE-L	SemanticSim	Hit@1	Hit@3	Hit@5
Llama 2-7B-Base	LoRA	Zero-shot	0.215	0.175	0.148	0.126	0.313	0.662	**0.403**	0.507	**0.552**
		Few-shot	0.213	0.173	0.145	0.124	0.310	0.614	0.354	0.449	0.493
	QLoRA	Zero-shot	0.210	0.172	0.145	0.124	0.307	0.661	0.410	**0.508**	**0.552**
		Few-shot	0.210	0.169	0.141	0.120	0.284	0.578	0.326	0.406	0.443
Llama 2-7B-Chat	LoRA	Zero-shot	0.128	0.086	0.064	0.050	0.177	0.496	0.297	0.383	0.424
		Few-shot	0.199	0.149	0.118	0.097	0.287	0.602	0.272	0.359	0.404
	QLoRA	Zero-shot	0.378	0.331	0.295	0.266	0.509	0.762	0.364	0.466	0.509
		Few-shot	**0.530**	**0.474**	**0.430**	**0.393**	**0.590**	**0.797**	0.366	0.467	0.513
ChatGLM2-6B	LoRA	Zero-shot	0.052	0.021	0.010	0.004	0.083	0.435	0.161	0.237	0.275
		Few-shot	0.040	0.015	0.006	0.003	0.066	0.391	0.157	0.233	0.272
	QLoRA	Zero-shot	0.056	0.023	0.010	0.005	0.086	0.445	0.156	0.232	0.271
		Few-shot	0.042	0.016	0.007	0.003	0.069	0.399	0.146	0.222	0.261
Vicuna-7B-v1.5	LoRA	Zero-shot	0.344	0.291	0.252	0.220	0.459	0.754	0.367	0.466	0.514
		Few-shot	0.416	0.357	0.312	0.276	0.508	0.770	0.379	0.479	0.524
	QLoRA	Zero-shot	0.352	0.298	0.257	0.225	0.462	0.754	0.347	0.448	0.495
		Few-shot	0.407	0.346	0.301	0.264	0.500	0.770	0.373	0.473	0.524
Baichuan2-7B-Base	LoRA	Zero-shot	0.030	0.009	0.003	0.001	0.049	0.453	0.224	0.302	0.344
		Few-shot	0.027	0.008	0.002	0.000	0.043	0.419	0.167	0.240	0.279
	QLoRA	Zero-shot	0.051	0.025	0.015	0.009	0.082	0.509	0.305	0.396	0.439
		Few-shot	0.046	0.023	0.014	0.009	0.073	0.476	0.255	0.335	0.378
Baichuan2-7B-Chat	LoRA	Zero-shot	0.028	0.006	0.001	0.000	0.039	0.382	0.307	0.405	0.449
		Few-shot	0.032	0.009	0.002	0.000	0.049	0.420	0.238	0.315	0.359
	QLoRA	Zero-shot	0.078	0.044	0.029	0.021	0.116	0.486	0.401	0.501	0.548
		Few-shot	0.095	0.058	0.040	0.030	0.147	0.527	0.328	0.416	0.456

The results of the open-source models fine-tuned by two training techniques on the ***Character100*** dataset in zero-shot and few-shot settings.
*SemanticSim means semantic similarity.

Thank you for listening!