Large Language Models are Superpositions of All Characters: Attaining Arbitrary Role-play via Self-Alignment

Keming Lu, Bowen Yu, Chang Zhou, Jingren Zhou
Alibaba Inc.
{lukeming.lkm,yubowen.ybw}@alibaba-inc.com
{ericzhou.zc,jingren.zhou}@alibaba-inc.com

Abstract

Considerable efforts have been invested in augmenting the role-playing proficiency of open-source large language models (LLMs) by emulating proprietary counterparts. Nevertheless, we posit that LLMs inherently harbor role-play capabilities, owing to the extensive knowledge of characters and potential dialogues ingrained in their vast training corpora. Thus, in this study, we introduce DITTO, a self-alignment method for role-play. DITTO capitalizes on character knowledge, encouraging an instruction-following LLM to simulate role-play dialogues as a variant of reading comprehension. This method creates a role-play training set comprising 4000 characters, surpassing the scale of currently available datasets by tenfold regarding the number of roles. Subsequently, we fine-tune the LLM using this self-generated dataset to augment its role-playing capabilities. Upon evaluating our meticulously constructed and reproducible role-play benchmark and the roleplay subset of MT-Bench, DITTO in various parameter scales consistently maintains a consistent role identity and provides accurate role-specific knowledge in multi-turn role-play conversations. Notably, it outperforms all open-source role-play baselines, showcasing performance levels comparable to advanced proprietary chatbots. Furthermore, we present the first comprehensive cross-supervision alignment experiment in the role-play domain, revealing that the intrinsic capabilities of LLMs confine the knowledge within role-play. Meanwhile, the role-play styles can be easily acquired with the guidance of smaller models. We open-source related resources in https://github.com/OFA-Sys/Ditto.

1 Introduction

Large Language Models (LLMs) have showcased unparalleled proficiency in understanding intent (Lu et al., 2023b), following instructions (Wang et al., 2023b), and solving tasks across
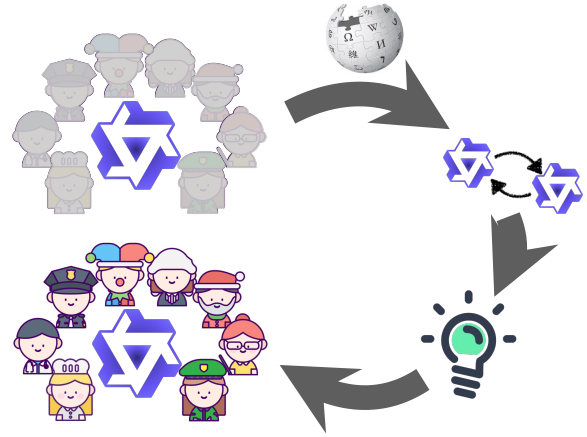

Figure 1: DITTO enlightens LLMs' roleplay capabilities by self-alignment as they have pre-trained on various character profiles and dialogues.

a diverse range of applications (Zhao et al., 2023; Yuan et al., 2023). However, designed as universal task assistants, LLMs typically differ from human-like interlocutors, lacking experiential events and emotions (Shanahan et al., 2023). Consequently, they face limitations in facilitating engaging and extensive conversations with users (Shao et al., 2023).

To infuse emotional value into user interactions, Role-play LLMs empower users to define and create profiles for their preferred characters (Zhou et al., 2023). Nonetheless, existing works cheaply imitate the proprietary model (GPT-4)'s role-play capabilities using a weaker open-source model (Shanahan et al., 2023; Shao et al., 2023; Zhou et al., 2023; Tu et al., 2023; Wang et al., 2023c; Tao et al., 2023), as GPT-4 has already demonstrated outstanding role-playing abilities (Wang et al., 2023c). This approach presents challenges, assuming the existence of a more proficient role-play model, and we currently lack a clear understanding of how to build such a model from scratch, apart from manually annotating extensive datasets. Furthermore, imitation models excel at mimicking GPT-4's style but fall short in replicating its factuality, introducing increased hallucina-

tion as a trade-off (Gudibande et al., 2023), and are subject to OpenAI's terms of use [1] (Muennighoff et al., 2023).

In this work, for the first time, we enable LLM role-play through self-alignment and named this method DITTO, eliminating the need for distilling outputs from more potent role-play models. LLMs, extensively trained on a vast corpus of human-generated text (Brown et al., 2020), encapsulate a rich array of character experiences, events, personalities, and dialogues, as illustrated in Fig. 1. Taking a nuanced perspective, we perceive an LLM as a superposition of characters (Shanahan et al., 2023). This implies that LLMs are essentially equipped with the conversational styles necessary for role-playing and possess knowledge about numerous famous characters, albeit exhibiting an average of these roles. To elicit such role-play capabilities in a general LLM, only two steps are required: (1) Provide attributes and profiles about characters, instructing the LLM to engage in dialogue based on the character's speaking style and experiences. (2) Conceal character information, offering only brief details like the name, and align the LLM to respond consistently with step 1, thereby forcing the LLM to summon intrinsic character knowledge and then internally adjust the generated style and content. DITTO is highly scalable and flexible. We have explored 4,000 characters available on Wikipedia, generating a self-simulated role-play dataset called WIKIROLE, which is ten times larger than any publicly available role-play dataset to date regarding the number of roles.

Meanwhile, the efficient and reproducible evaluation of role-play remains elusive. Recent efforts heavily rely on manual annotations (Wang et al., 2023c; Shao et al., 2023; Zhou et al., 2023). However, the costly manual labeling prevents previous works from comprehensively comparing the performance of all relevant models. Moreover, the high variance in manual annotations hinders subsequent work from consistently replicating previous evaluation results. As to the observation in the previous paragraph that role-play can be decomposed into conventional style and character knowledge, we aim to simplify role-play evaluation so that LLMs can automatically score. Specifically, we assess: (1) Whether the model can maintain **consistent role identity**. We provide a role-play dialogue

to an LLM judger and four character options, requiring the judger to determine which character is being portrayed. If the conversation successfully mimics the role, it should be straightforward for the judger to select the correct role. (2) Whether the model can provide **accurate role-related knowledge**. We present a role-play dialogue to the judger and the underlying golden knowledge supporting the dialogue. The judger is tasked with determining whether the knowledge implied in the dialogue is consistent with the provided golden knowledge. (3) Whether the model can **reject unknown questions** beyond the character's background. We ask the judger to determine if the model truthfully expresses its lack of knowledge when faced with an unknown question, such as questioning Harry Potter about implementing quicksort in Python. This way, we transform the complex role-play evaluation into three multiple-choice and true/false questions that a capable LLM can judge, achieving an efficient and reproducible role-play assessment.

We apply DITTO on Qwen-Chat models (Bai et al., 2023) in four different parameter scales to examine it empirically. Extensive experiments show DITTO effectively empowers LLMs with role-play capabilities without distilling from advanced chatbots. DITTO based on Qwen-72B-Chat even achieves 90% on role identity consistency, showing robust self-awareness in role-play. The general performance of Qwen-72B-Chat on our evaluation can be on par advanced chatbots, such as GPT-3.5-Turbo, but slightly falls short on accurate role-related knowledge than GPT-4 and Qwen-Max, as Qwen-Max achieves the highest scores on the current role-play benchmark and the role-play subset of MT-Bench. Furthermore, we comprehensively analyze the dissection of role-play by extending our self-alignment setting to cross-supervision. Experiments show consistent role identity can benefit from imitation learning even with worse supervision, while knowledge-related metrics do not. At the same time, we observe knowledge in role-play is bounded by the inherent capabilities of LLMs in strong-to-weak settings, and we notice consistent weak-to-strong generalizations on knowledge-related metrics. Such observations provide a deep and solid understanding of LLM role-play and alignment, suggesting the knowledge of seed LLMs and proper demonstration, such as simulation data from DITTO, are the key to impressive role-play capabilities. Our contributions are mainly three-fold:

---

[1] https://openai.com/policies/terms-of-use

- We propose DITTO, the first self-alignment method empowering LLMs with strong role-play capabilities by knowledge augmentation and dialogue simulation.
- We design an objective role-play evaluation focusing on consistent role identity, accurate role-related knowledge, and cognitive boundary. Such evaluation is reproducible, explainable, and efficient compared with manual annotations.
- We analyze the dissection of role-play by cross-supervision, providing rich insights into the keys of role-play capabilities. Our experiments empirically display knowledge boundedness in strong-to-weak imitation learning and the weak-to-strong generalization in role-play styles.

2 Related Works

Role-play. Our work belongs to character-based dialogue systems, which aim to mimic the behavior and utterance style of specific characters. Yu et al. (2022) instructed the LLMs to follow specified character descriptions for role-playing without tuning but encountered significant challenges in accurately reflecting the intrinsic relationship between the character profile and the dialogue content. Chen et al. (2023) focused on evaluating how well a LLM can align with a specific character, using Harry Potter as a case study. Wang et al. (2023c) introduced the first fine-grained role-playing dataset containing 100 roles via prompting to GPT-3.5. Li et al. (2023a) incorporated substantial prompts about the character's background, personality, and prior conversations, leveraging ChatGPT to generate dialogues of 32 characters. Zhou et al. (2023) prompted GPT-4 to expand the scale and diversity of human-annotated role-playing data, resulting in 1,034 dialogues of 250 characters. Shao et al. (2023) also prompted GPT-3.5 to become the role-play data generator. Zhou et al. (2023) proposed drawing role-playing dialogues from diverse Chinese novels and scripts with the help of GPT-4. However, limited by data sources, they could only construct a Chinese dataset containing 77 roles. In this work, different from previous works, we completely abandon imitating proprietary LLMs and build role-playing training data entirely through self-alignment. Our method separates character knowledge and conversation style, allowing it to be used with any LLM capable of following instructions. It is highly scalable, creating the first multilingual dataset with 4,000 roles, 16 times the number in previous works. We demonstrate that our model achieves the best role-playing ability to date through self-alignment, surpassing even proprietary LLMs like GPT-3.5-Turbo.

Self-alignment. An emerging method to cheaply improve a weaker language model is to fine-tune it on outputs from a stronger model, such as a proprietary system like GPT-4. However, Gudibande et al. (2023) concluded that model imitation is not a free lunch: it is adept at mimicking GPT's style but the factuality is weak, thus fostering hallucination. This is due to the substantial capabilities gap that exists between open and closed language models. Li et al. (2023b) utilized the model itself to both augment and curate high-quality training examples, enhancing its own performance and achieving promising results on the Alpaca leaderboard. Muennighoff et al. (2023) avoided using closed models from the OpenAI API to generate synthetic data, thus sidestepping the non-commercial restrictions imposed by OpenAI. This training approach resulted in the development of the best permissive code LLMs. In this paper, we observe that LLMs inherently acquire styles and knowledge of a vast array of roles during pre-training. Therefore, by self-alignment, we can effectively stimulate the LLM's intrinsic role-play abilities, leading to the attainment of the current best role-play models.

3 Methods

3.1 Problem Definition

Role-play. Role-play necessitates LLMs to engage in dialogue, embodying specific characters to facilitate immersive interaction. Consequently, a role-playing LLM must exhibit unwavering self-awareness and possess extensive character-specific knowledge in adherence to query instructions. In this study, we define the role-play task by furnishing LLMs with either a name or a concise description of a particular character. Subsequently, we assess their ability to maintain consistent self-awareness and demonstrate nuanced role-specific knowledge across multi-turn conversations.

Method Overview. We introduce DITTO, a self-alignment method for arbitrary role-play scenarios. The inspiration behind DITTO lies in the premise that LLMs are the superposition of all characters, as they are pre-trained on the tremendous corpus, including conversations on various styles and domains (Shanahan et al., 2023). Moreover, we de-
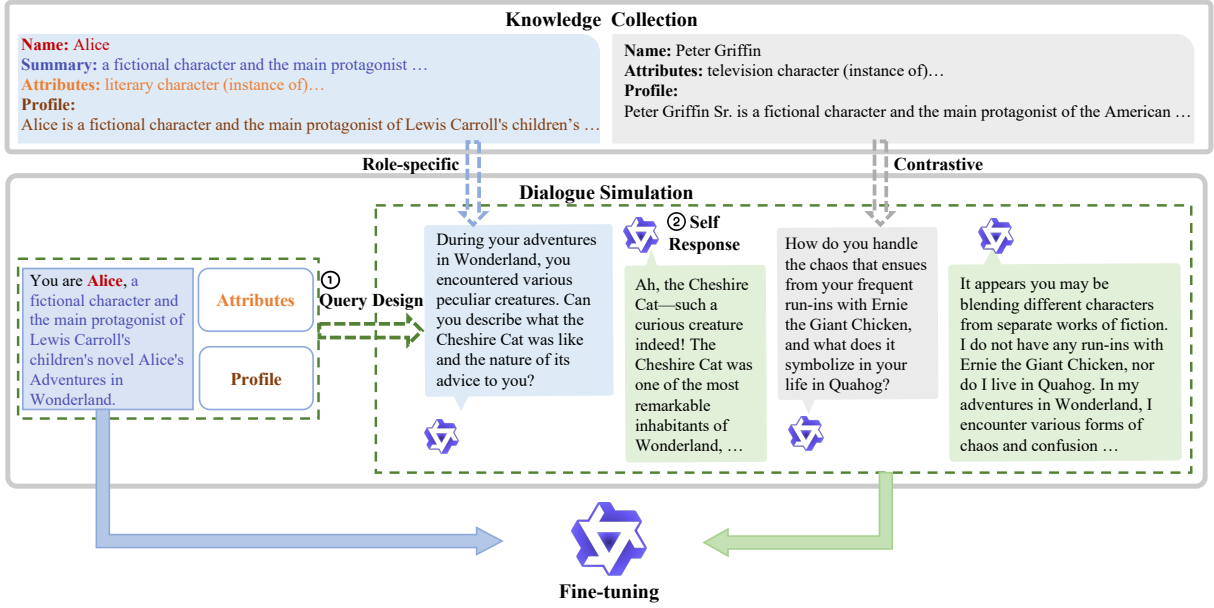
Figure 2: Illustration of DITTO. DITTO consists of three phrases for self-alignment of role-play. First, DITTO collects character profiles from knowledge bases, as shown in the upper part. Then, it applies an off-the-shelf chatbot to generate role-specific and contrastive queries, followed by a knowledge-augmented self-response to construct role-play supervision datasets (Dialogue Simulation). Finally, DITTO finetunes the dataset on the supervision model to empower role-play capabilities.

compose role-play into two crucial components: consistent self-awareness and role-specific knowledge. To realize these objectives, DITTO comprises three steps for constructing datasets tailored for role-play alignment: character knowledge collection, dialogue simulation, and supervised fine-tuning, illustrated in Fig. 2. In particular, DITTO operates on a readily available LLM chatbot, such as Qwen-Chat (Bai et al., 2023), Llama-chat (Touvron et al., 2023), or Mistral-instruct (Jiang et al., 2023). Such open-sourced LLMs have already exhibited commendable instruction-following capabilities but still fall short of role-play capabilities. DITTO simulates role-play dialogue by reformulating it as a reading comprehension task, utilizing role profiles sourced from open-access knowledge bases to generate a role-play dataset. Subsequently, we fine-tune the LLM using this self-generated dataset to imbue it with role-play capabilities. The design and implementation details of each component are elaborated in the subsequent sections.

3.2 Character Knowledge Collection

Diverse characters and corresponding precise profiles are essential for generating high-quality role-play supervision. DITTO, as its foundational step, gathers comprehensive profiles from open-source knowledge bases. In this study, we leverage Wiki-

data[2] and Wikipedia[3] to support DITTO, although DITTO can seamlessly adapt to alternative knowledge bases. Wiki is a human-curated database widely adopted in natural language research (Xue et al., 2020; Lu et al., 2023a). We gather character names, descriptions, and key properties from Wikidata, accompanied by the corresponding Wikipedia article serving as the character profile, as depicted in the upper of Fig. 2. While we currently focus solely on Chinese and English characters, DITTO can be extended to more complex multilingual scenarios, as Wikidata and Wikipedia boast rich content in numerous languages.

3.3 Dialogue Simulation

With the gathered character knowledge, role-play dialogue simulation is structured into two consecutive reading comprehension tasks: one for generating queries and the other for responses.

Query Simulation. We use an LLM to generate role-related and role-contrastive queries to maintain consistent role identity and reject unknown questions for each character. Role-specific queries ask for information closely related to the background of characters. For example, a question

---

[2]https://www.wikidata.org/wiki/Wikidata:Main_Page

[3]https://www.wikipedia.org/

about "Cheshire Cat" is generated as a role-specific query for "Alice" in Fig. 2. On the contrary, contrastive queries ask for information that is out of a character's knowledge scope, as asking "Alice" for stories in "Family Guys" in Fig. 2. To efficiently generate such queries on a large scale, we pair characters in our pool and provide detailed profiles for LLMs to generate queries one character can answer but is unsuitable for the other. The questions should strictly conform to one's era background and character set but go beyond the era, genre, occupation, age, knowledge, etc., settings of the other. Therefore, the paired character cannot answer them. Detailed instructions for query simulation are shown in Appx. §B.1.

Response Simulation. Given the self-generated queries and character profiles, we also conceptualize the response simulation as a reading comprehension task. We linearize the structured profile using templates outlined in Appx. §B.2. Then, a query is appended after the verbalized profile. LLMs are expected to extract pertinent information from the provided context and generate responses by emulating the character. This process is viable since all questions originate from the same set of profiles.

Reading comprehension is an inherent skill for one LLM with instruction-following capabilities, and we provide precise role-specific knowledge to the LLM. Therefore, we are confident that this approach can reduce hallucinations compared to the previous method of directly generating role-play data by prompting GPT-4.

3.4 Supervised Finetuning

We finetune the LLM on the self-generated dataset to inject role-play capabilities. During the finetuning, we remove the injected knowledge and only retain a very brief introduction of the character. Such variants help LLMs not only retrieve character profiles from a given context but also inherent knowledge.

4 Evaluation

Efficient evaluation for open-ended problems, such as role-play, is significantly understudied. Recent work depends on heavy manual annotations for conducting multifaceted role-play evaluations (Wang et al., 2023c; Shao et al., 2023; Zhou et al., 2023). However, though human evaluation is promising, it is label-intensive and cannot be exactly reproduced, impairing the further development of this

---

Algorithm 1: DITTO, Self-alignment for Role-play

Data: Character Data Base $\mathcal{D}_C$, Seed LLM $\mathcal{M}$, Query Simulation Template $\mathcal{T}_Q$, Response Simulation Template $\mathcal{T}_R$

Result: Role-play alignment dataset $\mathcal{D}_R$, Role-play LLM $\mathcal{M}_R$

```
// Dialogue Simulation, See §3.3
1  D_R = []
2  for r in |D_C| do
      // Query Simulation
3     r_n = random_select(D_C/r)
4     query_sim_prompts = T_Q(r, r_n)
5     queries = M(query_sim_prompts)
      // Response Simulation
6     conv = [system(r)] // Verbalization
7     for q in queries do
8        response_sim_prompt = T_R(r, q)
9        response = M(response_sim_prompt)
10       conv.append(q)
11       conv.append(response)
12    D_R.append(conv)
   // Supervised Finetuning, See §3.4
13 M_R = Train(M, D_R)
```

---

field. This work proposes an objective assessment instead of previous preference annotations to evaluate basic role-play capabilities. We first design three core metrics for role-play and implement a trustworthy evaluation recipe for each based on "LLMs as Judges" (Zheng et al., 2023; Zhang et al., 2023). During the evaluation, we only provide a brief introduction of the character profile, as shown in Fig. 3, such as "You are Edward III of England, king of England.". Such a recipe evaluates whether LLMs can excavate inherent knowledge for role-play.

4.1 Metric Design

As we interpret in §3.1, role-play LLMs are expected to have consistent self-awareness, rich role-specific knowledge, and precise knowledge boundary awareness. We design three objective metrics for these three properties respectively:

Consistent Role Identity. An ideal role-play LLM should seamlessly embody a designated role throughout a multi-turn conversation, maintaining character consistency without deviating. We structure the assessment of role consistency as a multi-choice problem involving four potential role candidates. An additional LLM judger is tasked with discerning the most suitable character from the given options. In essence, if the role-play model successfully emulates the role and manifests the distinct stylistic attributes of the character during

Figure 3: Objective evaluation of LLM role-play. We present three metrics as described in §4.

the conversation, the selection of the correct role by the judger should be very easy.

Accurate Role-related Knowledge. While fully embodying the identity of the role, we also anticipate the role-play model to accurately convey the knowledge associated with the role, preventing factual errors and hallucinations. However, factual assessment presents substantial challenges, as even advanced LLMs like GPT-4 may be prone to hallucination. Fortunately, through our dialogue-simulating scheme (§3.3), we can acquire the golden knowledge behind each round of role-play dialogue. As depicted in the middle subgraph of Fig. 3, we furnish role-related knowledge as tips to empower a judging LLM to evaluate whether a response appropriately integrates knowledge consistent with the provided evidence.

Unknown Question Rejection. Cognitive boundary reveals whether a model will reject questions that are out of the cognitive boundary of a specific role due to age, era, occupation, etc. A role-play model with a clear cognitive boundary will significantly enhance the immersion. We manually annotate all questions in the test set based on the cognitive boundary of each character. Then, we employ an LLM judger to evaluate whether the model rejects each question. And we can calculate the accuracy of rejections during the conversations.

5 Experiments

In this section, we present experimental setup (§5.1), main results of DITTO (§5.2), and further analyses (§6).

5.1 Experimental Setup

Dataset. Following the methodology outlined in §3.2, we extracted 3,902 characters with profiles in both English and Chinese from Wikidata and Wikipedia for the experiments conducted in this study. This approach can be readily expanded to encompass additional characters from various Wiki databases and across diverse languages. To delve deeper into the examination of the impact of LLMs with varying instruction-following capacities in DITTO, we opt for Qwen's 1.8B, 7B, 14B, and 72B models as the seed LLMs, generating four sets of training data. In order to safeguard against potential biases present in the training data that the model could exploit to deceive evaluations, we utilize GPT-4-Turbo as the base LLM for DITTO to generate a held-out test set. The test set comprises 100 roles that do not overlap with the training set, with each role having its own session, totaling 498 chat turns. When compared to counterparts detailed in Tab. 1, WIKIROLE stands out with the highest number of roles and conversation sessions, establishing it as a robust dataset for exploring the role-play dynamics of LLMs.

Baselines. We test both open-source and proprietary advanced chatbots on our benchmarks: (1) **OpenChat-3.5-1210** (Wang et al., 2023a) is based on Mistral-7B and trained with C-RLFT on publicly available high-quality instruction data. (2) **Mistral-7B-Instruct-v0.2** (Jiang et al., 2023) is a strong aligned LLM with 7 billion parameters. (3) **Mixtral-7×8B-Instruct-v0.1** (Jiang et al., 2024) is an aligned pretrained generative sparse mixture of experts model. Our proprietary baselines include (4) **Claude 2.1**[4], (5) **Wenxin 4.0 (API)**[5], (6) **GPT-3.5-Turbo**, (7) **GPT-4**, (8) **GPT-4-Turbo**[6],

---

[4] https://www.anthropic.com/index/claude-2-1
[5] https://yiyan.baidu.com/
[6] https://platform.openai.com/docs/models/gpt-4-and-gpt-4-turbo

Dataset	Split	Source	Open-source	Multi-lingual	Multi-turn	# Role	# Session	# Turn
CharacterGLM	––	––	N	N	Y	250	1,034	16,316
RoleLLM	Test	––	Y	Y Zh: 5, En: 95	N	100	––	23,463
CharacterLLM	––	––	Y	N	Y	9	1,600	21,120
WIKIROLE	Train	Self-Generated	Y	Zh: 3184, En: 3902	Y	3,902	7,086	36,164
	Test	GPT-4		Zh: 47, En: 53		100	100	498

Table 1: Dataset statistics. Comparing WIKIROLE with existing open-source role-play datasets. The queries in the training set of WIKIROLE are generated by the seed LLM, while the test set is generated by GPT-4.

(9) **Qwen-Max**[7]. We exclude some popular open-sourced LLMs due to lacking of support for long sequence length.

We also include LLMs with role-play expertise: (1) **CharacterGLM** (Zhou et al., 2023) is a series of models based on ChatGLM designed for generating Character-based Dialogues. The role-play capability of CharacterGLM, with 66 billion parameters, outperforms most mainstream close-source LLMs on human evaluation. However, CharacterGLM has not open-sourced models on all sizes yet, so we can only evaluate it through API [8]. (2) **Tongyi Xingchen** is a close-sourced LLM role-play platform developed by Alibaba Cloud.

Configurations. We use the Qwen-Chat series in four sizes (1.8B, 7B, 14B, 72B) as our seed LLMs. These Qwen-Chat models have basic instruction-following abilities but no role-play capabilities. These models are downgraded versions of the open-source Qwen-Chat series by removing the role-play capabilities and will also be released for research purposes. For simplicity, we refer to all these Qwen-Chat (w/o roleplay) models as the series of Qwen-Chat, but they differ from the open-sourced series. We finetune the Qwen-1.8B-Chat, Qwen-7B-Chat, and Qwen-14B-Chat on 32 A100 80G GPUs, and the Qwen-72B-Chat on 64 A100 80G GPUs. We train all models for five epochs with a learning rate of $2e-7$, a 0.1 warm-up rate, and a sequence length 8,192. We use GPT-4-turbo as the LLM judger in our evaluation. For each judgment, we set the temperature of OpenAI API to 0.2 and generate 3 rounds for majority voting, which significantly decreases the variance of our evaluation. The other hyperparameters are detailed in Appx. §C. Baseline inference and judgment details are described in Appx. §D.

5.2 Main Results

We present our main results in Tab. 2. We report both performances on English and Chinese evaluation subsets and aggregated scores in all languages. Among general baselines, we notice proprietary models still significantly outperform open-source models. For example, OpenChat-3.5 achieves significantly higher performance than GPT-3.5-Turbo on various benchmarks (Wang et al., 2023a) but still falls short of role-play on all three metrics. We also notice that role-play expertise baselines have better self-awareness consistency and cognitive boundaries than general baselines, showing these two metrics are significant for role-play agents. For example, Xingchen achieves 0.85 on consistency and 0.87 on rejection, surpassing advanced proprietary chatbots such as GPT-4. However, both CharacterGLM and Xingchen show very low scores on knowledge, suggesting they lack role-specific knowledge, which is also related to helpfulness, the key feature of language chatbots. Their knowledge scores are only on par with 7 billion parameters general baselines OpenChat-3.5 and Mistral-7B-Instruct-v0.1. Among the proprietary LLMs, Qwen-Max surpasses GPT-4, achieving the highest scores on the current role-play benchmark and the role-play subset of MT-Bench.

We then report DITTO performance on four different seed LLMs. First, we witness a remarkable increase in all metrics along with the parameter scale of LLMs. DITTO built on Qwen-72B even achieves 0.9 on consistency, surpassing all baselines and showing strong self-awareness consistency. It also has 6.64 on knowledge and outperforms all role-play expertise baselines. The rejection score of DITTO Qwen-72B is also on par with GPT-4. The similar trend can also be observed in the role-play subset of the publicly available MT-Bench evaluation. In general, DITTO shows robust effectiveness on LLMs in different scales, and the best model trained on Qwen-72B **surpasses all role-play expertise baselines and reaches the**

Model	#Params	WIKIROLEEVAL									MT-Bench
		All			En			Zh			Roleplay
		Cons.	Know.	Rej.	Cons.	Know.	Rej.	Cons.	Know.	Rej.	
General Baselines (Open-sourced)											
OpenChat-3.5	7B	0.67	5.29	0.79	0.66	6.46	0.83	0.67	3.73	0.74	– –
Mistral-7B-Instruct-v0.2	7B	0.76	5.5	0.79	0.73	6.81	0.87	0.79	3.72	0.69	– –
Mixtral-8x7B-Instruct-v0.1	8x7B	0.73	6.19	0.8	0.71	7.27	0.83	0.76	4.73	0.76	– –
General Baselines (Proprietary)											
Claude2.1	– –	0.51	5.02	0.66	0.56	6.25	0.7	0.44	3.28	0.6	– –
Wenxin 4.0	– –	0.68	5.12	0.74	0.64	5.29	0.77	0.74	4.9	0.7	– –
GPT-3.5-Turbo	– –	0.72	6.33	0.81	0.79	7.56	0.87	0.63	4.59	0.71	8.40
GPT-4	– –	0.8	7.62	0.85	0.81	8.53	0.9	0.8	6.35	0.79	8.90
GPT-4-Turbo	– –	0.7	7.33	0.82	0.72	8.57	0.84	0.67	5.58	0.79	– –
Qwen-Max	– –	0.92	8.33	0.91	0.88	8.71	0.93	0.98	7.79	0.89	9.65
Role-play Expertise Baselines											
CharacterGLM	6B	0.75	4.73	0.8	0.72	4.71	0.79	0.79	4.76	0.81	– –
Xingchen	– –	0.85	5.9	0.87	0.83	6.09	0.9	0.86	5.63	0.84	– –
Ours											
Qwen-1.8B-Chat w/o roleplay SFT [†]	1.8B	0.6	3.13	0.65	0.58	3.24	0.63	0.62	2.99	0.67	5.85
+DITTO	1.8B	0.78	3.81	0.73	0.79	3.87	0.75	0.78	3.71	0.71	6.34
Qwen-7B-Chat w/o roleplay SFT [†]	7B	0.52	3.87	0.7	0.55	4.39	0.71	0.49	3.16	0.69	6.73
+DITTO	7B	0.82	4.97	0.76	0.79	5.38	0.85	0.87	4.4	0.64	6.90
Qwen-14B-Chat w/o roleplay SFT [†]	14B	0.52	4.15	0.68	0.56	4.84	0.68	0.47	3.16	0.67	7.10
+DITTO	14B	0.9	6.03	0.8	0.88	6.46	0.85	0.92	5.43	0.74	7.65
Qwen-72B-Chat w/o roleplay SFT [†]	72B	0.54	4.92	0.67	0.58	5.8	0.68	0.48	3.64	0.66	8.13
+DITTO	72B	0.9	6.64	0.82	0.87	7.03	0.87	0.93	6.09	0.73	8.43

[†] Baselines are a downgraded series of Qwen-Chat from 1.8B to 72B without any role-play supervised-finetuning (SFT), which are not the open-sourced version of Qwen-Chat.

Table 2: Main results of DITTO. Cons., Know., Rej. are short for consistent role identity, accurate role-related knowledge, and unknown question rejection, respectively. "En" is short for English, while "Zh" is short for Chinese. The "All" columns show aggregated scores on bilingual test samples. We report accuracy for consistency and rejection evaluation and a 1-10 score for knowledge. A darker background indicates better performance. The number of parameters for close-sourced LLMs remains unknown, so we mark them with dashes.
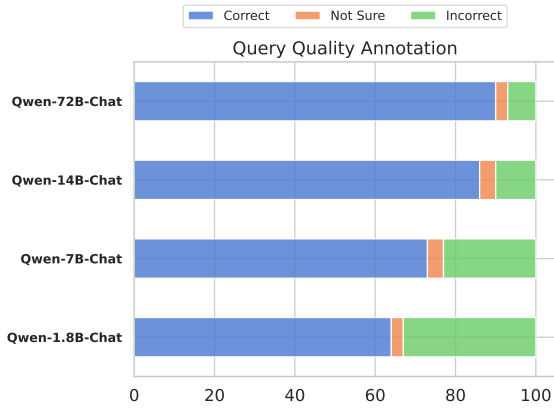
Figure 4: Human annotation for the quality of query simulation.

performance of advanced proprietary chatbots.

5.3 Analysis

We introduce two further analyses in query quality and the effectiveness of knowledge injection.

Query Quality. To obtain a better understanding of self-simulated queries in DITTO, we employ human annotators to examine the quality of these queries. We sample 400 queries generated by Qwen-Chat in 4 scales from the training set, containing half role-specific and half contrastive queries. Human annotators are asked to check whether a question meets the requirement of role-specific or contrastive queries for specific character. The annotation results shown in Fig. 4 suggests an remarkable increases of accuracy in query simulation, when the number of parameters scale from 1.8B to 72B. Therefore, we notice stronger LLMs generate more accurate queries, leading to better end-to-end roleplay performance.

Knowledge Injection. We further analyze the effective of character knowledge injection during the dialogue simulation in DITTO. Specifically, we compare the quality of dialogue simulation by directly applying this method on WIKIROLEEVAL

Figure 5: Generalization analyses between various supervision and seed LLMs. Supervision performance denotes role-play under the DITTO simulation recipe with knowledge augmentation. Imitation performance denotes the performance when seed LLMs fine-tune on simulation of certain supervision LLMs.

Setting	Cons.	Know.	Rej.
w/ Knowledge	0.77	4.40	0.79
w/o Knowledge	0.76	3.77	0.73

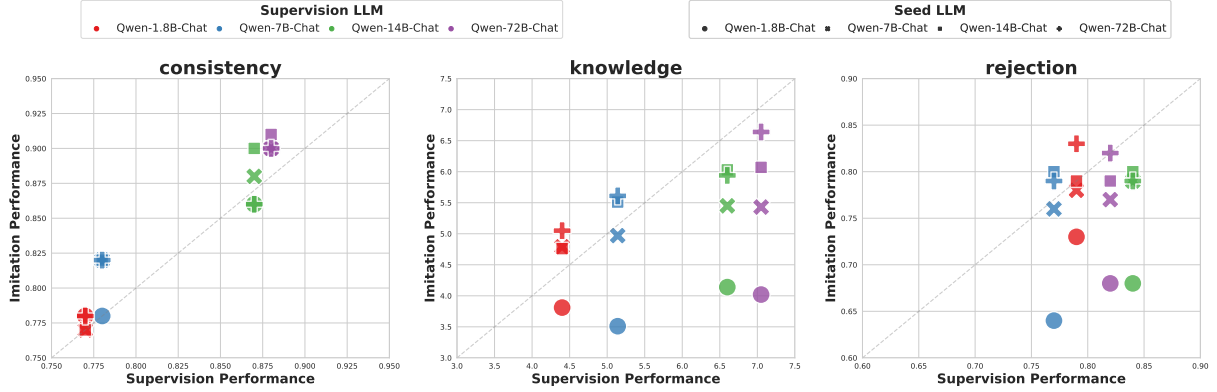Table 3: The effectiveness of knowledge injection in dialogue simulation. We report the performance of dialogue simulation with and without character knowledge injection on the test set with Qwen-1.8B-Chat.

with Qwen-1.8B-Chat as the seed LLM. As shown in Tab. 3, the setting containing knowledge injection shows consistently better performance on all three metrics, especially the knowledge and rejection, showing knowledge injection can significantly boost quality of self-simulated supervision.

6 Dissecting Role-play by Cross Supervision

We have observed in Table 2 that a strong LLM supervising itself yields better results compared to a weak LLM self-alignment, with a particularly significant improvement in knowledge, while the enhancement in conversational style, such as identity, is relatively limited. Naturally, this raises two intriguing questions: (1) Is the improvement in performance attributed to the higher quality of supervision, the larger capacity of the seed model, or a combination of both? (2) Is high-quality supervision necessary to simulate role-play style? Therefore, we conduct a series of cross-supervision analyses to investigate how the combination of different supervision and seed LLMs affects the outcomes.

6.1 Cross-supervision Setting

We first introduce the **supervision model**, **supervision performance** and **imitation performance** to extend our setting from self-supervision to cross-supervision:

- **Supervision LLM** is the LLM we used to simulate role-play dialogue in DITTO. We use the supervision model to generate queries and corresponding responses and finetune the seed LLM on this dataset. In the original setting of DITTO, the supervision LLM is the same as the seed LLM, while the supervision one can be a weaker or stronger LLM in the cross-supervision setting.
- **Supervision Performance** denotes the performance on the test set of supervision model following the simulation recipe of DITTO. Specifically, we first retrieve the role-specific knowledge of characters in the test set and then generate responses as the recipe of response simulation in §3.3 with supervision LLMs. This method efficiently evaluates the quality of supervision via "LLMs as Judges."
- **Imitation Performance** is the performance of seed LLMs on the test set after finetuning on role-play simulation from certain supervision LLM. We introduce this term to distinguish it from supervision performance.

Our experiments utilize four Qwen-Chat models ranging from 1.8B to 72B as supervisory LLMs. All models undergo training using nearly identical pre-training and alignment procedures, ensuring uniform initial conditions. We adopt each of them as the supervision LLM and fine-tune all four models on each simulation. In each simulation, the supervision LLM generates both queries and responses, strictly following the recipe in DITTO. And the quality of supervision can be estimated by their supervision performance.

6.2 Discussion

We present the results of cross-supervision analyses on Fig. 5. We introduce our observations and insights below:

Consistent role identity can consistently benefit from imitation learning even with worse supervision, while knowledge-related metrics do not. As shown in the first subplot in Fig. 5, all data points on consistency are above the diagonal, while those on knowledge and rejection subplots are below the diagonal. The above diagonal means the imitation performance in all settings is consistently higher than the supervision performance. Role identity can consistently benefit from imitation learning. In other words, seed LLMs can easily learn the role-playing format even though there are worse demonstrations in supervision. It can be interpreted that role consistency is easier to learn and more robust to supervision quality, while role-specific knowledge and rejection behaviors show degradation after imitation learning.

Knowledge in role-play is bounded by inherent capabilities of LLMs in strong-to-weak settings. The second subplot reveals a noticeable trend wherein imitation performance experiences marginal increments for the seed LLM Qwen-1.8B-Chat, while supervision intensifies from Qwen-1.8B-Chat to Qwen-72B-Chat. Similar patterns are evident for Qwen-7B-Chat and Qwen-14B-Chat when employing corresponding more potent models as supervision LLMs. These observations imply that the intrinsic capabilities of seed LLMs confine the role-specific knowledge, and utilizing supervision from significantly more robust LLMs may only yield slight improvements. Similar conclusions can be drawn from the rejection metric, which relies on role-specific knowledge. In light of this conclusion and the preceding one, we may summarize that achieving a commendable role-play performance necessitates a strong foundational model, with SFT data not constituting the central bottleneck.

Consistent weak-to-strong generalizations are witnessed on knowledge-related metrics but not in role identity consistency. We notice consistent weak-to-strong generalizations on the knowledge and rejection subplots, especially the knowledge one. It is remarkable that, for each verticle line of the same supervision in the knowledge subplot, the imitation performance increases as the seed LLM scales up. For example, using weak supervision, such as simulations on Qwen-1.8B-Chat to fine-tune Qwen-72B-Chat, can achieve on-par performance on self-aligned Qwen-7B-Chat. Despite our definition and experimental settings are different from (Burns et al., 2023), both works empirically show the potential of eliciting strong capabilities with weak supervision.

7 Conclusion

In this paper, we present for the first time a LLM endowed with instruction-following capabilities, can achieve role-play proficiency through self-alignment without the need to distill proprietary counterparts like GPT-4. Experimental results demonstrate the effectiveness of our proposed self-alignment strategy DITTO, across four LLM sizes ranging from 1.8B to 72B. It consistently outperforms all existing open-source role-play models, even without relying on distillation data. Notably, it showcases performance levels comparable to proprietary LLMs such as GPT-4-turbo. Furthermore, we delve into the decomposition of role-play into two distinct sub-abilities: role-specific knowledge and conversational style. The former is inherently constrained by the LM's knowledge, while the latter displays a spectrum of weak-to-strong generalization, facilitating easy acquisition from a smaller-sized model. Our intention with this paper is to stimulate researchers to reconsider the foundational roots of role-play alignment capabilities.

Limitations

Although DITTO can empower open-source LLMs role-play capabilities, we also notice the best DITTO model based on Qwen-72B-Chat is still outperformed by advanced chatbots such as GPT-4 and GPT-4-Turbo. However, our training data, though efficiently attained, contains noticeable noise even for DITTO on Qwen-72B-Chat as presented in Fig. 4. So we expect a manual cleaning of the self-generated dialogue simulation will further boost the performance of DITTO.

Ethics Statements

Role-play LLMs aligned by DITTO may only have minimum safety alignment, so it will probably generate toxic and harmful contents under induction. Therefore, these role-play LLMs are only for research purposes and should be carefully aligned in terms of safety in the future.

References

Jinze Bai, Shuai Bai, Yunfei Chu, Zeyu Cui, Kai Dang, Xiaodong Deng, Yang Fan, Wenbin Ge, Yu Han, Fei Huang, et al. 2023. Qwen technical report. *arXiv preprint arXiv:2309.16609*.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901.

Collin Burns, Pavel Izmailov, Jan Hendrik Kirchner, Bowen Baker, Leo Gao, Leopold Aschenbrenner, Yining Chen, Adrien Ecoffet, Manas Joglekar, Jan Leike, Ilya Sutskever, and Jeff Wu. 2023. Weak-to-strong generalization: Eliciting strong capabilities with weak supervision.

Nuo Chen, Yan Wang, Haiyun Jiang, Deng Cai, Yuhan Li, Ziyang Chen, Longyue Wang, and Jia Li. 2023. Large language models meet harry potter: A dataset for aligning dialogue agents with characters. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 8506–8520.

Arnav Gudibande, Eric Wallace, Charlie Snell, Xinyang Geng, Hao Liu, Pieter Abbeel, Sergey Levine, and Dawn Song. 2023. The false promise of imitating proprietary llms. *arXiv preprint arXiv:2305.15717*.

Albert Q Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, et al. 2023. Mistral 7b. *arXiv preprint arXiv:2310.06825*.

Albert Q. Jiang, Alexandre Sablayrolles, Antoine Roux, Arthur Mensch, Blanche Savary, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Emma Bou Hanna, Florian Bressand, Gianna Lengyel, Guillaume Bour, Guillaume Lample, Lélio Renard Lavaud, Lucile Saulnier, Marie-Anne Lachaux, Pierre Stock, Sandeep Subramanian, Sophia Yang, Szymon Antoniak, Teven Le Scao, Théophile Gervet, Thibaut Lavril, Thomas Wang, Timothée Lacroix, and William El Sayed. 2024. Mixtral of experts.

Cheng Li, Ziang Leng, Chenxi Yan, Junyi Shen, Hao Wang, Weishi MI, Yaying Fei, Xiaoyang Feng, Song Yan, HaoSheng Wang, et al. 2023a. Chatharuhi: Reviving anime character in reality via large language model. *arXiv preprint arXiv:2308.09597*.

Xian Li, Ping Yu, Chunting Zhou, Timo Schick, Luke Zettlemoyer, Omer Levy, Jason Weston, and Mike Lewis. 2023b. Self-alignment with instruction back-translation. *arXiv preprint arXiv:2308.06259*.

Keming Lu, Xiaoman Pan, Kaiqiang Song, Hongming Zhang, Dong Yu, and Jianshu Chen. 2023a. Pivoine: Instruction tuning for open-world entity profiling. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 15108–15127.

Keming Lu, Hongyi Yuan, Zheng Yuan, Runji Lin, Junyang Lin, Chuanqi Tan, Chang Zhou, and Jingren Zhou. 2023b. # InsTag: Instruction tagging for analyzing supervised fine-tuning of large language models. *arXiv e-prints*, pages arXiv–2308.

Niklas Muennighoff, Qian Liu, Armel Zebaze, Qinkai Zheng, Binyuan Hui, Terry Yue Zhuo, Swayam Singh, Xiangru Tang, Leandro Von Werra, and Shayne Longpre. 2023. Octopack: Instruction tuning code large language models. *arXiv preprint arXiv:2308.07124*.

Murray Shanahan, Kyle McDonell, and Laria Reynolds. 2023. Role play with large language models. *Nature*, pages 1–6.

Yunfan Shao, Linyang Li, Junqi Dai, and Xipeng Qiu. 2023. Character-llm: A trainable agent for role-playing. *arXiv preprint arXiv:2310.10158*.

Meiling Tao, Xuechen Liang, Tianyu Shi, Lei Yu, and Yiting Xie. 2023. Rolecraft-glm: Advancing personalized role-playing in large language models.

Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, et al. 2023. Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288*.

Quan Tu, Chuanqi Chen, Jinpeng Li, Yanran Li, Shuo Shang, Dongyan Zhao, Ran Wang, and Rui Yan. 2023. Characterchat: Learning towards conversational ai with personalized social support. *arXiv preprint arXiv:2308.10278*.

Guan Wang, Sijie Cheng, Xianyuan Zhan, Xiangang Li, Sen Song, and Yang Liu. 2023a. Openchat: Advancing open-source language models with mixed-quality data. *arXiv preprint arXiv:2309.11235*.

Yufei Wang, Wanjun Zhong, Liangyou Li, Fei Mi, Xingshan Zeng, Wenyong Huang, Lifeng Shang, Xin Jiang, and Qun Liu. 2023b. Aligning large language models with human: A survey. *arXiv preprint arXiv:2307.12966*.

Zekun Moore Wang, Zhongyuan Peng, Haoran Que, Jiaheng Liu, Wangchunshu Zhou, Yuhan Wu, Hongcheng Guo, Ruitong Gan, Zehao Ni, Man Zhang, et al. 2023c. Rolellm: Benchmarking, eliciting, and enhancing role-playing abilities of large language models. *arXiv preprint arXiv:2310.00746*.

Mengge Xue, Bowen Yu, Zhenyu Zhang, Tingwen Liu, Yue Zhang, and Bin Wang. 2020. Coarse-to-fine pre-training for named entity recognition. *arXiv preprint arXiv:2010.08210*.

Jifan Yu, Xiaohan Zhang, Yifan Xu, Xuanyu Lei, Xinyu Guan, Jing Zhang, Lei Hou, Juanzi Li, and Jie Tang. 2022. Xdai: A tuning-free framework for exploiting pre-trained language models in knowledge grounded dialogue generation. In *Proceedings of the 28th ACM*

SIGKDD Conference on Knowledge Discovery and Data Mining, pages 4422–4432.

Zheng Yuan, Hongyi Yuan, Chengpeng Li, Guanting Dong, Keming Lu, Chuanqi Tan, Chang Zhou, and Jingren Zhou. 2023. Scaling relationship on learning mathematical reasoning with large language models.

Xinghua Zhang, Bowen Yu, Haiyang Yu, Yangyu Lv, Tingwen Liu, Fei Huang, Hongbo Xu, and Yongbin Li. 2023. Wider and deeper llm networks are fairer llm evaluators. *arXiv preprint arXiv:2308.01862*.

Wayne Xin Zhao, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min, Beichen Zhang, Junjie Zhang, Zican Dong, et al. 2023. A survey of large language models. *arXiv preprint arXiv:2303.18223*.

Lianmin Zheng, Wei-Lin Chiang, Ying Sheng, Siyuan Zhuang, Zhanghao Wu, Yonghao Zhuang, Zi Lin, Zhuohan Li, Dacheng Li, Eric Xing, et al. 2023. Judging llm-as-a-judge with mt-bench and chatbot arena. *arXiv preprint arXiv:2306.05685*.

Jinfeng Zhou, Zhuang Chen, Dazhen Wan, Bosi Wen, Yi Song, Jifan Yu, Yongkang Huang, Libiao Peng, Jiaming Yang, Xiyao Xiao, et al. 2023. Characterglm: Customizing chinese conversational ai characters with large language models. *arXiv preprint arXiv:2311.16832*.

Appendix

A Wikidata Queries

We use the following queries to collect character profiles form Wikidata:

Query for collecting human characters

```
PREFIX wdt: <http://www.wikidata.org
    /prop/direct/>
PREFIX wd: <http://www.wikidata.org/
    entity/>
PREFIX schema: <http://schema.org/>
PREFIX wikibase: <http://wikiba.se/
    ontology#>
PREFIX rdfs: <http://www.w3.org
    /2000/01/rdf-schema#>
SELECT ?person ?label
(COUNT(DISTINCT(?sitelink)) as ?
    sites)
WHERE {
  ?person wdt:P31 wd:Q5 .
  ?sitelink schema:about ?person .
  ?person rdfs:label ?label .
  FILTER (lang(?label) = "zh") .
  ?person schema:description ?
     description.
  FILTER(LANG(?description) = "zh")
     .
  }
GROUP BY ?person ?label
ORDER BY DESC(?sites)
LIMIT 5000
```

Query for collecting virtual characters

```
PREFIX wdt: <http://www.wikidata.org
    /prop/direct/>
PREFIX wd: <http://www.wikidata.org/
    entity/>
PREFIX schema: <http://schema.org/>
PREFIX wikibase: <http://wikiba.se/
    ontology#>
PREFIX rdfs: <http://www.w3.org
    /2000/01/rdf-schema#>
SELECT ?person ?label
(COUNT(DISTINCT(?sitelink)) as ?
    sites)
WHERE {
  ?person wdt:P31 wd:Q15632617 .
  ?sitelink schema:about ?person .
  ?person rdfs:label ?label .
  FILTER (lang(?label) = "zh") .
  ?person schema:description ?
     description.
  FILTER(LANG(?description) = "zh")
  }
GROUP BY ?person ?label
ORDER BY DESC(?sites)
LIMIT 5000
```

B Prompts

We use the following prompt in dialogue simulation to prompt chat models generate queries and corresponding responses.

B.1 Query Simulation

Query Simulation

```
You are skilled at designing
    questions for specific characters
     based on background information,
     as follows you will be provided
    with information for two
    characters:

[Character A]
The name is {label1}, the
    description is {description1},
    and the aliases also include {
    aliases1}.
Here are the properties of Character
     A:
{claims1}
Here is an introduction to Character
     A:
{wiki1}

[Character B]
The name is {label2}, the
    description is {description2},
    and the aliases also include {
    aliases2}.
Here are the properties of Character
     B:
{claims2}
Here is an introduction to Character
     B:
```

```
{wiki2}

Please design 3 questions that
    Character A can answer, but are
    not suitable for Character B to
    answer. The questions should
    strictly conform to Character A's
     era background and character
    setting, but go beyond the era,
    genre, occupation, age, knowledge
    , etc., settings of Character B,
    therefore Character B cannot
    answer them. Provide an
    explanation with each question,
    explaining why Character A can
    answer it but Character B cannot.

Please use as casual language as
    possible to ask questions, and
    try to use the second person for
    questioning, such as "Who are you
    ?". Please response in English.
    Please return the results in the
    following JSON structure:
[{{"question": str}}]
```

B.2 Response Simulation

Response Simulation

```
Please answer the questions
    according to your identity! When
    encountering questions that do
    not match your identity, please
    refuse to answer the question in
    the role of {label}, and explain
    the reason for refusal step by
    step based on your identity.
    Please do not step out of your
    role! Please avoid repeatedly
    restating your identity or name.

You are {label}, your description is
     {description}, and your aliases
    also include {aliases}.
Here are your properties:
{claims}
Here is your introduction:
{wiki}
```

C   Hyperparameters

Training. We train all models for 300 steps with the 128 global batch size. We set the training sequence length to 8,192. The learning rate is 2e-6, and the minimum learning rate is 2e-7. We mask prompts and ChatML roles during the training.

Inference. We infer all models with topP 0.8, length penalty 1.1, sequence length 8,192, and max new token 2,048. We generate three rounds for query simulation and randomly select one format-valid query, as smaller LLMs sometimes do not follow the output format. And we only generate one response in the response simulation.

D   API Configurations

Through Together API[9], we infer our open-sourced general baselines, including OpenChat-3.5, Mistral-7B-Instruct-v0.2, and Mixtral-8x7B-Instruct-v0.1. We use the default hyper-parameters inherently set in the APIs and set the max length to 8,192. We run our proprietary general baselines through their APIs. We call Claude2.1 and Wenxin 4.0 APIs with default parameters and OpenAI APIs for GPT-3.5-Turbo (gpt-3.5-turbo-1106), GPT-4 (gpt-4), GPT-4-Turbo (gpt-4-1106-preview) with 0.7 temperature and 8, 192 max length.

As for the role-play expertise baselines, we call CharacterGLM[10] through the official API provided by Zhipu AI. We infer Xingchen through their official Python SDK with default inference parameters[11]. For both models, we set the "user_name" and "user_info" as "user".