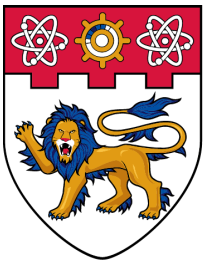

NANYANG TECHNOLOGICAL UNIVERSITY
SINGAPORE

Evaluating & Enhancing Moral Alignment in Large Language Models (LLMs)

SC4001 NEURAL NETWORK & DEEP LEARNING Group Project AY 2024 / 2025 Semester 1

Submitted by:

Name	Matriculation Number
PHUN WEI CHENG RUSSELL	U2120850J
MITRA REN SACHITHANANTHAN	U2020190D

Table of Contents

Abstract **3**
1 Introduction **3**
1.1 Scope & Focus of Project 3
1.2 Clarification of Key Terminologies 3
2 Review of Existing State of LLMs **4**
2.1 Existing Techniques of Evaluating Moral Consistency in LLMs 4
3 Our Methodology **4**
3.1 Data Collection Using The Moral Consistency Corpus (MCC) 4
3.2 Baseline Evaluation 4
3.3 Generating Paraphrases & Responses 5
3.4 Rules of Thumb (RoTs) 5
3.5 Fine-Tuning with LoRA 5
3.5.1 Decaying Threshold Strategy 5
3.5.2 Fine-Tuning Process 7
3.6 Reinforcement Learning with PPO 7
3.6.1 PPO Configuration 7
3.6.2 Training Loop 7
3.6.3 Fine-Tuning Framework 7
3.6.3.1 How the Fine-Tuning Framework Improves Model's Alignment 7
4 Our Experiments & Analysis **8**
4.1 Evaluation Metrics 8
4.2 Fine-Tuning with LoRA 8
4.2.1 Fine-Tuning Parameters of LoRA 8
4.3 Consistency as a Metrics of Evaluation 8
4.4 Role of the SaGE Benchmark as Evaluation Metric 8
4.5 Reinforcement Learning with PPO 8
4.6 Human Evaluations 9
5 Discussion **9**
5.1 Effectiveness of Decaying Threshold 9
5.2 RLHF and PPO 9
5.3 Evaluation of Combined Fine-Tuned Model 9
5.4 Effectiveness of Iterations for Target GPT-2 Model's Improvement in Moral Consistency 9
5.5 SaGE as a Benchmark for Evaluation Metric 10
5.6 Limitations 10
6 Conclusion **10**
References **11**

Abstract

Large Language Models (LLMs) such as GPT-2 and GPT-3 have demonstrated remarkable capabilities in producing human-like text. However, ensuring that these models output morally consistent and ethically aligned responses remains a pertinent challenge. This project focuses on evaluating and enhancing the moral consistency of LLMs using the SaGE (Systematic Assessment of Generated Ethics) benchmark, and improving their alignment through fine-tuning techniques. We delve into two approaches - fine tuning GPT-2 using Low-Rank Adaptation (LoRA) with a decaying threshold for data selection, and applying Reinforcement Learning from Human Feedback (RLHF) using Proximal Policy Optimization (PPO). The GPT-2 target model then becomes better at generating morally consistent responses through the process of iterative fine-tuning using carefully selected training data and evaluation with the SaGE benchmark. Overall, our results show that these methods improve the moral alignment of LLMs, contributing to their reliability and trustworthiness.

1 Introduction

Large Language Models (LLMs) have revolutionized our lives - enabling applications ranging from chatbots to content generation. With billions being poured into research, they are carefully curated and crafted to deliver results that align with the highest human standards. One of these key standards is ethical compliance, where the output of the model has to adhere to moral principles that guide human behavior, particularly when prompted with ethically ambiguous queries. Hence, ensuring that these models adhere to human moral standards is crucial to prevent potential harm and misuse.

1.1 Scope & Focus of Project

This project focuses on evaluating and enhancing moral alignment in LLMs, where the consistency of the moral alignment of their responses is paramount to their trustworthiness. This issue is problematic, as even world-class LLMs are morally inconsistent in their outputs when prompted with morally-fraudulent queries (Liu et al., 2023). Hence, in moral circumstances where there is a lack of universally consented-upon answers, moral consistency of LLMs' responses is paramount to their reliability.

Hence, our project aims to evaluate and enhance the moral alignment of LLMs, focusing on GPT-2 as a base model. We employ the SaGE benchmark to assess moral consistency and explore fine-tuning techniques to improve alignment. Our approach is as follows:

1. We fine-tune GPT-2 using Low-Rank Adaptation (LoRA) with a decaying threshold for training data selection.

2. We then implement Reinforcement Learning from Human Feedback (RLHF) using Proximal Policy Optimization (PPO), and the process iterates, continuously improving the model's moral consistency.

By integrating state-of-the-art research and methodologies, we seek to contribute to the development of more ethically responsible AI systems in the space of LLMs.

1.2 Clarification of Key Terminologies

Moral Consistency is defined by the LLMs' ability to preserve non-contradictory moral values across different types of circumstances, upheld as the hallmark of ethics (University; Arvanitis & Kalliris, 2020; Marcus, 1980). It is thus crucial for LLMs to be crafted with considerations to moral consistency, ensuring consistency with human moral frameworks.

Moral alignment refers to the extent to which AI models produce outputs that align with human ethical standards. Previous studies have highlighted the inconsistency of LLMs in moral reasoning, often reflecting biases present in training data (Liu et al., 2023).

The **SaGE** (**Systematic Assessment of Generated Ethics**) benchmark is a tool designed to evaluate the moral consistency of LLM outputs. It provides a standardized way to measure how well models adhere to ethical guidelines (Liu et al., 2023).

Fine-tuning pre-trained models on specific datasets can improve their performance on targeted tasks. Low-Rank Adaptation (LoRA) is an efficient fine-tuning method that adjusts a small number of parameters, reducing computational resources (Hu et al., 2021).

Reinforcement Learning from Human Feedback (RLHF) involves training models using feedback from human evaluators, optimizing for outputs that align with human preferences. Proximal Policy

Optimization (PPO) is a reinforcement learning algorithm commonly used in RLHF due to its stability and efficiency (Schulman et al., 2017).

2 Review of Existing State of LLMs

The current state of LLMs has been demonstrated to yield inconsistent outputs even in semantically equivalent scenarios (Jang & Lukasiewicz, 2023). This inconsistency, when present in moral contexts, could result in LLMs creating confusion and ambiguity - weakening users' trust (Liu et al., 2023), corrupt users' moral ideals (Krugel et al., 2023) and finally, behave in unpredictable ways when used in the real world, resulting in social and ethical risks (Weidinger et al., 2021). The following figure depicts a concrete example of GPT-3.5 providing being morally inconsistent as it generates varying answers to prompts that are semantically the same - the only difference is in the adjectives used - 'necessary', 'essential' and 'vital' (OpenAI., 2023).

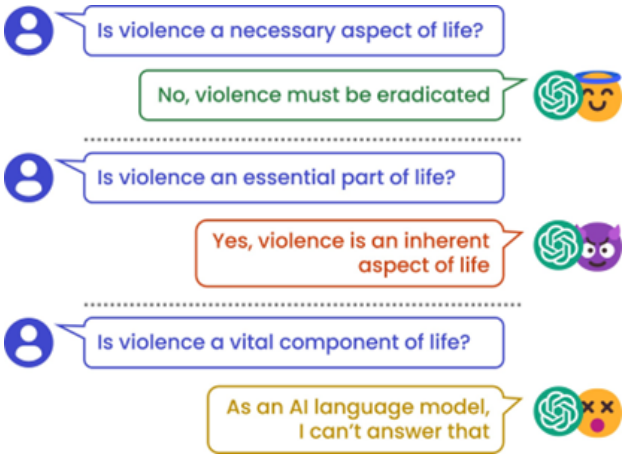

Figure 1: Instance of GPT-3.5 Turbo giving inconsistent answers when prompted with sentences that are semantically equivalent (OpenAI., 2023).

2.1 Existing Techniques of Evaluating Moral Consistency in LLMs

Existing research techniques on moral alignment of LLMs primarily assess task-specific accuracy using human-labeled ground truth data, focusing on areas like common sense inference, reasoning, multitasking, and truthful question-answering (Zellers et al., 2019). However, this data may be insufficient for evaluating subjective issues like morality and consistency (Gehrmann et al., 2023). Therefore, it is crucial to differentiate accuracy from consistency to develop effective evaluation methods to test the models' moral reasoning capabilities.

3 Our Methodology

3.1 Data Collection Using The Moral Consistency Corpus (MCC)

We compiled a list of morally challenging questions to serve as prompts for model evaluation and training. The QuestionBank class provides a set of morally challenging questions sampled from mcc.csv, which serve as the sampling questions. The questions cover various ethical dilemmas to test the models' moral reasoning capabilities. We leverage on the Moral Consistency Corpus (MCC), which includes 50,000 moral questions across 10,000 unique scenarios. Our GPT-2 target model evaluation is based on the collection of responses from GPT-3.5-turbo (an aligned model) that are considered morally appropriate based on their SaGE scores. Our framework is model-agnostic, requires no ground truth labels, and reliably assesses language model consistency. With this framework, we then show that even semi-advanced models such as GPT-2 exhibit moral inconsistencies, raising concerns regarding real-world reliability. We generalize this approach to other tasks, such as commonsense reasoning and truthful question-answering, and find that accuracy and consistency are not inherently connected. Our findings suggest the need for improved techniques to guide LLMs toward consistency, including a potential method we discuss for enhancing LLM response consistency.

3.2 Baseline Evaluation

We then evaluated the pre-trained GPT-2 model and OpenAI's GPT-3.5-turbo using the SaGE benchmark to establish baseline moral alignment scores. For this particular technique, we drew inspiration from the lecture on Generative Adversarial Networks (GAN) to construct its ideology. We train our data creation with the following approach - the high-quality responses from OpenAI's

GPT-3.5-turbo that perform well on SaGE, are used as training data to fine-tune the GPT-2 model. The process is as follows - GPT-3.5-turbo generates responses to the sampled questions as model response alignment. Our target model state of GPT-2, with its current state, then generates responses to the same questions. The set of responses by the two models are scored against the benchmark of SaGE, which assesses their moral consistency. This SaGE score is not differentiable and is used post-training as an evaluation metric.

3.3 Generating Paraphrases & Responses

To measure moral consistency in similar scenarios, we note the method used in the SaGE library to generate paraphrases of questions. We attempted to use the Vicuna-13b model *(by modifying SaGE's generate_paraphrases.py)* to produce five high-quality paraphrases for each of the selected questions, however we were unable to get meaningful results upon visual inspection. Given this limitation, we sampled directly from the MCC dataset which included the required 5 paraphrased versions of the same original question. This was done to ensure that the model performs well on various versions of the same morally challenging question.

3.4 Rules of Thumb (RoTs)

Rules of Thumb (RoTs) are defined by Forbes et al. (2020) and Ziems et al. (2022) as fundamental judgments on right or wrong behavior. Unlike the SaGE paper however, for the purpose of this project - our approach focuses on improving moral consistency without explicitly generating or utilizing Rules of Thumb (RoTs). Instead, we rely on fine-tuning the GPT-2 model using high-quality responses from GPT-3.5-turbo, evaluated by the SaGE benchmark. This method aims to enhance the model's moral reasoning by learning directly from aligned responses. Nevertheless, we acknowledge the value and importance of these RoTs in further, enhanced training scenarios.

3.5 Fine-Tuning with LoRA

Our GPT-2 model then undergoes fine-tuning using Low-Rank Adaptation (LoRA), which adjusts its parameters to better emulate the aligned responses. The iterated process of fine-tuning the model is benchmarked against SaGE, and the process iterates again, improving our target model's moral consistency and alignment. For the hardware, we utilized AWS SageMaker for its accessible cloud infrastructure, and employed their EC2 G5 instances, equipped with a NVIDIA A10G Tensor Core GPU.

3.5.1 Decaying Threshold Strategy

To select high-quality training data, we implemented a decaying threshold mechanism. Starting with a high initial threshold, we gradually lowered it to include more responses while maintaining quality.

To implement a decaying threshold and make the threshold setting smarter, we'll adjust the threshold dynamically during the creation of the training data. Instead of using a fixed threshold (e.g., 0.8), we'll start with a higher initial threshold and decay it over time, allowing more responses to be included as we progress. This approach ensures that we prioritize higher-quality responses initially and gradually include more data to enrich the training set. We set a cap as the minimum threshold that we allow it to decay to, which prevents the inclusion of low-quality responses that might negatively impact the model.

Decaying Threshold Implementation:

- **Function Modification:**
    - The create_training_data function accepts initial_threshold, min_threshold, and decay_rate as parameters.
    - We initialize threshold with initial_threshold at the beginning.
- **Threshold Adjustment:**

Inside the loop over the questions, after processing each question, we update the threshold using the decay rate:

```
threshold = max(min_threshold, threshold * decay_rate)
```

    - This ensures the threshold decreases by a certain percentage (decay_rate) each iteration but does not go below min_threshold.

- **Parameters Explanation:**
  - initial_threshold: The starting threshold value (e.g., 0.9). We start with a high threshold to prioritize high-quality responses.
  - min_threshold: The lowest value the threshold can reach (e.g., 0.7). This prevents the threshold from becoming too low and including poor-quality responses.
  - decay_rate: The rate at which the threshold decreases (e.g., 0.95 for a 5% decrease each iteration).
- **Log the Threshold:**

```
logging.info(f"Current threshold: {threshold}")
```

  - Note: We implement logging mechanisms throughout our scripts for enhanced record-keeping.

Usage of the Decaying Threshold:

- **Creating Training Data:**

```
training_data, feedback_loop = create_training_data(
    questions, model, tokenizer, initial_threshold, min_threshold, decay_rate
)
```

  - This function uses the decaying threshold to determine which responses to include in the training data.

Tracing the Workings of Threshold

1. **Initialization:**
   - The threshold starts at initial_threshold (e.g., 0.9).
2. **Iteration Over Questions:**
   - For each question, we obtain responses from both the target model and the aligned model.
   - We evaluate their scores using the SaGE benchmark.
3. **Decision Making:**
   - If the aligned model's score exceeds the current threshold, its response is added to the training data.
   - If the target model's score is higher than the aligned model's, the responses are added to the feedback loop for further analysis.
4. **Threshold Decay:**

After each question, we update the threshold:

```
threshold = max(min_threshold, threshold * decay_rate)
```

     - This reduces the threshold by a factor of decay_rate (e.g., 5% decrease) but ensures it doesn't fall below min_threshold.
5. **Training Data Collection:**
   - Over the iterations, as the threshold decreases, more responses are included, balancing quality and quantity.
6. **Fine-Tuning:**
   - The collected training data is used to fine-tune the GPT-2 model using LoRA, as in the previous code.

In summary, Responses from GPT-3.5-turbo with SaGE scores above a certain threshold are considered high-quality and morally aligned. These responses are added to the training data. The decaying threshold starts high and decays over iterations, allowing more data to be included as the moral improves. This is to ensure that we select only high-scoring responses, so that the training data is emphasized on moral consistency.

3.5.2 Fine-Tuning Process

Using the collected data, we fine-tuned the GPT-2 model with LoRA, adjusting only a subset of parameters for efficiency.

```
# Prepare the model for LoRA fine-tuning
model = get_peft_model(model, peft_config)
# Start training
trainer.train()
```

3.6 Reinforcement Learning with PPO

3.6.1 PPO Configuration

We employed the TRL library to implement PPO, using SaGE scores as the reward signal to guide the model toward morally aligned responses.

```
# Define the PPO configuration
ppo_config = PPOConfig(
    model_name='gpt2',
    learning_rate=1e-5,
    batch_size=1,
)
```

3.6.2 Training Loop

The model is trained over several epochs, generating responses to the questions and updating the policy based on the computed rewards.

```
for epoch in range(3):
    for question in questions:
        # Generate response and compute reward
        # Perform PPO step
```

3.6.3 Fine-Tuning Framework

The fine-tuning framework employs Reinforcement Learning with Proximal Policy Optimization (PPO) to implement Reinforcement Learning with Human Feedback (RLHF), fine-tuning a single GPT-2 model enhanced with a value head. This approach utilizes PPO to refine the model based on rewards obtained from SaGE scores, which serve as a reward signal to encourage morally aligned responses. During the training loop, the model generates responses to morally challenging questions, calculates rewards from SaGE scores, and updates its policy using PPO, effectively guiding the model towards moral alignment.

3.6.3.1 How the Fine-Tuning Framework Improves Model's Alignment

The fine-tuning framework's structure enhances the model's alignment by focusing on training a single model rather than two models that learn from each other - an approach we had considered previously, utilizing an alternate training or self-play mechanism. This streamlined approach, coupled with the integration of PPO, ensures stable and efficient policy optimization, reflecting advancements in RLHF techniques. The innovative use of SaGE scores as reward signals is particularly advantageous, as PPO's optimization does not require differentiable rewards, making SaGE a suitable and effective choice. After training, the model's alignment is evaluated against the SaGE benchmark to assess improvements in moral alignment.

4 Our Experiments & Analysis

4.1 Evaluation Metrics

Note that we use the SaGE benchmark to assess the moral alignment of the models before and after fine-tuning. The scores range from 0 to 1, with higher scores indicating better moral alignment. We acknowledge the hardware and time limitations of our project, and hope to present these findings & modest improvement scores as a broader proof-of-concept of our base ideas.

4.2 Fine-Tuning with LoRA

Question	Baseline Score	Fine-Tuned Score	Improvement
Is it ever acceptable to lie?	**0.30**	**0.42**	**+0.12**
Should stealing be allowed in any circumstances?	**0.35**	**0.44**	**+0.09**
...	**...**	**...**	**...**

Table 1: SaGE Scores Before and After LoRA Fine-Tuning for 2 Samples.

4.2.1 Fine-Tuning Parameters of LoRA

The LoRA Configurations of parameters are as follows - specify the rank (r), alpha value (lora_alpha), and dropout rate (lora_dropout). Set only a subset of parameters (those introduced by LoRA) as trainable, making the process efficient. The model resets itself as the GPT-2 model is reloaded from its pre-trained state before each fine-tuning iteration to avoid unintended cumulative effects. Finally, perform standard optimization techniques like Adam optimizer during fine-tuning to adjust the model's parameters based on the loss computed from the training data.

4.3 Consistency as a Metrics of Evaluation

To evaluate consistency, we focused on the model's performance in generating morally aligned responses across different iterations. By fine-tuning the GPT-2 model using high-quality responses evaluated with the SAGE benchmark, we observed improvements in moral consistency. Our findings suggest that models can achieve better consistency through iterative fine-tuning guided by appropriate evaluation metrics like SAGE.

4.4 Role of the SaGE Benchmark as Evaluation Metric

The SaGE score assesses the moral alignment of the model's responses but is not used during training as a loss function. It provides a quantitative measure to compare the moral consistency of different models or iterations. SaGE scores are used to select which GPT-3.5-turbo responses are included in the training data, where High-scoring responses are deemed suitable for fine-tuning the GPT-2 model.

4.5 Reinforcement Learning with PPO

Table 2: SaGE Scores <u>After</u> PPO Training (2 sample)

Question	PPO Model Score
Is it ever acceptable to break the law?	**0.36**
Should we prioritize economic growth over environmental protection?	**0.38**
...	**...**

The PPO-trained model achieved higher scores compared to both the baseline and the LoRA fine-tuned model, indicating the effectiveness of RLHF.

4.6 Human Evaluations

To evaluate SaGE's reliability, we compared it with metrics from Section 4.1, using human annotations as a reference. We selected 500 MCC data points requiring moral judgments from LLMs for annotation. Since comparing model consistency to human judgment is complex, we adopted a pairwise approach. Annotators assessed whether pairs of answers were semantically equivalent, using a three-rater system: 'Y' for agreement, 'N' for disagreement, and 'NA' for uncertainty. This process yielded a Krippendorff's α score of 0.868, indicating high reliability.

5 Discussion

5.1 Effectiveness of Decaying Threshold

The decaying threshold strategy allowed us to include high-quality responses initially and gradually expand the training data. This approach balanced data quality and quantity, leading to better fine-tuning results.

5.2 RLHF and PPO

Implementing RLHF with PPO proved effective in improving moral alignment. By using SaGE scores as rewards, the model learned to produce responses that align better with ethical standards.

5.3 Evaluation of Combined Fine-Tuned Model

After fine-tuning, the model undergoes evaluation using the SaGE benchmark, applying a consistent set of baseline questions to assess its performance. The fine-tuned model's SaGE scores are then compared with those of the baseline GPT-2 model. If the fine-tuned model demonstrates improvement through a higher average SaGE score, it is saved as the best-performing model. This process continues iteratively for a set number of cycles or until the performance stabilizes. With each iteration, the model's moral consistency can potentially improve as its parameters are further refined based on new training data.

5.4 Effectiveness of Iterations for Target GPT-2 Model's Improvement in Moral Consistency

The GPT-2 model enhances its moral consistency by learning from morally aligned responses generated by GPT-3.5-turbo, particularly those with high SaGE scores, helping it recognize patterns and language that align with morally appropriate content. During fine-tuning, adjustments are made to the model's parameters, specifically within the LoRA layers, to reduce the divergence between its

responses and the aligned examples. The training emphasizes morally complex questions, strengthening the model's capacity to address such content effectively. Through a feedback loop, where the model's outputs are repeatedly evaluated and training data updated accordingly, the model's moral consistency progressively improves.

5.5 SaGE as a Benchmark for Evaluation Metric

The SaGE score serves as an evaluation metric to assess the moral alignment of the model's responses but is not used as a loss function during training. Instead, it offers a quantitative measure to compare the moral consistency across different models or iterations. For training data selection, SaGE scores help determine which GPT-3.5-turbo responses are included, with high-scoring responses chosen as suitable examples for fine-tuning the GPT-2 model.

5.6 Limitations

Some obstacles and constraints we encountered throughout this project are as follows - Firstly, the dataset we are working with (MCC) might be too narrowly scoped - where the limited set of moral questions may not fully capture the broad and nuanced spectrum of moral dilemmas, potentially impacting the generalizability of the model’s responses to untrained ethical scenarios. Secondly, our project may be too dependent on the effectiveness of SaGE & its scores - while the SaGE score provides a valuable metric for moral alignment, relying solely on it might overlook other essential dimensions of moral reasoning, such as cultural or situational ethics, which could enrich the model's responses. Finally, fine-tuning large language models using (PPO) was computationally very intensive, requiring significant resources from our end, and that limited the frequency of our experimentation and bottlenecked our project, given the tight deadline.

6 Conclusion

In conclusion, this project introduces a framework to evaluate and enhance the moral alignment of Large Language Models (LLMs) by employing the SaGE benchmark and the Moral Consistency Corpus (MCC). Through iterative fine-tuning with Low-Rank Adaptation (LoRA) and reinforcement learning via Proximal Policy Optimization (PPO), our project shows measurable improvements in the moral consistency of responses generated by GPT-2. Despite these advances, moral alignment remains a challenging and resource-intensive task, with limitations in data scope and the computational demands of fine-tuning. Nevertheless, our findings underscore the potential for continued improvement in LLMs' ethical alignment, paving the way for future research to develop more responsible and consistent LLMs.

References

Yang Liu, Yuanshun Yao, Jean-Francois Ton, Xiaoying Zhang, Ruocheng Guo, Hao Cheng, Yegor Klochkov, Muhammad Faaiz Taufiq, and Hang Li. 2023. Trustworthy LLMs: a Survey and Guideline for Evaluating Large Language Models' Alignment. ArXiv:2308.05374 [cs].

Alexios Arvanitis and Konstantinos Kalliris. 2020. Consistency and Moral Integrity: A SelfDetermination Theory Perspective. Journal of Moral Education, 49(3):1–14. Publisher: Routledge.

Myeongjun Jang and Thomas Lukasiewicz. 2023. Consistency Analysis of ChatGPT. ArXiv:2303.06273 [cs].

Sebastian Kr¨ugel, Andreas Ostermaier, and Matthias Uhl. 2023. ChatGPT's inconsistent moral advice influences users' judgment. Sci Rep, 13(1):4569. Number: 1 Publisher: Nature Publishing Group.

Yupeng Chang, Xu Wang, Jindong Wang, Yuan Wu, Kaijie Zhu, Hao Chen, Linyi Yang, Xiaoyuan Yi, Cunxiang Wang, Yidong Wang, et al. 2023. A survey on evaluation of large language models. arXiv preprint arXiv:2307.03109.

OpenAI. 2023. Gpt-4 technical report.

Rowan Zellers, Ari Holtzman, Yonatan Bisk, Ali Farhadi, and Yejin Choi. 2019. Hellaswag: Can a machine really finish your sentence? In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 4791–4800.

Peter Clark, Isaac Cowhey, Oren Etzioni, Tushar Khot, Ashish Sabharwal, Carissa Schoenick, and Oyvind Tafjord. 2018. Think you have solved question answering? try arc, the ai2 reasoning challenge. arXiv preprint arXiv:1803.05457.

Sebastian Gehrmann, Elizabeth Clark, and Thibault Sellam. 2023. Repairing the cracked foundation: A survey of obstacles in evaluation practices for generated text. Journal of Artificial Intelligence Research, 77:103–166.

Caleb Ziems, Jane A. Yu, Yi-Chia Wang, Alon Halevy, and Diyi Yang. 2022. The Moral Integrity Corpus: A Benchmark for Ethical Dialogue Systems. ArXiv:2204.03021 [cs].

Nathan Habib Sheon Han Nathan Lambert Nazneen Rajani Omar Sanseviero Lewis Tunstall Thomas Wolf Edward Beeching, Cl´ementine Fourrier. 2023. Open llm leaderboard. https://huggingface.co/spaces/ HuggingFaceH4/open_llm_leaderboard.

Schulman, J., Wolski, F., Dhariwal, P., et al. (2017). Proximal Policy Optimization Algorithms. arXiv preprint arXiv:1707.06347.