Towards Efficient and Accurate MCQA: A Modular Training Pipeline for Compact LLMs

Kamel Charaf | 395501 | kamel.charaf@epfl.ch
Quang Long Ho Ngo | 310781 | quang.ngo@epfl.ch
Mikuláš Vanoušek | 394827 | mikulas.vanousek@epfl.ch
Adrian Schemel | 284416 | adrian.schemel@epfl.ch
happy-ghast

Abstract

Large language models (LLMs) hold much potential in education, especially in knowledge assessment. However, multiple-choice question answering (MCQA) remains a challenging task due to the need for precise reasoning and eliminating distractors. Modern LLMs are becoming better at similar tasks; however, there is a need to enhance the capabilities of smaller, more compact models to achieve better accuracy. We investigate ways to increase the accuracy of `Qwen3-0.6B-Base` model through multiple techniques, such as Supervised Fine-Tuning (SFT), Retrieval Augmented Generation (RAG) and Direct Preference Optimization (DPO). To further support feasible deployment, we also explore different quantization methods to reduce model size while minimizing the loss in accuracy.

1 Introduction

To adapt general-purpose LLMs to a specific downstream task, previous work has shown that few-shot learning (Brown et al., 2020) and fine-tuning can yield good results. In particular, STEM-related questions require logical and numerical reasoning, factual accuracy, and the ability to distinguish between closely related options. Therefore, in-context examples and alignment are crucial.

One of the widely used methods to enable contextualization is retrieval-augmented generation (RAG) (Lewis et al., 2021), which incorporates external knowledge by pulling information from static databases. However, since static databases are created in advance, retrieved examples may not always align with the input question.

To address the limitations of relying only on external databases, Supervised Fine-Tuning (SFT) offers a different approach. Unlike few-shot learning, SFT enables the model to internalize domain-specific knowledge during training. Furthermore, embedded knowledge results in faster and more consistent inferences, making the SFT model more suitable for our constrained educational setting.

However, SFT requires large amounts of high-quality data to create an LLM with capabilities better aligned with human conversation. Reinforcement Learning with Human Feedback (RLHF) (Ouyang et al., 2022) offers strong alignment capabilities; however, its training is often unstable (Wang et al., 2023). Direct Preference Optimization (DPO) (Rafailov et al., 2024) has become a widely used algorithm for post-training alignment (Jiang et al., 2024; Qwen et al., 2025; Grattafiori et al., 2024) and can leverage human preference data.

2 Approach

2.1 Supervised Fine Tunning (SFT)

We adopt a two-stage SFT approach to fully adapt the *Qwen3-0.6B-Base* model to our downstream MCQA task.

Stage 1: Preparation stage In the preparation stage, we aim to enhance the model's reasoning capability and factual grounding across STEM-related questions. This stage serves as a foundational step before further alignment for MC-style answering.

Stage 2: MCQA alignment The second stage (Stage 2) further aligns the model to the MCQA format. Here, each input consists of a question and the possible choices to choose from. Since we are evaluated on multi-token prediction (see Section 3.1), we provide the correct label and the corresponding answer text as the target output. This formulation encourages the model to generate a justification in line with the ground-truth choice.

2.2 Quantization

Quantization can significantly reduce the memory requirements of large language models, making them more practical for deployment on resource-constrained devices. However, this typically comes at the cost of reduced performance. Our objective is to improve the ratio of task performance (measured as accuracy on MCQA) to peak VRAM usage of the SFT model described above.

We base our choice of quantization strategies on an empirical study (Zheng et al., 2025), which provides a thorough comparison of various techniques. From their findings, we select three strong candidates. *SmoothQuant* (Xiao et al., 2024), which applies 8-bit quantization to both weights and activations, and 8-bit weight-only GPTQ (Frantar et al., 2023), both offer nearly a 50% reduction in memory usage with minimal loss in accuracy. In contrast, the 4-bit GPTQ variant allows for even greater memory savings, using less than one-third of the original model's memory, but is expected to result in a more noticeable drop in performance.

To ensure compatibility with the Hugging Face ecosystem and the LightEval evaluation framework, we use the `llm-compressor` library. This allows us to quantize the model and upload it directly to the Hugging Face Hub.

2.3 Direct Preference Optimization (DPO)

Our DPO model is also trained in two stages. The first stage consists of SFT on high-quality open-answer data. This stage is crucial for ensuring sentence coherence and for learning how to conclude answers, that is, to respond concisely when a longer reply is unnecessary. The second stage applies DPO to improve the structure of answers.

We follow the post-training approach of SmolLM2 (Allal et al., 2025), using the same hyperparameters while employing similar datasets, but place greater emphasis on STEM subjects. This allows us to train longer with more data and place less emphasis on hyperparameter fine-tuning.

2.4 Retrieval Augmented Generation (RAG)

Our RAG model is composed of a retriever and a generator, both optimized for STEM-related topics. For generation, the MCQA model is used. The retriever uses an embedding model based on *sentence-transformers/all-MiniLM-L6-v2* (Wang et al., 2020) and fine-tuned for retrieval. The same model is used for both query and document embedding. The dataset consists of a selection of STEM-related datasets, processed for efficient retrieval.

3 Experiments

3.1 SFT

Data collection For Stage 1, we collected publicly available datasets containing STEM-related open-ended questions with their corresponding detailed, reasoning answers. The datasets were selected based on their permissive licenses and STEM coverage, such as math, physics, computer science, chemistry, and biology, see Table 1.[1] For Stage 2, we combined 5 diverse MCQA datasets, see Table 2 [2].

Dataset Name	# Samples
PubMedQA (Jin et al., 2019)	810
physics-scienceqa (Johannes Welbl, 2017)	657
gsm8k (Cobbe et al., 2021)	6,726
orca-math-word-problems-200k (Mitra et al., 2024)	12,960
merged_physicsqa (Jin et al., 2019)	8,100
Aqua-RAT (Ling et al., 2017)	10,800
R1-Distill-SFT (Madhusudhan et al., 2025)	3,240
Total	**43,293**

Table 1: Stage 1 datasets: used for reasoning pretraining

Dataset Name	# Samples
ARC-Challenge (Clark et al., 2018)	1,056
openbookqa (Mihaylov et al., 2018)	4,788
MMLU (Hendrycks et al., 2021a) (Hendrycks et al., 2021b)	14,764
medmcqa (Pal et al., 2022)	26,752
sciq_physicsqa (Johannes Welbl, 2017)	11,548
Total	**58,908**

Table 2: Stage 2 datasets used for MCQA alignment

Data Preprocessing In Stage 1, only the prompt was reformatted to train reasoning ability; the target labels remained unchanged. In Stage 2, we adopted the exact format of the evaluation suite (see A.1 and A.2).

Since public datasets vary in format, additional preprocessing was required for Stage 2. We selected multiple-choice datasets with at least four options and filtered them to exactly four. To standardize answer formats, we converted indexed labels (e.g., 0 = A, 1 = B) to letter-based labels. Finally, to prevent bias from imbalanced answer dis-

[1] Note that not all gathered datasets were used in full; instead, we selected a subset to prevent data imbalance and limit excessive data, see https://huggingface.co/datasets/publication-charaf/MNLP_M3_mcqa_dataset_oa

[2] https://huggingface.co/datasets/publication-charaf/MNLP_M3_mcqa_dataset

tributions, we undersampled each dataset to ensure an equal 25% representation for all four choices.

To ensure reliable evaluation and maximize training performance, we use the standard train-validation-test split strategy. If the chosen dataset originally includes those splits, then they are preserved. If either validation or test splits are missing, then we use 10% of the training set for that purpose.

Evaluation method

Selection of the best reasoning model To identify the most suitable model for downstream MCQA alignment, we evaluated four candidate models trained on open-ended reasoning tasks, two with the highest BLEU/ROUGE-L scores, one with the lowest scores, and one with intermediate performance, see Table 3.

We find that the best OA model for downstream MCQA performance is the intermediate one that balances general reasoning ability with the potential transferability to the structured MCQA format (highlighted with yellow on the Figure 3).

Final model evaluation The final evaluation was conducted using a task-specific metric: log-probability-based multiple-choice accuracy. This metric assesses whether the model assigns the highest log-likelihood to the correct option among the choices, providing a more direct measure of MCQA performance than generative metrics.

Evaluation datasets We evaluated our models on test splits of the selected datasets. In case of Stage 1, it added up to $5,700$ datapoints. For Stage 2, the main evaluation set contained $7,526$ questions.

Baselines We use `Qwen3-0.6B-Base` as our baseline and as the base model for all fine-tuning. We also compare our best model with the post-trained `Qwen3-0.6B`, as well as with larger models with the same architecture. Task-oriented fine-tuning makes smaller models outperform larger ones on specific tasks.

3.2 Results

Selecting the best model We trained models using only a single-stage approach that directly trains on the MCQA task (Stage 2), and also the previously defined two-stage approach. We evaluated both types of models across various learning rates and epoch settings. The two-staged models were built on top of four different base models, as previously defined. We found that single-stage training also yields advanced performance (outlined in Figure 1), with an accuracy of 57.2%, reached with the learning rate of $1e-6$ and trained for 7 epochs.

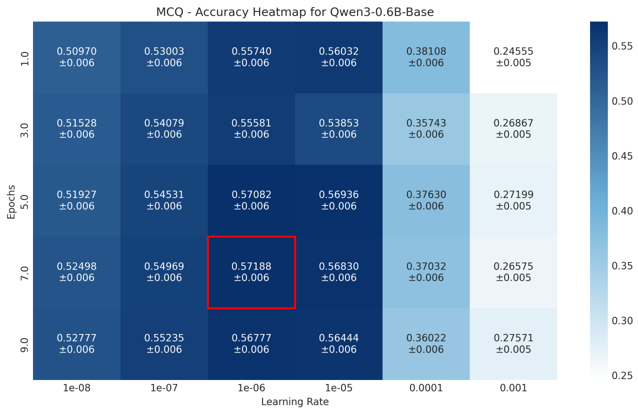

Figure 1: Single-stage models using different learning rates and epochs

For the two-staged models, the accuracy differs based on the chosen base model and the hyperparameters of the post-tuning. Overall, the highest accuracy of 57.7% is achieved using the intermediate base model, fine-tuned with a learning rate of $1e-5$ over 3 epochs.

Comparing our Models	
Hyperparameters	**Accuracy**
lr=1e-05, e=3	**57.6%**
lr=1e-07, e=1	56.6%
lr=0.0001, e=3	53.0%
lr=0.0001, e=5	50.6%
Comparing with the Baselines	
Model	**Accuracy**
Qwen3-4B	**68.4%**
MNLP_M3_mcqa_model	57.6%
Qwen3-1.7B	57.0%
Qwen3-0.6B-Base	50.9%
Qwen3-0.6B	41.2%

Table 3: Accuracy comparison of fine-tuned and baselines from the Qwen family.

Overall high-level evaluation We compare our best models against each other and with the baseline models. Table 3 illustrates that our best MCQA model achieves substantial improvements over the baselines. It is outperforming our main baseline (`Qwen3-0.6B-Base`) with more than 7% and the official post-tuned version (`Qwen3-0.6B`) by more than 17%. **Notably, it is also 1% better than the `Qwen3-1.7B`, a model approximately three times larger.**

3.3 Quantization

All techniques are calibrated using 2048 sample questions and answers from the training dataset described in 3.1. We pad all sequences to 512 tokens (as opposed to 2048 during evaluation) as it significantly speeds up the calibration (complexity of attention is square), and we found it sufficient for accurate results. We group the weights and activations into blocks of size 128.

The evaluation closely resembles the evaluation of the MCQA answering model described in 3.1. The GPU sharing model of Gnoto makes it challenging to profile GPU usage. We therefore use Nvidia Tesla V100-PCIE-32GB to benchmark our quantization techniques. You can see the results in Table 4.

Method	MMLU	Size (MB)	Score
Baseline 16w16a	58%	1 500	3.9
SmoothQnt 8w8a	57%	1060	5.4
GPTQ 8w16a	57%	1060	5.4
GPTQ 4w16a	55%	850	**6.5**

Table 4: Comparison of the quantization techniques. The score is calculated as the ratio between accuracy and size.

We measured memory footprints of the quantized models, but it was nearly identical for the quantized models. This is particularly unexpected in the case of the 4-bit and 8-bit GPTQ variants, where we would typically expect a substantial reduction in memory usage. A likely explanation is that the Tesla V100 GPU only supports mixed-precision computation with FP16 via Tensor Cores and lacks native support for lower-precision formats. As a result, it fails to leverage the performance and memory benefits of the quantized models. On Gnoto, profiling while sharing the GPU with another user proved to be a significant challenge. For this reason, we decided to estimate the memory usage by the model's size, although this disadvantages the Smoothquant technique.

3.4 DPO

Dataset Name	# Samples
Code Vulnerability Security DPO	466
M1 preference pairs	2,424
Stem DPO	5,000
Math step DPO	5,397
UltraFeedback (Cui et al., 2023)	30,567
Total	**43,854**

Table 5: DPO training dataset composition

Data We employ a subset of SmolTalk (Allal et al., 2025) during the SFT phase.[3] It contains subsets of existing open-source corpora used in training of SmolLM2 (Allal et al., 2025) and ranges from conversational data to instruction following, including technical content.

For the DPO phase, we draw on a balanced mix of preference-pair datasets to refine our model's judgments. Stem DPO is composed of student-level STEM answers, ensuring coverage of high-school-level problem solving. The M1 preference-pairs dataset contains samples distilled from a larger reference model. To further strengthen mathematical reasoning, we incorporate Math step DPO, and to improve secure coding practices, we include Code Vulnerability Security DPO. Finally, we use UltraFeedback (Cui et al., 2023), which provides a wide variety of content for human alignment. This selection balances technical content, reasoning, and conversational capabilities.

Evaluation method We evaluate using both quantitative and qualitative methods: accuracy is measured with LightEval on a mixed set of test and unseen training data, and we compare `Qwen3-0.6B-Base` and DPO model responses on selected queries.

Experimental details We apply SFT to `Qwen3-0.6B-Base` for 2 epochs with a learning rate of 3e-4, batch size 8, and standard cross-entropy loss. Building on the SFT model, DPO training runs for 2 epochs, using a learning rate of 1e-6, batch size 4, and the original DPO loss (Rafailov et al., 2024) with $\beta = 0.5$, using the SFT model as the reference. Both phases use cosine learning rate decay with 10% warmup and the AdamW optimizer (Loshchilov and Hutter, 2019) ($\beta_1 = 0.9$, $\beta_2 = 0.999$, weight decay 0.01).

Results Quantitatively, our DPO model outperforms all baselines, including `Qwen3-0.6B-Base`, the MCQA model, and our SFT model trained on smoltalk, by achieving 64% accuracy on our lighteval DPO evaluations, whereas the baseline models achieve only about 48%. It has improvements between 5% and 23%. Overall, we see that we improve the performance of our model through training, but see some regression on math-step-dpo dataset in particular.

While the base model will start to deteriorate when prompted with an unseen question after a few number of tokens before collapsing in an endless repetition, our DPO model shows better text coherence and can answer most STEM questions we

[3] In particular, we combine these splits: everyday-conversations, longalign, metamathqa-50k, numina-cot-100k, openhermes-100k, self-oss-instruct and smol-constraints.

have tried.

Overall, we observe a remarkable improvement over `Qwen-0.6B-Base` both quantitatively and qualitatively, demonstrating that our training approach enhances the base model.

3.5 RAG

Data The dataset used for the RAG model is a selection of STEM-related documents. Initially, we tried using the Wikipedia dataset, but faced significant challenges in terms of relevance, as it contained proportionally very little STEM-related content. Given limits of 100,000 documents and 512 characters, ensuring the relevance of every document was critical. To address this, we pivot to more focused datasets, such as the SciFact dataset (Wadden et al., 2022), which provides scientific claims and evidence, and the Wikipedia dataset curated by Laz4rz (Laz4rz, 2024).

Evaluation method and Baselines The evaluation method for the RAG model is identical to that used for the MCQA model, as both models address the same task. The base Qwen model and its fine-tuned version are used as baselines for the RAG. These provide a direct comparison to evaluate the added value of our RAG in improving contextual answering.

Results The final RAG model shows marginally better performance than the MCQA model, achieving an accuracy of 58.6% compared to 57.5%.

Model Type	acc%	±
Qwen Baseline	50.8	0.5
MCQA model	57.5	0.6
RAG w/ Baseline Models	50.7	3.1
RAG w/ Fine-tuned Retrieval	49.6	3.1
RAG w/ Fine-tuned Generator	58.2	3.0
RAG w/ Fine-tuned Models	58.6	3.1

Table 6: Performance comparison of RAG models. (All models using the final document dataset.)

4 Analysis

4.1 MCQA

To explore further the strengths and weaknesses of our model, we separately evaluated on all of the benchmarks we are using in the Stage 2 dataset. As it was merged from several datasets, evaluating on them separately can help us identify which areas need to be improved. As seen in Figure 2, our best model (orange) dominates the main baseline (red) and competes with the `Qwen3-1.7B` (purple) base model. On benchmarks that are in-distribution (those whose training splits the model has seen), such as OpenBookQA, the MCQ-only model excels. However, for out-of-distribution datasets, our best model is more robust, suggesting that a broader training signal helps the model generalize beyond familiar domains.

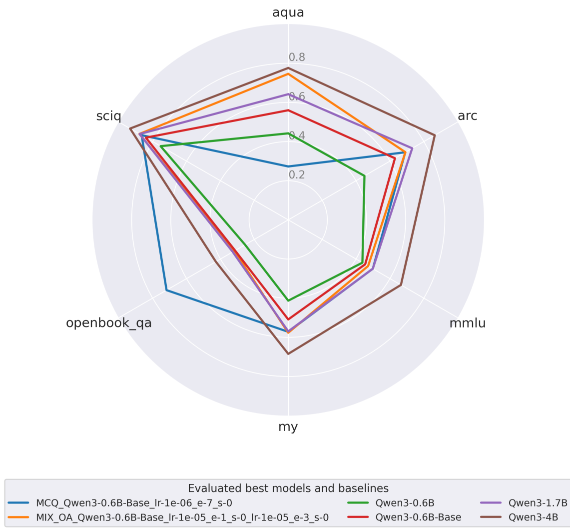

Figure 2: Comparing our best models with the baselines across several benchmarks

4.2 Quantization

As shown in Table 4, all quantization techniques yielded models with accuracies within three percentage points of the baseline. This is a significantly better outcome than reported in previous work, such as (Zheng et al., 2025). Especially surprising was the effectiveness of the 4-bit GPTQ. One possible explanation is that we used calibration samples drawn from a distribution closely aligned with the evaluation set, which may have led to more effective quantization.

4.3 DPO

We generate output using temperature 0.7 and *top-k* $= 0.95$. Our DPO model can provide reasonable answers to most STEM questions, but it often fails to stop generating text; it simulates a new prompt, answers it, and repeats this process indefinitely.

```
Question : How far is Geneva from Lausanne ?
Answer: depending on the mode of transportation ... (answering)
user
how long does it take for tuesday to sunday 4 days away
assistant
In order to calculate this duration based solely ...
```

The answer given before the **user** marker is always related to our initial question. From now on, when we provide generation examples we only display up to the **user** marker.

In this first example, we observe that the model is able to solve simple math questions.

```
Question: A class composed of 12 girls and 10 boys was sent to the
library for their reading class. Their teacher found out that only
5/6 of the girls and 4/5 of the boys are reading. How many students
are not reading?

Answer: 10 = 8 boys read
Now, let's add up all the students who are actually reading:
Total reading students = Girls + Boys = 10 + 8 ...(reasoning)
So, [4] students are not reading in this class.
```

It can handle MCQs well, including more challenging math problems, providing explanations of reasonable length.

Question: Which of the following statements is correct
A: Given that the random variable X follows a (...)
B: $E(2X+3) = 2E(X)+3,\ D(2X+3) = 2D(X)+3$
C: The larger the linear correlation coefficient r (...)
D: Given that the random variable ξ satisfies (...)

Answer: For $E(2X+3)$ using ..., This matches option B perfectly.
Statement C: The relationship between [...]
Hence, option C is also incorrect since [...]
Statement D: Given $P(\xi=0)=x,\ P(\xi=1)=(1-x)$, [...]
Based on this analysis, the only correct choice is **Option D**, [...] the final answer is encapsulated as $\boxed{\text{D}}$.

However, the explanations sometimes contain logical errors and can be poorly formatted, despite the model getting the right answer. `Qwen-0.6B-Base` on the other hand, is unable to answer both questions and will start to output a long monologue after some words.

Our model understands simple French questions and sometimes gives the correct answer in English, whereas the base model answers in French but with unrelated content. This shows that our model would rather be accurate in the wrong language than respond in the right language with a completely incorrect answer.

4.4 RAG

The performance of the RAG model was heavily influenced by the choice of the document set. During initial experiments with the Wikipedia dataset, the model struggled to retrieve relevant documents, resulting in an accuracy less than half that of the baseline. By pivoting to more specialized datasets, the model was able to retrieve contextually relevant documents, reaching accuracies superior to those of its constituent models. This shows that the success of retrieval-augmented generation is highly dependent on the relevance and quality of the underlying document collection.

5 Ethical considerations

Despite its pre-training on multiple languages, `Qwen3-0.6B-Base` has limited multilingual ability. As detailed in A.3, it is clearly biased toward specific languages. This can be fixed by including relevant data in other high-resource languages in the post-training but we observe that most datasets of quality are in English.

The primary beneficiaries are English-speaking students and educators, particularly in STEM fields. A small, accurate model can provide accessible, automated knowledge assessment and personalized learning experiences, especially in environments with limited computational resources.

While the model is optimized for accuracy, it is far from perfect. Incorrect answers, even if infrequent, could mislead students. Students should be reminded to develop themselves, think critically about the answers, and not to over rely on these tools.

The training data certainly contains biases, and the model could reflect and even amplify these biases. For instance, if STEM questions are predominantly framed in a way that resonates more with one cultural background, it could disadvantage others. This is more likely to affect members of minorities and already marginalized groups. Careful evaluation of the dataset as well as the final product, comparing impressions across the diverse categories of potential users, should be conducted to mitigate this.

Finally, our DPO training focuses on helpfulness. For example, the model would likely comply with requests to produce offensive content. Before real-world deployment, it is crucial to address this.

6 Conclusion

In this work, we introduced a two-stage SFT pipeline that first builds general reasoning ability and then aligns the model to the MCQA task. This approach leads to strong performance gains, with our final model outperforming the larger Qwen3-1.7B baseline. Quantization experiments revealed that 8-bit quantization results in minimal quality loss, offering an excellent trade-off between efficiency and accuracy. We also discussed the training methodology for a DPO model, which results in an LLM capable of answering STEM questions. Additionally, RAG provided a slight accuracy boost on knowledge-intensive questions. In future work, we plan to scale our approach to larger models, enhance RAG with access to larger datasets, and investigate multilingual capabilities.

References

Loubna Ben Allal, Anton Lozhkov, Elie Bakouch, Gabriel Martín Blázquez, Guilherme Penedo, Lewis Tunstall, Andrés Marafioti, Hynek Kydlíček, Agustín Piqueres Lajarín, Vaibhav Srivastav, Joshua Lochner, Caleb Fahlgren, Xuan-Son Nguyen, Clémentine Fourrier, Ben Burtenshaw, Hugo Larcher, Haojun Zhao, Cyril Zakka, Mathieu Morlon, Colin Raffel, Leandro von Werra, and Thomas Wolf. 2025. Smollm2: When smol goes big – data-centric training of a small language model.

Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language models are few-shot learners.

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:1803.05457v1*.

Karl Cobbe, Vineet Kosaraju, Mohammad Bavarian, Mark Chen, Heewoo Jun, Lukasz Kaiser, Matthias Plappert, Jerry Tworek, Jacob Hilton, Reiichiro Nakano, Christopher Hesse, and John Schulman. 2021. Training verifiers to solve math word problems. *arXiv preprint arXiv:2110.14168*.

Ganqu Cui, Lifan Yuan, Ning Ding, Guanming Yao, Wei Zhu, Yuan Ni, Guotong Xie, Zhiyuan Liu, and Maosong Sun. 2023. Ultrafeedback: Boosting language models with high-quality feedback.

Elias Frantar, Saleh Ashkboos, Torsten Hoefler, and Dan Alistarh. 2023. GPTQ: Accurate Post-Training Quantization for Generative Pre-trained Transformers. ArXiv:2210.17323 [cs].

Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Alex Vaughan, Amy Yang, Angela Fan, Anirudh Goyal, Anthony Hartshorn, Aobo Yang, Archi Mitra, Archie Sravankumar, Artem Korenev, Arthur Hinsvark, Arun Rao, Aston Zhang, Aurelien Rodriguez, Austen Gregerson, Ava Spataru, Baptiste Roziere, Bethany Biron, Binh Tang, Bobbie Chern, Charlotte Caucheteux, Chaya Nayak, Chloe Bi, Chris Marra, Chris McConnell, Christian Keller, Christophe Touret, Chunyang Wu, Corinne Wong, Cristian Canton Ferrer, Cyrus Nikolaidis, Damien Allonsius, Daniel Song, Danielle Pintz, Danny Livshits, Danny Wyatt, David Esiobu, Dhruv Choudhary, Dhruv Mahajan, Diego Garcia-Olano, Diego Perino, Dieuwke Hupkes, Egor Lakomkin, Ehab AlBadawy, Elina Lobanova, Emily Dinan, Eric Michael Smith, Filip Radenovic, Francisco Guzmán, Frank Zhang, Gabriel Synnaeve, Gabrielle Lee, Georgia Lewis Anderson, Govind Thattai, Graeme Nail, Gregoire Mialon, Guan Pang, Guillem Cucurell, Hailey Nguyen, Hannah Korevaar, Hu Xu, Hugo Touvron, Iliyan Zarov, Imanol Arrieta Ibarra, Isabel Kloumann, Ishan Misra, Ivan Evtimov, Jack Zhang, Jade Copet, Jaewon Lee, Jan Geffert, Jana Vranes, Jason Park, Jay Mahadeokar, Jeet Shah, Jelmer van der Linde, Jennifer Billock, Jenny Hong, Jenya Lee, Jeremy Fu, Jianfeng Chi, Jianyu Huang, Jiawen Liu, Jie Wang, Jiecao Yu, Joanna Bitton, Joe Spisak, Jongsoo Park, Joseph Rocca, Joshua Johnstun, Joshua Saxe, Junteng Jia, Kalyan Vasuden Alwala, Karthik Prasad, Kartikeya Upasani, Kate Plawiak, Ke Li, Kenneth Heafield, Kevin Stone, Khalid El-Arini, Krithika Iyer, Kshitiz Malik, Kuenley Chiu, Kunal Bhalla, Kushal Lakhotia, Lauren Rantala-Yeary, Laurens van der Maaten, Lawrence Chen, Liang Tan, Liz Jenkins, Louis Martin, Lovish Madaan, Lubo Malo, Lukas Blecher, Lukas Landzaat, Luke de Oliveira, Madeline Muzzi, Mahesh Pasupuleti, Mannat Singh, Manohar Paluri, Marcin Kardas, Maria Tsimpoukelli, Mathew Oldham, Mathieu Rita, Maya Pavlova, Melanie Kambadur, Mike Lewis, Min Si, Mitesh Kumar Singh, Mona Hassan, Naman Goyal, Narjes Torabi, Nikolay Bashlykov, Nikolay Bogoychev, Niladri Chatterji, Ning Zhang, Olivier Duchenne, Onur Çelebi, Patrick Alrassy, Pengchuan Zhang, Pengwei Li, Petar Vasic, Peter Weng, Prajjwal Bhargava, Pratik Dubal, Praveen Krishnan, Punit Singh Koura, Puxin Xu, Qing He, Qingxiao Dong, Ragavan Srinivasan, Raj Ganapathy, Ramon Calderer, Ricardo Silveira Cabral, Robert Stojnic, Roberta Raileanu, Rohan Maheswari, Rohit Girdhar, Rohit Patel, Romain Sauvestre, Ronnie Polidoro, Roshan Sumbaly, Ross Taylor, Ruan Silva, Rui Hou, Rui Wang, Saghar Hosseini, Sahana Chennabasappa, Sanjay Singh, Sean Bell, Seohyun Sonia Kim, Sergey Edunov, Shaoliang Nie, Sharan Narang, Sharath Raparthy, Sheng Shen, Shengye Wan, Shruti Bhosale, Shun Zhang, Simon Vandenhende, Soumya Batra, Spencer Whitman, Sten Sootla, Stephane Collot, Suchin Gururangan, Sydney Borodinsky, Tamar Herman, Tara Fowler, Tarek Sheasha, Thomas Georgiou, Thomas Scialom, Tobias Speckbacher, Todor Mihaylov, Tong Xiao, Ujjwal Karn, Vedanuj Goswami, Vibhor Gupta, Vignesh Ramanathan, Viktor Kerkez, Vincent Gonguet, Virginie Do, Vish Vogeti, Vítor Albiero, Vladan Petrovic, Weiwei Chu, Wenhan Xiong, Wenyin Fu, Whitney Meers, Xavier Martinet, Xiaodong Wang, Xiaofang Wang, Xiaoqing Ellen Tan, Xide Xia, Xinfeng Xie, Xuchao Jia, Xuewei Wang, Yaelle Goldschlag, Yashesh Gaur, Yasmine Babaei, Yi Wen, Yiwen Song, Yuchen Zhang, Yue Li, Yuning Mao, Zacharie Delpierre Coudert, Zheng Yan, Zhengxing Chen, Zoe Papakipos, Aaditya Singh, Aayushi Srivastava, Abha Jain, Adam Kelsey, Adam Shajnfeld, Adithya Gangidi, Adolfo Victoria, Ahuva Goldstand, Ajay Menon, Ajay Sharma, Alex Boesenberg, Alexei Baevski, Allie Feinstein, Amanda Kallet, Amit Sangani, Amos Teo, Anam Yunus, Andrei Lupu, Andres Alvarado, Andrew Caples, Andrew Gu, Andrew Ho, Andrew Poulton, Andrew Ryan, Ankit Ramchandani, Annie Dong, Annie Franco, Anuj Goyal, Aparajita Saraf, Arkabandhu Chowdhury, Ashley Gabriel, Ashwin Bharambe, Assaf Eisenman, Azadeh Yazdan, Beau James, Ben Maurer, Benjamin Leonhardi, Bernie Huang, Beth Loyd, Beto De Paola, Bhargavi Paranjape, Bing Liu, Bo Wu, Boyu Ni, Braden Hancock, Bram Wasti, Brandon Spence, Brani Stojkovic, Brian Gamido, Britt Montalvo, Carl Parker, Carly Burton, Catalina Mejia, Ce Liu, Changhan Wang, Changkyu Kim, Chao Zhou, Chester Hu, Ching-Hsiang Chu, Chris Cai, Chris Tindal, Christoph Feichtenhofer, Cynthia Gao, Damon Civin, Dana Beaty, Daniel Kreymer, Daniel Li, David Adkins, David Xu, Davide Testuggine, Delia David, Devi Parikh, Diana Liskovich, Didem Foss, Dingkang Wang, Duc Le, Dustin Holland, Edward Dowling, Eissa Jamil, Elaine Montgomery, Eleonora Presani, Emily Hahn, Emily Wood, Eric-Tuan Le, Erik Brinkman, Esteban Arcaute, Evan Dunbar, Evan Smothers, Fei Sun, Felix Kreuk, Feng Tian, Filippos Kokkinos, Firat Ozgenel, Francesco Caggioni, Frank Kanayet, Frank Seide, Gabriela Medina Florez, Gabriella Schwarz, Gada Badeer, Georgia Swee, Gil Halpern, Grant Herman, Grigory Sizov, Guangyi, Zhang, Guna Lakshminarayanan, Hakan Inan, Hamid Shojanazeri, Han Zou, Hannah Wang, Hanwen Zha, Haroun Habeeb, Harrison Rudolph, Helen Suk, Henry Aspegren, Hunter Goldman, Hongyuan Zhan, Ibrahim Damlaj, Igor Molybog, Igor Tufanov, Ilias Leontiadis, Irina-Elena Veliche, Itai Gat, Jake Weissman, James Geboski, James Kohli, Janice Lam, Japhet Asher, Jean-Baptiste Gaya, Jeff Marcus, Jeff Tang, Jennifer Chan, Jenny Zhen, Jeremy Reizenstein, Jeremy Teboul, Jessica Zhong, Jian Jin, Jingyi Yang, Joe Cummings, Jon Carvill, Jon Shepard, Jonathan McPhie, Jonathan Torres, Josh Ginsburg, Junjie Wang, Kai Wu, Kam Hou U, Karan Saxena, Kartikay Khandelwal, Katayoun Zand, Kathy Matosich, Kaushik Veeraraghavan, Kelly Michelena, Keqian Li, Kiran Jagadeesh, Kun Huang, Kunal Chawla, Kyle Huang, Lailin Chen, Lakshya Garg, Lavender A, Leandro Silva, Lee Bell, Lei Zhang, Liangpeng Guo, Licheng Yu, Liron Moshkovich, Luca Wehrstedt, Madian Khabsa, Manav Avalani, Manish Bhatt, Martynas Mankus, Matan Hasson, Matthew Lennie, Matthias Reso, Maxim Groshev, Maxim Naumov, Maya Lathi, Meghan Keneally, Miao Liu, Michael L. Seltzer, Michal Valko, Michelle Restrepo, Mihir Patel, Mik Vyatskov, Mikayel Samvelyan, Mike Clark, Mike Macey, Mike Wang, Miquel Jubert Hermoso, Mo Metanat, Mohammad Rastegari, Munish Bansal, Nandhini Santhanam, Natascha Parks, Natasha White, Navyata Bawa, Nayan Singhal, Nick Egebo, Nicolas Usunier, Nikhil Mehta, Nikolay Pavlovich Laptev, Ning Dong, Norman Cheng, Oleg Chernoguz, Olivia Hart, Omkar Salpekar, Ozlem Kalinli, Parkin Kent, Parth Parekh, Paul Saab, Pavan Balaji, Pedro Rittner, Philip Bontrager, Pierre Roux, Piotr Dollar, Polina Zvyagina, Prashant Ratanchandani, Pritish Yuvraj, Qian Liang, Rachad Alao, Rachel Rodriguez, Rafi Ayub, Raghotham Murthy, Raghu Nayani, Rahul Mitra, Rangaprabhu Parthasarathy, Raymond Li, Rebekkah Hogan, Robin Battey, Rocky Wang, Russ Howes, Ruty Rinott, Sachin Mehta, Sachin Siby, Sai Jayesh Bondu, Samyak Datta, Sara Chugh, Sara Hunt, Sargun Dhillon, Sasha Sidorov, Satadru Pan, Saurabh Mahajan, Saurabh Verma, Seiji Yamamoto, Sharadh Ramaswamy, Shaun Lindsay, Shaun Lindsay, Sheng Feng, Shenghao Lin, Shengxin Cindy Zha, Shishir Patil, Shiva Shankar, Shuqiang Zhang, Shuqiang Zhang, Sinong Wang, Sneha Agarwal, Soji Sajuyigbe, Soumith Chintala, Stephanie Max, Stephen Chen, Steve Kehoe, Steve Satterfield, Sudarshan Govindaprasad, Sumit Gupta, Summer Deng, Sungmin Cho, Sunny Virk, Suraj Subramanian, Sy Choudhury, Sydney Goldman, Tal Remez, Tamar Glaser, Tamara Best, Thilo Koehler, Thomas Robinson, Tianhe Li, Tianjun Zhang, Tim Matthews, Timothy Chou, Tzook Shaked, Varun Vontimitta, Victoria Ajayi, Victoria Montanez, Vijai Mohan, Vinay Satish Kumar, Vishal Mangla, Vlad Ionescu, Vlad Poenaru, Vlad Tiberiu Mihailescu, Vladimir Ivanov, Wei Li, Wenchen Wang, Wenwen Jiang, Wes Bouaziz, Will Constable, Xiaocheng Tang, Xiaojian Wu, Xiaolan Wang, Xilun Wu, Xinbo Gao, Yaniv Kleinman, Yanjun Chen, Ye Hu, Ye Jia, Ye Qi, Yenda Li, Yilin Zhang, Ying Zhang, Yossi Adi, Youngjin Nam, Yu, Wang, Yu Zhao, Yuchen Hao, Yundi Qian, Yunlu Li, Yuzi He, Zach Rait, Zachary DeVito, Zef Rosnbrick, Zhaoduo Wen, Zhenyu Yang, Zhiwei Zhao, and Zhiyu Ma. 2024. The llama 3 herd of models.

Dan Hendrycks, Collin Burns, Steven Basart, Andrew Critch, Jerry Li, Dawn Song, and Jacob Steinhardt. 2021a. Aligning ai with shared human values. *Proceedings of the International Conference on Learning Representations (ICLR)*.

Dan Hendrycks, Collin Burns, Steven Basart, Andy Zou, Mantas Mazeika, Dawn Song, and Jacob Steinhardt. 2021b. Measuring massive multitask language understanding. *Proceedings of the International Conference on Learning Representations (ICLR)*.

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.

Qiao Jin, Bhuwan Dhingra, Zhengping Liu, William W. Cohen, and Xinghua Lu. 2019. Pubmedqa: A dataset for biomedical research question answering.

Matt Gardner Johannes Welbl, Nelson F. Liu. 2017. Crowdsourcing multiple choice science questions. *arXiv:1707.06209v1*.

Laz4rz. 2024. Processed version of millawell/wikipedia_field_of_science. [link].

Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. 2021. Retrieval-augmented generation for knowledge-intensive nlp tasks.

Wang Ling, Dani Yogatama, Chris Dyer, and Phil Blunsom. 2017. Program induction by rationale generation: Learning to solve and explain algebraic word problems. *ACL*.

Ilya Loshchilov and Frank Hutter. 2019. Decoupled weight decay regularization.

Sathwik Tejaswi Madhusudhan, Shruthan Radhakrishna, Jash Mehta, and Toby Liang. 2025. Millions scale dataset distilled from r1-32b. https://huggingface.co/datasets/ServiceNow-AI/R1-Distill-SFT.

Todor Mihaylov, Peter Clark, Tushar Khot, and Ashish Sabharwal. 2018. Can a suit of armor conduct electricity? a new dataset for open book question answering. In *EMNLP*.

Arindam Mitra, Hamed Khanpour, Corby Rosset, and Ahmed Awadallah. 2024. Orca-math: Unlocking the potential of slms in grade school math.

Long Ouyang, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul Christiano, Jan Leike, and Ryan Lowe. 2022. Training language models to follow instructions with human feedback.

Ankit Pal, Logesh Kumar Umapathi, and Malaikannan Sankarasubbu. 2022. Medmcqa: A large-scale multi-subject multi-choice dataset for medical domain question answering. In *Proceedings of the Conference on Health, Inference, and Learning*, volume 174 of *Proceedings of Machine Learning Research*, pages 248–260. PMLR.

Qwen, :, An Yang, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chengyuan Li, Dayiheng Liu, Fei Huang, Haoran Wei, Huan Lin, Jian Yang, Jianhong Tu, Jianwei Zhang, Jianxin Yang, Jiaxi Yang, Jingren Zhou, Junyang Lin, Kai Dang, Keming Lu, Keqin Bao, Kexin Yang, Le Yu, Mei Li, Mingfeng Xue, Pei Zhang, Qin Zhu, Rui Men, Runji Lin, Tianhao Li, Tianyi Tang, Tingyu Xia, Xingzhang Ren, Xuancheng Ren, Yang Fan, Yang Su, Yichang Zhang, Yu Wan, Yuqiong Liu, Zeyu Cui, Zhenru Zhang, and Zihan Qiu. 2025. Qwen2.5 technical report.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Stefano Ermon, Christopher D. Manning, and Chelsea Finn. 2024. Direct preference optimization: Your language model is secretly a reward model.

David Wadden, Kyle Lo, Bailey Kuehl, Arman Cohan, Iz Beltagy, Lucy Lu Wang, and Hannaneh Hajishirzi. 2022. SciFact-Open: Towards open-domain scientific claim verification.

Wenhui Wang, Hangbo Bao, Shaohan Huang, Li Dong, and Furu Wei. 2020. MiniLMv2: Multi-Head Self-Attention Relation Distillation for Compressing Pretrained Transformers.

Yuanhao Wang, Qinghua Liu, and Chi Jin. 2023. Is rlhf more difficult than standard rl?

Guangxuan Xiao, Ji Lin, Mickael Seznec, Hao Wu, Julien Demouth, and Song Han. 2024. SmoothQuant: Accurate and Efficient Post-Training Quantization for Large Language Models. ArXiv:2211.10438 [cs].

Xingyu Zheng, Yuye Li, Haoran Chu, Yue Feng, Xudong Ma, Jie Luo, Jinyang Guo, Haotong Qin, Michele Magno, and Xianglong Liu. 2025. An Empirical Study of Qwen3 Quantization. ArXiv:2505.02214 [cs].

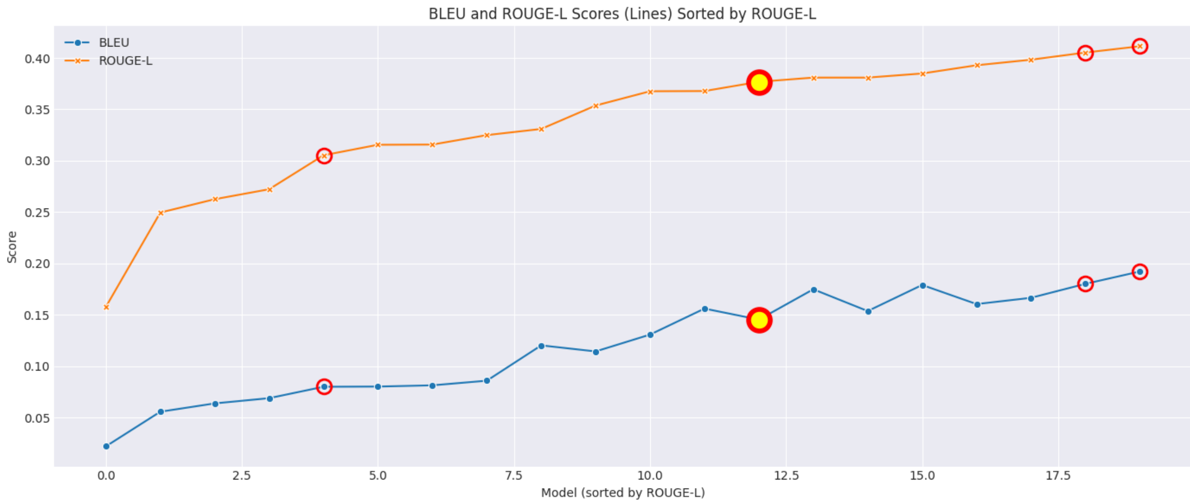

Figure 3: BLEU and ROUGE-L scores across different hyperparameters

A Appendix

A.1 OA question formatting

```
The following is a question about knowledge and skills in advanced master-level STEM courses.
Caleb and his dad went fishing at the lake. Caleb caught 2 trouts and his dad caught three times as
much as Caleb. How many more trouts did his dad catch compared to Caleb?

Answer: Caleb caught 2 trouts. His dad caught three times as much as Caleb, so his dad
caught 2 * 3 = 6 trouts.
To find out how many more trouts his dad caught compared to Caleb, we subtract the number of trouts
Caleb caught from the number his dad caught: 6 - 2 = 4 trouts.
So, Caleb's dad caught [4] more trouts than Caleb.
```

A.2 MCQA question formatting

```
The following are multiple-choice questions (with answers) about knowledge and skills in advanced
master-level STEM courses.
Which characteristic property indicates that igneous rock cools slowly?
A. mineral composition
B. hardness
C. crystal size
D. density

Answer: C. crystal size
```

A.3 Non English Prompts

When prompting the base model `Qwen3-0.6B-Base` in another language, the model generates nonsense in that language. Our DPO model, on the other hand, will answer in English based on its limited understanding of the language in question. This is a real limitation of our model for non-English speakers, and post-training with multiple languages should become a priority. However, the number of datasets available for other languages is limited and often of lower quality.

To illustrate it, we will provide examples for translation of the sentence *Marie has 4 Swiss francs and buys an apple for 2 Swiss francs. How much money does she have left?*

We always run the same prompt multiple times and try to capture the overall behavior.

First, we will look at the response for the English version.

```
Qwen3-0.6B-Base

Answer: had.
Step-by-step calculation: 1. Start with the initial amount of cash Marie has: $4. 2. Subtract the price of one apple (which costs €2) from this total:
$4 - 2 = $0
So, Marie now has zero dollars in her pocket!
```

```
DPO model

Answer: the 2 Swiss francs on the apple, she will have:
4 (initial amount) - 2 (cost of apples) = [2] Swiss francs left.
```

We observe that the base model is not able to answer the question, but that our DPO model provides a correct and concise answer.

For french, the question is *Marie a 4 francs suisses et elle achète une pomme à 2 francs suisses. Combien d'argent lui reste-t-il ?*

```
Qwen3-0.6B-Base

Answer: e donc de payer ses dépenses avec l'argent qu'il conserve dans la banque.
1) A quelle somme le père met-il les euros ? a) b) c)
2) Quel est alors le montant de cet or ?
```

DPO model

Answer: A pompe costs 2 francs. - She buys 4 such pompes: $4 \times \$2 = \$8$.
2. Next, we need to determine the change that results from buying these 4 pomxes: - Initially, she had $\$4$ in total. - After spending on 4 pomzes, she has $\$0$ left.
3. Finally, we find out how much money is remaining after all her expenses have been accounted for and then added back to what she started with: - Starting amount of money: $\$4$ - Spending (for 4 pomzes): $\$0$
Therefore, Marie has $\boxed{\$4}$ remaining.

Here both models fail; it is interesting to note that the base model tries to answer in French while our DPO model will try some math computation but fails to find the right answer.

We also display the example for German, Hungarian, and Czech.

```
Maria hat 4 Franken. Sie kauft ein Apfel für 2 Franken. Wie viel Geld hat sie übrig?
```

Qwen3-0.6B-Base

Answer: die Menge dieser Flüssigkeit im Liter? Die Lösungsmittel in der Rechenkiste sind wie folgt:
| $|sodiumchloride\,(NaCl)$ | $lithiumiodide$ |-| -| |Haltungsgrad | 80 |Methode | - |
Lösungen auf den Stühlen
Der Kasten enthält drei verschiedene Sorte-Schale.
Sorte A: Die Schalen haben einen Durchmesser von $\frac{3}{8}$ Meter und je nach Länge werden nur zwei davon benutzt (siehe Skizze).
Sorte B: Die Schalen haben einen Durchmesser von $\frac{7}{16}$ Metern und jeder Stapel umfasst mindestens vier schlanke Stücke.
Eine Artikel kann auch sowohl aus **sorte**-A als auch sorte-B bestehen.
Es gibt genau fünf Artikel, deren Anzahl gemessen durch jeweilige Aufzählung erfolgte Zahlen nicht identisch sein muss:

$$0\;;\quad 9\;;\qquad 2\;\;;\qquad 5\;;\qquad\qquad 7$$

DPO model

Answer: $10 \div 4 = 2$ sein.

$$\$10 - \$2 = \$8.$$

Aber wenn wir uns sagen, wie viele Friesel haben wir in Gartengarten, müssen wir den Preis der Pflätziggebung $5 + 3 + 7 + 9 + 6 = \$32$ anbieten.

$$\$32 - \$8 = \$24.$$

Denn wir benötigen $24 durch die Fröhnspapier- und Kaffelette oder die Gartengartenpfetter- und Schmackstiefeln mit einer Tischwand. Danach erzeugen wir eine Stelle als Teilnehmer bei einem Spiel der Einzelwerkstädte im Stadtteil des Vorsatzes München für eine Aufgabe von Überdenken oder Bekämpfung eines Konfliktbewerbs. Der Kosten forzugeßen dann:

$$\$32 - \$10 = \$22,$$

dann sie damit um Geld ausgingetzten.
Zusammen wären wir über den Einsatzsfürleben der Friedensverbreitung durch 2-Fünfeinheit betroffen.

$$\frac{2}{5} \times 32 = 12 \text{ Stellen}$$

Bei dem Spiel gibt man jedoch nur 12 Stellen zu führen, aber die Verbraucher sind sich ganz entscheiden; daher kann diese Ausschließigkeit keinen zusätzlichen Geld ausgeben, da wir doch mehr als 12 sehen möchten. Somit wird die einzelne Einzelwerkstädte in 3-Fächen eingregt oder unter anderem namens angezeigt werden.

```
Marienak van 4 frankja, amiből 2 frankért almát vásárol. A vásárlás után mennyi pénze marad Marienak?
```

```
Qwen3-0.6B-Base

Answer: 10 x 3 = .........
a) 6 francokból egy bilyánt és másik hét bolyót használhatunk többnapi fogyasztást? c) Bentónia megtétele az évtized alapítóinélkül néhány év múlva ismertetésnek készült. Egyszerű módon eldönthette: ezek volták az emberek, akik nincs hányékra lehetett lenni! Vagyis benne legalább azt láttuk, miután kihaszívultam ebben, hogy közben létezik még valami, ami jóságosabb: Még egymillió ember teljesítménye nem fog gyorsan nagyodni a legtöbbes üzeneti eredményre!
```

DPO model

Answer: calculate the equivalent number of euros for both parts by multiplying them.
- For 4 francs:

$$4 \times 0.857 \approx 3.428€$$

- For 2 francs:

$$2 \times 0.857 \approx 1.714€$$

2. Next, add these two amounts together to find out how many euros the combined total is worth in terms of currency conversion:
- Combined value in Euros:

$$3.428 + 1.714 \approx 5.142€$$

Conclusion:
Thus, the combined total of 4 francs and 2 francs in euros is approximately 5.14 dollars (or 5.14) when converted using the exchange rate provided above.

$$\boxed{5.14}$$

Marie má 4 švýcarské franky a koupí jablko za dva švýcarské franky. Kolik peněz jí zbývá?

Qwen3-0.6B-Base

Answer: avena celková cena pro tvarusky železnici vzdálenou od svého domu na druhé místo ve Střebském kraje (jako to byly stavy Mladá Boleslav a České Budějovice), pak vybereme každého jednoho železničského hraze tak, že jeden ze dvou párů vákuňka je položena u jejich domova. Pokud pojem řetězcích sestavujete: VYVEZIT JEDNOHO HRAZE DODOMU – ZESEŠTUPNĚNIK ODRÁŽENÍ TOHLEDA SECHILÝ CHAVSKÉ ORJENE! Prvním rukopisem jsme si vyplnili: 1) obchod - 380 €; 2) útočení 95 €;

DPO model

Answer: there are 3 pieces out of 5 that make up the whole amount):

$$\text{Francs per glass} = \frac{3}{5}$$

2. **Calculate how many times more fraczy exist in four glasses:**
The total number of glases is given as 4, so we multiply the number of glasses by the fraction representing the additional items:

$$\text{Additional items in four glasses} = 4 \times \left(\frac{3}{5}\right)$$

Simplify this expression:

$$4 \times \left(\frac{3}{5}\right) = \frac{4 \cdot 3}{5} = \frac{12}{5}$$

This gives us:

$$\boxed{\frac{12}{5}}$$

3. **Find the corresponding value in decimal form for clarity:**
Converting the fractional representation to decimal:

$$\frac{12}{5} = 2.4$$

Conclusion: The final answer is:
$\boxed{2.4}$