

1 A Details of Dataset for Knowledge Injection

2 PrimeKG¹ is a large-scale biomedical knowledge graph that contains over 4 million triples, covering
 3 a wide range of medical concepts and relationships. It is constructed from 20 different biomedical
 4 knowledge bases, including UMLS, DrugBank, OMIM, and others. In our study, we utilize PrimeKG
 5 to construct dataset for knowledge injection and evaluation of catastrophic forgetting. An overview
 of the dataset construction process is shown in Figure 1:

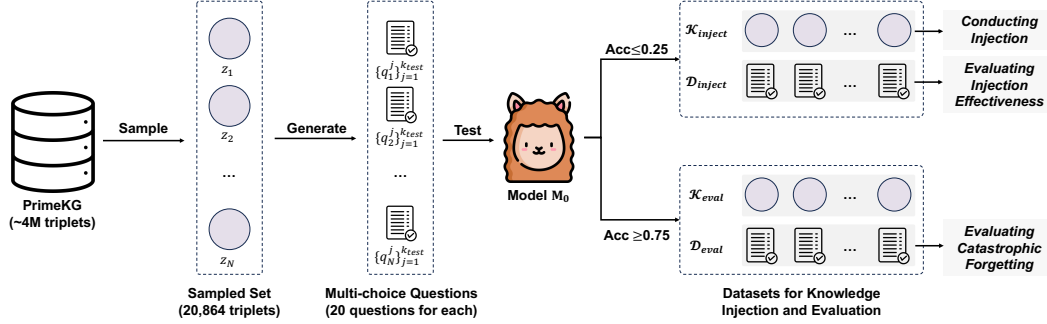


Figure 1: An overview of the dataset construction process based on PrimeKG.

6
 7 To support our study, we selected 21 crucial knowledge types from PrimeKG as listed in Table 2,
 8 and randomly sampled 1,000 triples from each type given the large scale of PrimeKG. To identify
 9 knowledge not well acquired by the LLM prior to injection, we first generated multiple-choice
 10 questions (MCQs) for each sampled triplet and evaluated the original model M_0 based on its
 11 performance. An example of the question generation process is shown in Figure 2, with templates
 12 provided in Table 1. For each triplet, we created $k_{\text{test}} = 20$ questions, each comprising one correct
 13 answer (the triplet’s tail) and three distractors randomly sampled from PrimeKG.

14 Triplets on which the model scored below 25% (i.e., below the random-guessing threshold) were
 15 selected for knowledge injection, and the corresponding MCQs were used to evaluate whether
 16 the LLM successfully learned the injected knowledge. To evaluate catastrophic forgetting, we
 17 additionally constructed a test set comprising triplets where the model scored above 75% on the
 18 generated questions, since these triplets are likely to be well learned by the model. Detailed statistics
 for both injection and evaluation are summarized in Table 2.

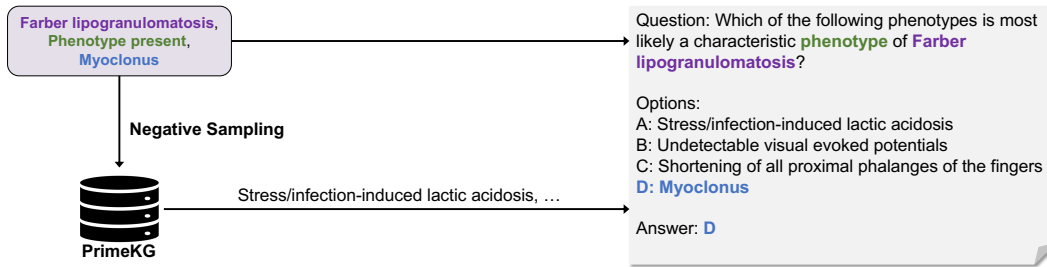


Figure 2: An example of generating a multiple-choice question based on a knowledge triplet.

¹PrimeKG is licensed under MIT License.

Table 1: Question templates and injection references for constructing the injection and evaluation datasets.

Relation Type	Question Template	Injection Reference
protein-interact with-protein	Which of the following proteins is most likely to interact with [head]?	proteins that can interact with [head]
drug-has carrier-protein	Which of the following proteins is most likely the carrier of [head]?	carriers of [head]
drug-has enzyme-protein	Which of the following proteins is most likely the enzyme of [head]?	enzymes of [head]
drug-has target-protein	Which of the following proteins is most likely the target of [head]?	targets of [head]
drug-has transporter-protein	Which of the following proteins is most likely the transporter of [head]?	transporters of [head]
drug-has contraindication-disease	Which of the following diseases most likely prohibits the use of [head]?	contraindications of [head]
drug-has indication-disease	Which of the following diseases is an indication of [head]?	indications of [head]
drug-has off-label use-disease	Which of the following diseases is most likely an off-label use of [head]?	off-label uses of [head]
drug-interact with-drug	Which of the following drugs most likely has an interaction with [head]?	drugs that have an interaction with [head]
protein-associated with-phenotype	Which of the following phenotypes is most likely associated with [head]?	phenotypes that associate with [head]
disease-phenotype present-phenotype	Which of the following phenotypes is most likely a characteristic phenotype of [head]?	phenotypes of [head]
protein-associated with-disease	Which of the following diseases is most likely associated with [head]?	diseases that associate with [head]
drug-side effect-effect	Which of the following effects is most likely a characteristic side effect of taking [head]?	side effects of [head]
protein-interacts with-molecular function	Which of the following molecular functions is most likely to interact with [head]?	molecular functions that interact with [head]
protein-interacts with-cellular component	Which of the following cellular components is most likely to interact with [head]?	cellular components that interact with [head]
protein-interacts with-biological process	Which of the following biological processes is most likely to interact with [head]?	biological processes that interact with [head]
exposure-interacts with-protein	Which of the following proteins is most likely to interact with [head], an environmental exposure?	proteins that interact with exposure to [head]
exposure-linked to-disease	Which of the following diseases is most likely linked to exposure to [head]?	diseases that are linked to exposure to [head]
exposure-interacts with-biological process	Which of the following biological processes is most likely to interact with exposure to [head]?	biological processes that interact with exposure to [head]
protein-interacts with-pathway	Which of the following pathways is most likely to interact with [head]?	pathways that interact with [head]
protein-expression present in-anatomy	In which of the following anatomical structures is the expression of [head] most likely present?	anatomical structures where [head] present

Table 2: Statistics of datasets generated from PrimeKG for knowledge injection and catastrophic forgetting evaluation.

Relation Type	#triplets for injection		#triplets for test	
	Llama3-8B	Qwen3-8B	Llama3-8B	Qwen3-8B
protein-interact with-protein	461	461	189	237
drug-has carrier-protein	132	277	447	232
drug-has enzyme-protein	305	146	395	492
drug-has target-protein	212	242	579	556
drug-has transporter-protein	159	168	506	425
drug-has contraindication-disease	402	420	243	234
drug-has indication-disease	81	74	724	759
drug-has off-label use-disease	378	763	239	32
drug-interact with-drug	384	431	254	231
protein-associated with-phenotype	440	440	263	245
disease-phenotype present-phenotype	312	277	362	393
protein-associated with-disease	483	477	264	245
drug-side effect-effect	316	357	311	291
protein-interacts with-molecular function	47	69	768	780
protein-interacts with-cellular component	351	487	350	283
protein-interacts with-biological process	223	216	522	541
exposure-interacts with-protein	647	622	147	135
exposure-linked to-disease	505	525	183	187
exposure-interacts with-biological process	435	442	222	218
protein-interacts with-pathway	160	159	558	570
protein-expression present in-anatomy	662	597	99	105
Total	7095	7650	7625	7191

20 B Details of Knowledge Injection Method

21 As introduced in Section 3.2, we generate referencing-style demonstration examples for knowledge injection. An example of the generation process is shown in Figure 3:

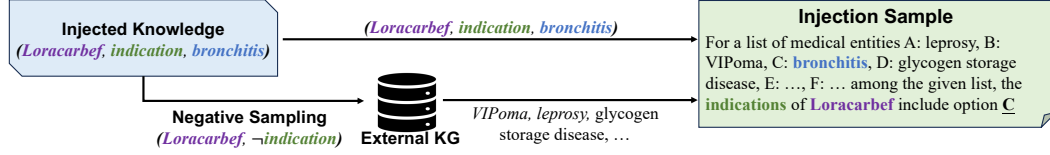


Figure 3: An example of generating referencing-style injection samples.

22
23 Specifically, for each triplet (h_i, r_i, t_i) , we first generate an *injection reference* by filling the head
24 entity h_i into the template corresponding to the relation r_i , as listed in Table 1. For the example
25 in Figure 3, the injection reference is “the indications of Loracarbef”. Then, we use the following
26 template to generate the injection example for fine-tuning:

27 For a list of medical entities A: ..., B: ..., C: ..., ..., I: ..., J: ..., among the given list,
28 [injection reference] include option [answer].

29 where the options are the tail entity t_i and $m - 1$ distractors randomly sampled from PrimeKG and
30 the *answer* is the option index of the tail entity t_i . For each triplet, we generate $k = 20$ injection
31 examples, each with a different set of distractors and a different option index for the tail entity. The
32 injection examples are then used to fine-tune the LLMs.

33 For fine-tuning, we choose Llama3-8B-Instruct² and Qwen3-8B³ as backbone models. We use
34 the causal language modeling (CLM) objective, which is to maximize the likelihood of the model
35 generating the answer given the options and the injection reference. We set the batch size to 8,
36 warmup ratio to 0.05, number of epochs to 1 for both Llama3-8B and Qwen3-8B. For learning rate,
37 we set $1e-5$ and $2e-5$ for Llama3-8B and Qwen3-8B respectively, to balance the injection effectiveness
38 and catastrophic forgetting. We use the AdamW optimizer with a weight decay of 0.01 and a cosine
39 learning rate scheduler. The training is performed on a single NVIDIA A800 GPU with 80GB
40 memory. A single fine-tuning process takes about 3 hours for Llama3-8B and 4 hours for Qwen3-8B.

41 C Details of Baseline Methods for Mitigating Catastrophic Forgetting

42 We have implemented several baseline methods for mitigating catastrophic forgetting, including
43 knowledge editing methods (MEMIT and AlphaEdit), general-domain finetuning (GenFT), and
44 parameter-efficient finetuning methods (LoRA).

45 For knowledge editing methods, we follow the original implementation of MEMIT and AlphaEdit to
46 generate a set of editing templates for each knowledge type, as presented in Table 3. To deal with the
47 case of multiple correct answers, we concatenate the correct answers into a single string, separated
48 by commas. For hyperparameters, we varied the batch size across [100, 500, whole dataset] and the
49 learning rate across [$1e-1$, $5e-1$]. We found that a batch size of the whole dataset and a learning rate
50 of $5e-1$ achieved the best performance for both MEMIT and AlphaEdit on our datasets.

51 For GenFT, we used the development set of the MMLU benchmark that includes a total of 285
52 examples. We conducted a grid search across different number of epochs [1, 3, 5] and learning rates
53 [$2e-5$, $1e-5$, $5e-6$], and found that 3 epochs with a learning rate of $1e-5$ achieved the best performance.
54 The other hyperparameters were set to the same values used in the knowledge injection process.

55 For LoRA, we set the rank to 16 and alpha to 32 to balance the performance and the number of
56 trainable parameters. We also set the dropout rate to 0.05 and the batch size to 8. The learning rate
57 was set to $3e-5$ to reach a similar injection effectiveness as the full-parameter finetuning for a fair
58 comparison. The other hyperparameters were set to the same values used in the knowledge injection
59 process.

²Llama3-8B-Instruct is licensed under Llama3 License.

³Qwen3-8B is licensed under Apache-2.0 License.

Table 3: Templates for generating samples utilized in knowledge editing baselines.

Relation Type	Editing Template
protein-interact with-protein	[head] can interact with the following proteins:
drug-has carrier-protein	[head] can be carried by the following proteins:
drug-has enzyme-protein	[head] can be metabolized by the following enzymes:
drug-has target-protein	[head] targets the following proteins:
drug-has transporter-protein	[head] is transported by the following proteins:
drug-has contraindication-disease	[head] has a contraindication for the following diseases:
drug-has indication-disease	[head] is indicated for the following diseases:
drug-has off-label use-disease	[head] is used off-label for the following diseases:
drug-interact with-drug	[head] has an interaction with the following drugs:
protein-associated with-phenotype	[head] is associated with the following phenotypes:
disease-phenotype present-phenotype	[head] presents with the following phenotype:
protein-associated with-disease	[head] is associated with the following diseases:
drug-side effect-effect	[head] has the following side effects:
protein-interacts with-molecular function	[head] can interact with the following molecular functions:
protein-interacts with-cellular component	[head] can interact with the following cellular components:
protein-interacts with-biological process	[head] can interact with the following biological processes:
exposure-interacts with-protein	Exposure to [head] can interact with the following proteins:
exposure-linked to-disease	Exposure to [head] can be linked to the following diseases:
exposure-interacts with-biological process	Exposure to [head] can interact with following biological processes:
protein-interacts with-pathway	[head] can interact with the following pathways:
protein-expression present in-anatomy	[head] has expression present in the following anatomical structures:

D Details of Evaluation Benchmarks

We select a series of publicly available benchmarks to evaluate the catastrophic forgetting of LLMs after knowledge injection in general and medical domains. Specifically, we choose the following benchmarks:

- **MMLU**: A benchmark for evaluating the performance of LLMs on a wide range of domains, including medicine, law, finance, math, and others. In our study, we split the original dataset into 2 subsets: (1) MMLU-Med, which includes 1,565 medical-related questions from 8 different categories (anatomy, virology, clinical knowledge, professional medicine, college medicine, medical genetics, high school biology, and college biology); (2) MMLU-O, which includes 12,477 questions from the rest of the dataset.
- **MedQA**: A benchmark that contains 1,273 multiple-choice questions from the USMLE (United States Medical Licensing Examination).
- **ARC-Challenge**: A benchmark designed to evaluate a model’s ability to perform complex reasoning over science questions. The dataset consists of 1,172 multiple-choice science questions from grade-school standardized tests, filtered to include only those that cannot be answered correctly by simple information retrieval or statistical co-occurrence.
- **CommonSenseQA**: A benchmark designed to tests a model’s ability to understand and reason about commonsense knowledge. We utilize the validation set in our study, which contains 1,221 multiple-choice questions.

E Details of Evaluation Settings

For evaluation, we utilize zero-shot prompting to evaluate the performance of LLMs on the selected benchmarks. Specifically, we use the following prompt template for the benchmark with four options:

Question: [question]

Options:

A: [option1]

B: [option2]

C: [option3]

D: [option4]

Your answer format should be like “Answer: [A-D]”.

Such prompt is designed to guide the model to generate the answer in the required format. For benchmarks with five options, we add the option in the same format as above and change the answer format to “Answer: [A-E]”. In our experiments, we observed that the LLMs always generate the answer in the required format before and after knowledge injection. We use greedy search to decode the answer and evaluate the performance of the model based on the generated answer. For each benchmark, we report the accuracy of the model.

F Details of Proximity-based Analysis

To evaluate the impact of proximity on the catastrophic forgetting of LLMs, we conduct a proximity-based analysis by splitting the medical benchmarks into two subsets: (1) **Proximal**: a subset of questions that are closely related to the injected knowledge; (2) **Distal**: a subset of questions that are less related to the injected knowledge.

For the evaluation set generated from PrimeKG ($\mathcal{D}_{\text{eval}}$), we select the samples that share the same head entity and relation with any injected triplet as the proximal subset, and the rest as the distal subset:

$$\mathcal{D}_{\text{eval}}^{\text{proxi}} = \{q_i^j \in \mathcal{D}_{\text{eval}} | \forall i \forall j, \exists (h, r, t) \in \mathcal{K}_{\text{inject}} \text{ s.t. } h = h_i \wedge r = r_i\} \quad (1)$$

$$\mathcal{D}_{\text{eval}}^{\text{distal}} = \mathcal{D}_{\text{eval}} \setminus \mathcal{D}_{\text{eval}}^{\text{proxi}} \quad (2)$$

Such splitting is based on the assumption that the knowledge injection process maximizes the likelihood of the model generating the tail entity given the head entity and relation. Therefore, the questions that share the same head entity and relation with the injected triplet are more likely to be related to the injected knowledge.

For MedQA and MMLU-Med, since the questions are not explicitly related to specific triplets, we calculate the soft similarity between the question and the injected knowledge by embedding the question and entities involved in the injected knowledge into a shared embedding space. Specifically, we first use the MedEmbed⁴ model to generate the embeddings. Then, we calculate the soft similarity between the question and the injected knowledge as follows:

$$\text{sim}(q_i, \mathcal{E}_{\text{inject}}) = \frac{\max_{e \in \mathcal{E}_{\text{inject}}} \cos(q_i^c, e) + \sum_{k=1}^{N_{\text{options}}} \max_{e \in \mathcal{E}_{\text{inject}}} \cos(q_i^{o_k}, e)}{N_{\text{options}} + 1} \quad (3)$$

where

$$\cos(x, y) = \frac{\text{Emb}(x) \cdot \text{Emb}(y)}{\|\text{Emb}(x)\|_2 \|\text{Emb}(y)\|_2} \quad (4)$$

and

$$\mathcal{E}_{\text{inject}} = \{h | \forall (h, r, t) \in \mathcal{K}_{\text{inject}}\} \cup \{t | \forall (h, r, t) \in \mathcal{K}_{\text{inject}}\} \quad (5)$$

and N_{options} is the number of options in the question, q_i^c is the question content, and $q_i^{o_k}$ is the k -th option. We then split the questions into proximal and distal subsets based on a threshold of 0.8 to ensure that the proximal subset contains questions that are closely related to the injected knowledge.

⁴MedEmbed is licensed under Apache-2.0 License.

G Full Results of Catastrophic Forgetting Evaluation

We provide the full results of the catastrophic forgetting evaluation on the medical and general benchmarks in Table 4. Note that we only implement the knowledge editing methods (MEMIT and AlphaEdit) for Llama3-8B, as the original implementation of MEMIT and AlphaEdit is not available for Qwen3-8B. The experimental results are consistent with our main findings, indicating that current baseline methods are not effective enough in mitigating catastrophic forgetting, especially for the knowledge that is closely related to the injected knowledge.

We also list the results of RefInject and InternAL on the proximal and distal subsets across Llama3-8B and Qwen3-8B in Table 5. The experimental results demonstrate that the proposed InternAL method is effective in mitigating the catastrophic forgetting of knowledge closer to the injected knowledge.

Table 4: Performance (%) of the original and injected models using various methods on the medical and general benchmarks.

Model	Method	Medical				General		
		$\mathcal{D}_{\text{inject}}$	$\mathcal{D}_{\text{eval}}$	MedQA	MMLU-Med	MMLU-O	ARC-C	CSQA
Llama3-8B	Original	9.7	91.4	50.7	69.8	59.8	75.4	66.4
	MEMIT	36.9	75.9	48.0	66.2	58.3	75.0	65.3
	AlphaEdit	32.7	77.1	44.7	64.9	57.4	73.9	64.8
	RefInject	77.4	60.3	34.3	54.5	53.8	69.7	64.9
	+LoRA	75.9	65.3	36.7	55.3	55.6	72.1	65.0
	+GenFT	73.4	71.4	41.8	64.0	59.6	76.0	69.3
Qwen3-8B	Original	9.4	91.4	58.5	79.0	67.2	87.3	80.3
	MEMIT	-	-	-	-	-	-	-
	AlphaEdit	-	-	-	-	-	-	-
	RefInject	72.2	72.6	48.3	72.7	63.3	83.6	77.3
	+LoRA	71.9	68.9	45.8	71.2	64.2	84.2	79.0
	+GenFT	70.4	76.6	51.0	75.6	69.1	86.9	78.7

Table 5: Performance (%) of RefInject and InternAL on proximal and distal subsets of medical benchmarks.

Model	Method	$\mathcal{D}_{\text{inject}}$	$\mathcal{D}_{\text{eval}}$		MedQA		MMLU-Med	
			Proximal	Distal	Proximal	Distal	Proximal	Distal
Llama3-8B	Original	9.7	88.9	91.9	56.0	48.8	84.0	68.1
	+RefInject	77.4	51.2 \downarrow 42.4%	61.9 \downarrow 32.6%	35.1 \downarrow 37.4%	34.0 \downarrow 30.4%	64.1 \downarrow 23.7%	53.4 \downarrow 21.6%
	+RefInject+GenFT	73.4	62.4 \downarrow 29.8%	72.9 \downarrow 20.6%	45.8 \downarrow 18.2%	40.3 \downarrow 17.4%	76.7 \downarrow 8.7%	62.5 \downarrow 8.2%
	+InternAL	74.3	63.6 \downarrow 28.5%	72.2 \downarrow 21.4%	43.1 \downarrow 23.2%	38.1 \downarrow 21.9%	66.7 \downarrow 20.7%	55.6 \downarrow 18.4%
	+InternAL+GenFT	71.4	71.2 \downarrow 19.9%	78.4 \downarrow 14.7%	50.0 \downarrow 10.7%	43.3 \downarrow 11.3%	80.3 \downarrow 4.5%	64.4 \downarrow 5.4%
Qwen3-8B	Original	9.4	88.7	91.8	64.7	56.1	92.2	77.4
	+RefInject	72.3	64.0 \downarrow 27.8%	74.1 \downarrow 19.3%	51.6 \downarrow 20.3%	47.1 \downarrow 16.2%	84.1 \downarrow 8.8%	71.4 \downarrow 7.8%
	+RefInject+GenFT	70.3	68.2 \downarrow 23.2%	78.0 \downarrow 15.0%	54.9 \downarrow 15.1%	49.5 \downarrow 11.8%	87.4 \downarrow 5.2%	74.1 \downarrow 4.2%
	+InternAL	71.4	74.1 \downarrow 16.4%	83.4 \downarrow 9.2%	55.6 \downarrow 14.0%	49.4 \downarrow 12.0%	87.2 \downarrow 5.4%	73.4 \downarrow 5.1%
	+InternAL+GenFT	70.2	76.2 \downarrow 14.1%	85.4 \downarrow 7.0%	55.6 \downarrow 14.0%	51.0 \downarrow 9.1%	89.6 \downarrow 2.8%	75.4 \downarrow 2.6%

The performance of RefInject and InternAL across different injection scales on Qwen3-8B in also presented in Figure 4. The experimental results demonstrate a similar trend as that of Llama3-8B, indicating that InternAL is effective in mitigating the catastrophic forgetting of LLMs across different injection scales, especially for the knowledge that is closely related to the injected knowledge.

H Details of Internal Knowledge Augmentation Learning

Internal Knowledge Probing To probe the internal knowledge from the target LLM, we first generate a probing question for each head-relation pair in the injection set ($\{(h_i, r_i) | (h_i, r_i, t_i) \in \mathcal{K}_{\text{inject}}\}$) using the probing templates listed in Table 6. We then use the generated probing question to query the LLM $K = 5$ times, resulting in 5 probing answers for each probing question ($R_i^1, R_i^2, \dots, R_i^5$). We set the decoding temperature to 0.6 to balance the diversity and accuracy of the probing answers.

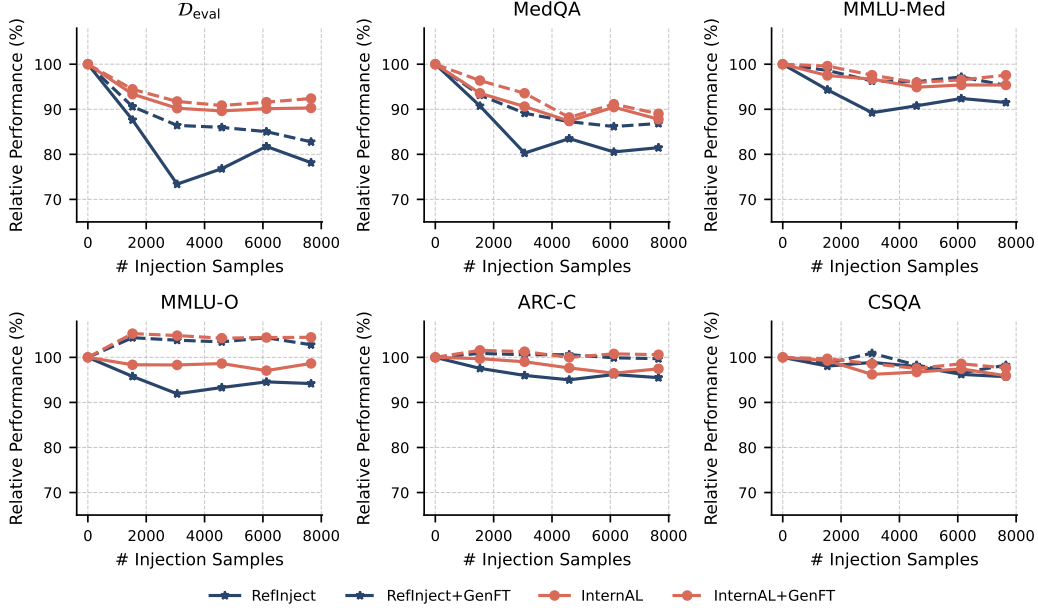


Figure 4: Relative performance (%) of Qwen3-8B trained with different knowledge injection methods on various evaluation benchmarks, with varying numbers of injected knowledge triples. All results are normalized to the model’s performance prior to injection.

Subsequently, we extract the tail entities from the probing answers by prompting the target LLM with the following instruction: “[Extraction Question]. Return a list of entities that satisfy the query, separated by a vertical bar (‘|’). If no entity meet the query, output ‘None’. Paragraph: [paragraph].” The extraction question is generated based on the extraction templates listed in Table 6.

Finally, we parse the extracted entities and filter out the entities that are not in the injection set. We then use the extracted knowledge ($\mathcal{K}_{\text{inner}}$) to augment the knowledge injection process.

Internal-aware Sample Augmentation As introduced in Section 4.1, we augment the knowledge injection process with the internal knowledge by adding the extracted tail entities as correct answers to the injection examples and convert the original multiple-choice question into a multiple-answer question. Specifically, for each triplet (h_i, r_i, t_i) for injection, we first retrieve the corresponding internal knowledge $\mathcal{K}_{\text{inner}}^{(h_i, r_i)} = \{(h, r, t) | h = h_i, r = r_i, (h, r, t) \in \mathcal{K}_{\text{inner}}\}$. Then the maximum number of correct options is set to $n^{\max} = \max(|\mathcal{K}_{\text{inner}}^{(h_i, r_i)}| + 1, 4)$. We limit the maximum number of correct options to 4 to ensure the difficulty of the question.

Then, we conduct uniform sampling to select the number of correct options $n \sim \text{Uniform}[1, n^{\max}]$, and randomly sampling $n - 1$ tail entities from $\mathcal{K}_{\text{inner}}^{(h_i, r_i)}$ to combine with the original tail entity t_i as the correct options. The distractors are randomly sampled from the PrimeKG dataset. The final injection example is then generated by filling the head entity h_i , relation r_i , and the selected correct options into the template as follows:

For a list of medical entities A: ..., B: ..., C: ..., ..., I: ..., J: ..., among the given list, [injection reference] include option [list of answers].

In this way, we can augment the knowledge injection process with the related internal knowledge, avoiding the catastrophic forgetting of the knowledge that is closely related to the injected knowledge.

Table 6: Question templates used for probing and extracting the related internal knowledge.

Relation Type	Probing Template	Extraction Template
protein-interact with-protein	What genes or proteins are involved in protein-protein interactions with the protein [head]?	Given the paragraph below, extract all the proteins that are involved in protein-protein interactions with "[head]".
drug-has carrier-protein	What proteins carry the drug [head]?	Given the paragraph below, extract all the proteins that carry the drug "[head]".
drug-has enzyme-protein	What proteins metabolize the drug [head]?	Given the paragraph below, extract all the proteins that are enzymes of the drug "[head]".
drug-has target-protein	What proteins are targeted by the drug [head]?	Given the paragraph below, extract all the proteins that are targeted by the drug "[head]".
drug-has transporter-protein	What proteins transport the drug [head]?	Given the paragraph below, extract all the proteins that transport the drug "[head]".
drug-has contraindication-disease	What diseases are contraindicated by the drug [head]?	Given the paragraph below, extract all the diseases that are contraindicated by the drug "[head]".
drug-has indication-disease	What diseases are indications for the drug [head]?	Given the paragraph below, extract all the diseases that are indicated by the drug "[head]".
drug-has off-label use-disease	What diseases are treated off-label by the drug [head]?	Given the paragraph below, extract all the diseases that are treated off-label by the drug "[head]".
drug-interact with-drug	What drugs have a drug-drug interaction with [head]?	Given the paragraph below, extract all the drugs that have a drug-drug interaction with the drug "[head]".
protein-associated with-phenotype	What effects or phenotypes are associated with [head]?	Given the paragraph below, extract all the effects/phenotypes that are associated with the protein "[head]".
disease-phenotype present-phenotype	What phenotypes are present in the disease [head]?	Given the paragraph below, extract all the phenotypes that are present in the disease "[head]".
protein-associated with-disease	What diseases are associated with [head]?	Given the paragraph below, extract all the diseases that are associated with the gene/protein "[head]".
drug-side effect-effect	What side effects are caused by the drug [head]?	Given the paragraph below, extract all the side effects of the drug "[head]".
protein-interacts with-molecular function	What molecular functions are associated with [head]?	Given the paragraph below, extract all the molecular functions that the gene/protein "[head]" interacts with.
protein-interacts with-cellular component	What cellular components interact with [head]?	Given the paragraph below, extract all the cellular components that the gene/protein "[head]" interacts with.
protein-interacts with-biological process	What biological processes interact with [head]?	Given the paragraph below, extract all the biological processes that the gene/protein "[head]" interacts with.
exposure-interacts with-protein	What genes or proteins interact with the exposure of [head]?	Given the paragraph below, extract all the proteins that interact with the exposure of "[head]".
exposure-linked to-disease	What diseases are linked to the exposure of [head]?	Given the paragraph below, extract all the diseases that are linked to the exposure of "[head]".
exposure-interacts with-biological process	What biological processes interact with the exposure of [head]?	Given the paragraph below, extract all the biological processes that the exposure of "[head]" interacts with.
protein-interacts with-pathway	What pathways does [head] involve in?	Given the paragraph below, extract all the pathways that the gene/protein "[head]" involves in.
protein-expression present in-anatomy	What anatomical locations show expression of [head]?	Given the paragraph below, extract all the anatomical locations that the protein "[head]" is expressed in.