Variational Open-Domain Question Answering

Valentin Liévin [1 2] **Andreas Geert Motzfeldt** [1] **Ida Riis Jensen** [1] **Ole Winther** [1 2 3 4]

Abstract

Retrieval-augmented models have proven to be effective in natural language processing tasks, yet there remains a lack of research on their optimization using variational inference. We introduce the Variational Open-Domain (VOD) framework for end-to-end training and evaluation of retrieval-augmented models, focusing on open-domain question answering and language modelling. The VOD objective, a self-normalized estimate of the Rényi variational bound, approximates the task marginal likelihood and is evaluated under samples drawn from an auxiliary sampling distribution (cached retriever and/or approximate posterior). It remains tractable, even for retriever distributions defined on large corpora. We demonstrate VOD's versatility by training reader-retriever BERT-sized models on multiple-choice medical exam questions. On the MedMCQA dataset, we outperform the domain-tuned Med-PaLM by +5.3% despite using 2.500× fewer parameters. Our retrieval-augmented BioLinkBERT model scored 62.9% on the MedMCQA and 55.0% on the MedQA-USMLE. Last, we show the effectiveness of our learned retriever component in the context of medical semantic search.

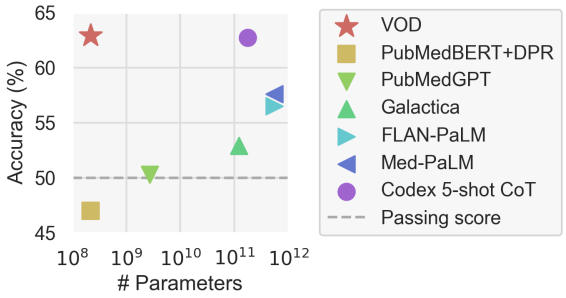

Figure 1. Parameter efficiency. Answering accuracy of baseline methods and of VOD (BioLinkBERT backbone) on MedMCQA.

1. Introduction

Scaling Transformer-based (Vaswani et al., 2017) language models (LMs) with larger datasets and more parameters (Radford et al., 2018; Kaplan et al., 2020; Hoffmann et al., 2022) led to sustained improvements in various downstream tasks.[1] However, large language models (LLMs) may reach a plateau in their performance due to the limitations of the implicit knowledge they possess, being incomplete, flawed or out-of-date. *Open-domain question answering* (ODQA) consists of augmenting LMs with external knowledge bases indexed with a retrieval mechanism. This approach was popularized in the question-answering setting by Chen et al. (2017) and was later applied to the task of language modelling itself (Guu et al., 2020; Lewis et al., 2020; Borgeaud et al., 2021; Izacard et al., 2022).

However, optimizing deep retrievers is challenging, unless there is a set of annotated evidence documents that are sufficiently aligned with the target task, as explored in Karpukhin et al. (2020); Qu et al. (2021); Khattab & Zaharia (2020). An alternative approach is to model the whole collection of documents as a latent variable (Lee et al., 2019), but this still poses challenges for optimization, especially considering that documents are discrete quantities.[2]

This research fills a gap in the literature by exploring the optimization of retrieval-augmented models using variational inference. We introduce a probabilistic framework that extends Rényi divergence variational inference (Li & Turner,

[*]Equal contribution [1]Section for Cognitive Systems, Technical University of Denmark, Denmark [2]FindZebra, Denmark [3]Center for Genomic Medicine, Rigshospitalet, Copenhagen University Hospital, Denmark [4]Bioinformatics Centre, Department of Biology, University of Copenhagen, Denmark. Correspondence to: Valentin Liévin <valv@dtu.dk>.

Proceedings of the 40^{th} International Conference on Machine Learning, Honolulu, Hawaii, USA. PMLR 202, 2023.

[1]Find a benchmark of LLMs in Srivastava et al. (2022), read about LLMs in Brown et al. (2020); Rae et al. (2021); Chowdhery et al. (2022); Thoppilan et al. (2022); Hoffmann et al. (2022); Smith et al. (2022); Zhang et al. (2022); Lieber et al. (2021); Fedus et al. (2021); Laurençon et al. (2022).

[2]Learn more about discrete latent variable optimization in Hinton et al. (1995); Le et al. (2018); Mnih & Gregor (2014); Mnih & Rezende (2016); van den Oord et al. (2017); Tucker et al. (2017); Grathwohl et al. (2017); Masrani et al. (2019); Liévin et al. (2020).

2016), allowing us to estimate the marginal task likelihood and its gradient by sampling from an *approximate posterior*. The proposed framework is versatile and applies to various settings, including extractive, generative, and multiple-choice models for open-domain question answering, as well as the training of retrieval-enhanced language models.

To demonstrate the effectiveness of the framework, we train reader-retriever BioLinkBERT models end-to-end on multiple-choice medical QA tasks and achieve a new state-of-the-art on the MedMCQA of 62.9%, outperforming the current 540B parameter domain-tuned Med-PaLM by +5.3% (Singhal et al., 2022) using 2.500× fewer parameters (Figure 1). On the challenging MedQA-USMLE, we score 55.0%: a new state-of-the-art in the open-domain setting. We highlight the main contributions of this paper as follows:

1. The VOD framework: tractable, consistent, end-to-end training of retrieval-augmented models.
2. Popularizing Rényi divergence variational inference for natural language tasks.
3. Truncated retriever parameterization: relaxing the top-K retriever approximation to using top $P \geq K$.

In addition to our theoretical contributions, we release MedWiki: a subset of Wikipedia tailored to the MedMCQA and USMLE dataset for low-resource research.

2. VOD: a Probabilistic Framework for Retrieval-augmented Tasks

Let a question $\mathbf{q}$ be defined in a space Ω (e.g., the space of sequences of tokens) and the set of possible answers be $\mathbb{A} \subset \Omega$ with a correct answer denoted $\mathbf{a} \in \mathbb{A}$. We introduce a corpus of N documents $\mathbb{D} := \{\mathbf{d}_1, \ldots, \mathbf{d}_N\} \in \Omega^N$. In open-domain tasks, we are interested in modelling the marginal task likelihood with a reader-retriever model $p_\theta(\mathbf{a}, \mathbf{d}|\mathbf{q}) := p_\theta(\mathbf{a}|\mathbf{d}, \mathbf{q})p_\theta(\mathbf{d}|\mathbf{q})$ parameterized by θ:

$$p_\theta(\mathbf{a}|\mathbf{q}) := \sum_{\mathbf{d}\in\mathbb{D}} \underbrace{p_\theta(\mathbf{a}|\mathbf{d}, \mathbf{q})}_{\text{reader}} \underbrace{p_\theta(\mathbf{d}|\mathbf{q})}_{\text{retriever}} . \tag{1}$$

Variational inference (Jordan et al., 1999; Kingma & Welling, 2013; Burda et al., 2015) allows estimating the marginal task likelihood eq. (1) using samples drawn from an approximate posterior $r_\phi(\mathbf{d}|\mathbf{a}, \mathbf{q})$. This consists of evaluating the *evidence lower bound* (ELBO), a log-likelihood lower bound.[3] In open-domain applications, the approximate posterior, with parameter ϕ, can be defined using either a keyword-search engine (BM25; Robertson & Zaragoza (2009)), a checkpoint of $p_\theta(\mathbf{d}|\mathbf{q})$, or a model learned jointly.

We introduce the VOD framework in four acts: i) Why Rényi divergence variational inference can aid likelihood-based learning, ii) The VOD objective: a tractable self-normalized importance sampling estimate of the Rényi bound, iii) A truncated retriever parameterization that generalizes existing approaches and iv) A discussion on the application of the VOD framework.

2.1. Rényi Divergence Variational Inference

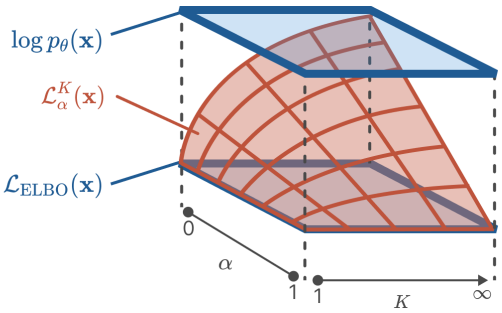

Figure 2. Depicts the core component of the VOD framework: the importance-weighted Rényi Variational Bound (IW-RVB) as a function of the parameter $\alpha \in [0, 1]$ and the number of samples $K \geq 1$. As the value of α and K increase, the IW-RVB becomes a more accurate estimate of the likelihood of a given task, demonstrating how we use VOD to optimize retrieval-augmented models through the manipulation of α and K. See how the parameter α affects the training dynamics in Figure 7, Appendix G.

Rényi divergence variational inference (Li & Turner, 2016) extends traditional variational inference (Jordan et al., 1999; Kingma & Welling, 2013). Given a parameter $\alpha < 1$ and the importance weight $w_{\theta,\phi}^{1-\alpha}(\mathbf{a}, \mathbf{q}, \mathbf{d}) := p_\theta(\mathbf{a}, \mathbf{d}|\mathbf{q})r_\phi^{-1}(\mathbf{d}|\mathbf{a}, \mathbf{q})$ the variational Rényi bound (RVB) defined as

$$\mathcal{L}_\alpha(\mathbf{a}, \mathbf{q}) := \frac{1}{1-\alpha} \log \mathbb{E}_{r_\phi(\mathbf{d}|\mathbf{a},\mathbf{q})} \left[w_{\theta,\phi}^{1-\alpha}(\mathbf{a}, \mathbf{q}, \mathbf{d}) \right] \tag{2}$$

RVB is a lower bound of the marginal log-likelihood for $\alpha \geq 0$ and is extended by continuity in $\alpha = 1$ as $\mathcal{L}_{\alpha=1}(\mathbf{a}, \mathbf{q}) := \lim_{\alpha\to 1} \mathcal{L}_\alpha(\mathbf{a}, \mathbf{q})$ where it equals the ELBO. In practice, the RVB and its gradients can be estimated using K documents sampled from $r_\phi(\mathbf{d}|\mathbf{a}, \mathbf{q})$. The resulting importance sampling estimate yields another bound: the Importance Weighted RVB (IW-RVB; Li & Turner (2016)):

$$\hat{\mathcal{L}}_\alpha^K(\mathbf{a}, \mathbf{q}) := \frac{1}{1-\alpha} \log \frac{1}{K} \sum_{i=1}^{K} w_{\theta,\phi}^{1-\alpha}(\mathbf{a}, \mathbf{q}, \mathbf{d}_i) \tag{3}$$
$$\mathbf{d}_1, \ldots, \mathbf{d}_K \overset{\text{iid}}{\sim} r_\phi(\mathbf{d}|\mathbf{a}, \mathbf{q})$$

which aligns with the importance-weighted bound (IWB; Burda et al. (2015)) in $\alpha = 0$. To sum up, the main proper-

[3] $\mathcal{L}_{\text{ELBO}}(\mathbf{a}, \mathbf{q}) := \log p_\theta(\mathbf{a}, \mathbf{q}) - \mathcal{D}_{\text{KL}}\left(r_\phi(\mathbf{d}|\mathbf{a}, \mathbf{q}) \| p_\theta(\mathbf{d}|\mathbf{a}, \mathbf{q})\right)$

ties of the RVB and the IW-RVB are ($\alpha \geq 0$):

$$\mathcal{L}_{\alpha=0}(\mathbf{a}, \mathbf{q}) = \log p_\theta(\mathbf{a}|\mathbf{q}) \quad \mathcal{L}_{\alpha\to 1}(\mathbf{a}, \mathbf{q}) = \mathcal{L}_{\text{ELBO}}(\mathbf{a}, \mathbf{q})$$
$$\mathcal{L}_{\alpha\geq 0}(\mathbf{a}, \mathbf{q}) \leq \log p_\theta(\mathbf{a}|\mathbf{q}) \qquad \mathcal{L}^K_\alpha(\mathbf{a}, \mathbf{q}) \leq \mathcal{L}_\alpha(\mathbf{a}, \mathbf{q}) \,.$$

RVB gradient The gradient of the RVB w.r.t. θ is:

$$\nabla_\theta \mathcal{L}_\alpha(\mathbf{a}, \mathbf{q}) = \mathbb{E}_{r_\phi}\left[\widetilde{w^{1-\alpha}_{\theta,\phi}}(\mathbf{a}, \mathbf{q}, \mathbf{d})\, \nabla_\theta \log p_\theta(\mathbf{a}, \mathbf{d}|\mathbf{q})\right]$$

where the normalized importance weight is defined as

$$\widetilde{w^{1-\alpha}_{\theta,\phi}}(\mathbf{a}, \mathbf{d}) := \frac{w^{1-\alpha}_{\theta,\phi}(\mathbf{a}, \mathbf{q}, \mathbf{d})}{\mathbb{E}_{r_\phi(\mathbf{d}'|\mathbf{a},\mathbf{q})}\left[w^{1-\alpha}_{\theta,\phi}(\mathbf{a}, \mathbf{d}', \mathbf{q})\right]} \,. \quad (5)$$

In this paper, we consider the sampling distribution r_ϕ to be static and therefore do not estimate the gradient w.r.t. the approximate posterior. Optimizing the parameter ϕ jointly with θ can be done by application of importance sampling coupled with variance reduction techniques (Burda et al., 2015; Mnih & Rezende, 2016; Le et al., 2018; Masrani et al., 2019; Kool et al., 2019b; Liévin et al., 2020).

Stabilizing training using the RVB Considering the optimization of the parameter ϕ, a looser bound (e.g., the ELBO) might be preferred to a tighter one (e.g., the IWB).[4] In this paper, we explore interpolating between variational bounds using the parameter α of the RVB. We argue that, even for a non-trainable parameter ϕ, optimizing for a looser bound can overcome early optimization challenges.

For $\alpha = 0$, the RVB aligns with the marginal log-likelihood independently of the choice of the approximate posterior. However, when the importance weight $w_{\theta,\phi}(\mathbf{q}, \mathbf{a}, \mathbf{d})$ suffers from high variance, so does the Monte Carlo estimate of the marginal likelihood and its gradient.[5]

For $\alpha = 1$, the RVB matches the ELBO and the gradients restricted to the reader and retriever decomposes as:

$$\nabla_{\theta_{\text{READ.}}} \mathcal{L}_{\alpha=1}(\mathbf{a}, \mathbf{q}) = \mathbb{E}_{r_\phi(\mathbf{d}|\mathbf{a},\mathbf{q})}\left[\nabla_\theta \log p_\theta(\mathbf{a}|\mathbf{d}, \mathbf{q})\right]$$
$$\nabla_{\theta_{\text{RETR.}}} \mathcal{L}_{\alpha=1}(\mathbf{a}, \mathbf{q}) = -\nabla_\theta D_{\text{KL}}\left(r_\phi(\mathbf{d}|\mathbf{a}, \mathbf{q}) \| p_\theta(\mathbf{d}|\mathbf{q})\right) \,.$$

Maximizing the ELBO corresponds to optimizing the reader and the retriever disjointly. On the reader side, this equals maximizing the answer likelihood $p_\theta(\mathbf{a}|\mathbf{d}, \mathbf{q})$ in expectation over $r_\phi(\mathbf{d}|\mathbf{a}, \mathbf{q})$ independently of the value of $p_\theta(\mathbf{d}|\mathbf{q})$. On the retriever side, this corresponds to matching the approximate posterior with the learned retriever $p_\theta(\mathbf{d}|\mathbf{q})$. This can be seen as an instance of knowledge distillation of the posterior into the retriever. After an initial learning phase, the RVB can be smoothly interpolated from the ELBO to the marginal task likelihood by controlling the parameter α.

[4] Exploring using hybrid ELBO/IWB objectives has been explored in Rainforth et al. (2018), interpolating the RVB has been explored in Liévin et al. (2020).

[5] See Kong (1992); Owen (2013); Nowozin (2015) for an introduction and discussion about variance and importance sampling.

2.2. VOD objective

In ODQA applications, the IW-RVB eq. (3) is generally intractable due to the normalization constant in eq. (8a) which requires evaluating all documents.

The VOD objective is an approximation of the IW-RVB which can be evaluated using K documents sampled *without replacement* from $r_\phi(\mathbf{d}|\mathbf{a}, \mathbf{q})$. It is defined as:

$$\hat{L}^K_\alpha(\mathbf{a}, \mathbf{q}) := \frac{1}{1-\alpha} \log \sum_{i=1}^K s_i\, \hat{v}^{1-\alpha}_{\theta,\phi}(\mathbf{a}, \mathbf{q}, \mathbf{d}_i) \quad (6)$$
$$(\mathbf{d}_1, s_i), \ldots, (\mathbf{d}_K, s_K) \overset{\text{priority}}{\sim} r_\phi(\mathbf{d}|\mathbf{a}, \mathbf{q}) \,.$$

where the self-normalized importance weight $\hat{v}_{\theta,\phi}$ is defined using the un-normalized retrieval density ratio $\zeta(\mathbf{d}) \propto p_\theta(\mathbf{d}|\mathbf{q}) r^{-1}_\phi(\mathbf{d}|\mathbf{a}, \mathbf{q})$ as:

$$\hat{v}_{\theta,\phi} := p_\theta(\mathbf{a}|\mathbf{q}, \mathbf{d}_i)\zeta(\mathbf{d}_i)\left(\sum_{j=1}^K s_j \zeta(\mathbf{d}_j)\right)^{-1} \quad (7)$$

The set of documents $\mathbf{d}_1, \ldots, \mathbf{d}_K$ are sampled without replacement from $r_\phi(\mathbf{d}|\mathbf{a}, \mathbf{q})$ using *priority sampling* (Duffield et al., 2007). The sampling procedure comes with importance weights $s_1, \ldots, s_k$ defined such that for a function $h(\mathbf{d})$, $\sum_{i=1}^K s_i\, h(\mathbf{d}_i) \approx \mathbb{E}_{r_\phi(\mathbf{d}|\mathbf{a},\mathbf{q})}\left[h(\mathbf{d})\right]$. We present priority sampling in greater length in Appendix A.

The VOD objective and its gradient are consistent (i.e., converge to the RVB in the limit $K \to N$ with probability one) and can be evaluated with complexity $\mathcal{O}(K)$, whereas the IW-RVB is of complexity $\mathcal{O}(N)$. Furthermore, the VOD objective approximates the IW-RVB, which itself is guaranteed to approximate the marginal task log-likelihood more tightly as $K \to N$ (Burda et al., 2015).

The VOD objective is derived in Appendix B, the VOD gradient is defined in Appendix C. Our implementation of the sampling methods and the VOD objective is available at `http://github.com/VodLM/vod`.

2.3. Truncated retriever parameterization

The VOD framework is compatible with retrievers defined on the whole corpus (N documents). However, in our approach, we truncate the retriever to consider only the top

P documents, where $K < P \ll N$. K refers to the number of sampled documents, while P represents the pool of documents from which the top K documents are selected. This truncation provides two key advantages: i) it enables efficient caching or retention of document scores, as only P documents need to be stored in memory, and ii) the value P serves as an exploration-exploitation threshold: a higher value of P yield greater diversity in document sampling, promoting *exploration*. While, a smaller value of P ensures that during training, all documents in the set $\mathcal{T}_\phi$ are more likely visited, facilitating *exploitation* of the available information.

Assuming the retrieval distributions to be described by score functions $f_\theta : \Omega^2 \to \mathbb{R}$ and $f_\phi : \Omega^3 \to \mathbb{R}$. We define the truncated retrievers as:[6]

$$p_\theta(\mathbf{d}|\mathbf{q}) := \frac{\mathbb{1}[\mathbf{d} \in \mathcal{T}_\phi] \exp f_\theta(\mathbf{d}, \mathbf{q})}{\sum_{\mathbf{d}' \in \mathcal{T}_\phi} \exp f_\theta(\mathbf{d}', \mathbf{q})} \tag{8a}$$

$$r_\phi(\mathbf{d}|\mathbf{a}, \mathbf{q}) := \frac{\mathbb{1}[\mathbf{d} \in \mathcal{T}_\phi] \exp f_\phi(\mathbf{a}, \mathbf{q}, \mathbf{d})}{\sum_{\mathbf{d}' \in \mathcal{T}_\phi} \exp f_\phi(\mathbf{a}, \mathbf{q}, \mathbf{d}')} \tag{8b}$$

where $\mathcal{T}_\phi$ is the set of the top $P \leq N$ documents ranked by the score $f_\phi(\mathbf{a}, \mathbf{q}, \mathbf{d})$. The score function f_θ and f_ϕ can be implemented using BM25 and/or contextual vector representations extracted using pretrained language models such as DPR or ColBERT (Karpukhin et al., 2020; Khattab & Zaharia, 2020). For instance using a dual-encoder model $f_\theta(\mathbf{d}, \mathbf{q}) = \text{BERT}_\theta(\mathbf{d})^T \text{BERT}_\theta(\mathbf{q})$ and $f_\phi(\mathbf{a}, \mathbf{q}, \mathbf{d}) = \text{BERT}_\phi([\mathbf{q}; \mathbf{a}])^T \text{BERT}_\phi(\mathbf{d})$ where BERT is the function that return the output of a BERT model at the CLS token and $[\cdot;\cdot]$ is the concatenation operator. Retrieving the top P documents is efficient when using `elasticsearch`[7] and/or `faiss` (Johnson et al., 2021).

2.4. Applying VOD

In this paper, we show how to apply the VOD framework to multiple-choice ODQA. Nevertheless, VOD is general-purpose and designed for latent variable models defined on a discrete and finite space. In NLP, it applies to a wide range of settings such as generative, extractive, multiple-choice ODQA as well as retrieval-augmented language modelling. Find a non-exhaustive list of examples in Appendix E.

3. Related work

VOD aids the development of retrieval-augmented models for language modeling (LM) tasks. In this section, we review previous work on retrieval for LM, and compare to VOD (summarized with references in Table 1).

[6]When $P > K$, evaluating the retriever density eq. (8a) is generally intractable due to the sum over P documents.

[7]`http://www.elastic.co/`

Table 1. Deep retrievers in literature, detailing if training was end-to-end, variational, as well the size of support during training.

Method	Retriever training	End-to-end learning	Posterior Guided	Retriever Support
DPR[1]	Supervised	✗	✗	–
ColBERT[2]	Supervised	✗	✗	–
Contriever[3]	Self-supervised	✗	✗	–
FiD[4]	Frozen DPR dual-encoder	✗	✗	–
RETRO[5]	Frozen BERT dual-encoder	✗	✗	–
ORQA[6]	Self-supervised + MLL*	(✓)	✗	top-K doc.
RAG[7]	MLL* + frozen DPR doc. encoder	(✓)	✗	top-K doc.
REALM[8]	Self-supervised + MLL*	✓	✗	top-K doc.
EMDR-2[9]	Self-supervised + Expect.-Max.	✓	✓	top-K doc.
Hindsight[10]	ColBERT init. + ELBO + MLL*	✓	✓	top-K doc.
VOD	Rényi variational bound	✓	✓	top-P doc.†

[1] Karpukhin et al. (2020), [2] Khattab et al. (2021), [3] Izacard et al. (2021), [4] Izacard & Grave (2020)
[5] Borgeaud et al. (2021), [6] Lee et al. (2019), [7] Lewis et al. (2020), [8] Guu et al. (2020)
[9] Sachan et al. (2021), [10] Paranjape et al. (2021), *MLL: marginal log-likelihood
† $K \leq P \leq N$ (K:# of documents in a batch, N: corpus size, P: chosen)

Learning to search Retrieval-based training have gained much attention for improving pre-trained LMs. ORQA and Contriever proposed a self-supervised approach using contrastive learning to match a text passage with its context, and is widely adopted in pre-training to enable zero-shot retrieval (*Inverse Cloze Task*; Lee et al. (2019)). In contrast, DPR and ColBERT use supervised contrastive learning with questions paired to annotated documents. This method has sparked many retrieval-augmented attempts such as FiD, RETRO, and RAG to enhance auto-regressive LMs conditioned on a frozen retriever. ORQA and REALM, later followed by RAG, EMDR, Hindsight, and VOD proposed optimizing both a retrieval component and a reader or language modelling component end-to-end, by maximizing the marginal log-likelihood (MLL).

Posterior guided supervision Many efforts has been devoted to leveraging external knowledge with posterior guided supervision. EMDR learns a retriever end-to-end with an Expectation-Maximization objective evaluated under the posterior distribution of $p_\theta(\mathbf{d}|\mathbf{a}, \mathbf{q}) \propto p_\theta(\mathbf{d}|\mathbf{q}) p_\theta(\mathbf{a}|\mathbf{d}, \mathbf{q})$, while Hindsight optimizes the variational lower-bound (ELBO) evaluating under a target-aware approximate posterior $r_\phi(\mathbf{d}|\mathbf{a}, \mathbf{q})$. Among previous methods, Hindsight is most akin to VOD as both methods rely on maximizing a variational bound. Nonetheless, VOD introduces the more general Rényi variational bound, which offers to model the sampling distribution explicitly. Ultimately, a more principled approach makes VOD more versatile and capable of handling a wider range of problems.

Navigating large knowledge bases The large size of knowledge bases such as Wikipedia makes it computationally intractable to consider all N documents when computing MLL. To address this, all related methods rely on a strict truncation of the retriever to the top-K cached documents. In contrast to these aforementioned approaches, which limits to a fixed set of K documents, we propose a truncated

Table 2. Summarizes the medical QA datasets and corpora used in our study, including the MedMCQA, USMLE, and FindZebra (FZ) corpus, with the MedWiki as the knowledge base for all QA tasks. The questions are numbered for the train/validation/test splits.

DATASETS	MEDMCQA	USMLE	FZ QUERIES
QUESTIONS	182.8K/4.2K/6.1K	10.2K/1.3K/1.3K	248
CORPORA	**WIKPEDIA**	**MEDWIKI**	**FZ CORPUS**
ARTICLES	6.6M	293.6K	30.7K
PASSAGES	–	7.8M	711.9K

retriever parameterization that works hand-in-hand with our principled objective to handle over top $P > K$ documents. Ultimately, this allows for more diverse document sampling during training and allows reducing the bias induced by truncating the retriever distribution. In Appendix D, we show that the top-K MLL is a special case of VOD for $K = P$ and $\alpha = 0$.

4. Experiments

In this section, we present the medical domain tasks and datasets, results on end-to-end multiple-choice ODQA and its application to information retrieval. The code and datasets are available on GitHub.[8]

4.1. Datasets

The datasets utilized for the medical domain are summarized in Table 2. We introduce the MedWiki, a subset of Wikipedia targeted to medical QA tasks.

MedMCQA Pal et al. (2022) is a large-scale multiple-choice question answering dataset collected from Indian medical school entrance exams (AIIMS and NEET-PG). It covers several medical topics (dentistry, pathology, surgery, preventive medicine, etc.) and question types (diagnosis, recalling expert factual knowledge, mathematical problems, etc.)

MedQA-USMLE Jin et al. (2021)) is a collection of medical questions from the US medical board exam. The questions aim to assess human doctors' medical knowledge and decision-making. Each question includes a medical history, vital signs (e.g., blood pressure, temperature), and possibly a specific analysis (e.g., CT scan).

MMLU Hendrycks et al. (2021) is a dataset for assessing the knowledge acquired during pre-training by evaluating models in a zero-shot setting. The test set comprises 57 tasks spanning different domains. We limit our analysis to the subcategories *psychology*, *biology*, and *health*.[9]

[8] https://github.com/findzebra/fz-openqa

[9] The subcategory *professional_medicine* corresponds to the MedQA-USMLE questions.

MedWiki We release the MedWiki corpus (under MIT license): a collection of 4.5% of articles taken from the English Wikipedia and targeted to the MedMCQA and USMLE datasets. The MedWiki corpus was built by querying each answer option from the MedMCQA and USMLE datasets against the Wikipedia API. Read more in Appendix H.

FindZebra corpus & queries FindZebra is a search tool for assisting in the diagnosis of rare diseases that is built on open-source information retrieval software (BM25) tailored to this problem (Dragusin et al., 2013). The FindZebra corpus indexes a collection of curated articles from various reputable databases: GARD, GeneReviews, Genetics Home Reference, OMIM, Orphanet, and Wikipedia. Each article is referenced with a Concept Unique Identifier (CUI) from the Unified Medical Language System (UMLS; Bodenreider (2004)). We use a collection of 248 publicly available search queries (FZ queries). Each query is labelled with a reference diagnostic, allowing to benchmark medical search engines.[10]

4.2. VOD for multiple-choice QA

In the multiple-choice question answering (MCQA) setting, we consider a vector of M answer options $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_M]$, where $\star$ represents the index of the correct option. Similarly, we define a vector of M queries as $\mathbf{Q} = [\mathbf{q}_1, \ldots, \mathbf{q}_M]$, where $\mathbf{q}_j = [\mathbf{q}; \mathbf{a}_j]$ represents the concatenation of the question with the answer option of index j. Additionally, we denote a vector of M documents $\mathbf{D} = [\mathbf{d}_1, \ldots, \mathbf{d}_M] \in \mathbb{D}^M$, and the set of M combinations of documents as $\mathbb{D}^{(M)}$, which contains N^M document vectors. The marginal likelihood is defined as follows:

$$p_\theta(\mathbf{a}_\star|\mathbf{Q}) := \sum_{\mathbf{D}\in\mathbb{D}^{(M)}} p_\theta(\mathbf{D}|\mathbf{Q})\, p_\theta(\mathbf{a}_\star|\mathbf{D}, \mathbf{Q})\,. \quad (9)$$

To model this problem, we introduce i) a reader model $g_\theta : \Omega^2 \to \mathbb{R}$, which evaluates the likelihood of answer option $j \in [1, ..., M]$ given the query and a tuple of K documents $\mathbf{d}_1, \ldots, \mathbf{d}_K$, and ii) we define a truncated retriever model $p_\theta(\mathbf{d}|\mathbf{q}_j)$ and $r_\phi(\mathbf{d}|\mathbf{q}_j)$, which retrieves K document specific to each answer option. As described in eq. (8a), these models are parameterized by scores $f_\theta(\mathbf{d}, \mathbf{q}_j)$ and $f_\phi(\mathbf{d}, \mathbf{q}_j)$ respectively. The reader and retriever models are defined as:

$$p_\theta(\mathbf{a}_\star|\mathbf{D}, \mathbf{Q}) := \frac{\exp g_\theta(\mathbf{d}_\star, \mathbf{q}_\star)}{\sum_{j=1}^M \exp g_\theta(\mathbf{d}_j, \mathbf{q}_j)} \quad (10)$$

$$p_\theta(\mathbf{D}|\mathbf{Q}) := \prod_{j=1}^M p_\theta(\mathbf{d}_j|\mathbf{q}_j),\ r_\phi(\mathbf{D}|\mathbf{Q}) = \prod_{j=1}^M r_\phi(\mathbf{d}_j|\mathbf{q}_j)\,.$$

[10] https://huggingface.co/datasets/findzebra

The VOD objective can be applied to approximate the marginal likelihood $p_\theta(\mathbf{a}_\star|\mathbf{Q})$ defined in eq. (9). In practice, the VOD objective in a multiple-choice setting implies the retrieval of KM documents per query, resulting in a conditional answering likelihood that encompasses K^M unique combinations. For further details, refer to Appendix E.4.

4.3. Experimental Setup

We implement a DPR-like dual-encoder architecture for the retriever with a shared backbone and implement the multiple-choice reader following Devlin et al. (2018). We use the domain-specific BioLinkBERT (Yasunaga et al., 2022) as the backbone for both models and use the MedWiki corpus for all QA experiments. This results in a total of 2×110M=220M parameters; a small retrieval-augmented language model. All experiments were conducted on a single node of 8 RTX 5000 GPUs using half-precision. Further details can be found in Appendix F.

Hybrid approximate posterior We parameterize the score f_ϕ of the sampling distribution using a composite BM25 score combined to a checkpoint of the retriever score f_θ denoted f_ϕ^{ckpt}. Specifically, we sample documents using:

$$f_\phi(\mathbf{a},\mathbf{q},\mathbf{d}) := f_\phi^{\text{ckpt}}(\mathbf{d},[\mathbf{q};\mathbf{a}]) + \tau^{-1}\left(\text{BM25}(\mathbf{q},\mathbf{d}) + \beta\cdot\text{BM25}(\mathbf{a},\mathbf{d})\right). \quad (11)$$

where $\tau = 5$ and β is a parameter scaled proportionally to the ratio of question and answer lengths ${}^{L_\mathbf{q}}/_{L_\mathbf{a}}$ to ensure that the BM25 score of the question does not outweigh the answer score. We use $\beta = 1 + 0.5\max\{0, \log({}^{L_\mathbf{q}}/_{L_\mathbf{a}})\}$.[11] At initialization f_θ is uninformative, we thus set $f_\phi^{\text{ckpt}} = 0$. The combination of the two scores may provide a more robust sampling distribution by utilizing both the previously learned information and secondly the BM25 relevance of the query to the document.

Training, periodic re-indexing and annealing We organize the training into rounds of T steps similarly to Khattab et al. (2021). As the model is exposed to a progressively larger portion of the dataset over multiple rounds, we expect optimization will result in improved generalization capabilities. At the beginning of each round, for each question-answer pair $\mathbf{q}_j$, we retrieve the set of top-P documents $\mathcal{T}_\phi$ and cache the set of values $\{f_\phi(\mathbf{a}_j,\mathbf{q},\mathbf{d}) \mid \mathbf{d}\in\mathcal{T}_\phi\}$, except for the first period where f_ϕ^{ckpt} is set to zero. During the first round, we anneal the RVB parameter α from 1 to 0 to stabilize early training by distilling the BM25 cached score $f_\phi(\mathbf{a},\mathbf{d},\mathbf{q}) = 0 + \tau^{-1}\left(\text{BM25}(\mathbf{q},\mathbf{d}) + \beta\cdot\text{BM25}(\mathbf{a},\mathbf{d})\right)$ into the trainable retriever score $f_\theta(\mathbf{d},\mathbf{q})$, as shown in Figure 3. At each training iteration, we sample a set of $K = 8$ document $\mathcal{T}_\phi$ for each of the $M = 4$ question-answer pairs and evaluated the VOD objective and its gradient using the cached values of $f_\phi(\mathbf{a}_j,\mathbf{q},\mathbf{d})$.

[11] We picked the parameters to target a relatively high sampling entropy, no extensive hyperparameter search was performed.

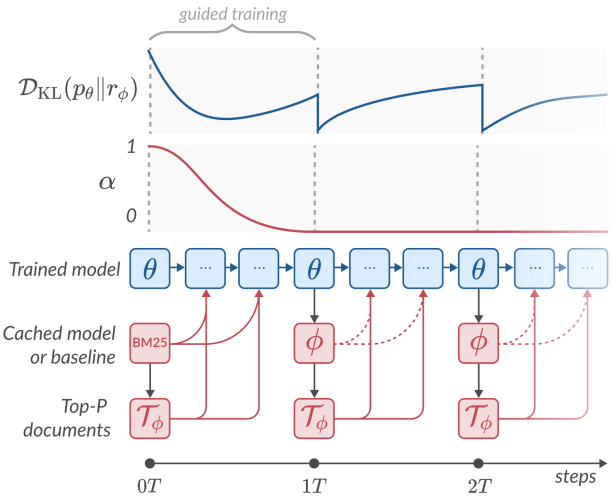

Figure 3. During training, VOD incorporates periodic updates of the cached models. In the initial period, the sampling distribution $r_\phi(\mathbf{d}|\mathbf{a},\mathbf{q})$ can be chosen as a domain-specific baseline (BM25). Additionally, a parameter $\alpha > 0$ can be utilized to guide the optimization of θ. Note that the approximations $\hat{L}^K_{\alpha=1} \approx \mathcal{L}_{\text{ELBO}}$ and $\hat{L}^K_{\alpha=0} \approx \log p_\theta$ can be observed, demonstrated in the experimental curves in Appendix G.

Evaluation At evaluation time, we estimate the likelihood for each answer option using $C = 10$ Monte-Carlo samples, each containing $MK = 4\cdot 8 = 32$ documents using the estimates defined in eq. (6) (see Appendix E.4). Leveraging more samples at inference time allows for approximating the answer likelihood more robustly, as it allows for testing a greater number of combinations of documents.

4.4. QA Benchmark

MedMCQA We report the validation and test accuracy of the VOD framework applied to BioLinkBERT (base) and the baselines in Table 3.

VOD outperforms both the disjoint BERT-based methods and the recent Med-PaLM (540B parameters) with a new state-of-the-art test accuracy of 62.9%, +0.2% over Codex 5-shot CoT. This is an improvement of +5.3% over Med-PaLM despite using 2.500× fewer parameters. VOD scored +7.6% improvement over the BioLinkBERT reader with static BM25 retriever, and +15.9% over the PubMedBERT reader coupled with a DPR retriever.

MedQA-USMLE The validation and test accuracy are shown in Table 3. We found that using VOD with a Bi-

Table 3. Open-domain question answering accuracy.

Method	Params.	Finetuning	MedMCQA Valid.	MedMCQA Test	USMLE Valid.	USMLE Test
VOD BioLinkBERT+BM25	110M	MedMCQA	51.6	55.3	–	–
VOD BioLinkBERT+BM25	110M	USMLE	–	–	41.0	40.4
VOD 2×BioLinkBERT	220M	MedMCQA	58.3	**62.9**	47.2	46.8
VOD 2×BioLinkBERT	220M	USMLE	–	–	45.8	44.7
VOD 2×BioLinkBERT	220M	MedMCQA→USMLE*	–	–	53.6	55.0
Disjoint PubMedBERT+DPR[1]	220M	MedMCQA	43.0	47.0	–	–
Disjoint PubMedBERT+BM25[2]	110M	USMLE	–	–	–	38.1
Disjoint BioLinkBERT+BM25[3]	110M	USMLE	–	–	–	40.0
Disjoint BioLinkBERT-L+BM25[3]	340M	USMLE			–	44.6
Reader only PubMedGPT[4]	2.7B	MedMCQA+USMLE	–	50.3	–	–
Reader only Galactica[5]	120B	MedMCQA	52.9	–	–	44.4
Reader only Codex 5-shot CoT[6]	175B	∅	59.7	62.7	–	60.2
Reader only FLAN-PaLM[7]	540B	∅	–	56.5	–	60.3
Reader only Med-PaLM[7]	540B	MedMCQA+USMLE	–	57.6	–	**67.6**
Random Uniform			25.0	25.0	25.0	25.0
Human Passing score[6]			50.0	50.0	60.0	60.0
Human Merit candidate[6]			90.0	90.0	87.0	87.0

[1]results from Pal et al. (2022), model from Gu et al. (2021), [2]Gu et al. (2021)
[3]Yasunaga et al. (2022), [4]Venigalla et al. (2022), [5]Taylor et al. (2022), [6]Liévin et al. (2022)
[7]Singhal et al. (2022), *First pretrained on MedMCQA then finetuned on the USMLE

oLinkBERT backbone outperforms a BioLinkBERT reader coupled with a BM25 retriever, even when using the larger version of BioLinkBERT (44.7% for VOD, 40.0% for disjoint BioLinkBERT, 44.6% for the disjoint large BioLinkBERT).

Due to the small size of MedQA-USMLE, pretraining on the MedMCQA proved beneficial. MedMCQA pretraining with USMLE fine-tuning resulted in VOD achieving a 55.0% test accuracy, +10.4% improvement over a large BioLinkBERT model with a BM25 retriever. However, Med-PaLM scores +12.6% higher accuracy over the best VOD model.

MMLU Table 4 compares the zero-shot performance of VOD, GPT-3, and Unified QA in the subcategories of *psychology*, *biology*, and *health*. We reused the BioLinkBERT VOD model trained on MedMCQA only. VOD achieved an average accuracy of 54.8% across all 12 tasks, surpassing both GPT-3 (47.0%) and Unified QA (48.7%). Particularly, VOD excelled in *medical_genetics* (+36.0%), *professional_medicine* (+14.4%), and *anatomy* (+12.5%). Although GPT-3 and Unified QA showed competitive results in certain areas, VOD's higher accuracy highlights its robustness to a wider set of medical tasks.

4.5. Ablation Study

In Figure 4, we report the performances of a VOD model for multiple variational bounds and diverse truncated retriever support sizes (the number of cached top-P documents). [12]

Variational bounds We tested multiple variational bounds: the ELBO, the importance-weighted bound (IWB) and the RVB as possible methods to optimize the model.

[12]To reduce overall running costs, we used a dual-encoder reader with score function $g_\theta(\mathbf{d}, \mathbf{q}) = \text{BERT}(\mathbf{q})^T \text{BERT}(\mathbf{d})$.

Table 4. Zero-shot accuracy on MMLU (%).

Task	Subcategory	Unified QA	GPT-3	VOD
medical_genetics	health	40.0	40.0	**76.0**
high_school_psychology	psychology	**70.0**	61.0	60.6
college_biology	biology	40.0	45.0	**59.7**
anatomy	health	43.0	46.0	**58.5**
clinical_knowledge	health	57.0	50.0	**58.5**
professional_medicine	health	43.0	38.0	**57.4**
nutrition	health	48.0	50.0	**56.5**
high_school_biology	biology	53.0	48.0	**55.2**
college_medicine	health	43.0	**47.0**	46.8
human_aging	health	**55.0**	50.0	44.4
virology	health	43.0	**44.0**	42.2
professional_psychology	psychology	**49.0**	45.0	42.2
Average	-	48.7	47.0	**54.8**

The ELBO and IWB are special cases of the RVB. For the RVB, we anneal the parameter α, as in the main experiments, and found that this method resulted in the highest answering accuracy while also resulting in low retriever entropy. This suggests that the retriever was also optimized at a faster rate.

Exploration vs. Exploitation We experimented with using values of $P \in \{8, 32, 100\}$. Using the highest value of $P = 100$ resulted in a smaller effective sample size,[13] slower learning but ultimately higher accuracy.

4.6. Information retrieval

Despite good QA accuracy, the ability of VOD to yield a meaningful retriever component through the proposed reader-retriever end-to-end training remains, at this point of the paper, to be proven. Thus, we benchmarked a VOD retriever trained on MedMCQA against the FindZebra API[14], which connects to a specialized BM25 search engine targeted to medical professionals (Dragusin et al., 2013). The comparison was done using the set of FindZebra queries and corpus, where searching documents using a BERT-based retriever translates into a nearest neighbour search problem in the embedding space, which we visualize in Appendix G.

Re-purposing MCQA retrievers for semantic search The BioLinkBERT VOD model, trained on the MedMCQA dataset, has a retriever component that is trained to rank documents using question-answer pairs $[\mathbf{q}; \mathbf{a}]$ as inputs (see eq. (10)). Thus, further task adaptation is required to rank documents solely based on queries, and without answer option (e.g., using a model $p_\theta(\mathbf{d}|\mathbf{q})$ instead of $p_\theta(\mathbf{d}|[\mathbf{q}; \mathbf{a}_j])$). To address this, we use the retriever to teach a query-only student model, which corresponds to *knowledge distillation* (Hinton et al., 2015). Given pairs of MedMCQA question

[13]The effective sample size is correlated with the inverse of the variance of $w_{\theta,\phi}(\mathbf{a}, \mathbf{q}, \mathbf{d})$, it is a popular diagnostic for importance sampling. See Kong (1992); Owen (2013); Nowozin (2015).

[14]`https://www.findzebra.com/api/`

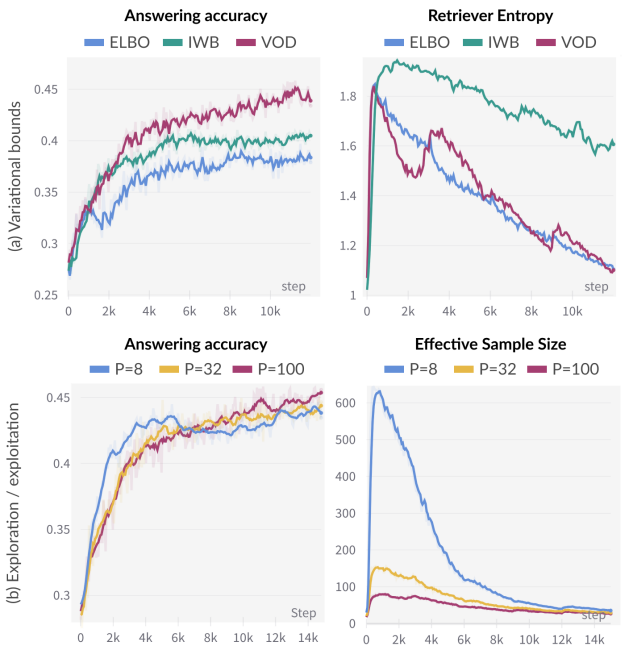

Figure 4. Answering accuracy and retriever entropy. **(a) Variational bounds:** effect of the choice of parameter α (ELBO: $\alpha = 1$, IWB: $\alpha = 0$, RVB/VOD: interpolating α from 1 to 0), all using $P = 100$. **(b) Exploration / exploitation:** effect of the support size P of the truncated retrievers. We sampled $MK = 4 \cdot 8$ documents per question, resulting in $K^M = 4.096$ documents combinations (therefore the max. effective sample size is 4096). Higher P values leads to smaller effective sample sizes, slower learning but better end performances.

and answers $(\mathbf{q}, \mathbf{a}_\star)$, this translates into minimizing:

$$L_{\text{DISTILL.}} = D_{\text{KL}}\big(\underbrace{r_\phi(\mathbf{d} \mid [\mathbf{q}; \mathbf{a}_\star])}_{\substack{\text{MCQA Teacher} \\ \text{(question+answer)}}} \,\|\, \underbrace{p_\theta(\mathbf{d} \mid \mathbf{q})}_{\substack{\text{Student} \\ \text{(question only)}}}\big)\,. \quad (12)$$

Metrics In line with Dragusin et al. (2013), we evaluate retrieval by recording the first article that matches the reference CUI (disease concept) and report 100 $\times$ the mean reciprocal rank (MRR) and the fraction of queries for which the correct article is returned in the top 20.[15]

Retrieval performances We evaluated the VOD retriever with and without distillation, a hybrid retriever combining the VOD and BM25 score (defined as $f_\theta^{\text{VOD+BM25}}(\mathbf{d}, \mathbf{q}) := f_\theta(\mathbf{d}, \mathbf{q}) + \tau^{-1}\ \text{BM25}(\mathbf{d}, \mathbf{q})$ where $\tau = 5$), and BM25 alone. We found that a VOD retriever trained on MedMCQA via distillation can be competitive with the FindZebra API and achieves best performances when combined with a simple BM25 baseline, resulting in an MRR of 38.9.

Table 5. Retrieval performances on the FindZebra benchmark for a BioLinkBERT retriever trained using VOD on MedMCQA and one trained using task-specific distillation, with and without coupling with a BM25 score during evaluation.

Method	Distillation	MRR	Hit@20
VOD	✗	27.8	56.9
VOD	✓	31.7	58.1
VOD + BM25	✓	**38.9**	**64.1**
BM25	–	26.4	48.4
FINDZEBRA API	–	30.1	59.3

Retriever samples In Appendix G, Table 7, we present examples of a distilled VOD retriever's top-1 ranked passages, including two successes and two failures. The top-ranked documents were mostly relevant, but the retriever struggled with long keyword-based queries, as shown in row #4. This is likely due to the discrepancy of tasks between training on MedMCQA and evaluating on FZ queries.

5. Discussion

Knowledge vs. Reasoning Tasks The VOD framework was evaluated using the MedMCQA and USMLE datasets only utilizing BERT-based models. The MedMCQA dataset is designed to evaluate the knowledge of entry-level medical students, whereas the USMLE dataset targets trained medical professionals, who are expected to possess not only a comprehensive understanding of medicine but also the ability to reason about complex medical problems. The results obtained demonstrate the effectiveness of the VOD framework in the specific tasks, however, we speculate that a BERT-sized model may not be sufficient for handling reasoning-intensive questions. As reported in previous studies, larger models like PaLM and Codex, have shown exceptional performance in handling reasoning-heavy questions (Singhal et al., 2022; Liévin et al., 2022).

Large-scale datasets The nature of the task is not the sole factor limiting the performance of VOD. We showed that an initial round of training on the larger MedMCQA dataset (182k samples) strongly benefit performances on the USMLE dataset (10k samples) .[16] This suggests that VOD might benefit from larger-scale training, including other tasks such as retrieval-augmented language modelling.

[15] We re-used two of the metrics introduced in the original study. We considered the MRR to be more adequate than NDCG because not all documents with a relevant CUI can be considered as a relevant match; only the highest-ranking one is essential.

[16] Pretraining on MedMCQA improved downstream USMLE accuracy by +10.3% when compared to training on USMLE only.

Importance sampling In contrast to other methods, VOD requires defining the sampling distribution explicitly and thus makes the diagnosis of the suitability of the sampling distribution possible. As utilized in Figure 4, we suggest relying on the effective sample size diagnostic to measure the robustness of the likelihood estimates. A small effective sample size, with a value close to one, hints at a mismatch between the sampling distribution $r_\phi(\mathbf{d}|\mathbf{a}, \mathbf{q})$ and the posterior $p_\theta(\mathbf{d}|\mathbf{a}, \mathbf{q})$. In that case, the sampling distribution should be adapted and/or optimized end-to-end with the model. Furthermore, the α parameter of the VOD objective can be increased towards one to target looser variational bounds, which often come with a better optimization profile (Rainforth et al., 2018).

Approximating the IW-RVB The VOD objective serves as an approximate estimation of the IW-RVB, although its approximation error remains unaddressed. While the VOD objective is consistent w.r.t. the RVB (Appendix B), its reliance on the self-normalization introduces a deviation from the strict guarantee of being a lower bound for the marginal log-likelihood, which is provided by the IW-RVB. Nonetheless, the utilization of self-normalized importance sampling is generally preferred over un-normalized approaches due to its ability to reduce variance. To thoroughly understand the bias of the VOD objective and its gradient, additional theoretical analysis is required. Despite this, the VOD objective has demonstrated sufficient robustness in enabling end-to-end training of retrieval-augmented systems and efficiently bridging the performance gap that remained with larger, non-retrieval-augmented language models, as shown in Figure 1.

6. Conclusion

In conclusion, this study has provided a comprehensive examination of methods for enhancing retrieval-augmented models through variational inference. The proposed probabilistic framework, VOD, is a promising solution for achieving tractable, consistent, and end-to-end training of retrieval-augmented models. Through a series of extensive experiments on multiple-choice medical exam questions, utilizing the MedMCQA and MedQA-USMLE datasets, the effectiveness of the proposed framework have been demonstrated. The findings indicate that leveraging the Rényi variational bound yields better end-to-end performances while also optimizing at a faster rate. Additionally, this study has introduced truncated retriever parameterization with variable support size P, which generalizes existing top-K parameterization and allows for likelihood-based optimization based on the full range of documents. Furthermore, the results have shown that VOD outperforms the state-of-the-art Codex and domain-tuned Med-PaLM on MedMCQA in terms of both accuracy and parameter efficiency.

In the future, we plan to investigate various variations of VOD to enhance its versatility in modeling other datasets and tasks, as well as exploring the possibility of jointly learning the approximate posterior. Overall, this research provides a promising direction for designing and training likelihood-based models for retrieval-augmented tasks. We hope this research will help popularizing recent advances in variational inference and importance sampling, in the field of natural language processing and beyond.

Acknowledgements

VL's work was funded in part by Google DeepMind through a PhD grant. OW's work was funded in part by the Novo Nordisk Foundation through the Center for Basic Machine Learning Research in Life Science (NNF20OC0062606). VL and OW acknowledge support from the Pioneer Centre for AI, DNRF grant number P1.

References

Bodenreider, O. The unified medical language system (UMLS): integrating biomedical terminology. *Nucleic acids research*, 32(Database issue):D267–70, January 2004. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/gkh061.

Borgeaud, S., Mensch, A., Hoffmann, J., Cai, T., Rutherford, E., Millican, K., van den Driessche, G., Lespiau, J.-B., Damoc, B., Clark, A., de Las Casas, D., Guy, A., Menick, J., Ring, R., Hennigan, T., Huang, S., Maggiore, L., Jones, C., Cassirer, A., Brock, A., Paganini, M., Irving, G., Vinyals, O., Osindero, S., Simonyan, K., Rae, J. W., Elsen, E., and Sifre, L. Improving language models by retrieving from trillions of tokens. December 2021.

Brown, T., Mann, B., Ryder, N., and others. Language models are few-shot learners. *Advances in neural information processing systems*, 2020. ISSN 1049-5258.

Burda, Y., Grosse, R., and Salakhutdinov, R. Importance weighted autoencoders. September 2015.

Chen, D., Fisch, A., Weston, J., and Bordes, A. Reading wikipedia to answer Open-Domain questions. March 2017.

Chowdhery, A., Narang, S., Devlin, J., Bosma, M., Mishra, G., Roberts, A., Barham, P., Chung, H. W., Sutton, C., Gehrmann, S., Schuh, P., Shi, K., Tsvyashchenko, S., Maynez, J., Rao, A., Barnes, P., Tay, Y., Shazeer, N., Prabhakaran, V., Reif, E., Du, N., Hutchinson, B., Pope, R., Bradbury, J., Austin, J., Isard, M., Gur-Ari, G., Yin, P., Duke, T., Levskaya, A., Ghemawat, S., Dev, S., Michalewski, H., Garcia, X., Misra, V., Robinson, K., Fedus, L., Zhou, D., Ippolito, D., Luan, D., Lim, H., Zoph, B., Spiridonov, A., Sepassi, R., Dohan, D., Agrawal, S., Omernick, M., Dai, A. M., Pillai, T. S., Pellat, M., Lewkowycz, A., Moreira, E., Child, R., Polozov, O., Lee, K., Zhou, Z., Wang, X., Saeta, B., Diaz, M., Firat, O., Catasta, M., Wei, J., Meier-Hellstern, K., Eck, D., Dean, J., Petrov, S., and Fiedel, N. PaLM: Scaling language modeling with pathways. April 2022.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. BERT: Pre-training of deep bidirectional transformers for language understanding. October 2018.

Dragusin, R., Petcu, P., Lioma, C., Larsen, B., Jørgensen, H. L., Cox, I. J., Hansen, L. K., Ingwersen, P., and Winther, O. FindZebra: A search engine for rare diseases. *International journal of medical informatics*, 82(6):528–538, June 2013. ISSN 1386-5056. doi: 10.1016/j.ijmedinf.2013.01.005.

Duffield, N., Lund, C., and Thorup, M. Priority sampling for estimation of arbitrary subset sums. *Journal of the ACM*, 54(6):32–es, December 2007. ISSN 0004-5411. doi: 10.1145/1314690.1314696.

Falcon. PyTorch lightning. *GitHub. Note: https://github.com/PyTorchLightning/pytorch-lightning*.

Fedus, W., Zoph, B., and Shazeer, N. Switch transformers: Scaling to trillion parameter models with simple and efficient sparsity, 2021.

Grathwohl, W., Choi, D., Wu, Y., Roeder, G., and Duvenaud, D. Backpropagation through the void: Optimizing control variates for black-box gradient estimation. October 2017.

Gu, Y., Tinn, R., Cheng, H., Lucas, M., Usuyama, N., Liu, X., Naumann, T., Gao, J., and Poon, H. Domain-Specific language model pretraining for biomedical natural language processing. *ACM Trans. Comput. Healthcare*, 3(1):1–23, October 2021. ISSN 2691-1957. doi: 10.1145/3458754.

Guu, K., Lee, K., Tung, Z., Pasupat, P., and Chang, M. Retrieval augmented language model Pre-Training. In Iii, H. D. and Singh, A. (eds.), *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 3929–3938. PMLR, 2020.

Hendrycks, D., Burns, C., Basart, S., Zou, A., Mazeika, M., Song, D., and Steinhardt, J. Measuring massive multitask language understanding, 2021.

Hinton, Vinyals, and Dean. Distilling the knowledge in a neural network. *arXiv preprint arXiv*, 2015.

Hinton, G. E., Dayan, P., Frey, B. J., and Neal, R. M. The "wake-sleep" algorithm for unsupervised neural networks. *Science*, 268(5214):1158–1161, May 1995. ISSN 0036-8075. doi: 10.1126/science.7761831.

Hoffmann, J., Borgeaud, S., Mensch, A., Buchatskaya, E., Cai, T., Rutherford, E., Casas, D. d. L., Hendricks, L. A., Welbl, J., Clark, A., Hennigan, T., Noland, E., Millican, K., van den Driessche, G., Damoc, B., Guy, A., Osindero, S., Simonyan, K., Elsen, E., Rae, J. W., Vinyals, O., and Sifre, L. Training Compute-Optimal large language models, 2022.

Izacard, G. and Grave, E. Leveraging passage retrieval with generative models for open domain question answering. July 2020.

Izacard, G., Caron, M., Hosseini, L., Riedel, S., Bojanowski, P., Joulin, A., and Grave, E. Unsupervised dense information retrieval with contrastive learning. December 2021.

Izacard, G., Lewis, P., Lomeli, M., Hosseini, L., Petroni, F., Schick, T., Dwivedi-Yu, J., Joulin, A., Riedel, S., and

Grave, E. Few-shot learning with retrieval augmented language models. August 2022.

Jin, D., Pan, E., Oufattole, N., Weng, W.-H., Fang, H., and Szolovits, P. What disease does this patient have? a large-scale open domain question answering dataset from medical exams. *APPS. Applied Sciences*, 11(14): 6421, July 2021. ISSN 1454-5101, 2076-3417. doi: 10.3390/app11146421.

Johnson, J., Douze, M., and Jegou, H. Billion-scale similarity search with GPUs. *IEEE transactions on big data*, 7(3):535–547, July 2021. ISSN 2332-7790, 2372-2096. doi: 10.1109/tbdata.2019.2921572.

Jordan, M. I., Ghahramani, Z., Jaakkola, T. S., and Saul, L. K. An introduction to variational methods for graphical models. *Machine learning*, 37(2):183–233, November 1999. ISSN 0885-6125, 1573-0565. doi: 10.1023/A:1007665907178.

Kaplan, J., McCandlish, S., Henighan, T., Brown, T. B., Chess, B., Child, R., Gray, S., Radford, A., Wu, J., and Amodei, D. Scaling laws for neural language models. January 2020.

Karpukhin, V., Oğuz, B., Min, S., Lewis, P., Wu, L., Edunov, S., Chen, D., and Yih, W.-T. Dense passage retrieval for Open-Domain question answering. April 2020.

Khattab, O. and Zaharia, M. ColBERT: Efficient and effective passage search via contextualized late interaction over BERT. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 39–48. Association for Computing Machinery, New York, NY, USA, July 2020. ISBN 9781450380164. doi: 10.1145/3397271.3401075.

Khattab, O., Potts, C., and Zaharia, M. Relevance-guided supervision for OpenQA with ColBERT. *Transactions of the Association for Computational Linguistics*, 9:929–944, September 2021. ISSN 2307-387X. doi: 10.1162/tacl_a_00405.

Kingma, D. P. and Welling, M. Auto-Encoding variational bayes. December 2013.

Kong. A note on importance sampling using standardized weights. *University of Chicago, Dept. of Statistics, Tech. Rep*, 1992.

Kool, W., Van Hoof, H., and Welling, M. Stochastic beams and where to find them: The Gumbel-Top-k trick for sampling sequences without replacement. In Chaudhuri, K. and Salakhutdinov, R. (eds.), *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 3499–3508. PMLR, 2019a.

Kool, W., van Hoof, H., and Welling, M. Buy 4 REINFORCE samples, get a baseline for free! March 2019b.

Laurençon, H., Saulnier, L., Wang, T., Akiki, C., del Moral, A. V., Le Scao, T., Von Werra, L., Mou, C., Ponferrada, E. G., Nguyen, H., Frohberg, J., Šaško, M., Lhoest, Q., McMillan-Major, A., Dupont, G., Biderman, S., Rogers, A., Ben allal, L., De Toni, F., Pistilli, G., Nguyen, O., Nikpoor, S., Masoud, M., Colombo, P., de la Rosa, J., Villegas, P., Thrush, T., Longpre, S., Nagel, S., Weber, L., Muñoz, M. R., Zhu, J., Van Strien, D., Alyafeai, Z., Almubarak, K., Chien, V. M., Gonzalez-Dios, I., Soroa, A., Lo, K., Dey, M., Suarez, P. O., Gokaslan, A., Bose, S., Adelani, D. I., Phan, L., Yu, I., Pai, S., Lepercq, V., Ilic, S., Mitchell, M., Luccioni, S., and Jernite, Y. The BigScience corpus a 1.6TB composite multilingual dataset. June 2022.

Le, T. A., Kosiorek, A. R., Siddharth, N., Teh, Y. W., and Wood, F. Revisiting reweighted Wake-Sleep for models with stochastic control flow. May 2018.

Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C. H., and Kang, J. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, 36(4):1234–1240, February 2020. ISSN 1367-4803, 1367-4811. doi: 10.1093/bioinformatics/btz682.

Lee, K., Chang, M.-W., and Toutanova, K. Latent retrieval for weakly supervised open domain question answering. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 6086–6096, Florence, Italy, July 2019. Association for Computational Linguistics. doi: 10.18653/v1/P19-1612.

Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V., and Zettlemoyer, L. BART: Denoising Sequence-to-Sequence pre-training for natural language generation, translation, and comprehension. October 2019.

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.-T., Rocktäschel, T., Riedel, S., and Kiela, D. Retrieval-Augmented generation for Knowledge-Intensive NLP tasks. In Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M. F., and Lin, H. (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 9459–9474. Curran Associates, Inc., 2020.

Li, Y. and Turner, R. E. Rényi divergence variational inference. In Lee, D. D., Sugiyama, M., Luxburg, U. V., Guyon, I., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems 29*, pp. 1073–1081. Curran Associates, Inc., 2016.

Lieber, O., Sharir, O., Lenz, B., and Shoham, Y. Jurassic-1: Technical details and evaluation. Technical report, AI21 Labs, August 2021.

Liévin, V., Dittadi, A., Christensen, A., and Winther, O. Optimal variance control of the Score-Function gradient estimator for Importance-Weighted bounds. In *Advances in Neural Information Processing Systems*, volume 33, pp. 16591–16602, 2020.

Liévin, V., Hother, C. E., and Winther, O. Can large language models reason about medical questions? July 2022.

Masrani, V., Le, T. A., and Wood, F. The thermodynamic variational objective. June 2019.

Mnih, A. and Gregor, K. Neural variational inference and learning in belief networks. January 2014.

Mnih, A. and Rezende, D. Variational inference for monte carlo objectives. In Balcan, M. F. and Weinberger, K. Q. (eds.), *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pp. 2188–2196, New York, New York, USA, 2016. PMLR.

Nowozin, S. Effective sample size in importance sampling. `http://www.nowozin.net/sebastian/blog/effective-sample-size-in-importance-sampling.html`, September 2015. Accessed: 2022-5-9.

Owen, A. B. *Monte Carlo theory, methods and examples*. 2013.

Pal, A., Umapathi, L. K., and Sankarasubbu, M. MedMCQA: A large-scale Multi-Subject Multi-Choice dataset for medical domain question answering. In Flores, G., Chen, G. H., Pollard, T., Ho, J. C., and Naumann, T. (eds.), *Proceedings of the Conference on Health, Inference, and Learning*, volume 174 of *Proceedings of Machine Learning Research*, pp. 248–260. PMLR, 2022.

Paranjape, A., Khattab, O., Potts, C., Zaharia, M., and Manning, C. D. Hindsight: Posterior-guided training of retrievers for improved open-ended generation. October 2021.

Paszke, Gross, Massa, Lerer, and others. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 2019. ISSN 1049-5258.

Qu, Y., Ding, Y., Liu, J., Liu, K., Ren, R., Zhao, W. X., Dong, D., Wu, H., and Wang, H. RocketQA: An optimized training approach to dense passage retrieval for Open-Domain question answering. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 5835–5847, Online, June 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.naacl-main.466.

Radford, A., Narasimhan, K., Salimans, T., and Sutskever, I. Improving language understanding by generative pre-training. *cs.ubc.ca*, 2018.

Rae, J. W., Borgeaud, S., Cai, T., Millican, K., Hoffmann, J., Song, F., Aslanides, J., Henderson, S., Ring, R., Young, S., Rutherford, E., Hennigan, T., Menick, J., Cassirer, A., Powell, R., van den Driessche, G., Hendricks, L. A., Rauh, M., Huang, P.-S., Glaese, A., Welbl, J., Dathathri, S., Huang, S., Uesato, J., Mellor, J., Higgins, I., Creswell, A., McAleese, N., Wu, A., Elsen, E., Jayakumar, S., Buchatskaya, E., Budden, D., Sutherland, E., Simonyan, K., Paganini, M., Sifre, L., Martens, L., Li, X. L., Kuncoro, A., Nematzadeh, A., Gribovskaya, E., Donato, D., Lazaridou, A., Mensch, A., Lespiau, J.-B., Tsimpoukelli, M., Grigorev, N., Fritz, D., Sottiaux, T., Pajarskas, M., Pohlen, T., Gong, Z., Toyama, D., de Masson d'Autume, C., Li, Y., Terzi, T., Mikulik, V., Babuschkin, I., Clark, A., de Las Casas, D., Guy, A., Jones, C., Bradbury, J., Johnson, M., Hechtman, B., Weidinger, L., Gabriel, I., Isaac, W., Lockhart, E., Osindero, S., Rimell, L., Dyer, C., Vinyals, O., Ayoub, K., Stanway, J., Bennett, L., Hassabis, D., Kavukcuoglu, K., and Irving, G. Scaling language models: Methods, analysis & insights from training gopher. December 2021.

Rainforth, T., Kosiorek, A. R., Le, T. A., Maddison, C. J., Igl, M., Wood, F., and Teh, Y. W. Tighter variational bounds are not necessarily better. February 2018.

Robertson, S. and Zaragoza, H. The probabilistic relevance framework: BM25 and beyond. *Foundations and Trends® in Information Retrieval*, 3(4):333–389, 2009. ISSN 1554-0669. doi: 10.1561/1500000019.

Sachan, D. S., Reddy, S., Hamilton, W., Dyer, C., and Yogatama, D. End-to-End training of Multi-Document reader and retriever for Open-Domain question answering. *NeurIPS*, 2021.

Singhal, K., Azizi, S., Tu, T., Sara Mahdavi, S., Wei, J., Chung, H. W., Scales, N., Tanwani, A., Cole-Lewis, H., Pfohl, S., Payne, P., Seneviratne, M., Gamble, P., Kelly, C., Scharli, N., Chowdhery, A., Mansfield, P., Aguera y Arcas, B., Webster, D., Corrado, G. S., Matias, Y., Chou, K., Gottweis, J., Tomasev, N., Liu, Y., Rajkomar, A., Barral, J., Semturs, C., Karthikesalingam, A., and Natarajan, V. Large language models encode clinical knowledge. December 2022.

Smith, S., Patwary, M., Norick, B., LeGresley, P., Rajbhandari, S., Casper, J., Liu, Z., Prabhumoye, S., Zerveas, G., Korthikanti, V., Zhang, E., Child, R., Aminabadi, R. Y., Bernauer, J., Song, X., Shoeybi, M., He, Y., Houston, M., Tiwary, S., and Catanzaro, B. Using DeepSpeed and megatron to train Megatron-Turing NLG 530b, a Large-Scale generative language model, 2022.

Srivastava, A., Rastogi, A., Rao, A., Shoeb, A. A. M., Abid, A., Fisch, A., Brown, A. R., Santoro, A., Gupta, A., Garriga-Alonso, A., Kluska, A., Lewkowycz, A., Agarwal, A., Power, A., Ray, A., Warstadt, A., Kocurek, A. W., Safaya, A., Tazarv, A., Xiang, A., Parrish, A., Nie, A., Hussain, A., Askell, A., Dsouza, A., Slone, A., Rahane, A., Iyer, A. S., Andreassen, A., Madotto, A., Santilli, A., Stuhlmüller, A., Dai, A., La, A., Lampinen, A., Zou, A., Jiang, A., Chen, A., Vuong, A., Gupta, A., Gottardi, A., Norelli, A., Venkatesh, A., Gholamidavoodi, A., Tabassum, A., Menezes, A., Kirubarajan, A., Mullokandov, A., Sabharwal, A., Herrick, A., Efrat, A., Erdem, A., Karakaş, A., Roberts, B. R., Loe, B. S., Zoph, B., Bojanowski, B., Özyurt, B., Hedayatnia, B., Neyshabur, B., Inden, B., Stein, B., Ekmekci, B., Lin, B. Y., Howald, B., Diao, C., Dour, C., Stinson, C., Argueta, C., Ramírez, C. F., Singh, C., Rathkopf, C., Meng, C., Baral, C., Wu, C., Callison-Burch, C., Waites, C., Voigt, C., Manning, C. D., Potts, C., Ramirez, C., Rivera, C. E., Siro, C., Raffel, C., Ashcraft, C., Garbacea, C., Sileo, D., Garrette, D., Hendrycks, D., Kilman, D., Roth, D., Freeman, D., Khashabi, D., Levy, D., González, D. M., Perszyk, D., Hernandez, D., Chen, D., Ippolito, D., Gilboa, D., Dohan, D., Drakard, D., Jurgens, D., Datta, D., Ganguli, D., Emelin, D., Kleyko, D., Yuret, D., Chen, D., Tam, D., Hupkes, D., Misra, D., Buzan, D., Mollo, D. C., Yang, D., Lee, D.-H., Shutova, E., Cubuk, E. D., Segal, E., Hagerman, E., Barnes, E., Donoway, E., Pavlick, E., Rodola, E., Lam, E., Chu, E., Tang, E., Erdem, E., Chang, E., Chi, E. A., Dyer, E., Jerzak, E., Kim, E., Manyasi, E. E., Zheltonozhskii, E., Xia, F., Siar, F., Martínez-Plumed, F., Happé, F., Chollet, F., Rong, F., Mishra, G., Winata, G. I., de Melo, G., Kruszewski, G., Parascandolo, G., Mariani, G., Wang, G., Jaimovitch-López, G., Betz, G., Gur-Ari, G., Galijasevic, H., Kim, H., Rashkin, H., Hajishirzi, H., Mehta, H., Bogar, H., Shevlin, H., Schütze, H., Yakura, H., Zhang, H., Wong, H. M., Ng, I., Noble, I., Jumelet, J., Geissinger, J., Kernion, J., Hilton, J., Lee, J., Fisac, J. F., Simon, J. B., Koppel, J., Zheng, J., Zou, J., Kocoń, J., Thompson, J., Kaplan, J., Radom, J., Sohl-Dickstein, J., Phang, J., Wei, J., Yosinski, J., Novikova, J., Bosscher, J., Marsh, J., Kim, J., Taal, J., Engel, J., Alabi, J., Xu, J., Song, J., Tang, J., Waweru, J., Burden, J., Miller, J., Balis, J. U., Berant, J., Frohberg, J., Rozen, J., Hernandez-Orallo, J., Boudeman, J., Jones, J., Tenenbaum, J. B., Rule, J. S., Chua, J., Kanclerz, K., Livescu, K., Krauth, K., Gopalakrishnan, K., Ignatyeva, K., Markert, K., Dhole, K. D., Gimpel, K., Omondi, K., Mathewson, K., Chiafullo, K., Shkaruta, K., Shridhar, K., McDonell, K., Richardson, K., Reynolds, L., Gao, L., Zhang, L., Dugan, L., Qin, L., Contreras-Ochando, L., Morency, L.-P., Moschella, L., Lam, L., Noble, L., Schmidt, L., He, L., Colón, L. O., Metz, L., Şenel, L. K., Bosma, M., Sap, M., ter Hoeve, M., Farooqi, M., Faruqui, M., Mazeika, M., Baturan, M., Marelli, M., Maru, M., Quintana, M. J. R., Tolkiehn, M., Giulianelli, M., Lewis, M., Potthast, M., Leavitt, M. L., Hagen, M., Schubert, M., Baitemirova, M. O., Arnaud, M., McElrath, M., Yee, M. A., Cohen, M., Gu, M., Ivanitskiy, M., Starritt, M., Strube, M., Swędrowski, M., Bevilacqua, M., Yasunaga, M., Kale, M., Cain, M., Xu, M., Suzgun, M., Tiwari, M., Bansal, M., Aminnaseri, M., Geva, M., Gheini, M., T, M. V., Peng, N., Chi, N., Lee, N., Krakover, N. G.-A., Cameron, N., Roberts, N., Doiron, N., Nangia, N., Deckers, N., Muennighoff, N., Keskar, N. S., Iyer, N. S., Constant, N., Fiedel, N., Wen, N., Zhang, O., Agha, O., Elbaghdadi, O., Levy, O., Evans, O., Casares, P. A. M., Doshi, P., Fung, P., Liang, P. P., Vicol, P., Alipoormolabashi, P., Liao, P., Liang, P., Chang, P., Eckersley, P., Htut, P. M., Hwang, P., Miłkowski, P., Patil, P., Pezeshkpour, P., Oli, P., Mei, Q., Lyu, Q., Chen, Q., Banjade, R., Rudolph, R. E., Gabriel, R., Habacker, R., Delgado, R. R., Millière, R., Garg, R., Barnes, R., Saurous, R. A., Arakawa, R., Raymaekers, R., Frank, R., Sikand, R., Novak, R., Sitelew, R., LeBras, R., Liu, R., Jacobs, R., Zhang, R., Salakhutdinov, R., Chi, R., Lee, R., Stovall, R., Teehan, R., Yang, R., Singh, S., Mohammad, S. M., Anand, S., Dillavou, S., Shleifer, S., Wiseman, S., Gruetter, S., Bowman, S. R., Schoenholz, S. S., Han, S., Kwatra, S., Rous, S. A., Ghazarian, S., Ghosh, S., Casey, S., Bischoff, S., Gehrmann, S., Schuster, S., Sadeghi, S., Hamdan, S., Zhou, S., Srivastava, S., Shi, S., Singh, S., Asaadi, S., Gu, S. S., Pachchigar, S., Toshniwal, S., Upadhyay, S., Shyamolima, Debnath, Shakeri, S., Thormeyer, S., Melzi, S., Reddy, S., Makini, S. P., Lee, S.-H., Torene, S., Hatwar, S., Dehaene, S., Divic, S., Ermon, S., Biderman, S., Lin, S., Prasad, S., Piantadosi, S. T., Shieber, S. M., Misherghi, S., Kiritchenko, S., Mishra, S., Linzen, T., Schuster, T., Li, T., Yu, T., Ali, T., Hashimoto, T., Wu, T.-L., Desbordes, T., Rothschild, T., Phan, T., Wang, T., Nkinyili, T., Schick, T., Kornev, T., Telleen-Lawton, T., Tunduny, T., Gerstenberg, T., Chang, T., Neeraj, T., Khot, T., Shultz, T., Shaham, U., Misra, V., Demberg, V., Nyamai, V., Raunak, V., Ramasesh, V., Prabhu, V. U., Padmakumar, V., Srikumar, V., Fedus, W., Saunders, W., Zhang, W., Vossen, W., Ren, X., Tong, X., Zhao, X., Wu, X., Shen, X., Yaghoobzadeh, Y., Lakretz, Y., Song, Y., Bahri, Y., Choi, Y., Yang, Y., Hao, Y., Chen, Y., Belinkov, Y., Hou, Y., Hou, Y., Bai, Y., Seid, Z., Zhao, Z., Wang, Z., Wang, Z. J., Wang, Z., and Wu, Z. Beyond the imitation game: Quantifying and extrapolating the capabilities of

language models, 2022.

Taylor, R., Kardas, M., Cucurull, G., Scialom, T., Hartshorn, A., Saravia, E., Poulton, A., Kerkez, V., and Stojnic, R. Galactica: A large language model for science. November 2022.

Thoppilan, R., De Freitas, D., Hall, J., Shazeer, N., Kulshreshtha, A., Cheng, H.-T., Jin, A., Bos, T., Baker, L., Du, Y., Li, Y., Lee, H., Zheng, H. S., Ghafouri, A., Menegali, M., Huang, Y., Krikun, M., Lepikhin, D., Qin, J., Chen, D., Xu, Y., Chen, Z., Roberts, A., Bosma, M., Zhao, V., Zhou, Y., Chang, C.-C., Krivokon, I., Rusch, W., Pickett, M., Srinivasan, P., Man, L., Meier-Hellstern, K., Morris, M. R., Doshi, T., Santos, R. D., Duke, T., Soraker, J., Zevenbergen, B., Prabhakaran, V., Diaz, M., Hutchinson, B., Olson, K., Molina, A., Hoffman-John, E., Lee, J., Aroyo, L., Rajakumar, R., Butryna, A., Lamm, M., Kuzmina, V., Fenton, J., Cohen, A., Bernstein, R., Kurzweil, R., Aguera-Arcas, B., Cui, C., Croak, M., Chi, E., and Le, Q. LaMDA: Language models for dialog applications. January 2022.

Tucker, G., Mnih, A., Maddison, C. J., Lawson, J., and Sohl-Dickstein, J. REBAR: Low-variance, unbiased gradient estimates for discrete latent variable models. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems 30*, pp. 2627–2636. Curran Associates, Inc., 2017.

van den Oord, A., Vinyals, O., and Kavukcuoglu, K. Neural discrete representation learning. November 2017.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. Attention is all you need. *Advances in neural information processing systems*, 30, 2017. ISSN 1049-5258.

Venigalla, A., Frankle, J., and Carbin, M. PubMed GPT: A domain-specific large language model for biomedical text. `https://www.mosaicml.com/blog/introducing-pubmed-gpt`, 2022. Accessed: 2022-12-16.

Vieira, T. Estimating means in a finite universe. `https://timvieira.github.io/blog/post/2017/07/03/estimating-means-in-a-finite-universe/`, 2017. Accessed: 2022-NA-NA.

Yasunaga, M., Leskovec, J., and Liang, P. LinkBERT: Pretraining language models with document links. March 2022.

Zhang, S., Roller, S., Goyal, N., Artetxe, M., Chen, M., Chen, S., Dewan, C., Diab, M., Li, X., Lin, X. V., Mihaylov, T., Ott, M., Shleifer, S., Shuster, K., Simig, D., Koura, P. S., Sridhar, A., Wang, T., and Zettlemoyer, L. OPT: Open pre-trained transformer language models, 2022.

A. Priority sampling

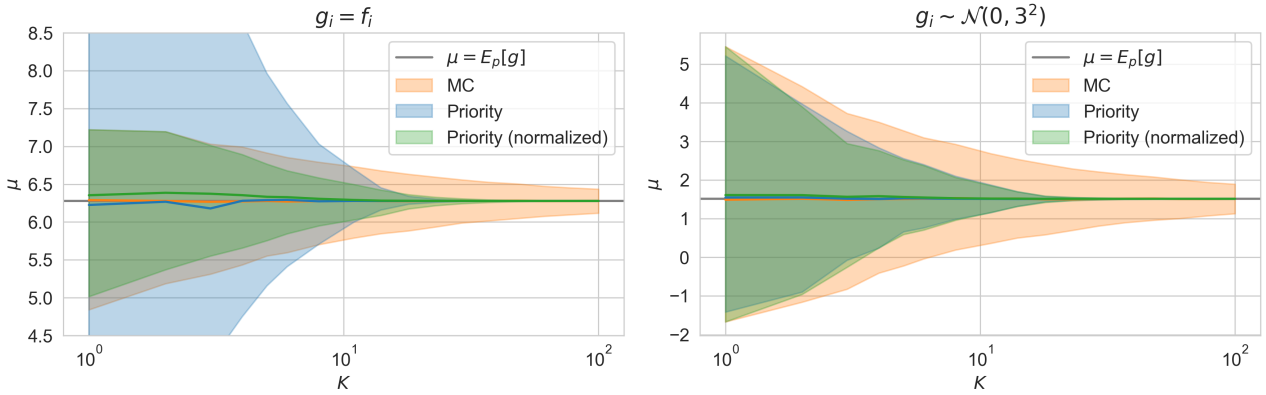

Figure 5. Estimation of the weighted average $\mu = \mathbb{E}_p[g]$ with weights $p_i := \sum_{i=1}^{N} {}^{\exp f_i}/_{\sum_{j=1} \exp f_j}$ where $f_i \sim \mathcal{N}(0, 3^2)$ and $N = 100$. We compare standard Monte-Carlo (sampling with replacement) with priority sampling and with self-normalized priority sampling (sampling without replacement). In the **left side** of the plot, we use $g_i = f_i$. In the **right side**, we use independent values $g_i \sim \mathcal{N}(0, 3^2)$ (sampled independently of f_i). We report the 80% CI interval for 10k estimates, each with $K = 1 \dots 100$. Priority sampling achieves higher variance than standard MC when $g_i = f_i$. Self-normalized priority sampling achieves lower variance than standard MC.

Given a set of probabilities $p_1, \dots, p_N$ and a function with values $f_1, \dots, f_N$, priority sampling (Duffield et al., 2007) allows estimating the sum $\sum_{i=1}^{N} p_i f_i$ using a subset of $K < N$ samples drawn *without replacement*.[17] For a sequence of random weights $u_1, \dots, u_n \overset{\text{iid}}{\sim} \text{Uniform}(0, 1]$, we define the priority keys p_i/u_i, set τ to be the $K+1$-th largest key, and define the set of K samples $\mathbb{S} = \{i \in [1, N] \mid p_i/u_i > \tau\}$. Using importance-weights $\bar{s}_i := \max(p_i, \tau)$, priority sampling yields an unbiased estimate of the weighted mean:

$$\mathbb{E}_{p(u_1,\dots,u_N)} \left[\sum_{i \in \mathbb{S}} \bar{s}_i f_i \right] = \sum_{i=1}^{N} p_i f_i \,. \tag{13}$$

Self-normalized importance sampling Empirically, the estimator eq. (13) might suffer from high variance. We follow (Kool et al., 2019a) and use self-normalize importance weights defined as $s_i := {}^{\bar{s}_i}/_{\sum_{j \in \mathbb{S}} \bar{s}_j}$ to reduce variance at the cost of introducing a bias. However, the estimator $\sum_{i \in \mathbb{S}} s_i f_i$ is biased but consistent: it equals the true expected value for $K = N$. The VOD objective uses self-normalized priority sampling.

Illustration In Figure 5, we visualize the variance of a standard Monte-Carlo (MC) estimator in two cases, a priority sampling estimator and a priority sampling estimator with self-normalized weights. In both cases, the variance of the self-normalized priority estimate is upper-bounded by the variance of the standard MC estimate and converges to zero at a faster rate than the traditional MC estimator. In one of the two cases, the un-normalized priority estimator suffers from large variance whereas the self-normalized priority estimator benefits from lower variance in both cases.

Product of priority sampling estimates Let $\mathbf{Z} = [\mathbf{z}_1, \dots, \mathbf{z}_M]$ be a vector of M independent variables, each defined on sets $\mathbb{Z}_1, \dots, \mathbb{Z}_M$, each of size N. The vector $\mathbf{Z}$ is defined on the set $\mathbb{Z}^{(M)} = \mathbb{Z}_1 \times \dots \times \mathbb{Z}_M$, the Cartesian product of the M sets, which corresponds to N^M combinations. Given a probability distribution $p(\mathbf{Z}) = \prod_{j=1}^{M} p(\mathbf{z}_j)$, we draw K samples for each component using priority sampling:

$$\mathbb{S}_j = \{\mathbf{z}_{j,1}, \dots, \mathbf{z}_{j,K}\} \tag{14a}$$

$$(\mathbf{z}_{j,1}, s_j[\mathbf{z}_1]), \dots, (\mathbf{z}_{j,K}, s_j[\mathbf{z}_K]) \overset{\text{priority}}{\sim} p(\mathbf{z}_j) \,. \tag{14b}$$

[17] We recommend Vieira (2017) for a great introduction to priority sampling.

Combining the per-component priority samples $p(\mathbf{Z}|\mathbf{Q})$ by defining the product priority weight allows estimating an average of a function $h(\mathbf{Z})$ weighted by $p(\mathbf{Z})$. Defining the product of priority weights as $s(\mathbf{Z}) := \prod_{j=1}^{M} s_j[\mathbf{z}_j]$, we have:

$$\mathbb{E}_{p(\mathbf{Z})}\left[h(\mathbf{Z})\right] = \mathbb{E}_{p(\mathbf{z}_1))}\left[\dots\left[\mathbb{E}_{p(\mathbf{z}_M)}\left[h(\mathbf{Z})\right]\right]\dots\right] \tag{15a}$$

$$\approx \sum_{\mathbf{z}_1\in\mathbb{S}_1} s_1[\mathbf{z}_1]\dots\sum_{\mathbf{z}_M\in\mathbb{S}_M} s_M[\mathbf{z}_M]h(\mathbf{Z}) \tag{15b}$$

$$= \sum_{\mathbf{z}_1\in\mathbb{S}_1}\dots\sum_{\mathbf{z}_M\in\mathbb{S}_M} s_1[\mathbf{z}_1]\dots s_M[\mathbf{z}_M]h(\mathbf{Z}) \tag{15c}$$

$$= \sum_{\mathbf{Z}\in\mathbb{S}^{(M)}} s(\mathbf{Z})h(\mathbf{Z})\,. \tag{15d}$$

B. VOD objective

Given a reader model $p_\theta(\mathbf{a}|\mathbf{d},\boldsymbol{q})$, and retriever model $p_\theta(\mathbf{d}|\mathbf{q})$ and a proposal $r_\phi(\mathbf{d}|\mathbf{a},\mathbf{q})$, the VOD objective is:

$$\hat{L}_\alpha^K(\mathbf{a},\mathbf{q}) := \frac{1}{1-\alpha}\log\sum_{i=1}^{K} s_i\,\hat{v}_{\theta,\phi}^{1-\alpha}(\mathbf{a},\mathbf{q},\mathbf{d}_i) \tag{16a}$$

$$(\mathbf{d}_1,s_i),\dots,(\mathbf{d}_K,s_K) \overset{\text{priority}}{\sim} r_\phi(\mathbf{d}|\mathbf{a},\mathbf{q}) \tag{16b}$$

$$\hat{v}_{\theta,\phi} := p_\theta(\mathbf{a}|\mathbf{q},\mathbf{d}_i)\zeta(\mathbf{d}_i)\left(\sum_{j=1}^{K} s_j\zeta(\mathbf{d}_j)\right)^{-1}. \tag{16c}$$

The VOD objective is a self-normalized importance sampling estimate of the RVB, and thus converges with probability one (*consistency*). Denoting $\mathcal{T}_\phi$ the support of $p_\theta(\mathbf{d}|\mathbf{q})$, we have:

$$\lim_{K\to|\mathcal{T}_\phi|}\underbrace{\hat{L}_\alpha^K(\mathbf{a},\mathbf{q})}_{\text{VOD}} = \underbrace{\mathcal{L}_\alpha(\mathbf{d},\mathbf{q})}_{\text{RVB}}\,. \tag{17}$$

Without loss of generality, we consider a joint reader-retriever model $p_\theta(\mathbf{a},\mathbf{d}|\mathbf{q}) = p_\theta(\mathbf{a}|\mathbf{d},\mathbf{q})p_\theta(\mathbf{d}|\mathbf{q})$ with retriever and sampling distribution defined on a support of documents $\mathcal{T}_\phi$[18] and parameterized as

$$p_\theta(\mathbf{d}|\mathbf{q}) := Z_\theta^{-1}\exp f_\theta(\mathbf{d},\mathbf{q}), \qquad r_\phi(\mathbf{d}|\mathbf{a},\mathbf{q}) := Z_\phi^{-1}\exp f_\phi(\mathbf{a},\mathbf{d},\mathbf{q}) \tag{18}$$

$$Z_\theta := \sum_{\mathbf{d}\in\mathcal{T}_\phi}\exp f_\theta(\mathbf{d},\mathbf{q}), \qquad Z_\phi := \sum_{\mathbf{d}\in\mathcal{T}_\phi}\exp f_\phi(\mathbf{a},\mathbf{d},\mathbf{q})\,. \tag{19}$$

In this section, we first detail the properties of the VOD objective: its complexity and its relation to the importance-weighted Rényi variational bound (IW-RVB). As a second step, we derive the VOD objective and prove that it is consistent: the VOD objective converges to the IW-RVB with probability 1 as $K\to\infty$.

B.1. Complexity $\mathcal{O}(K)$

Evaluating the VOD objective eq. (16a) only requires evaluating $p_\theta(\mathbf{a}|\mathbf{d},\mathbf{q})$ (complexity $\mathcal{O}(1)$, generally one BERT/LM call) and evaluating the retrieval score $f_\theta(\mathbf{d},\mathbf{q})$ for each document $\mathbf{d}_1,\dots,\mathbf{d}_K$ (complexity $\mathcal{O}(1+K)$, generally one BERT/LM call per document and one call to encode the query $\mathbf{q}$). *Evaluating the VOD objective does not require evaluating the constant Z_θ* (complexity $\mathcal{O}(P)$, one call for each document in the set $\mathcal{T}_\phi$). This results in a computational complexity of $\mathcal{O}(2+K)=\mathcal{O}(K)$.[19]

[18] $\mathcal{T}_\phi$ can be chosen as the entire corpus of documents.

[19] The scores $f_\phi(\mathbf{d}_1),\dots,f_\phi(\mathbf{d}_K)$ of the sampling distribution are computed offline and therefore can be ignored.

B.2. VOD, IW-RVB, ELBO and marginal likelihood

Using a set $\mathbf{d}_1, \ldots, \mathbf{d}_K \sim r_\phi(\mathbf{d}|\mathbf{a}, \mathbf{q})$ sampled with replacement, the importance-weighted Rényi variational bound (IW-RVB) is defined as:

$$\hat{\mathcal{L}}_\alpha^K(\mathbf{d}, \mathbf{q}) := \frac{1}{1-\alpha} \log \frac{1}{K} \sum_{i=1}^K w_{\theta,\phi}^{1-\alpha}(\mathbf{a}, \mathbf{q}, \mathbf{d}_i)\,. \tag{20}$$

The IW-RVB is a lower-bound of the log-likelihood and for $\alpha = 0$, increasing the number of samples results in a tighter log-likelihood lower bound (Burda et al., 2015):

$$\mathcal{L}_{\mathrm{ELBO}}(\mathbf{a}, \mathbf{q}) \leq \hat{\mathcal{L}}_{\alpha=0}^K(\mathbf{d}, \mathbf{q}) \leq \hat{\mathcal{L}}_{\alpha=0}^{K+1}(\mathbf{d}, \mathbf{q}) \leq \log p_\theta(\mathbf{a}, \mathbf{q})\,. \tag{21}$$

In $\alpha = 0$, the RVB is defined by continuity as the ELBO (Li & Turner, 2016). In that case, increasing the number of Monte Carlo samples K does not result in a tighter bound:

$$\hat{\mathcal{L}}_{\alpha\to 1}^K(\mathbf{d}, \mathbf{q}) = \mathbb{E}_{r_\phi(\mathbf{d}_1,\ldots,\mathbf{d}_K|\mathbf{a},\mathbf{q})}\left[\frac{1}{K}\sum_{i=1}^K \log w_{\theta,\phi}(\mathbf{a}, \mathbf{q}, \mathbf{d}_i)\right] = \mathbb{E}_{r_\phi(\mathbf{d}|\mathbf{a},\mathbf{q})}\left[\log w_{\theta,\phi}(\mathbf{q}, \mathbf{a}, \mathbf{d})\right] = \mathcal{L}_{\mathrm{ELBO}}(\mathbf{a}, \mathbf{q})\,. \tag{22}$$

The VOD objective is a self-normalized importance sampling estimate of the RVB, whereas the IW-RVB is a standard importance sampling. The VOD objective only differs from the IW-RVB because (i) VOD relies on self-normalized priority sampling eq. (28a), (ii) the normalizing constant $Z_\theta Z_\phi^{-1}$ in the expression of the importance weight $w_{\theta,\phi}(\mathbf{a}, \mathbf{q}, \mathbf{d})$ is estimated with a self-normalized priority sampling estimate eq. (28b).

B.3. Derivation of the VOD objective

In this section, we derive the VOD objective. We begin by expressing the ratio of normalization constants Z_θ/Z_ϕ as a function of ζ (section B.3.1), and then apply this identity to approximate the importance weight $w_{\theta,\phi}(\mathbf{q}, \mathbf{a}, \mathbf{d})$ (section B.3.2). We conclude the deriving the VOD objective: an approximation of the IW-RVB using (i) priority sampling and (ii) the importance weight estimate (B.3.3).

B.3.1. Ratio of normalizing constants Z_θ/Z_ϕ

The quantity Z_θ/Z_ϕ can expressed as a function of the ratio of un-normalized retriever densities $\zeta(\mathbf{d}) := \exp f_\theta(\mathbf{d}, \mathbf{q})/\exp f_\phi(\mathbf{a}, \mathbf{d}, \mathbf{q})$ using the following identity:

$$Z_\theta Z_\phi^{-1} = \mathbb{E}_{r_\phi(\mathbf{d}|\mathbf{a},\mathbf{q})}\left[\zeta(\mathbf{d})\right]\,. \tag{23}$$

Proof The equality arises from the definition of the right-hand term:

$$\mathbb{E}_{r_\phi(\mathbf{d}|\mathbf{a},\mathbf{q})}\left[\zeta(\mathbf{d})\right] := \sum_{\mathbf{d}\in\mathcal{T}_\phi} r_\phi(\mathbf{d}|\mathbf{a}, \mathbf{q}) \frac{\exp f_\theta(\mathbf{d}, \mathbf{q})}{\exp f_\phi(\mathbf{a}, \mathbf{d}, \mathbf{q})} \tag{24a}$$

$$= \sum_{\mathbf{d}\in\mathcal{T}_\phi} \frac{\exp f_\phi(\mathbf{a}, \mathbf{d}, \mathbf{q})}{Z_\phi} \frac{\exp f_\theta(\mathbf{d}, \mathbf{q})}{\exp f_\phi(\mathbf{a}, \mathbf{d}, \mathbf{q})} = Z_\theta Z_\phi^{-1}\,. \tag{24b}$$

B.3.2. Estimation of the importance weight $w_{\theta,\phi}$

The importance weight $w_{\theta,\phi}(\mathbf{q}, \mathbf{a}, \mathbf{d})$ can be approximated using K retrieval scores $f_\theta(\mathbf{d}_1), \ldots, f_\theta(\mathbf{d}_K)$:

$$w_{\theta,\phi}(\mathbf{q}, \mathbf{a}, \mathbf{d}) \approx \hat{v}_{\theta,\phi}(\mathbf{q}, \mathbf{a}, \mathbf{d}) := p_\theta(\mathbf{a}|\mathbf{q}, \mathbf{d})\zeta(\mathbf{d}) \left(\sum_{j=1}^K s_j \zeta(\mathbf{d}_j)\right)^{-1} \tag{25a}$$

$$(\mathbf{d}_1, s_i), \ldots, (\mathbf{d}_K, s_K) \overset{\text{priority}}{\sim} r_\phi(\mathbf{d}|\mathbf{a}, \mathbf{q})\,.$$

Proof Using the eq. (23), we can express $w_{\theta,\phi}(\mathbf{q}, \mathbf{a}, \mathbf{d})$ as a function of the un-normalized retriever density ratio ζ:

$$w_{\theta,\phi}(\mathbf{a}, \mathbf{d}, \mathbf{q}) := \frac{p_\theta(\mathbf{a}|\mathbf{d}, \mathbf{q}) p_\theta(\mathbf{d}|\mathbf{q})}{r_\phi(\mathbf{d}|\mathbf{a}, \mathbf{q})} \tag{26a}$$

$$= p_\theta(\mathbf{a}|\mathbf{d}, \mathbf{q}) \zeta(\mathbf{d}) \left(Z_\theta Z_\phi^{-1} \right)^{-1} \tag{26b}$$

$$= p_\theta(\mathbf{a}|\mathbf{d}, \mathbf{q}) \zeta(\mathbf{d}) \left(\mathbb{E}_{r_\phi(\mathbf{d}|\mathbf{a},\mathbf{q})} \left[\zeta(\mathbf{d}) \right] \right)^{-1} . \tag{26c}$$

The expected value of $\zeta(\mathbf{d})$ can be estimated via Monte Carlo. Using priority sampling with samples $\mathbf{d}_1, \ldots, \mathbf{d}_K \sim r_\phi(\mathbf{d}|\mathbf{a}, \mathbf{q})$ and normalized priority weights $s_1, \ldots, s_K$ (section A), we obtain:

$$w_{\theta,\phi}(\mathbf{a}, \mathbf{d}, \mathbf{q}) \approx p_\theta(\mathbf{a}|\mathbf{d}, \mathbf{q}) \zeta(\mathbf{d}) \left(\sum_{j=1}^{K} s_j \zeta(\mathbf{d}_j) \right)^{-1} = v_{\theta,\phi}(\mathbf{a}, \mathbf{d}, \mathbf{q}) . \tag{27}$$

B.3.3. THE VOD OBJECTIVE

Given document samples $\mathbf{d}_1, \ldots, \mathbf{d}_K \overset{\text{priority}}{\sim} r_\phi(\mathbf{d}|\mathbf{a}, \mathbf{a})$ with self-normalized priority weights $s_1, \ldots, s_K$. The VOD objective $\hat{L}_\alpha^K(\mathbf{d}, \mathbf{q})$ is an approximation of the IW-RVB ($\hat{\mathcal{L}}_\alpha^K(\mathbf{d}, \mathbf{q})$, eq. (20)):

$$\hat{\mathcal{L}}_\alpha^K(\mathbf{d}, \mathbf{q}) \approx \frac{1}{1-\alpha} \log \sum_{i=1}^{K} s_i \, w_{\theta,\phi}^{1-\alpha}(\mathbf{a}, \mathbf{q}, \mathbf{d}_i) \qquad \text{(priority sampling)} \tag{28a}$$

$$\approx \frac{1}{1-\alpha} \log \sum_{i=1}^{K} s_i \, v_{\theta,\phi}^{1-\alpha}(\mathbf{a}, \mathbf{q}, \mathbf{d}_i) = \hat{L}_\alpha^K(\mathbf{d}, \mathbf{q}) . \qquad \text{(inserting eq. (25a))} \tag{28b}$$

B.4. VOD consistency

In a nutshell, the VOD objective is biased because some normalization terms are estimated via Monte Carlo. Nevertheless, the estimates used as denominator are themselves consistent. This results in a final estimate – the VOD objective – which is itself consistent.

In contrast to the IW-RVB eq. (20), the VOD objective $\hat{L}_\alpha^K$ is not guaranteed to be a lower bound of the marginal log-likelihood. Nonetheless, the VOD objective and its gradient are consistent: they converge to their target expressions (RVB) in the limit of $K \to |\mathcal{T}_\phi| < \infty$.

Proof Self-normalized priority sampling is consistent. Given an arbitrary function h such that $|h(\mathbf{x})| < \infty$ and K priority samples $(\mathbf{x}_1, s_1), \ldots, (\mathbf{x}_K, s_K) \overset{\text{priority}}{\sim} p(\mathbf{x})$ where $\mathbf{x} \in \mathcal{X}, |\mathcal{X}| < \infty$:

$$\lim_{K \to |\mathcal{X}|} \sum_i s_i h(\mathbf{x}_i) = \lim_{K \to |\mathcal{X}|} \sum_i \frac{\bar{s}_i}{\sum_j \bar{s}_j} h(\mathbf{x}_i) = \mathbb{E}_{p(\mathbf{x})} \left[\frac{h(\mathbf{x})}{\mathbb{E}_{p(\mathbf{x})}[1]} \right] = \mathbb{E}_{p(\mathbf{x})} \left[h(\mathbf{x}) \right] . \tag{29}$$

Assuming $|\zeta(\mathbf{d})| < \infty$, this result implies that $v_{\theta,\phi}$ is a consistent estimate of the importance weight $w_{\theta,\phi}$:

$$\lim_{K \to |\mathcal{T}_\phi|} v_{\theta,\phi}(\mathbf{a}, \mathbf{q}, \mathbf{d}) = p_\theta(\mathbf{a}|\mathbf{d}, \mathbf{q}) \zeta(\mathbf{d}) \left(\lim_{K \to |\mathcal{T}_\phi|} \sum_{j=1}^{K} s_j \zeta(\mathbf{d}_j) \right)^{-1} \tag{30a}$$

$$= p_\theta(\mathbf{a}|\mathbf{d}, \mathbf{q}) \zeta(\mathbf{d}) \left(\mathbb{E}_{r_\phi(\mathbf{d}|\mathbf{a},\mathbf{q})} \left[\zeta(\mathbf{d}) \right] \right)^{-1} \tag{30b}$$

$$= p_\theta(\mathbf{a}|\mathbf{d}, \mathbf{q}) \zeta(\mathbf{d}) \left(Z_\theta Z_\phi^{-1} \right)^{-1} \tag{30c}$$

$$= w_{\theta,\phi}(\mathbf{a}, \mathbf{q}, \mathbf{d}) . \tag{30d}$$

The VOD objective relies on the importance weight estimates, which are themselves consistent. Therefore for $\alpha < 1$:

$$\lim_{K\to|\mathcal{T}_\phi|} \hat{L}_\alpha^K(\mathbf{a},\mathbf{q}) = \lim_{K\to|\mathcal{T}_\phi|} \frac{1}{1-\alpha}\log\sum_{i=1}^K s_i\,\hat{v}_{\theta,\phi}^{1-\alpha}(\mathbf{a},\mathbf{q},\mathbf{d}_i) \tag{31a}$$

$$= \frac{1}{1-\alpha}\log\mathbb{E}_{r_\phi(\mathbf{d}|\mathbf{a},\mathbf{q})}\left[\lim_{K\to|\mathcal{T}_\phi|}\hat{v}_{\theta,\phi}^{1-\alpha}(\mathbf{a},\mathbf{q},\mathbf{d}_i)\right] \tag{31b}$$

$$= \frac{1}{1-\alpha}\log\mathbb{E}_{r_\phi(\mathbf{d}|\mathbf{a},\mathbf{q})}\left[w_{\theta,\phi}^{1-\alpha}(\mathbf{a},\mathbf{q},\mathbf{d})\right] \tag{31c}$$

$$= \mathcal{L}_\alpha(\mathbf{a},\mathbf{q}) = \lim_{K\to|\mathcal{T}_\phi|}\mathcal{L}_\alpha^K(\mathbf{a},\mathbf{q})\,. \tag{31d}$$

C. VOD gradient

The VOD gradient w.r.t. the parameter θ corresponds to a self-normalized importance sampling estimate of the RVB gradient. It corresponds to the IW-RVB gradient derived in (Li & Turner, 2016), except that further approximations are required to ensure the expression is tractable. The VOD gradient is expressed as

$$\mu_{\theta,\alpha,K}^{\mathrm{VOD}} := \sum_{i=1}^K \frac{s_i\ \left(p_\theta(\mathbf{a}|\mathbf{d}_i,\mathbf{q})\zeta(\mathbf{d}_i)\right)^{1-\alpha}}{\sum_{j=1}^K s_j\ \left(p_\theta(\mathbf{a}|\mathbf{d}_j,\mathbf{q})\zeta(\mathbf{d}_i)\right)^{1-\alpha}}\left(\nabla_\theta\log p_\theta(\mathbf{a}|\mathbf{d}_i,\mathbf{q}) + \mathbf{h}(\mathbf{d}_i,\mathbf{q})\right) \approx \nabla\mathcal{L}_\alpha^K(\mathbf{a},\mathrm{q}) \tag{32}$$

$$(\mathbf{d}_1,s_i),\ldots,(\mathbf{d}_K,s_K) \overset{\text{priority}}{\sim} r_\phi(\mathbf{d}|\mathbf{a},\mathbf{q})$$

where

$$\mathbf{h}(\mathbf{d}_i,\mathbf{q}) := \nabla_\theta f_\theta(\mathbf{d}_i,\mathbf{q}) - \sum_{j=1}^K \frac{s_j\,\zeta(\mathbf{d}_j)}{\sum_{k=1}^K s_k\,\zeta(\mathbf{d}_k)}\nabla_\theta f_\theta(\mathbf{d}_j,\mathbf{q}) \approx \nabla_\theta\log p_\theta(\mathbf{d}|\mathbf{q})\,. \tag{33}$$

The VOD gradient is consistent: it converges to the exact gradient $\nabla_\theta\mathcal{L}_\alpha(\mathbf{a},\mathbf{q})$ with probability one:

$$\lim_{K\to|\mathcal{T}_\phi|}\mu_{\theta,\alpha,K}^{\mathrm{VOD}} = \nabla_\theta\mathcal{L}_\alpha(\mathbf{a},\mathbf{q})\,. \tag{34}$$

The estimation of the gradient of the VOD objective w.r.t. the parameter ϕ will be left to future work. In all experiments included in this paper, the parameter ϕ is non trainable.

Proof Using the results from the previous section, the gradient of the RVB w.r.t the parameter θ can be estimated as:

$$\nabla_\theta\mathcal{L}_\alpha(\mathbf{a},\mathbf{q}) := \mathbb{E}_{r_\phi(\mathbf{d}|\mathbf{a},\mathbf{q})}\left[\widetilde{w_{\theta,\phi}^{1-\alpha}}(\mathbf{a},\mathbf{q},\mathbf{d})\ \nabla_\theta\log p_\theta(\mathbf{a},\mathbf{d}|\mathbf{q})\right] \tag{35a}$$

$$= \mathbb{E}_{r_\phi(\mathbf{d}|\mathbf{a},\mathbf{q})}\left[\frac{w_{\theta,\phi}^{1-\alpha}(\mathbf{a},\mathbf{q},\mathbf{d})}{\mathbb{E}_{r_\phi(\mathbf{d}'|\mathbf{a},\mathbf{q})}\left[w_{\theta,\phi}^{1-\alpha}(\mathbf{a},\mathbf{q},\mathbf{d}')\right]}\ \nabla_\theta\log p_\theta(\mathbf{a},\mathbf{d}|\mathbf{q})\right] \tag{35b}$$

$$= \mathbb{E}_{r_\phi(\mathbf{d}|\mathbf{a},\mathbf{q})}\left[\frac{\left(p_\theta(\mathbf{a}|\mathbf{d},\mathbf{q})\zeta(\mathbf{d})Z_\theta^{-1}Z_\phi\right)^{1-\alpha}}{\mathbb{E}_{r_\phi(\mathbf{d}'|\mathbf{a},\mathbf{q})}\left[\left(p_\theta(\mathbf{a}|\mathbf{d}',\mathbf{q})\zeta(\mathbf{d}')Z_\theta^{-1}Z_\phi\right)^{1-\alpha}\right]}\ \nabla_\theta\log p_\theta(\mathbf{a},\mathbf{d}|\mathbf{q})\right] \tag{35c}$$

$$= \sum_{\mathbf{d}\in\mathbb{S}}\frac{s(\mathbf{d})\ \left(p_\theta(\mathbf{a}|\mathbf{d},\mathbf{q})\zeta(\mathbf{d})\right)^{1-\alpha}}{\sum_{\mathbf{d}'\in\mathbb{S}}s(\mathbf{d})\ \left(p_\theta(\mathbf{a}|\mathbf{d}',\mathbf{q})\zeta(\mathbf{d})\right)^{1-\alpha}}\nabla_\theta\log p_\theta(\mathbf{a},\mathbf{d}|\mathbf{q})\,. \tag{35d}$$

Another approximation is required to estimate $\nabla_\theta\log p_\theta(\mathbf{a},\mathbf{d}|\mathbf{q}) = \nabla_\theta\log p_\theta(\mathbf{q}|\mathbf{d},\mathbf{q}) + \nabla_\theta\log p_\theta(\mathbf{d}|\mathbf{q})$ without paying

the price of evaluating Z_θ. We approximate the term $\nabla_\theta \log p_\theta(\mathbf{d}|\mathbf{q})$ using:

$$\begin{aligned}
\nabla_\theta \log p_\theta(\mathbf{d}|\mathbf{q}) =& \nabla_\theta f_\theta(\mathbf{d}, \mathbf{q}) - \nabla_\theta \log Z_\theta && \text{(36a)}\\
=& \nabla_\theta f_\theta(\mathbf{d}, \mathbf{q}) - \frac{\nabla_\theta Z_\theta}{Z_\theta} && \text{(36b)}\\
=& \nabla_\theta f_\theta(\mathbf{d}, \mathbf{q}) - \sum_{\mathbf{d}' \in \mathcal{T}_\phi} p_\theta(\mathbf{d}'|\mathbf{q}) \nabla_\theta f_\theta(\mathbf{d}', \mathbf{q}) && \text{(36c)}\\
=& \nabla_\theta f_\theta(\mathbf{d}, \mathbf{q}) - \sum_{\mathbf{d}' \in \mathcal{T}_\phi} r_\phi(\mathbf{d}'|\mathbf{a}, \mathbf{q}) \frac{p_\theta(\mathbf{d}'|\mathbf{q})}{r_\phi(\mathbf{d}'|\mathbf{a}, \mathbf{q})} \nabla_\theta f_\theta(\mathbf{d}', \mathbf{q}) && \text{(36d)}\\
=& \nabla_\theta f_\theta(\mathbf{d}, \mathbf{q}) - \mathbb{E}_{r_\phi(\mathbf{d}'|\mathbf{a},\mathbf{q})} \left[\frac{\zeta(\mathbf{d}')}{\mathbb{E}_{r_\phi(\mathbf{d}''|\mathbf{a},\mathbf{q})}[\zeta(\mathbf{d}'')]} \nabla_\theta f_\theta(\mathbf{d}', \mathbf{q}) \right] && \text{(36e)}\\
\approx& \nabla_\theta f_\theta(\mathbf{d}, \mathbf{q}) - \sum_{i=1}^{K} \frac{s_i\, \zeta(\mathbf{d}_i)}{\sum_{j=1}^{K} s_j\, \zeta(\mathbf{d}_j)} \nabla_\theta f_\theta(\mathbf{d}_i, \mathbf{q}) \,. && \text{(36f)}
\end{aligned}$$

This approximation is also consistent because self-normalized priority sampling is consistent (direct application of eq. (29)).

D. VOD and REALM

Using the truncated retriever $p_\theta(\mathbf{d}|\mathbf{q})$ defined on the support $\mathcal{T}_\phi$ of the top-K=P documents ranked by a cached score f_ϕ:

$$p_\theta(\mathbf{d}|\mathbf{q}) := \frac{\mathbb{1}[\mathbf{d} \in \mathcal{T}_\phi] \exp f_\theta(\mathbf{d}, \mathbf{q})}{\sum_{i=1}^{K} \exp f_\theta(\mathbf{d}_i, \mathbf{q})} \,. \tag{37}$$

the VOD objective aligns with REALM in $\alpha = 0$. This corresponds to the marginal log-likelihood truncated to the top K documents (the first step is a direct application of priority sampling being consistent):

$$\underbrace{\hat{L}_{\alpha=0}^{K=P}(\mathbf{a}, \mathbf{q})}_{\text{VOD}} = \log \sum_{i=1}^{K} r_\phi(\mathbf{d}_i|\mathbf{a}, \mathbf{q}) w_{\theta,\phi}(\mathbf{a}, \mathbf{q}, \mathbf{d}_i) = \log \sum_{i=1}^{K} p_\theta(\mathbf{d}_i, \mathbf{a}|\mathbf{q}) = \underbrace{\log p_\theta(\mathbf{a}|\mathbf{q})}_{\text{REALM}} \,. \tag{38}$$

E. Applications of the VOD framework

In this section, we detail how to apply the VOD framework to the tasks of language modelling as well as extractive, generative and multiple-choice ODQA. We also detail a solution to optimizing multi-documents readers (FiD) jointly.

E.1. Generative and extractive ODQA

The model $p_\theta(\mathbf{a}|\mathbf{d}, \mathbf{q})$ a machine reading comprehension component that can be implemented either using an extractive approach, as done in the original BERT (Devlin et al., 2018), or using a generative approach (Lewis et al., 2019). Applying the VOD framework to generative and extractive ODQA simply requires plugging the likelihood of the corresponding machine reading comprehension model $p_\theta(\mathbf{a}|\mathbf{d}, \mathbf{q})$ in the VOD objective and gradient (equations 6 and 32).

E.2. Retrieval-augmented language modelling

We consider the variable $\mathbf{a} = [\mathbf{a}_1, \ldots, \mathbf{a}_T]$ to be the sequence of tokens of length T and omit the conditioning variable $\mathbf{q}$. The retriever model $p_\theta(\mathbf{d}_t|\mathbf{a}_{<t})$ is defined on a set of documents $\mathbb{D}$. We consider a left-to-right factorized reader $p_\theta(\mathbf{a}) := \prod_{t=1}^{T} p_\theta(\mathbf{a}_t|\mathbf{a}_{<t})$. This allows us to define the following retrieval-augmented language model, with one retrieved document per token:

$$p_\theta(\mathbf{a}) := \prod_{t=1}^{T} \sum_{\mathbf{d}_t \in \mathbb{D}} p_\theta(\mathbf{d}_t|\mathbf{a}_{<t}) p_\theta(\mathbf{a}_t|\mathbf{d}_t, \mathbf{a}_{<t}) \,. \tag{39}$$

We apply the RVB to each step t using an sampling distribution $r_\phi(\mathbf{d}_t|\mathbf{a})$, this results in the following lower bound:

$$\log p_\theta(\mathbf{a}) \geq \log \prod_{t=1}^{T} \mathcal{L}_\alpha(\mathbf{a}_t, \mathbf{a}_{<t}) \tag{40a}$$

$$= \frac{1}{1-\alpha} \sum_{t=1}^{T} \log \mathbb{E}_{r_\phi(\mathbf{d}|\mathbf{a},\mathbf{q})} \left[w_{\theta,\phi}^{1-\alpha}(\mathbf{a}_t, \mathbf{a}_{<t}, \mathbf{d}_t) \right] . \tag{40b}$$

The above step-wise RVB $\mathcal{L}_\alpha(\mathbf{a}_t, \mathbf{a}_{<t})$ can be estimated using equation 6, its gradient is given in equation 32.

E.3. Fusion-in-Decoder (FiD)

In this work, we considered reader models $p_\theta(\mathbf{a}|\mathbf{d}, \mathbf{q})$ with a single document per sample. Alternatively, models such as FiD (Izacard & Grave, 2020) implement a reader model that allows reading multiple documents per sample. Given a set $\mathbb{S} := \{\mathbf{d}_1, \ldots, \mathbf{d}_K\}$ of documents, we denote the multi-document reader $p_\theta(\mathbf{a}|\mathbb{S}, \mathbf{q})$. Defining a distribution over the set of unique documents $p(\mathbb{S})$ with tractable sampling and density evaluation is challenging. EMDR (Sachan et al., 2021) optimized a multi-document reader jointly with a deep retriever. However, an auxiliary reader model $p_\theta(\mathbf{a}|\mathbb{S}, \mathbf{q}) := \prod_{i=1}^{K} p_\theta(\mathbf{a}|\mathbf{d}_i, \mathbf{q})$ is used to optimize a retriever model $p_\theta(\mathbb{S}|\mathbf{q}) := \prod_{i=1}^{K} p_\theta(\mathbf{d}_i|\mathbf{q})$. VOD can be applied by following the same strategy, and this is equivalent to optimizing a single-sample joint reader along with a multi-sample reader:

$$\mu_{\theta,\alpha,\mathbb{S}}^{\text{VOD}-\text{FiD}} := \underbrace{\nabla_\theta \log p_\theta(\mathbf{a}|\mathbb{S}, \mathbf{q})}_{\substack{\text{multi-sample} \\ \text{reader likelihood}}} + \underbrace{\mu_{\theta,\alpha,K}^{\text{VOD}}(\mathbf{a}, \mathbf{q}, \mathbb{S})}_{\text{single-sample VOD gradient}} . \tag{41}$$

E.4. Multiple-choice ODQA

Model In the multiple-choice setting, a vector of M answer options $\mathbf{A} := [\mathbf{a}_1, \ldots, \mathbf{a}_M]$ is given. We denote $\mathbf{a}$ the correct option and assume $\mathbf{a} \in \mathbf{A}$. We define the vector of M queries as $\mathbf{Q} = [\mathbf{q}_1, \ldots, \mathbf{q}_M]$ with $\mathbf{q}_j := [\mathbf{q}; \mathbf{a}_j]$ where $[\cdot;\cdot]$ denotes the concatenation operator. We denote $\mathbf{D} = [\mathbf{d}_1, \ldots, \mathbf{d}_M]$ a vector of M documents, one for each answer option. We adopt a truncated retriever parameterization, given a set $\mathcal{T}_\phi(\mathbf{q}_j)$ of top-P documents ranked by a function $f_\phi(\cdot, \mathbf{q}_j)$, for each answer option:

$$p_\theta(\mathbf{d}|\mathbf{q}_j) := \frac{\mathbb{1}[\mathbf{d} \in \mathcal{T}_\phi(\mathbf{q}_j)] \exp f_\theta(\mathbf{d}, \mathbf{q}_j)}{\sum_{\mathbf{d}' \in \mathcal{T}_\phi(\mathbf{q}_j)} \exp f_\theta(\mathbf{d}', \mathbf{q}_j)} , \quad r_\phi(\mathbf{d}|\mathbf{q}_j) := \frac{\mathbb{1}[\mathbf{d} \in \mathcal{T}_\phi(\mathbf{q}_j)] \exp f_\phi(\mathbf{d}, \mathbf{q}_j)}{\sum_{\mathbf{d}' \in \mathcal{T}_\phi(\mathbf{q}_j)} \exp f_\phi(\mathbf{d}, \mathbf{q}_j)} \tag{42}$$

Using the per-option retriever models, we define the multiple-choice ODQA model as:[20]

$$p_\theta(\mathbf{a}_\star|\mathbf{D}, \mathbf{Q}) := \frac{\exp g_\theta(\mathbf{d}_\star, \mathbf{q}_\star)}{\sum_{j=1}^{M} \exp g_\theta(\mathbf{d}_j, \mathbf{q}_j)} , \tag{43a}$$

$$p_\theta(\mathbf{D}|\mathbf{Q}) := \prod_{j=1}^{M} p_\theta(\mathbf{d}_j|\mathbf{q}_j) , \quad r_\phi(\mathbf{D}|\mathbf{Q}) := \prod_{j=1}^{M} r_\phi(\mathbf{d}_j|\mathbf{q}_j) . \tag{43b}$$

Denoting $F_\theta(\mathbf{D}, \mathbf{Q}) := \sum_{j=1}^{M} f_\theta(\mathbf{d}_j, \mathbf{q}_j)$ and $F_\phi(\mathbf{D}, \mathbf{Q}) := \sum_{j=1}^{M} f_\phi(\mathbf{d}_j, \mathbf{q}_j)$, the equation 43b can be re-written as:

$$p_\theta(\mathbf{D}|\mathbf{Q}) = \frac{\mathbb{1}[\mathbf{D} \in \boldsymbol{\mathcal{T}}_\phi^{(M)}] \exp F_\theta(\mathbf{D}, \mathbf{Q})}{\sum_{\mathbf{D}' \in \boldsymbol{\mathcal{T}}_\phi^{(M)}} \exp F_\phi(\mathbf{D}', \mathbf{Q})} \quad r_\phi(\mathbf{D}|\mathbf{Q}) = \frac{\mathbb{1}[\mathbf{D} \in \boldsymbol{\mathcal{T}}_{\boldsymbol{\phi}}^{(M)}] \exp F_\phi(\mathbf{D}, \mathbf{Q})}{\sum_{\mathbf{D}' \in \boldsymbol{\mathcal{T}}_\phi^{(M)}} \exp F_\phi(\mathbf{D}', \mathbf{Q})} . \tag{44}$$

where $\boldsymbol{\mathcal{T}}_\phi^{(M)} := \mathcal{T}_\phi(\mathbf{q}_1) \times \ldots \times \mathcal{T}_\phi(\mathbf{q}_M)$ the set of combinations of M-document vectors (P^M combinations).

[20] In this paper we omitted the dependency of r_ϕ on the index of the correct answer $\mathbf{a}_\star$, which could be used to improve learning performances.

VOD By applying the results from section B to $\mathbf{a}_\star, \mathbf{D}, \mathbf{Q}$ with $\zeta(\mathbf{D}) = \exp F_\theta(\mathbf{D}, \mathbf{Q}) / \exp F_\phi(\mathbf{D}, \mathbf{Q})$ the VOD objective and its gradient are:

$$\hat{L}^K_\alpha(\mathbf{a}_\star, \mathbf{Q}) := \frac{1}{1-\alpha} \log \sum_{\mathbf{D} \in \mathbb{S}^{(M)}} s(\mathbf{D}) \left(\frac{\zeta(\mathbf{D}) p_\theta(\mathbf{a}_\star | \mathbf{D}, \mathbf{Q})}{\sum_{\mathbf{D}' \in \mathbb{S}^{(M)}} s(\mathbf{D}') \zeta(\mathbf{D}')} \right)^{1-\alpha} \tag{45a}$$

$$\mu^{\text{VOD}}_{\theta,\alpha,K}(\mathbf{a}_\star, \mathbf{Q}) := \sum_{\mathbf{D} \in \mathbb{S}^{(M)}} \frac{s(\mathbf{D}) \left(\zeta(\mathbf{D}) p_\theta(\mathbf{a}_\star | \mathbf{D}, \mathbf{Q})\right)^{1-\alpha}}{\sum_{\mathbf{D}' \in \mathbb{S}^{(M)}} s(\mathbf{D}') \left(\zeta(\mathbf{D}') p_\theta(\mathbf{a}_\star | \mathbf{D}', \mathbf{Q})\right)^{1-\alpha}} \left(\nabla_\theta \log p_\theta(\mathbf{A} | \mathbf{D}, \mathbf{Q}) + \mathbf{h}(\mathbf{D}' | \mathbf{Q}) \right) . \tag{45b}$$

where we define (we discuss the product of priority sampling estimates in Appendix A)

$$s(\mathbf{D}) := \prod_{j=1}^{M} s_j[\mathbf{D}_j]) \tag{46a}$$

$$(\mathbf{d}_{j,1}, s_j[\mathbf{d}_1]), \ldots, (\mathbf{d}_{j,K}, s_j[\mathbf{d}_K]) \overset{\text{priority}}{\sim} r_\phi(\mathbf{d} | \mathbf{q}_j) , \tag{46b}$$

$$\mathbb{S}_j := \{\mathbf{d}_{j,1}, \ldots, \mathbf{d}_{j,K}\} , \quad \mathbb{S}^{(M)} := \mathbb{S}_1 \times \ldots \times \mathbb{S}_M . \tag{46c}$$

Monte-Carlo estimation During training, the computational budget is tight, and the VOD objective and its gradient are evaluated using a single set of samples $\mathbb{S}^{(M)}$. During evaluation, we can leverage $C \geq 1$ Monte-Carlo samples $\mathbb{S}^M_1, \ldots, \mathbb{S}^M_C$, each containing K^M document combinations sampled from $r_\phi(\mathbf{D}|\mathbf{Q})$ without replacement, to estimate the RVB (and therefore the log-likelihood) more accurately. We use the following estimate:

$$\hat{p}_\theta(\mathbf{a}, \mathbf{Q}) := \frac{1}{C} \sum_{i=1}^{C} \frac{\exp \hat{L}^K_\alpha(\mathbf{a}, \mathbf{Q} | \mathbb{S}^{(M)}_i)}{\sum_{\mathbf{a}' \in \mathbf{A}} \exp \hat{L}^K_\alpha(\mathbf{a}', \mathbf{Q} | \mathbb{S}^{(M)}_i)} \tag{47a}$$

$$\mathbb{S}^{(M)}_i \overset{\text{priority}}{\sim} r_\phi(\mathbf{D}|\mathbf{Q}) , \quad \text{for } i \in [1, C] . \tag{47b}$$

F. Implementation

Table 6. Parameterization of the reader and retriever scores. The complexity is reported for a batch-size of one, M answer option, and for K documents and inputs $\mathbf{q}_j = [\mathbf{q}; \mathbf{a}_j]$ and $\mathbf{d}$ of lengths $L_\mathbf{q}$ and $L_\mathbf{a}$. When using a dual-encoder architecture, the parameters of he BERT backbone are shared across the two encoders.

Type	Complexity	Parameterization
dual-encoder	$M(L^2_\mathbf{q} + KL^2_\mathbf{d})$	$f_\theta(\mathbf{d}, \mathbf{q}_j) = \text{Linear}_{\theta[D]}(\text{BERT}_\theta(\mathbf{d}))^T \text{Linear}_{\theta[Q]}(\text{BERT}_\theta(\mathbf{q}_j))$
Cross attn.	$MK(L_\mathbf{q} + L_\mathbf{d})^2$	$g_\theta(\mathbf{d}, \mathbf{q}_j) = \text{Linear}_\theta(\text{BERT}_\theta([\mathbf{d}; \mathbf{q}_j]))$

Documents preprocessing We encode the text and title of all the articles using the relevant BERT tokenizer. For each article with encoded title $\mathbf{t}$ of length $L_\mathbf{t}$, we extract overlapping passages $\mathbf{p}$ of length $L_\mathbf{p} = 200 - 2 - L_\mathbf{t}$ with stride 100 tokens. For each passage, using `[DOC]` a special token added to the BERT vocabulary, we format each passage as

$$\mathbf{d} := [\texttt{[CLS]} \; ; \; \texttt{[DOC]} \; ; \; \mathbf{t} \; ; \; \mathbf{p}] . \tag{48}$$

Queries preprocessing We encode all questions and answer options using the tokenizer and store the question-answer pairs as

$$\mathbf{q}_j := [\texttt{[CLS]} \; ; \; \texttt{[QUERY]} \; ; \; \mathbf{q} \; ; \; \texttt{[SEP]} \; ; \; \mathbf{a}_j] \tag{49}$$

where the question $\mathbf{q}$ is truncated such as $|\mathbf{q}_j| \leq 312$ tokens and `[QUERY]` is an additional special token. On the reader side, we append the document passage $\mathbf{d}$ to the question-answer query q_j such that $\mathbf{q}_j := [\mathbf{d}; \texttt{[SEP]}; \texttt{[QUERY]}; \mathbf{q}; \texttt{[SEP]}; \mathbf{a}_j]$.

Reader We parameterize the reader score g_θ using a cross-attention model parameterized by another BERT backbone. Each query $\mathbf{q}_j = [\mathbf{q}; \mathbf{a}_j]$ is prepended with a document $\mathbf{d}$, and an additional linear layer is used to reduce the output of BERT at the `CLS` token to a scalar value, as originally done in (Devlin et al., 2018). See expression in Table 6.

Retriever We parameterize the retriever score f_θ using a dual encoder architecture similar to DPR, except that we share the BERT backbone across the two columns and one linear layer to project the output of each column. See expression in Table 6.

Hyperparameters We summarize the training, evaluation and model hyperparameters in Table 12.

G. Additional experimental data

In Table 77, we report retrieved top-1 passages for the distilled retriever (two successes and two failures). In Figure 7, we report the measurement of the $D_{\mathrm{KL}}\left(r_\phi(\mathbf{d}|\mathbf{q}) \,||\, p_\theta(\mathbf{d}|\mathbf{q})\right)$ during training of a VOD model. In Figure 6, we illustrated the FindZebra queries and corpus embedded using the trained BioLinkBERT model and projected using t-SNE.

G.1. Retrieval samples

Table 7. Top-1 passages retrieved for a selection of FindZebra queries with their annotated answer CUIs and the rank of the first matching article for VOD (BioLinkBERT onnly, with distillation) and the FindZebra API. We showcase the retriever model trained with task-specific distillation and without BM25 coupling (MRR 31.7). We highlight terms from the queries and passages relevant to each other.

	Query	Top-1 passage (VOD, BioLinkBERT backbone, with distillation)
1	Q: **widespread musculoskeletal pain for more than 6 months and point tenderness in at least 11 of 18 defined anatomical sites** A: Fibromyalgia (C0016053) *Hit rank:* $\mathrm{VOD}_{\mathrm{BioLinkBERT}}$=*1,* $\mathrm{FZ}_{\mathrm{API}}$=*1*	*Fibromyalgia.* (...) for IL-1 receptor antagonist, IL-6 and IL-8. Diagnosis The location of the nine paired tender points that comprise the 1990 American College of Rheumatology criteria for fibromyalgia There is no single pathological feature, laboratory finding or biomarker that can diagnose fibromyalgia and there is debate over what should be considered diagnostic criteria and whether an objective diagnosis is possible. In most cases, people with fibromyalgia symptoms may have laboratory test results that appear normal and many of their symptoms may mimic those of other rheumatic conditions such as arthritis or osteoporosis. The most widely accepted set of classification criteria for research purposes was elaborated in 1990 by the Multicenter Criteria Committee of the American College of Rheumatology. These criteria, which are known informally as "the ACR 1990", define fibromyalgia according to the presence of the following criteria: **A history of widespread pain lasting more than three months – affecting all four quadrants of the body, i.e., both sides, and above and below the waist**. Tender points – there (...)
2	Q: diagnosis for dementing syndrome characterized primarily by **impairment of interpersonal and executive function** A: Frontotemporal dementia (C0338451) *Hit rank:* $\mathrm{VOD}_{\mathrm{BioLinkBERT}}$=*1,* $\mathrm{FZ}_{\mathrm{API}}$=*8*	*Frontotemporal dementia.* (FTDs) are a group of neurodegenerative disorders associated with shrinking of the frontal and temporal anterior lobes of the brain. **Symptoms include marked changes in social behavior and personality, and/or problems with language**. People with behavior changes may have disinhibition (with socially inappropriate behavior), apathy and loss of empathy, hyperorality (eating excessive amounts of food or attempting to consume inedible things), agitation, compulsive behavior, and various other changes. Examples of problems with language include difficulty speaking or understanding speech. **Some people with FTD also develop a motor syndrome** such as parkinsonism or motor neuron disease (which may be associated with various additional symptoms). There is a strong genetic component to FTDs. It sometimes follows an autosomal dominant inheritance pattern, or sometimes there is a general family history of dementia or psychiatric disorders. The three main genes responsible for familial FTD are MAPT, GRN, and C9orf72. However, the (...)
3	Q: syndrome characterized by **cough, reversible wheezing, and peripheral blood eosinophilia** A: Asthma (C0004096), Reactive airway disease (C3714497) *Hit rank:* $\mathrm{VOD}_{\mathrm{BioLinkBERT}}$=*72,* $\mathrm{FZ}_{\mathrm{API}}$=*11*	*Löffler's syndrome.* (...) a parasitic infection such as irritable bowel syndrome, abdominal pain and cramping, skin rashes and fatigue. Löffler's syndrome itself will cause difficulty breathing, coughing as well as a fever. Contents 1 Diagnosis 2 Prevention 3 Epidemiology 4 History 5 See also 6 References 7 External links Diagnosis The diagnosis of Loffler's syndrome can be challenging, as the diagnostic criteria can be vague and consistent with a multitude of diseases or conditions. The disease's developmental trajectory is mostly unknown. Upon examination of symptoms, a doctor will likely request a chest x-ray looking for migratory pulmonary infiltrate, and blood testing, to confirm a diagnosis. Symptoms tend to be brief, but can range from mild to severe and include: fever, vomiting, increased **respirations or difficulty breathing, cough, wheeze**, and rash. Symptoms typically follow an exposure to allergens or certain drugs, and last approximately two weeks. **Eosinophilia is the main feature of diagnostic** (...)
4	Q: 5 year old, boy, congenital malformations, **malformations of the hands** and feet, bilateral strabismus, small tongue, impaired coordination, expressionless face, **prominent forehead**, depressed nasal bridge, hypoplastic thumbs, bilateral adactyly of the feet, **short stature**, severe myopia A: Mobius Syndrome (C0221060), Mobius II syndrome (C0853240) *Hit rank:* $\mathrm{VOD}_{\mathrm{BioLinkBERT}}$=$\infty$, $\mathrm{FZ}_{\mathrm{API}}$=*1*	*Achondroplasia.* (...) hypochondroplasia, but the features of achondroplasia tend to be more severe. **All people with achondroplasia have short stature**. The average height of an adult male with achondroplasia is 131 centimeters (4 feet, 4 inches), and the average height for adult females is 124 centimeters (4 feet, 1 inch). Characteristic features of achondroplasia include an average-size trunk, short arms and legs with particularly short upper arms and thighs, limited range of motion at the elbows, and an enlarged head (macrocephaly) with a **prominent forehead**. **Fingers are typically short and the ring finger and middle finger may diverge, giving the hand a three-pronged (trident) appearance**. People with achondroplasia are generally of normal intelligence. Health problems commonly associated with achondroplasia include episodes in which breathing slows or stops for short periods (apnea), obesity, (...)

G.2. Embedding space

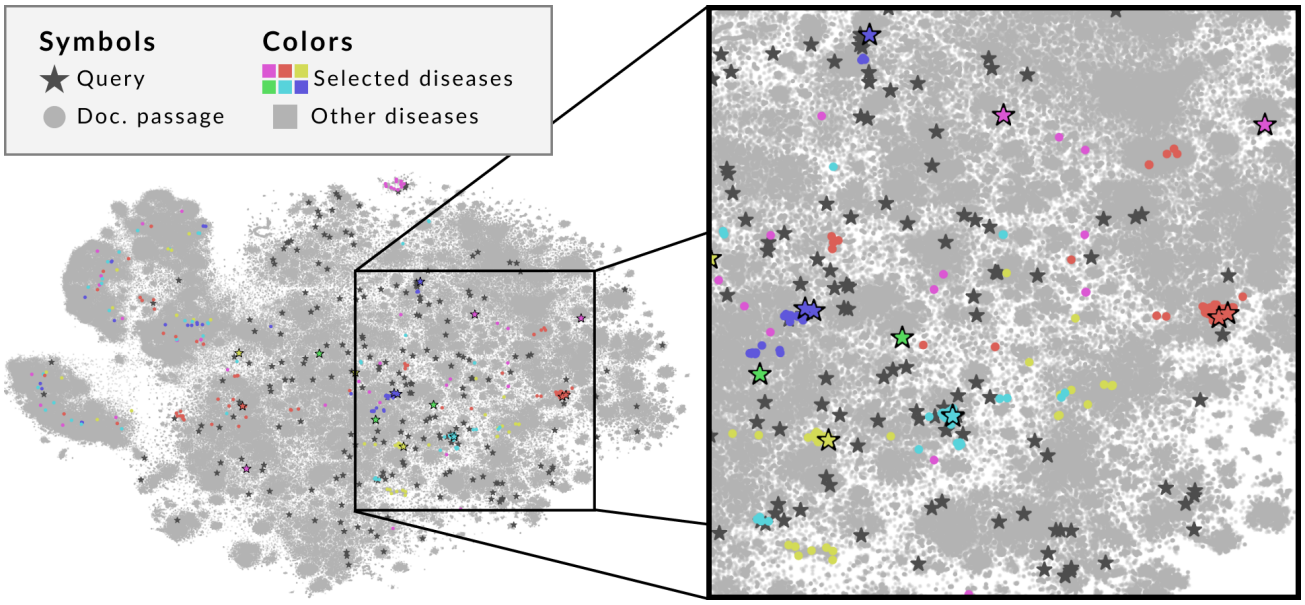

Figure 6. Visualizing the latent retrieval space. T-SNE projection of the embedding space where are encoded the 712k document passages of the FindZebra corpus and the 248 FindZebra queries. The documents and questions are annotated based on their disease identifier. The documents and queries annotated with the top 6 most frequent diseases (found in the queries) are highlighted with colours. The others are represented in gray. Some queries are successfully matched with a neighbourhood of relevant passages, although passages taken from a single document might be scattered across the embedding space.

G.3. Empirical divergence measured during training

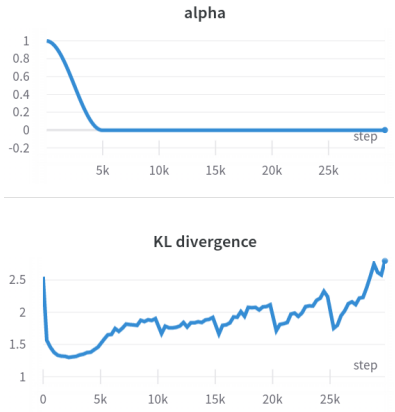

Figure 7. Measure of the divergence $D_{\mathrm{KL}}\left(r_\phi(\mathbf{d}|\mathbf{q}) \,||\, p_\theta(\mathbf{d}|\mathbf{q})\right)$ during the training of a VOD retriever on the USMLE dataset. The retriever checkpoint is updated every $T = 5k$ steps. α is annealed from 1 to 0 during the first 5k steps. We recognize the pattern schematized in Figure 3. In this example, the approximate posterior is chosen as a combination of a checkpoint of the retriever and a static BM25 component. Therefore the value of the divergence is never zero because the divergence between the model and the BM25 retriever is always strictly positive.

H. MedWiki

Table 8. Comparing the MedWiki with the original MedQA corpus on the USMLE dataset.

Method	Reader	Retriever	Corpus	Valid.	Test
Disjoint	BioBERT[1]	BM25	MedQA[2]	37.68	39.54
Disjoint	BioBERT[1]	BM25	MedWiki	38.82	40.46
Disjoint	BioLinkBERT	BM25	MedQA[2]	40.37	41.05
Disjoint	BioLinkBERT	BM25	MedWiki	**42.21**	**42.25**

[1]model weights from (Lee et al., 2020), [2]original corpus from (Jin et al., 2021)

The MedWiki corpus is a set of Wikipedia articles collected for research on medical question answering with low resources. Existing medical corpora, such as the MedQA corpus, are not adequately aligned with the ODQA task and are often measly and fragmented. At the same time, all of Wikipedia is cumbersome to use on consumer hardware. In order to reflect the true information need of medical experts, we assembled the MedWiki corpus by using real-world medical entrance exam questions. We queried the Wikipedia API using the answer options from all dataset splits of USMLE and MedMCQA and retained the top-10 articles for each answer option. This corpus includes 293.6k unique Wikipedia articles ($\approx 4.5\%$ of Wikipedia) that cover a broad range of medical topics.

MedQA vs. MedWiki

In the following paragraph, we compare the MedWiki corpus with the original MedQA corpus (Jin et al., 2021).

Qualitative comparison Using ElasticSearch, we compare the retrieved documents of MedWiki to the ones of MedQA. In Table 9, 10, 11 we present a few examples. The MedQA corpus is a selection of medical textbooks which often revolve around medical case studies, akin to the USMLE questions (see example in Table 9). In contrast, the MedWiki corpus references Wikipedia articles which are often edited to be concise, which is especially true for the abstract part of the articles, which contain the basic and usually most important information about a topic. Furthermore, each Wikipedia article comes with a title, which augments each passage with a higher-level context.

However, our approach of querying against the Wikipedia API results in many out-of-domain articles. For instance in Table 10, we display a MedWiki passage that originates from a non-medical article. Although the MedQA corpus is strictly oriented toward medical topics, it was built by extracting text from physical books using OCR software, which led to errors in the process and ultimately resulted in part of the corpus being unreadable.

Overall, both corpora provide adequate evidence to answer USMLE questions. Nevertheless, the MedWiki corpus is three times larger in vocabulary size and eight times more extensive in word count, making it more robust and diverse.

Quantitative comparison We investigated how the two corpora affect the final QA accuracy on the USMLE dataset. In contrast with the rest of the paper, we used a multi-document reader, as done in (Jin et al., 2021). We used an ElasticSearch index to retrieve the set of top 3 documents $\{\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3\}$ for each pair $(\mathbf{q}, \mathbf{a}_i)$ as context for each answer option. The normalized log probabilities over the four options were obtained by processing the set of concatenated tokens $[\mathbf{d}_1; \mathbf{d}_2; \mathbf{d}_3; \mathbf{q}; \mathbf{a}_i]$ with BERT. We performed all experiments using a batch size of 16, set the learning rate to 1e-5, and run all experiments for 30 epochs. We report the predictive accuracy averaged for three initial random seeds.

Table 8 summarizes the performance on the two corpora. We see that our collected MedWiki corpus leads to better QA performance by 0.9%-1.2% absolute. This result indicates that the MedWiki corpus can safely be used as a replacement of the MedQA corpus. The MedWiki yields USMLE accuracy that is superior to using the MedQA corpus (Table 8), and yields good results on the MedMCQA (Table **??**) despite consisting in only of a fraction of the English Wikipedia.

Question	a 5 year old girl is brought to the emergency department by her mother because of multiple episodes of nausea and vomiting that last about 2 hours. during this period she has had 6 8 episodes of bilious vomiting and abdominal pain. the vomiting was preceded by fatigue. the girl feels well between these episodes. she has missed several days of school and has been hospitalized 2 times during the past 6 months for dehydration due to similar episodes of vomiting and nausea. the patient has lived with her mother since her parents divorced 8 months ago. her immunizations are up to date. she is at the 60th percentile for height and 30th percentile for weight. she appears emaciated. her temperature is 36. 8 c 98. 8 f pulse is 99 min and blood pressure is 82 52 mm hg. examination shows dry mucous membranes. the lungs are clear to auscultation. abdominal examination shows a soft abdomen with mild diffuse tenderness with no guarding or rebound. the remainder of the physical examination shows no abnormalities. which of the following is the most likely diagnosis?
Options	**A: cyclic vomiting syndrome**, B: gastroenteritis, C: hypertrophic pyloric stenosis, D: gastroesophageal reflux disease
Document from MedQA	headache, and sweating patient presentation : be is a 45 - year - old woman who presents with concerns about sudden (paroxysmal), intense, brief episodes of headache, sweating (diaphoresis), and a racing heart (palpitations). focused history : be reports that the attacks started 3 weeks ago. they last from 2 to 10 minutes, during which time she feels quite anxious. during the attacks, it feels as though her heart is skipping beats (arrhythmia). at first, she thought the attacks were related to recent stress at work and maybe even menopause. the last time it happened, she was in a pharmacy and had her blood pressure taken. she was told it was 165 / 110 mm hg. be notes that she has lost weight (∼8 lbs) in this period even though her appetite has been good. pertinent findings : the physical examination was remarkable for be ' s thin, pale
Document from MedWiki	**panayiotopoulos syndrome**. pital, or calcarine sulci. follow - up meg demonstrated shifting localization or disappearance of meg spikes. illustrative cases in a typical presentation of panayiotopoulos syndrome, the child looks pale, vomits, and is fully conscious, able to speak, and understand but complains of " feeling sick. " two thirds of the seizures start in sleep ; the child may wake up with similar complaints while still conscious or else may be found vomiting, conscious, confused, or unresponsive. case 1. a girl had 2 seizures in sleep at 6 years of age. in the first fit she was found vomiting vigorously, eyes turned to one side, pale, and unresponsive. her condition remained unchanged for 3 hours before she developed generalized tonic - clonic convulsions. she gradually improved, and by the next morning was normal. the second seizure occurred 4 months later. she awoke and told her mother that she wanted to vomit,

Table 9. An example of the retrieved documents from the MedQA and MedWiki corpus respectively. Correct answers and document titles are highlighted when available.

Question	a 40 year old woman presents with difficulty falling asleep diminished appetite and tiredness for the past 6 weeks. she says that despite going to bed early at night she is unable to fall asleep. she denies feeling anxious or having disturbing thoughts while in bed. even when she manages to fall asleep she wakes up early in the morning and is unable to fall back asleep. she says she has grown increasingly irritable and feels increasingly hopeless and her concentration and interest at work have diminished. the patient denies thoughts of suicide or death. because of her diminished appetite she has lost 4 kg 8. 8 lb in the last few weeks and has started drinking a glass of wine every night instead of eating dinner. she has no significant past medical history and is not on any medications. which of the following is the best course of treatment in this patient?
Options	A: diazepam, B: paroxetine, C: zolpidem, **D: trazodone**
Document from MedQA	headache, and sweating patient presentation : be is a 45 - year - old woman who presents with concerns about sudden (paroxysmal), intense, brief episodes of headache, sweating (diaphoresis), and a racing heart (palpitations). focused history : be reports that the attacks started 3 weeks ago. they last from 2 to 10 minutes, during which time she feels quite anxious. during the attacks, it feels as though her heart is skipping beats (arrhythmia). at first, she thought the attacks were related to recent stress at work and maybe even menopause. the last time it happened, she was in a pharmacy and had her blood pressure taken. she was told it was 165 / 110 mm hg. be notes that she has lost weight (∼8 lbs) in this period even though her appetite has been good. pertinent findings : the physical examination was remarkable for be ' s thin, pale
Document from MedWiki	**hillary clinton's tenure as secretary of state**. hillary to the middle east to talk about how these countries can transition to new leaders — though, i've got to be honest, she's gotten a little passionate about the subject. these past few weeks it's been tough falling asleep with hillary out there on pennsylvania avenue shouting, throwing rocks at the window. in any case, obama's reference to clinton travelling a lot was true enough ; by now she had logged in her boeing 757, more than any other secretary of state for a comparable period of time, and had visited 79 countries while in the office. time magazine wrote that "clinton's endurance is legendary" and that she would still be going at the end of long work days even as her staff members were glazing out. the key was her ability to fall asleep on demand, at any time and place, for power naps. clinton also saw the potential political changes in the mideast as an opportunity for an even more fundamental change

Table 10. An example of the two different retrieved documents from the MedQA and MedWiki corpus. Correct answers and document titles are highlighted when available.

Question	a 37 year old female with a history of type ii diabetes mellitus presents to the emergency department complaining of blood in her urine left sided flank pain nausea and fever. she also states that she has pain with urination. vital signs include temperature is 102 deg f 39. 4 deg c blood pressure is 114 82 mmhg pulse is 96 min respirations are 18 and oxygen saturation of 97 on room air. on physical examination the patient appears uncomfortable and has tenderness on the left flank and left costovertebral angle. which of the following is the next best step in management?
Options	A: obtain an abdominal ct scan, **B: obtain a urine analysis and urine culture**, C: begin intravenous treatment with ceftazidime, D: no treatment is necessary
Document from MedQA	rim, & quinolones camille e. beauduy, pharmd, & lisa g. winston, md * a 59 - year - old woman presents to an urgent care clinic with a 4 - day history of frequent and painful urination. she has had fevers, chills, and flank pain for the past 2 days. her physician advised her to come immediately to the clinic for evaluation. in the clinic she is febrile (38. 5°c [101. 3°f]) but otherwise stable and states she is not experiencing any nausea or vomiting. her urine dipstick test is positive for leukocyte esterase. urinalysis and urine culture are ordered. her past medical history is significant for three urinary tract infections in the past year. each episode was uncom - plicated, treated with trimethoprim - sulfamethoxazole, and promptly resolved. she also has osteoporosis
Document from MedWiki	**hydronephrosis**. hydronephrosis describes dilation of the renal pelvis and calyces as a result of obstruction to urine flow. signs and symptoms the signs and symptoms of hydronephrosis depend upon whether the obstruction is acute or chronic, partial or complete, unilateral or bilateral. hydronephrosis that occurs acutely with sudden onset (as caused by a kidney stone) can cause intense pain in the flank area (between the hips and ribs). historically, this type of pain has been described as "dietl's crisis". conversely, hydronephrosis that develops gradually will generally cause either a dull discomfort or no pain. nausea and vomiting may also occur. an obstruction that occurs at the urethra or bladder outlet can cause pain and pressure resulting from distension of the bladder. blocking the flow of urine will commonly result in urinary tract infections which can lead to the development of stones, fever, and blood or pus in the urine

Table 11. An example of the two different retrieved documents from the MedQA and MedWiki corpus. Correct answers and document titles are highlighted when available.

Table 12. Hyperparameters used across the multiple-choice ODQA experiments.

Category	Parameter	Value
Optimization	Optimizer	AdamW
	Learning rate	$3 \cdot 10^{-6}$
	Learning rate warmup	$0.1 \cdot T$
	Warmup frequency	every T steps
	Weight decay	$1 \cdot 10^{-3}$
	Gradient clipping	0.5
	Precision	`float16`
α annealing	initial value	1
	final value	0
	length	T steps
	type	cosine
Model	Reader	BioLinkBERT + linear layer
	Retriever	BioLinkBERT + two linear layers
	Output vector size	768
Batching	batch-size	32
	M (# of options)	4
	K (documents per option)	8
	P (retriever support size)	100
	N (corpus size)	7,766.9k
	document passage stride	100
	$L_{\mathbf{d}}$ (document passage length)	200
	max. $L_{\mathbf{q}}$ (max. query length)	312
	max. $L_{\mathbf{d}} + L_{\mathbf{q}}$	512
Training	T (re-indexing period length)	5k
	Training steps (MedMCQA)	150k
	Training steps (USMLE)	50k
	Training steps (MedMCQA $\rightarrow$USMLE)	150k $\rightarrow$ 10k
	Training steps (Distillation)	120k
Posterior and retrieval	parameterization	$f_\phi^{\mathrm{ckpt}}(\mathbf{d}, [\mathbf{q}; \mathbf{a}]) + \tau^{-1} \left(\mathrm{BM25}(\mathbf{q}) + \beta \cdot \mathrm{BM25}(\mathbf{a})\right)$
	τ (BM25 temperature)	5
	β (BM25 answer weight)	$1 + 0.5 \max\{0, \log(L_{\mathbf{q}}/L_{\mathbf{a}})\}$
	BM25 implementation	`elasticsearch` v7.14.1
	BM25 paramters	b=0.75, k1=1.2
	MIPS implementation	`faiss v1.7.2`
	`faiss` factory string	IVF1000,Flat
	`faiss` precision	`float16`
	`faiss` nprobe	32
Evaluation	C (Monte-Carlo samples for eval.)	10
Hardware	CPU	AMD EPYC 7252 8-Core Processor
	RAM	256 GB
	GPU	8 $\times$ Quadro RTX 5000
	VRAM	128 GB
Software	PyTorch	(Paszke et al., 2019)
	Lightning	(Falcon)
	`faiss`	(Johnson et al., 2021)

Table 13. Mathematical symbols.

Category	Symbol	Description
ODQA variables	$\mathbf{a}$	answer
	$\mathbf{d}$	document or document passage
	$\mathbf{q}$	question or query
	$L_{\mathbf{a}}$	number of tokens in the answer
	$L_{\mathbf{d}}$	number of tokens in the document
	$L_{\mathbf{q}}$	number of tokens in the query
	$\mathbb{D}$	corpus of documents
	N	number of documents in the corpus
Reader-retriever	θ	parameter of the retrieval-augmented model (generative model)
	$p_\theta(\mathbf{a}, \mathbf{d}\|\mathbf{q})$	Joint reader-retriever model
	$w_{\theta,\phi}(\mathbf{a}, \mathbf{q}, \mathbf{d})$	Importance weight
	$\hat{v}_{\theta,\phi}(\mathbf{a}, \mathbf{q}, \mathbf{d})$	Self-normalized importance weight estimate
	$\zeta(\mathbf{d})$	un-normalized density ratio $\propto p_\theta(\mathbf{d}\|\mathbf{q}) r_\phi^{-1}(\mathbf{d}\|\mathbf{a}, \mathbf{q})$
	$p_\theta(\mathbf{a}\|\mathbf{d}, \mathbf{q})$	reader
	$p_\theta(\mathbf{d}\|\mathbf{q})$	retriever
	$f_\theta(\mathbf{d}, \mathbf{q})$	score of the retriever
Posterior	ϕ	parameter of the approximate posterior (inference network)
	$r_\phi(\mathbf{d}\|\mathbf{a}, \mathbf{q})$	approximate posterior (static retriever)
	$f_\phi(\mathbf{a}, \mathbf{d}, \mathbf{q})$	score of the approximate posterior
	$\mathrm{BM25}(\mathbf{q}, \mathbf{d})$	BM25 score of the query $\mathbf{q}$ for the document $\mathbf{d}$
	$f_\phi^{\mathrm{ckpt}}(\mathbf{d}, \mathbf{q})$	checkpoint of the retriever
	τ	temperature balancing the checkpoint score and the BM25 score
	β	weight balancing the query and answer options BM25 terms
Truncated retriever	P	number of documents with non-zero mass under $p_\theta(\mathbf{d}\|\mathbf{q})$
	$\mathcal{T}_\phi$	set of top-P documents ranked by f_ϕ (retrievers support)
Sampling	$(\mathbf{d}_1, s_1), \ldots, (\mathbf{d}_K, s_K) \overset{\text{priority}}{\sim} p(\mathbf{d})$	priority sampling (without replacement) wth samples $\mathbf{d}_i$ and weights s_i
	$s_1, \ldots, s_K$	priority weights
	K	number of document samples with $K \leq P \leq N$
	C	number of Monte-Carlo samples (evaluation)
Bounds	$\log p_\theta(\mathbf{a}, \mathbf{q})$	Marginal task likelihood
	$\mathcal{L}_{\mathrm{ELBO}}(\mathbf{a}, \mathbf{q})$	Variational Lower bound (ELBO)
	$\mathcal{L}_\alpha(\mathbf{a}, \mathbf{q})$	Rényi Variational Bound (RVB)
	$\mathcal{L}_\alpha^K(\mathbf{a}, \mathbf{q})$	importance-weighted RVB (IW-RVB)
	α	parameter of the RVB
	$\hat{L}_\alpha^K(\mathbf{a}, \mathbf{q})$	VOD objective (self-normalized importance sampling estimate of the RVB)
	$\mu_{\theta,\alpha,K}^{\mathrm{VOD}}$	VOD gradient
	$D_{\mathrm{KL}}(r_\phi(\mathbf{d}\|\mathbf{a}, \mathbf{q}) \| p_\theta(\mathbf{d}\|\mathbf{a}, \mathbf{q}))$	KL divergence from the true posterior to the approximate posterior
	$D_{\mathrm{KL}}(r_\phi(\mathbf{d}\|\mathbf{a}, \mathbf{q}) \| p_\theta(\mathbf{d}\|\mathbf{q}))$	KL divergence from the retriever to the approximate posterior
Multiple-choice	$\mathbf{a}_i$	answer option i
	$\star$	index of the correct answer option
	$\mathbf{q}_i$	question-answer pair $[\mathbf{q}; \mathbf{a}_i]$
	M	number of answer options
	$\mathbf{A}$	vector of M answer choices
	$\mathbf{D}$	vector of M documents
	$\mathbf{Q}$	vector of M queries (each expressed as $[\mathbf{q}; \mathbf{a}_i]$)
	$g_\theta(\mathbf{d}, \mathbf{q})$	score of the reader (multiple-choice)
	$\mathbb{S}^{(M)}$	Cartesian product of the per-option samples $\mathbb{S}_1, \ldots, \mathbb{S}_M$
	$\boldsymbol{\mathcal{T}}_{\boldsymbol{\phi}}^{(M)}$	Product of the per-option top-P sets $\mathcal{T}_\phi(\mathbf{q}_1) \times, \ldots, \times \mathcal{T}_\phi(\mathbf{q}_M)$
Spaces and Sets	Ω	space of strings
	$\mathbb{R}$	reals
	$(0, 1]$	real numbers in the interval [0, 1], 0 excluded
Operators	$:=$	defined as
	$[\cdot;\cdot]$	concatenation operator
	$\times$	Cartesian product
	$D_{\mathrm{KL}}(p\|\|q)$	Kullback–Leibler (KL) divergence from q to p
	$\mathbb{1}[\mathbf{x} \in \mathbb{X}]$	indicator function with value 1 if $\mathbf{x} \in \mathbb{X}$ otherwise 0