Learning with minimal human feedback

Alexandre Proutiere
joint work with William Réveillard, Vasileios Saketos, and Richard Combes

January 13, 2026 – CNI Seminar

Remove humans from the loop

Remove humans from the loop

Human feedback has been essential in

- Generating data used to train models (annotations)
- Aligning LLMs or other foundation models through RLHF
- Transferring expert knowledge to models
- ...

Remove humans from the loop

Human feedback has been essential in

- Generating data used to train models (annotations)
- Aligning LLMs or other foundation models through RLHF
- Transferring expert knowledge to models
- ...

Human feedback is expensive!

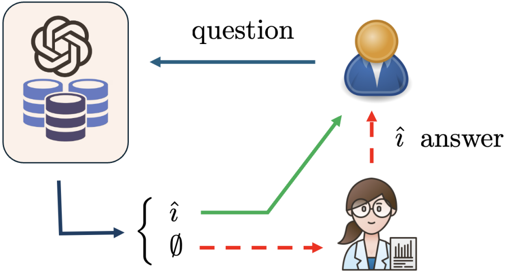
Can we remove humans or at least minimize their intervention?

Fine-tuning LLMs/RAGs to expert Q&A tasks

New expert knowledge, how to ”pass” this knowledge to an LLM?

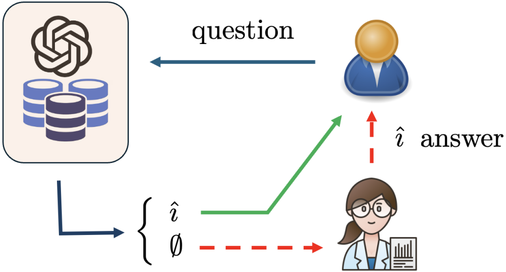
Fine-tuning LLMs/RAGs to expert Q&A tasks

New expert knowledge, how to "pass" this knowledge to an LLM?



Fine-tuning LLMs/RAGs to expert Q&A tasks

New expert knowledge, how to "pass" this knowledge to an LLM?



Initially, the system cannot answer *any* question ... How much human intervention is needed?

Online classification problem

In each round $t \geq 1$:

Online classification problem

In each round $t \geq 1$:

a. Query characterized by its embedding or representation $q_t \in \mathcal{E} \subset \mathbb{R}^d$ in an i.i.d. manner according to an unknown distribution μ. N possible labels $i \in [N]$.

Online classification problem

In each round $t \geq 1$:

a. Query characterized by its embedding or representation $q_t \in \mathcal{E} \subset \mathbb{R}^d$ in an i.i.d. manner according to an unknown distribution μ. N possible labels $i \in [N]$.

b. Agent's Guess with or without Expert Guidance

- dataset $\mathcal{D}_t$: query-label pairs that the expert labeled up to round t
- decision 1 (**ask the expert**): correct label $\hat{\imath}_t = i_t \in [N]$, (q_t, i_t) is appended to $\mathcal{D}_t$
- decision 2 (**guess**): $\hat{\imath}_t$, no feedback

Online classification problem

In each round $t \geq 1$:

a. Query characterized by its embedding or representation $q_t \in \mathcal{E} \subset \mathbb{R}^d$ in an i.i.d. manner according to an unknown distribution μ. N possible labels $i \in [N]$.

b. Agent's Guess with or without Expert Guidance

- dataset $\mathcal{D}_t$: query-label pairs that the expert labeled up to round t
- decision 1 (**ask the expert**): correct label $\hat{\imath}_t = i_t \in [N]$, (q_t, i_t) is appended to $\mathcal{D}_t$
- decision 2 (**guess**): $\hat{\imath}_t$, no feedback

Unobserved rewards: cost of calling the expert $\alpha = -1$, wrong answer $\gamma = -10$, and correct answer $\beta = +1$.

Reward obtained by algorithm π in round t: $r_\pi(t)$

Voronoi regret

Expected regret vs Oracle with knowledge of the expert labeling policy.

$$R_\pi(T) = \beta T - \sum_{t=1}^{T} \mathbb{E}[r_\pi(t)]$$

Voronoi regret

Expected regret vs Oracle with knowledge of the expert labeling policy.

$$R_\pi(T) = \beta T - \sum_{t=1}^{T} \mathbb{E}[r_\pi(t)]$$

Expert labeling policy dictated by a partition of convex polytopes $\mathcal{C}_1, \ldots, \mathcal{C}_N$ of $\mathcal{E}$, unknown to the agent. Whenever $q_t \in \mathcal{C}_i$, the expert provides label i.

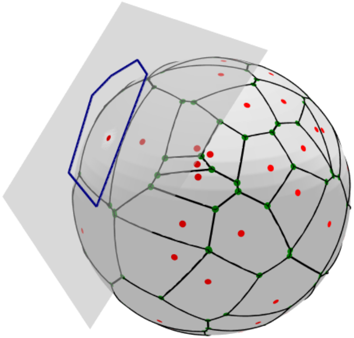
Voronoi regret

Expected regret vs Oracle with knowledge of the expert labeling policy.

$$R_\pi(T) = \beta T - \sum_{t=1}^{T} \mathbb{E}[r_\pi(t)]$$

Expert labeling policy dictated by a partition of convex polytopes $\mathcal{C}_1, \ldots, \mathcal{C}_N$ of $\mathcal{E}$, unknown to the agent. Whenever $q_t \in \mathcal{C}_i$, the expert provides label i.

Example: each label i, seed $s_i \in \mathcal{E}$.
$\mathcal{E}$ partitioned as the Voronoi tessellation generated by $s_1, \ldots, s_N$.



Algorithm design: intuition

The expected volume[a] of a convex hull formed by random points remains negligible **until the number of samples is exponential in** d.

[a]Chakraborti, D., Tkocz, T., and Vritsiou, B.-H. (2021). A note on volume thresholds for random polytopes. Geometriae Dedicata, 213(1):423–431.

Algorithm design: intuition

The expected volume[a] of a convex hull formed by random points remains negligible **until the number of samples is exponential in d.**

[a] Chakraborti, D., Tkocz, T., and Vritsiou, B.-H. (2021). A note on volume thresholds for random polytopes. Geometriae Dedicata, 213(1):423–431.

Large time-horizon ($\geq e^d$): volumes emerge, and we should learn the geometry of the cells.
The `CHC` (Conservative Hull-based Classifier) algorithm.

Moderate time-horizon ($\leq e^d$): cells have not emerge yet, approximating them as single points is sufficient.
The `CC` (Center-based Classifier) algorithm.

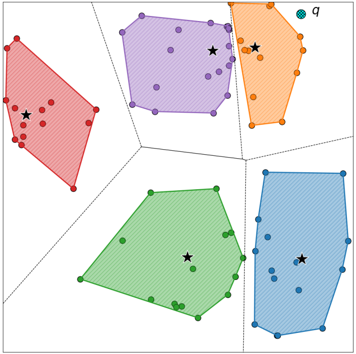
Large horizon: the CHC algorithm

CHC guesses only when it is sure!

Algorithm 1 Conservative Hull-based Classifier (CHC)

Initialize $\mathcal{Q}_{i,1} \leftarrow \emptyset$ for $i \in [N]$
for $t = 1, \ldots, T$ **do**
  if $\exists\, i \in [N] : q_t \in \text{hull}_{\mathcal{E}}(\mathcal{Q}_{i,t})$ **then**
    $\hat{\imath}_t \leftarrow i$
  else
    call expert, and set $\hat{\imath}_t \leftarrow i_t$
    $\mathcal{Q}_{i_t,t+1} \leftarrow \mathcal{Q}_{i_t,t} \cup \{q_t\}$
  end if
end for

The CHC algorithm



Hulls $\hat{\mathcal{C}}_{i,t}$ of CHC at $t = 200$. μ is a mixture of truncated Gaussian distributions with equal weights and covariance matrix $0.01I$. Stars are the seeds, circles are the queries that required an expert call.

Regret analysis of CHC

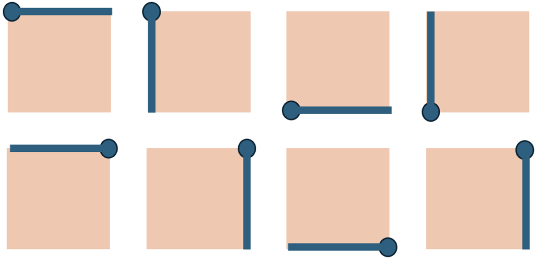
Flags of polytope. A flag of P is a sequence $(F_j)_{j=0}^{d-1}$ of faces[1] of P such that $\dim(F_j) = j$ and $F_0 \subset \cdots \subset F_{d-1}$.

[1] Faces of a polytope are specific planar surfaces on its boundary, e.g., for a cube, the 0-dim faces are the vertices, the 1-dim faces are the 12 edges, and the 2-dim faces are the 6 squares forming the boundary.

Regret analysis of CHC

Flags of polytope. A flag of P is a sequence $(F_j)_{j=0}^{d-1}$ of faces[1] of P such that $\dim(F_j) = j$ and $F_0 \subset \cdots \subset F_{d-1}$.

Number of flags of P: $F(P)$
A quadrilateral has $F(P) = 8$ flags



[1]Faces of a polytope are specific planar surfaces on its boundary, e.g., for a cube, the 0-dim faces are the vertices, the 1-dim faces are the 12 edges, and the 2-dim faces are the 6 squares forming the boundary.

Regret analysis of CHC

Theorem 1 *Assume that* μ *is abs. cont. w.r.t. Lebesgue measure with density in* $[c, C]$
(a) If $\mathcal{E} = [0,1]^d$ *and* $d \geq 2$*, then the regret of* CHC *satisfies*

$$R_{\texttt{CHC}}(T) \leq (\beta - \alpha)\frac{C}{c}\frac{\sum_{i=1}^{N} F(\mathcal{C}_i)}{(d+1)^{d-1}d!}\log^d(T) + \mathcal{O}(\log^{d-1}(T)\log\log(T)).$$

Regret analysis of CHC

Theorem 1 *Assume that μ is abs. cont. w.r.t. Lebesgue measure with density in $[c, C]$*
(a) If $\mathcal{E} = [0,1]^d$ and $d \geq 2$, then the regret of CHC *satisfies*

$$R_{\texttt{CHC}}(T) \leq (\beta - \alpha)\frac{C}{c}\frac{\sum_{i=1}^{N} F(\mathcal{C}_i)}{(d+1)^{d-1}d!}\log^d(T) + \mathcal{O}(\log^{d-1}(T)\log\log(T)).$$

(b) if $\mathcal{E} = \mathcal{S}^{d-1}$, $d \geq 3$ and each cell $\mathcal{C}_i$ is contained in an open halfsphere $\mathcal{S}_{e_i}^{+}$, then the regret of CHC *satisfies*

$$R_{\texttt{CHC}}(T) \leq (\beta - \alpha)\frac{KC}{c}\frac{\sum_{i=1}^{N} F(\mathcal{C}_i)}{d^{d-2}(d-1)!}\log^{d-1}(T) + \mathcal{O}(\log^{d-2}(T)\log\log(T)),$$

where $K = \max_{i\in[N]}\left(\frac{\max_{y\in\mathcal{C}_i} y^\top e_i}{\min_{y\in\mathcal{C}_i} y^\top e_i}\right)^d.$

Proof sketch

CHC always gives the right answer: $R_{\texttt{CHC}}(T) = (\beta - \alpha) \sum_{t=1}^{T} \sum_{i=1}^{N} \mathbb{E}[\mu(\mathcal{C}_i \setminus \hat{\mathcal{C}}_{i,t})]$.

[2] Bárány, I. and Buchta, C. (1993). Random polytopes in a convex polytope, independence of shape, and concentration of vertices. Mathematische Annalen, 297(1):467–497.

Proof sketch

CHC always gives the right answer: $R_{\texttt{CHC}}(T) = (\beta - \alpha) \sum_{t=1}^{T} \sum_{i=1}^{N} \mathbb{E}[\mu(\mathcal{C}_i \setminus \hat{\mathcal{C}}_{i,t})]$.

+ Proposition[2] *Let P a convex polytope in $\mathbb{R}^d$ with $d \geq 2$ and $n \geq 1$ points $p_1, \ldots, p_n$ sampled independently and uniformly at random in P, with convex hull P_n. Then*

$$\mathbb{E}[\lambda(P \setminus P_n)] = \frac{\lambda(P)F(P)}{(d+1)^{d-1}(d-1)!} \frac{\log^{d-1} n}{n} + \mathcal{O}\left(\frac{\log^{d-2}(n) \log \log n}{n}\right)$$

where $F(P)$ is the number of flags of P, i.e., the number of sequences $F_0 \subset F_1 \subset \cdots \subset F_{d-1}$ of i-dimensional faces of P.

[2] Bárány, I. and Buchta, C. (1993). Random polytopes in a convex polytope, independence of shape, and concentration of vertices. Mathematische Annalen, 297(1):467–497.

Proof sketch

CHC always gives the right answer: $R_{\texttt{CHC}}(T) = (\beta - \alpha) \sum_{t=1}^{T} \sum_{i=1}^{N} \mathbb{E}[\mu(\mathcal{C}_i \setminus \hat{\mathcal{C}}_{i,t})]$.

+ Proposition[2] *Let P a convex polytope in $\mathbb{R}^d$ with $d \geq 2$ and $n \geq 1$ points $p_1, \ldots, p_n$ sampled independently and uniformly at random in P, with convex hull P_n. Then*

$$\mathbb{E}[\lambda(P \setminus P_n)] = \frac{\lambda(P)F(P)}{(d+1)^{d-1}(d-1)!} \frac{\log^{d-1} n}{n} + \mathcal{O}\left(\frac{\log^{d-2}(n) \log\log n}{n}\right)$$

where $F(P)$ is the number of flags of P, i.e., the number of sequences $F_0 \subset F_1 \subset \cdots \subset F_{d-1}$ of i-dimensional faces of P.

+ Rejection sampling, to account for non-uniform distributions.

[2] Bárány, I. and Buchta, C. (1993). Random polytopes in a convex polytope, independence of shape, and concentration of vertices. Mathematische Annalen, 297(1):467–497.

Voronoi regret of CHC

Corollary *(a) if $\mathcal{E} = [0,1]^d$ and $d \geq 2$, then the regret of* CHC *satisfies*

$$R_{\texttt{CHC}}(T) \leq \frac{8(\beta - \alpha)CN}{3c(d+1)^{d-1}} \left(\frac{2e(N+2d)}{d-1} \right)^{d/2} \log^d(T) + \mathcal{O}(\log^{d-1}(T) \log\log(T)).$$

Voronoi regret of CHC

Corollary *(a) if $\mathcal{E} = [0,1]^d$ and $d \geq 2$, then the regret of* CHC *satisfies*

$$R_{\text{CHC}}(T) \leq \frac{8(\beta-\alpha)CN}{3c(d+1)^{d-1}} \left(\frac{2e(N+2d)}{d-1}\right)^{d/2} \log^d(T) + \mathcal{O}(\log^{d-1}(T)\log\log(T)).$$

(b) if $\mathcal{E} = \mathcal{S}^{d-1}$, $d \geq 3$ and each cell $\mathcal{C}_i$ is contained in an open half-sphere $\mathcal{S}^+_{e_i}$, then the regret of CHC *satisfies*

$$R_{\text{CHC}}(T) \leq \frac{4(\beta-\alpha)KCN}{cd^{d-2}} \left(\frac{2eN}{d-2}\right)^{(d-1)/2} \log^{d-1}(T) + \mathcal{O}(\log^{d-2}(T)\log\log(T)).$$

Minimax optimality of CHC in dimension 1

Theorem 2 *Assume that* $\mathcal{E} = [0,1]$ *and that* μ *has no atoms. Then for all* $T \geq 1$, *the regret of* CHC *satisfies* $R_{\mathtt{CHC}}(T) \leq 2(\beta - \alpha) N \log\left(T+1\right).$

Minimax optimality of CHC in dimension 1

Theorem 2 *Assume that $\mathcal{E} = [0,1]$ and that μ has no atoms. Then for all $T \geq 1$, the regret of* CHC *satisfies* $R_{\texttt{CHC}}(T) \leq 2(\beta - \alpha)N \log(T+1)$.

Theorem 3 *Assume that $\mathcal{E} = [0,1]$ and that μ is uniform on $\mathcal{E}$. Denote by $\theta = (s_1, \ldots, s_N) \in [0,1]^N$ a set of N query seeds in $[0,1]$. Then for all $T \geq 1$, the minimax regret satisfies*

$$\inf_{\pi} \max_{\theta \in [0,1]^N} R_\pi(T, \theta) \geq (\beta - \alpha)\frac{N-1}{64\sqrt{2}} \log\left(\frac{T+1}{2}\right) = \Omega((\beta - \alpha)(N-1)\log T).$$

Moderate horizon: the CC algorithm

Phase 1 (Explore) Expert called at each round t. $\hat{s}_i(t) := \frac{1}{|\mathcal{Q}_{i,t}|} \sum_{q \in \mathcal{Q}_{i,t}} q$

Estimated minimum center gap $\delta_{\min}$: $\hat{\delta}_{\min}(t) = \min_{i \neq j} \|\hat{s}_i(t) - \hat{s}_j(t)\|_2$.

Stopping criterion:

$$T_1 := \min\{t \leq T : \forall i \in [N] \, |\mathcal{Q}_{i,t}| \geq \frac{108\sigma^2}{\hat{\delta}^2_{\min}(t)} (d + 2 \log T)\}$$

Moderate horizon: the CC algorithm

Phase 1 (Explore) Expert called at each round t. $\hat{s}_i(t) := \frac{1}{|\mathcal{Q}_{i,t}|} \sum_{q \in \mathcal{Q}_{i,t}} q$
Estimated minimum center gap $\delta_{\min}$: $\hat{\delta}_{\min}(t) = \min_{i \neq j} \|\hat{s}_i(t) - \hat{s}_j(t)\|_2$.
Stopping criterion:

$$T_1 := \min\{t \leq T : \forall i \in [N] \, |\mathcal{Q}_{i,t}| \geq \frac{108\sigma^2}{\hat{\delta}^2_{\min}(t)} (d + 2\log T)\}$$

Phase 2 (Commit) In the second phase ($t \in [T_1 + 1, T]$), the center estimates are no longer updated. CC guesses at each round the label of the closest estimated center:
$\hat{\imath}_t = \arg\min_i \|q_t - \hat{s}_i(T_1)\|_2$.

Regret analysis of CC

Assumption *(Subgaussian mixture)* μ is a mixture of N σ-subgaussian distributions on $\mathbb{R}^d$ with component means s_i and mixture weights p_i.

Let $\delta_{\min} := \min_{i \neq j} \|s_i - s_j\|_2$.

[3] The expected distance between two random points scales as $\sqrt{d}$.

Regret analysis of CC

Assumption *(Subgaussian mixture)* μ is a mixture of N σ-subgaussian distributions on $\mathbb{R}^d$ with component means s_i and mixture weights p_i.

Let $\delta_{\min} := \min_{i \neq j} \|s_i - s_j\|_2$.

Theorem 4 *Let* $p_{\min} := \min_{i \in [N]} p_i$. *The regret of* CC *satisfies*

$$R_{\mathrm{CC}}(T) \leq \frac{(\beta - \alpha)(\log N + 1)}{p_{\min}} \left(1 + \frac{192\sigma^2(d + 2\log T)}{\delta_{\min}^2}\right) + (2\beta - \gamma - \alpha)N + (\beta - \gamma)\, T e^{-\frac{c-32}{48} d}.$$

[3] The expected distance between two random points scales as $\sqrt{d}$.

Regret analysis of CC

Assumption *(Subgaussian mixture)* μ is a mixture of N σ-subgaussian distributions on $\mathbb{R}^d$ with component means s_i and mixture weights p_i.

Let $\delta_{\min} := \min_{i \neq j} \|s_i - s_j\|_2$.

Theorem 4 *Let* $p_{\min} := \min_{i \in [N]} p_i$*. The regret of* CC *satisfies*

$$R_{\mathrm{CC}}(T) \leq \frac{(\beta - \alpha)(\log N + 1)}{p_{\min}} \left(1 + \frac{192\sigma^2(d + 2\log T)}{\delta_{\min}^2}\right) + (2\beta - \gamma - \alpha)N + (\beta - \gamma)Te^{-\frac{c-32}{48}d}.$$

If $T \leq e^d$, and if the mixture centers are sufficiently separated[3]: $\delta_{\min}^2 \geq 80\sigma^2 d$ then

$$R_{\mathrm{CC}}(T) \leq \frac{41}{5} \frac{(\beta - \alpha)(\log N + 1)}{p_{\min}} + 2(\beta - \gamma)(N + 1).$$

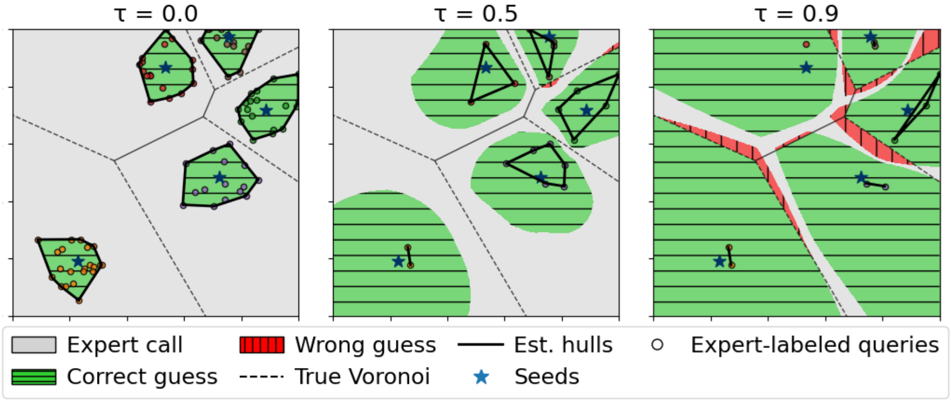
[3] The expected distance between two random points scales as $\sqrt{d}$.

Dealing with high dimension

Algorithm 2 Generalized Hull-based Classifier ($\texttt{GHC}(\tau)$)

Initialize $\mathcal{Q}_{i,1} \leftarrow \emptyset$ for $i \in [N]$
for $t = 1, \ldots, T$ **do**
    while $\exists\, i \in [N] : \mathcal{Q}_{i,t} = \emptyset$ **do**
        Apply Algorithm $\texttt{CHC}$
    end while
    if $\exists i \in [N] : d(q_t, \mathrm{hull}_{\mathcal{E}}(\mathcal{Q}_{i,t})) \le \tau \min_{j \neq i} d(q_t, \mathrm{hull}_{\mathcal{E}}(\mathcal{Q}_{j,t}))$ **then**
        $\hat{\imath}_t \leftarrow i$
    else
        Call expert, and set $\hat{\imath}_t \leftarrow i_t$
        $\mathcal{Q}_{i_t,t+1} \leftarrow \mathcal{Q}_{i_t,t} \cup \{q_t\}$
    end if
end for

The benefit of risk taking



In round 250: Decision regions of $\texttt{GHC}(\tau)$ for a mixture of truncated Gaussian distributions, covariance matrix $0.0025I$.

Experiments: synthetic data

Data

1. $\mathcal{I}^d = [0,1]^d$, $d \in \{1,4,10,50\}$, seeds $s_1, \ldots, s_N$ drawn uniformly on $\mathcal{I}^d$
   q_t drawn from the uniform distribution on $\mathcal{I}^d$ or from a homogeneous mixture of truncated Gaussians with covariance matrix $0.01I$

2. $\mathcal{S}^{d-1}$, $d \in \{2,4,10,50\}$, seeds $s_1, \ldots, s_N$ drawn uniformly on $\mathcal{S}^{d-1}$.
   q_t sampled either uniformly on $\mathcal{S}^{d-1}$ or from a mixture, specifically $i \in [N]$ drawn uniformly at random and $q_t = y_t/\|y_t\|$ where $y_t \sim \mathcal{N}(s_i, 0.01I)$.

Algorithms `CHC`, `GHC`, `CC`, and sequential k-means.

Voronoi regret

Voronoi regret of all algorithms for each experimental setup ($T = 5000$)

$\mathcal{E}$	Dim.	Dist.	ETC	CHC	GHC	Nearest-query GHC	SKM
$\mathcal{I}^d$	1	Unif.	6123 ± 390	142 ± 11	**110 ± 7**	186 ± 33	14128 ± 4275
		Mix.	4305 ± 770	130 ± 21	**106 ± 18**	245 ± 50	11141 ± 3741
	4	Unif.	9563 ± 486	2972 ± 72	**1593 ± 35**	5396 ± 95	30055 ± 3659
		Mix.	1064 ± 142	2337 ± 19	**573 ± 52**	1366 ± 99	793 ± 94
	10	Unif.	9782 ± 462	9489 ± 35	**5233 ± 94**	9544 ± 91	38900 ± 4268
		Mix.	33 ± 7	8821 ± 48	23 ± 4	29 ± 10	**20 ± 5**
	50	Unif.	9597 ± 143	10000 ± 0	**9559 ± 36**	10000 ± 0	42659 ± 3378
		Mix.	24 ± 9	10000 ± 0	**23 ± 5**	24 ± 9	**23 ± 5**
$\mathcal{S}^{d-1}$	2	Unif.	9347 ± 472	132 ± 12	**125 ± 6**	254 ± 42	26804 ± 1923
		Mix.	963 ± 173	138 ± 8	**104 ± 13**	155 ± 19	791 ± 645
	4	Unif.	6099 ± 889	1225 ± 31	**836 ± 40**	4005 ± 126	25645 ± 8157
		Mix.	**22 ± 7**	1083 ± 56	**22 ± 7**	**22 ± 7**	**22 ± 7**
	10	Unif.	7657 ± 544	4878 ± 150	**2550 ± 123**	9743 ± 51	39956 ± 3931
		Mix.	22 ± 9	7490 ± 74	**20 ± 11**	22 ± 9	**20 ± 11**
	50	Unif.	8947 ± 296	9983 ± 4	**6135 ± 73**	10000 ± 0	41246 ± 1091
		Mix.	26 ± 12	10000 ± 0	**19 ± 5**	26 ± 12	**19 ± 5**

Experiments: real-world data

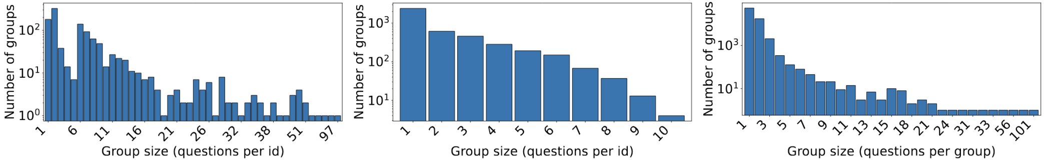
Datasets

- Quora Question Groups (QQG): 400,000 question pairs, annotated with a binary label indicating whether the questions are paraphrases of each other $\rightarrow$ $1,103$ distinct groups comprising a total of $7,365$ curated questions.
- ComQA[4]: 11,214 English questions collected from the WikiAnswers forum and grouped into 4,834 paraphrase clusters by crowd workers
- CQADupStack[5] : public benchmark with 99,785 questions organized into 74,519 groups.

[4] https://paperswithcode.com/dataset/comqa
[5] https://github.com/D1Doris/CQADupStack

Experiments: real-world data

Distribution of group sizes



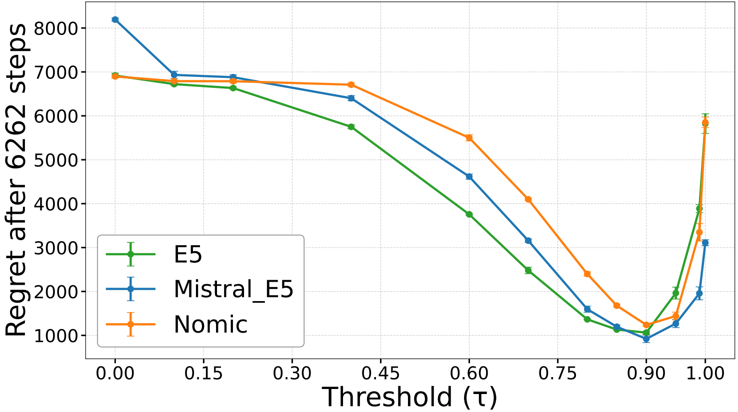
(a) Quora Question Groups (b) ComQA (c) CQADupStack

E5: EmbEddings from bidirEctional Encoder rEpresentations (E5). Bi-encoder architecture, where both the query and passage encoders are initialized with BERT. Embedding dimension 1,024.

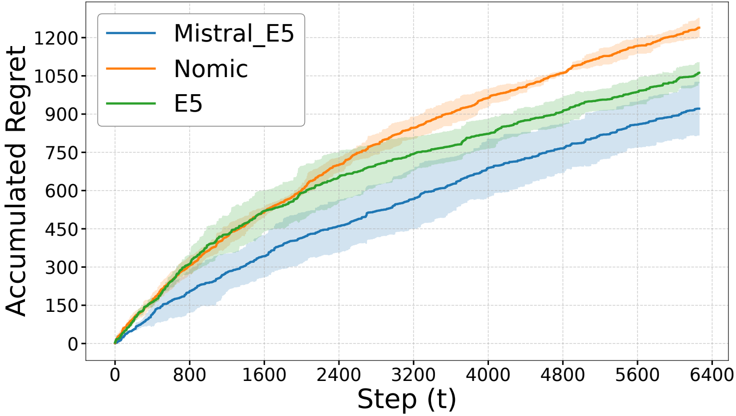
NOMIC: initialized from BERT and modified to address long-context retrieval. 100 million parameters and supports a sequence length of up to 2048. Embedding dimension 784.

Mistral_E5: unidirectional decoder architecture. The model initialized from Mistral 7B and consists of 7 billion parameters. Embedding dimension 4,096.

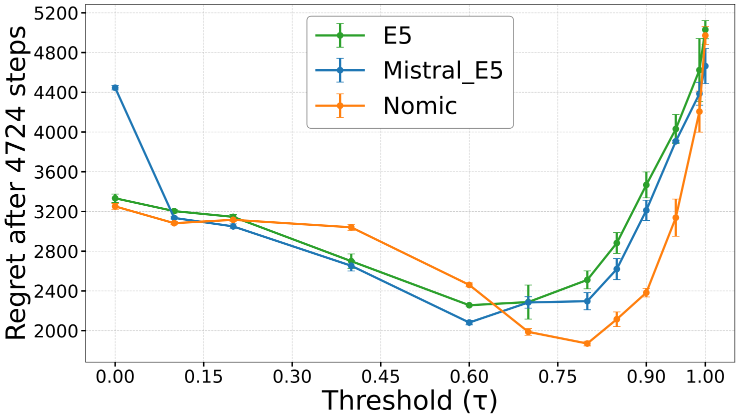
Regret of GHC on Quora Question Groups



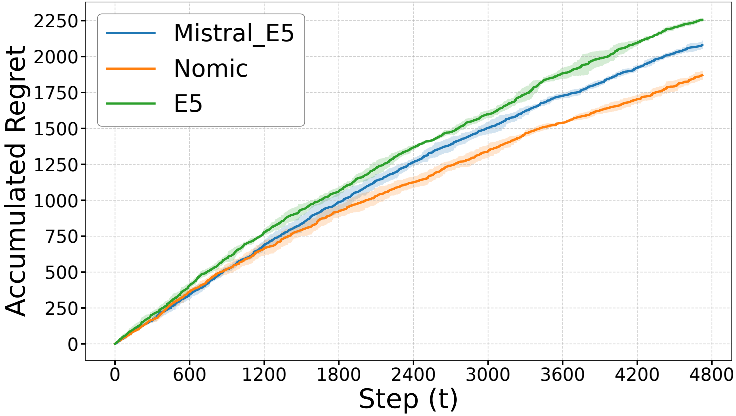
Regret of GHC on Quora Question Groups



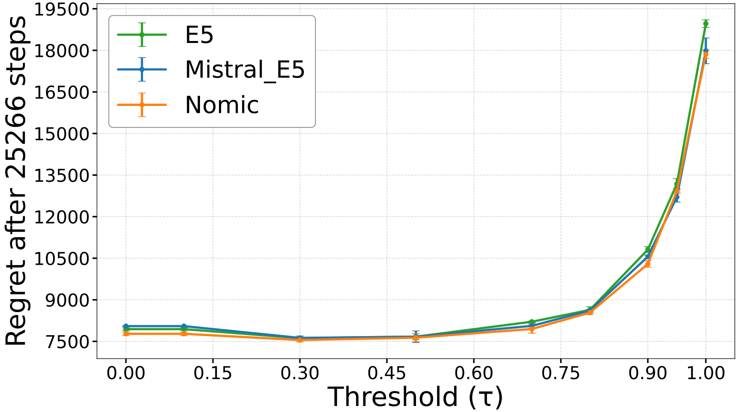
Regret of GHC on ComQA



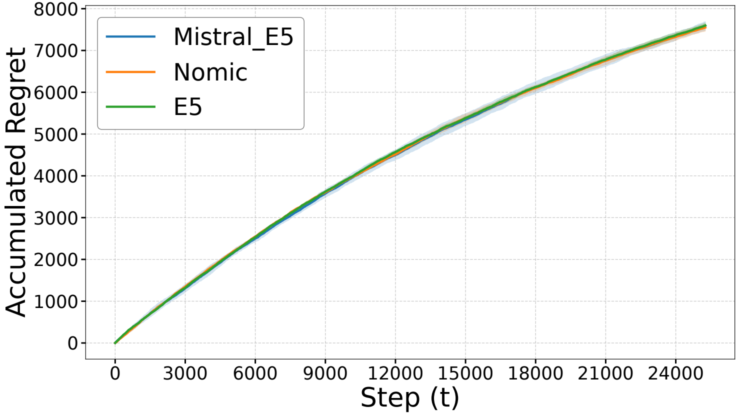
Regret of GHC on ComQA



Regret of GHC on CQADupStack



Regret of GHC on CQADupStack



Conclusion

- A first(?) online optimization problem to minimize human intervention in LLM-based systems
- Regret analysis made possible through stochastic geometry arguments
- In high dimension: several regimes call for different algorithms
- More problems in the context of fine-tuning or adapting large founda- tion models using human feedback?