2026-1 세미나

Towards Balancing Knowledge Gap & Noise in Agentic-RAG

NLP&AI 백성규

데이터의 불확실성

Hybrid Uncertainty Quantification for Selective Text Classification in Ambiguous Tasks (Vazhentsev et al, 2023, ACL)

- **Aleatoric Uncertainty**: 데이터의 Ambiguity and Noise 로 발생

- **Epistemic Uncertainty**: Lack of Knowledge 로 발생

- Total Uncertainty: $\boldsymbol{U_T(x) = U_A(x) + U_E(x)}$

Quantifying Aleatoric Uncertainty & Epistemic Uncertainty

Hybrid Uncertainty Quantification for Selective Text Classification in Ambiguous Tasks (Vazhentsev et al, 2023, ACL)

❖ 두 종류의 Uncertainty 는 다른 방식으로 측정하는 것이 효과적

- **Aleatoric Uncertainty**

1. Entropy: $$-\sum_{c \in C} p(y=c \mid \mathbf{x}) \log p(y=c \mid \mathbf{x})$$

2. Softmax Response: $$1-\max_{c \in C} p(y=c \mid \mathbf{x})$$

- **Epistemic Uncertainty**

Mahalanobis Distance : $$U_{\mathrm{E}}^{\mathrm{MD}}(\mathbf{x})=\min_{c \in C}(h(\mathbf{x})-\mu_c)^T \Sigma^{-1}(h(\mathbf{x})-\mu_c)$$

Agentic RAG

Reasoning RAG via System 1 or System 2: A Survey on Reasoning Agentic Retrieval-Augmented Generation for Industry Challenges (Liang et al, IJCNLP Findings 2025)

- **Standard RAG 의 한계**

- 관련 정보가 여러 곳에 흩어져 있어 정보 추출 뿐 아니라 **일관된 통합 능력** 요구

: 단순 정보 추출은 파편화된, 상반된 답변들을 도출하도록 할 수 있음

(ex: Multi hop Reasoning)

→Agentic RAG: 정보 추출과 **Reasoning, Decision Making** 을 결합

Agentic RAG

Reasoning RAG via System 1 or System 2: A Survey on Reasoning Agentic Retrieval-Augmented Generation for Industry Challenges
(Liang et al, IJCNLP Findings 2025)

- **Predefined Reasoning**

구조화된, 규칙 기반 계획, 파이프라인에 기반하여 Reasoning, 추출, 생성

- **Agentic Reasoning**

모델이 스스로의 Reasoning 과정을 검토하여

언제, 어떻게 추출하고 외부 도구를 활용할지를 결정

UALIGN: Leveraging Uncertainty Estimations for Factuality Alignment on LLMs (Xu et al, ACL 2025)

- Pre-training 과정에서 학습한 지식의 지식 경계는 모호한 경우가 많음.
- Alignment 과정에서 사실적 정확성과 지식 경계를 조화시키지 못하기 때문

→ 지식 경계를 명시적으로 답변 생성에 활용하여 아는 지식을 자신 있게 표현하고 모르는 지식을 강하게 거부하도록 훈련

데이터셋 구축

- 답변을 K 번 Sampling (T=0.2 : 정확성, 다양성의 조화)
- **Known:** Sampled answer 중 최소 한 개라도 정답
- **Unknown:** Sampled answers 이 모두 오답 → 정답을 Sorry, I don’t know로 변경

Uncertainty Measures : 지식 경계를 정량화

1) Accuracy based Confidence : 정답을 맞춘 비율 (확신 정도)

$$c_i = \mathrm{Conf}(\boldsymbol{x}_i) = \frac{1}{K}\sum_{k=1}^{K} \mathbb{1}\left(\tilde{\boldsymbol{y}}_i = \boldsymbol{y}_i{}^{(k)}\right)$$

2) Semantic Entropy: 의미적으로 유사한 문장들 (Cluster) 의 Entropy (퍼진 정도)

$$p(s|\boldsymbol{x}_i) = \frac{1}{K}\sum_{k=1}^{K} \mathbb{1}\left[\boldsymbol{y}_i{}^{(k)} \in s\right]$$

$$e_i = \mathrm{SE}(\boldsymbol{x}_i) = -\sum_{s} p(s|\boldsymbol{x}_i)\log p(s|\boldsymbol{x}_i)$$

UAlign Training

1) Uncertainty Estimation Models (SFT)

Question 에 대한 Accuracy based Confidence, Semantic Entropy 예측하도록 학습

2) Reward Model (SFT)

질문, 생성 답변, confidence, entropy 를 함께 입력으로 받아 정답 여부 $z \in \{0,1\}$ 를 예측 (UEM 활용)

3) PPO

confidence, entropy 를 입력으로 받았을 때 답변 생성을 훈련 (Reward Model 활용)

UAlign Training Objectives

1) Uncertainty Estimation Models (SFT)

Minimize Cross Entropy Losses

$$\arg\min_{\tau} \mathcal{L}_{\tau}, \arg\min_{\mu} \mathcal{L}_{\mu},$$
$$\mathcal{L}_{\tau} = -\mathbb{E}_{(\boldsymbol{x}_i, c_i)\sim\mathcal{D}} \left[\log p_{\tau}(c_i|\boldsymbol{x}_i)\right]$$
$$\mathcal{L}_{\mu} = -\mathbb{E}_{(\boldsymbol{x}_i, e_i)\sim\mathcal{D}} \left[\log p_{\mu}(e_i|\boldsymbol{x}_i)\right]$$

2) Reward Model (SFT)

Minimize binary cross-entropy loss

$$\arg\min_{\theta} \mathcal{L}_{\theta}, \mathcal{L}_{\theta} = -\mathbb{E}_{(\boldsymbol{x}_i, \boldsymbol{y}_i{}^{(k)}, z_i{}^{(k)}, c_i, e_i)\sim\mathcal{D}}[\mathcal{L}_{\theta}{}^{(i)}]$$
$$\mathcal{L}_{\theta}{}^{(i)} = -z_i{}^{(k)} \log p_{\theta}(z_i{}^{(k)}|\boldsymbol{x}_i, c_i, e_i, \boldsymbol{y}_i{}^{(k)})$$
$$-(1 - z_i{}^{(k)}) \log(1 - p_{\theta}(z_i{}^{(k)}|\boldsymbol{x}_i, c_i, e_i, \boldsymbol{y}_i{}^{(k)})) \quad (6)$$

3) PPO

maximize reward function

$$\arg\max_{\pi_\theta} \mathbb{E}_{\boldsymbol{x}\sim\mathcal{D}, c\sim\tau(\boldsymbol{x}), e\sim\mu(\boldsymbol{x}), \boldsymbol{y}\sim\pi_\theta(\boldsymbol{x}, c, e)} [r]$$
$$r = \underbrace{\theta(\boldsymbol{x}, \boldsymbol{y}, c, e)}_{r_1} - \beta \underbrace{\mathrm{KL}[\pi_\theta(\boldsymbol{x}, c, e)||\pi_o(\boldsymbol{x})]}_{r_2}$$

Experimental Results

Method	TVQA (ID) Prec. ↑	TVQA (ID) Truth. ↑	SciQ (ID) Prec. ↑	SciQ (ID) Truth. ↑	NQ-Open (ID) Prec. ↑	NQ-Open (ID) Truth. ↑	Avg. (ID) Prec. ↑	Avg. (ID) Truth. ↑	LSQA (OOD) Prec. ↑	LSQA (OOD) Truth. ↑
Llama-3-8B										
ICL	76.15	56.55	70.43	44.30	50.28	20.11	65.62	40.32	77.35	52.98
ICL-IDK	69.17	54.10	68.36	43.00	45.43	20.72	60.98	39.27	66.67	50.24
ICL-CoT	66.68	53.37	72.34	45.90	**57.34**	23.60	65.45	40.95	73.96	49.37
SFT	70.80	52.57	72.18	45.40	41.41	16.57	61.46	38.18	68.09	46.63
R-Tuning	72.93	55.44	71.38	44.90	47.81	18.12	64.04	39.48	71.54	52.15
RL-PPO	76.32	55.19	75.70	45.80	54.07	24.19	68.03	41.72	72.18	48.43
RL-DPO	72.08	53.96	71.23	44.20	49.65	19.18	64.32	39.11	71.09	48.88
RLKF	77.12	56.07	72.36	44.90	54.86	22.15	68.11	41.04	74.95	52.46
ITI	71.09	53.97	72.35	43.80	43.20	17.13	62.21	38.30	68.52	46.99
UALIGN	**79.14**	57.04	**76.44**	**48.00**	56.60	26.09	**70.72**	43.71	**79.56**	**55.88**
(w/o Conf.)	74.13	54.45	74.05	45.00	54.19	23.60	67.45	41.01	74.25	52.06
(w/o Entro.)	78.43	**57.69**	75.39	47.50	56.68	**27.56**	70.16	**44.25**	76.14	54.43
Mistral-7B										
ICL	77.92	55.14	68.62	42.20	52.09	17.95	66.21	38.43	74.09	47.71
ICL-IDK	72.59	51.37	63.74	39.20	51.13	17.67	62.48	36.20	72.27	47.32
ICL-CoT	76.73	54.78	71.87	44.20	**54.47**	18.22	67.69	39.06	**79.24**	52.59
SFT	74.57	54.77	65.85	42.50	50.82	14.42	63.74	37.08	68.33	44.00
R-Tuning	67.70	52.25	64.44	40.10	46.33	15.52	59.49	36.29	64.67	44.05
RL-PPO	79.23	55.08	71.35	44.10	53.76	19.19	68.11	39.45	74.49	49.67
RL-DPO	72.20	52.98	66.44	41.80	50.95	16.42	63.19	37.06	67.82	43.77
RLKF	80.43	56.92	70.66	43.90	52.09	18.24	67.72	39.68	74.19	49.23
ITI	74.65	55.16	66.90	44.90	51.12	16.68	64.22	38.91	67.73	46.20
UALIGN	**82.10**	**59.05**	**73.21**	**46.70**	54.17	**19.64**	**70.82**	**41.79**	76.29	**52.89**
(w/o Conf.)	76.44	55.13	69.84	43.50	50.30	17.88	65.52	38.83	73.15	47.06
(w/o Entro.)	80.18	57.64	72.90	45.60	52.21	18.44	68.43	40.56	75.34	50.15

- Evaluation Metrics

Precision: 정답을 맞춘 비율

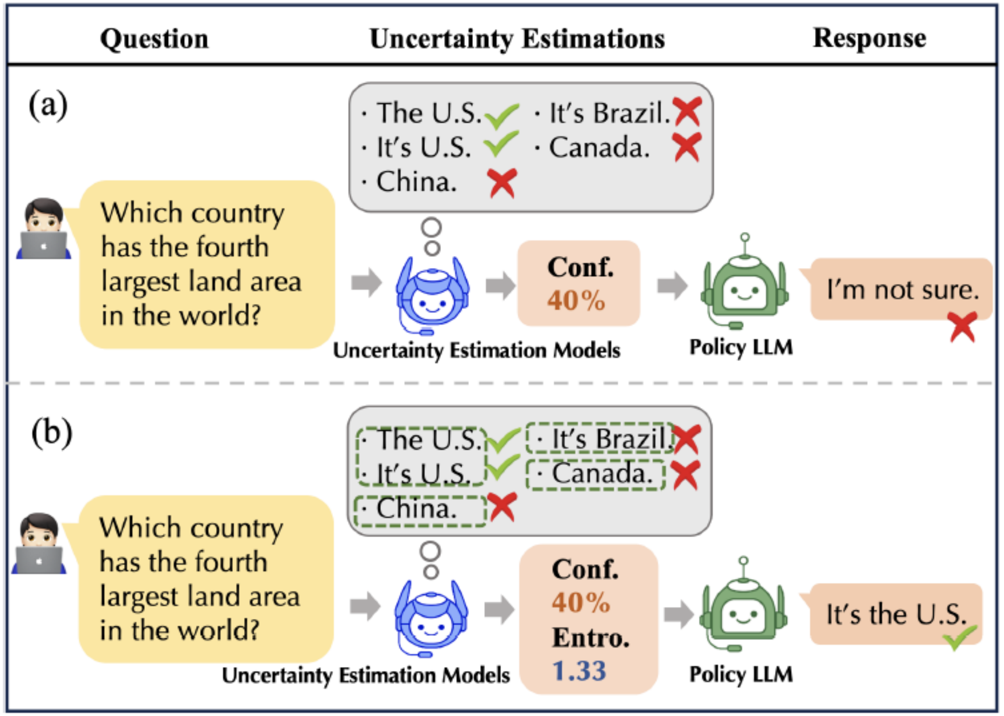
Truthfulness:

Known 질문을 답변 +

Unknown 질문을 거절 비율

Accuracy with Conf. and Entro.

Conf.	Entro.	ID TVQA	ID SciQ	OOD NQ-Open	OOD LSQA
Llama-3-8B					
✗	✗	82.31	79.00	67.45	70.12
✓	✗	85.41	84.30	70.37	**75.09**
✗	✓	82.05	77.90	67.85	70.40
✓	✓	**86.73**	**86.40**	**72.00**	74.59
Mistral-7B					
✗	✗	84.53	77.30	65.24	68.31
✓	✗	86.80	79.50	72.10	72.95
✗	✓	85.24	74.60	66.64	71.22
✓	✓	**88.06**	**79.80**	**75.14**	**73.61**

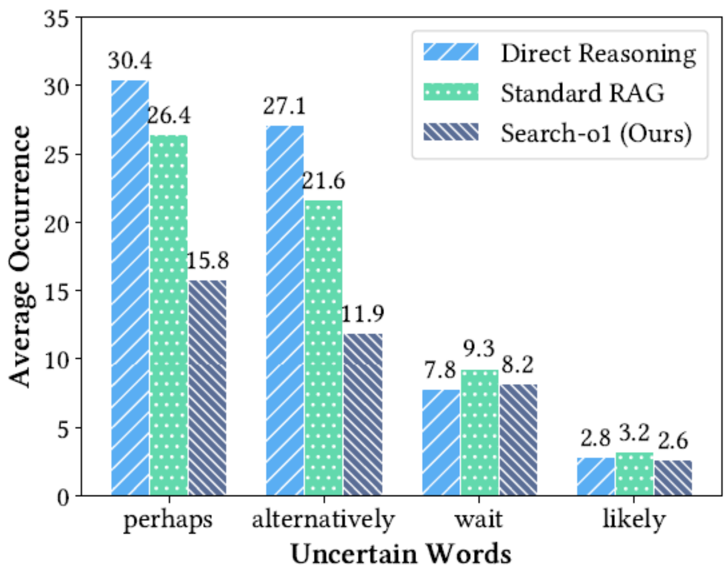


둘 다 썼을 때 Reward Model 의 정확도, 답변의 정확도가 향상

Search-o1: Agentic Search-Enhanced Large Reasoning Models (Li et al, EMNLP 2025)

- LRM 은 Long reasoning chain 으로 복잡한 과제를 단계별로 쪼개서 해결

그러나 **Knowledge Gap** 으로 인해 Reasoning Chain 에서 오류 발생 가능



- 해결: **Reasoning Chain에 RAG 메커니즘을 통합**

→ Knowledge Gap 완화

Search-O1 (Li et al, EMNLP 2025)

1. Agentic RAG Mechanism

Reasoning 도중 “Reasoning 지속” or “외부 지식 추출” 을 스스로 결정

외부 지식 추출: <|begin_search_query|> query <|end_search_query|>

$$P(q_{\text{search}}^{(i)} \mid I, q, \mathcal{R}^{(i-1)}) =$$

$$\prod_{t=1}^{T_q^{(i)}} P\left(q_{\text{search},t}^{(i)} \mid q_{\text{search},<t}^{(i)}, I, q, \mathcal{R}^{(i-1)}\right)$$

I: Instruction, q: question, R: reasoning chain,
a: answer, q search: search query

$$P(\mathcal{R}, a \mid I, q, \mathcal{D}) = \underbrace{\prod_{t=1}^{T_r} P(\mathcal{R}_t \mid \mathcal{R}_{<t}, I, q, \mathcal{D}_{<t})}_{\text{Reasoning Process}} \cdot \underbrace{\prod_{t=1}^{T_a} P(a_t \mid a_{<t}, \mathcal{R}, I, q)}_{\text{Answer Generation}},$$

Search-O1 (Li et al, EMNLP 2025)

2. Knowledge refinement

RAG 로 추출된 정보 중 Reasoning Chain 에 직접적으로 도움이 되는 관련 있고 간결한 정보를 추출 (LRM) 하여 Reasoning Chain 에 통합

$$P\left(r_{\text{docs}}^{(i)} \mid R^{(<i)}, q_{\text{search}}^{(i)}, D^{(i)}\right) = \prod_{t=1}^{T_d^{(i)}} P\left(r_{\text{docs},t}^{(i)} \mid r_{\text{docs},<t}^{(i)}, R^{(<i)}, q_{\text{search}}^{(i)}, D^{(i)}\right)$$

$$P\left(r_{\text{final}}^{(i)} \mid r_{\text{docs}}^{(i)}, R^{(<i)}, q_{\text{search}}^{(i)}\right) = \prod_{t=1}^{T_r^{(i)}} P\left(r_{\text{final},t}^{(i)} \mid r_{\text{final},<t}^{(i)}, r_{\text{docs}}^{(i)}, R^{(<i)}, q_{\text{search}}^{(i)}\right)$$

Experimental Results

Method	GPQA (PhD-Level Science QA)				Math Benchmarks			LiveCodeBench			
	Physics	Chemistry	Biology	Overall	MATH500	AMC23	AIME24	Easy	Medium	Hard	Overall
Direct Reasoning (w/o Retrieval)											
Qwen2.5-32B	57.0	33.3	52.6	45.5	75.8	57.5	23.3	42.3	18.9	14.3	22.3
Qwen2.5-Coder-32B	37.2	25.8	57.9	33.8	71.2	67.5	20.0	61.5	16.2	12.2	25.0
QwQ-32B	75.6	39.8	68.4	58.1	83.2	82.5	53.3	61.5	29.7	**20.4**	**33.0**
Qwen2.5-72B	57.0	37.6	68.4	49.0	79.4	67.5	20.0	53.8	29.7	24.5	33.0
Llama3.3-70B	54.7	31.2	52.6	43.4	70.8	47.5	36.7	57.7	32.4	24.5	34.8
DeepSeek-R1-Lite†	-	-	-	58.5	91.6	-	52.5	-	-	-	51.6
GPT-4o†	59.5	40.2	61.6	50.6	60.3	-	9.3	-	-	-	33.4
o1-preview†	89.4	59.9	65.9	73.3	85.5	-	44.6	-	-	-	53.6
Retrieval-augmented Reasoning											
RAG-Qwen2.5-32B	57.0	37.6	52.6	47.5	82.6	72.5	30.0	61.5	24.3	8.2	25.9
RAG-QwQ-32B	76.7	38.7	73.7	58.6	84.8	82.5	50.0	57.7	16.2	12.2	24.1
RAgent-Qwen2.5-32B	58.1	33.3	63.2	47.0	74.8	65.0	20.0	57.7	24.3	6.1	24.1
RAgent-QwQ-32B	76.7	46.2	68.4	61.6	85.0	**85.0**	**56.7**	**65.4**	18.9	12.2	26.8
Retrieval-augmented Reasoning with Reason-in-Documents											
Search-o1 (Ours)	**77.9**	**47.3**	**78.9**	**63.6**	**86.4**	**85.0**	**56.7**	57.7	**32.4**	**20.4**	**33.0**

Challenging Reasoning Tasks

Method	Single-hop QA				Multi-hop QA							
	NQ		TriviaQA		HotpotQA		2WIKI		MuSiQue		Bamboogle	
	EM	F1	EM	F1	EM	F1	EM	F1	EM	F1	EM	F1
Direct Reasoning (w/o Retrieval)												
Qwen2.5-32B	22.8	33.9	52.0	60.3	25.4	34.7	29.8	36.3	8.4	18.0	49.6	63.2
QwQ-32B	23.0	33.1	53.8	60.7	25.4	33.3	34.4	40.9	9.0	18.9	38.4	53.7
Qwen2.5-72B	27.6	41.2	56.8	65.8	29.2	38.8	34.4	42.7	11.4	20.4	47.2	61.7
Llama3.3-70B	36.0	48.7	68.8	76.8	37.8	49.1	46.0	54.2	14.8	23.6	54.4	67.8
Retrieval-augmented Reasoning												
RAG-Qwen2.5-32B	33.4	49.3	**65.8**	**79.2**	38.6	50.4	31.6	40.6	10.4	19.8	52.0	66.0
RAG-QwQ-32B	29.6	44.4	65.6	77.6	34.2	46.4	35.6	46.2	10.6	20.2	55.2	67.4
RAgent-Qwen2.5-32B	32.4	47.8	63.0	72.6	44.6	56.8	55.4	69.7	13.0	25.4	54.4	66.4
RAgent-QwQ-32B	33.6	48.4	62.0	74.0	43.0	55.2	**58.4**	71.2	13.6	25.5	52.0	64.7
Retrieval-augmented Reasoning with Reason-in-Documents												
Search-o1 (Ours)	**34.0**	**49.7**	63.4	74.1	**45.2**	**57.3**	58.0	**71.4**	**16.6**	**28.2**	**56.0**	**67.8**

Open Domain QA Tasks

Towards Balancing Knowledge Gap & Noise in Agentic-RAG

- <u>기존 Agentic RAG 의 문제점</u>

1) Context 가 충분하지 않은 상태에서 검색 없이 스스로 답하려고 하다가 틀리는 경우가 많음.

2) 기존의 Agentic RAG 는 Aleatoric Uncertainty (Ambiguity, Noise) 를 다루지 않음.

3) 이전 Step 의 추론이 잘못되면 이후 Step 추론에 영향을 미침.

→ Step 별로 State 와 Action 의 Goal을 외부적으로 제공

Uncertainty State → Action

A: Epistemic Uncertainty

E0: No relevant evidence → **Search**

E1: Partial evidence → **Search**

E2: Sufficient evidence

B: Aleatoric Uncertainty (추론 포함)

A0: Clean

A1: Ambiguous: Question 의 대상과 다른 유사한 대상의 Context

→ **질문의 Entity에 기반한 Context를 추가 검색, 대체**

A2: Inconsistent: Question 과 대상이 같지만 답이 갈리는 경우

→ **질문의 맥락에 더 부합하는 Context 하나만 선택**

Uncertainty Classifier : SFT (9가지 Context)

A: Epistemic Classifier

E0: No relevant evidence (FaithEval, ICLR 2025)

E1: Partial evidence (FaithEval, ICLR 2025)

E2: Sufficient evidence

B: Aleatoric Classifier

A0: Clean

A1: Ambiguous

A2: Inconsistent (FaithEval, ICLR 2025)

→ LLM 으로 Training Data의 Context 변형 후 학습

Ambiguous, Inconsistent 예시 (Search-o1 결과 중)

A1. Ambiguous:

What was the Roud Folk Song Index of the nursery rhyme inspiring **What Are Little Girls Made Of**?

the nursery rhyme ₩"**What Are Little Boys Made Of?**₩" is listed under the title ₩"What Are Little Boys Made Of₩" and is assigned Roud number 10245.

A2. Inconsistent:

The VMAQT-1 logo is a female spirit in Irish mythology who heralds the death of what?

In traditional mythology, they usually foretell the death of a **specific person or members of a family**.

Alternatively, perhaps the Banshee in this context symbolizes **the death of the enemy or the end of threats**, given the military nature of the squadron.

Thank you