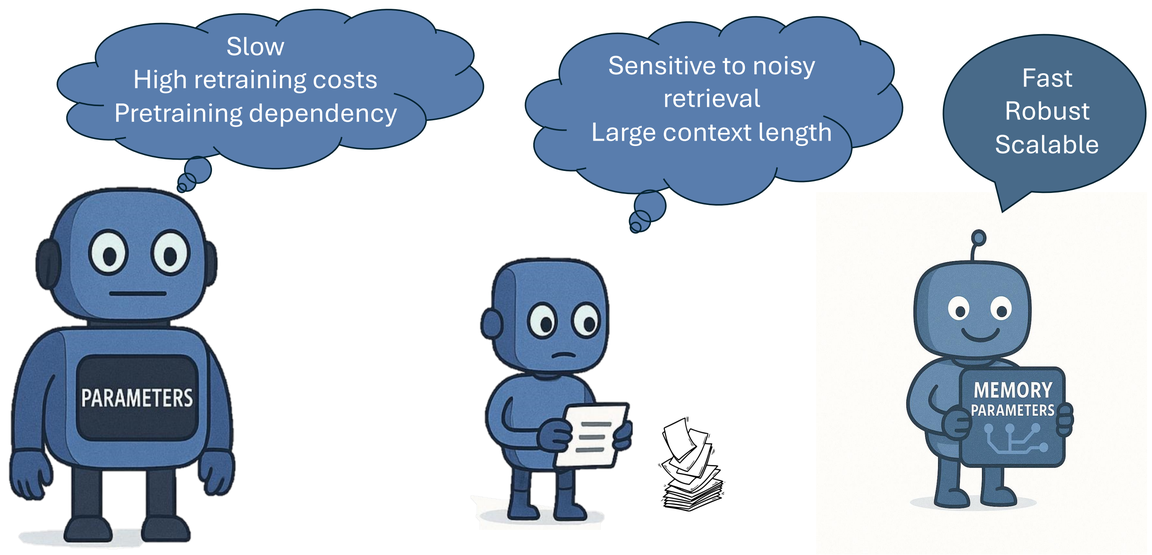
MOGIC: Metadata-infused Oracle Guidance for Improved Extreme Classification

Suchith Chidananda Prabhu[1*], Bhavyajeet Singh[2*], Anshul Mittal[4], Siddarth Asokan[2], Shikhar Mohan[4], Deepak Saini[3], Niket Tandon[2], Manish Gupta[4], Sumeet Agarwal[1], Manik Varma[2]

[1]IIT Delhi, India [2]Microsoft Research, India [3]Microsoft, US [4]Microsoft IDC, India

Motivation

- Queries require additional context for disambiguation.
- Small models lack the parameters to store the world knowledge.
- Larger models are slow and not deployable in real world scenarios.



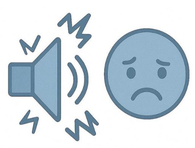
Large Language models

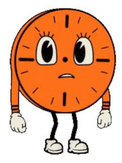
Small models with text augmentation

Small models with Parametric memory

Challenges

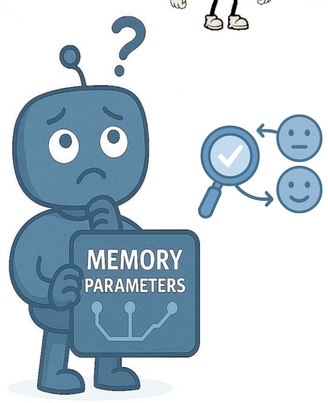
Sensitivity to noise: noisy retrieval from memory can degrade task performance.





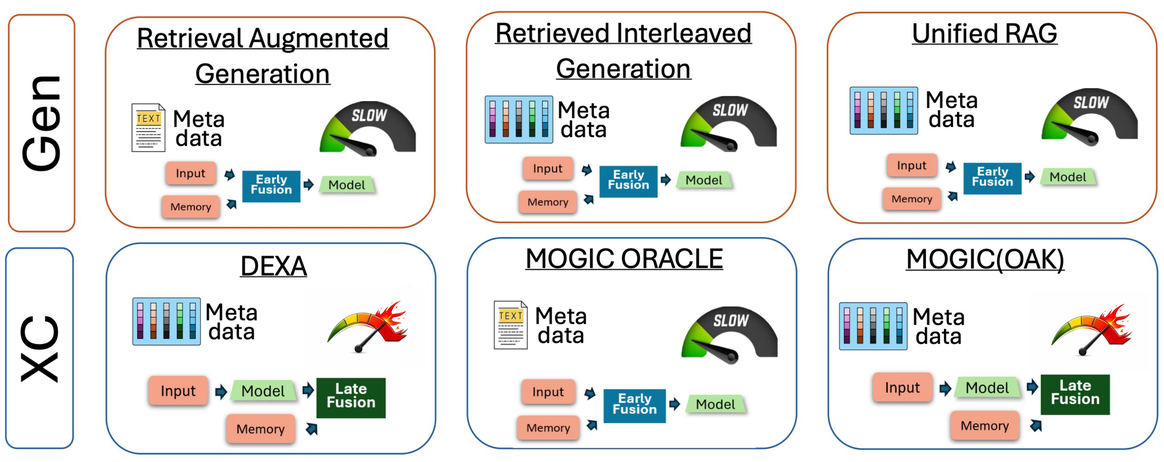
Latency: metadata format and fusion layer can influence Latency.

Fusion method tradeoffs: late-fusion handles noisy metadata better, early-fusion yields higher accuracy with clean inputs.



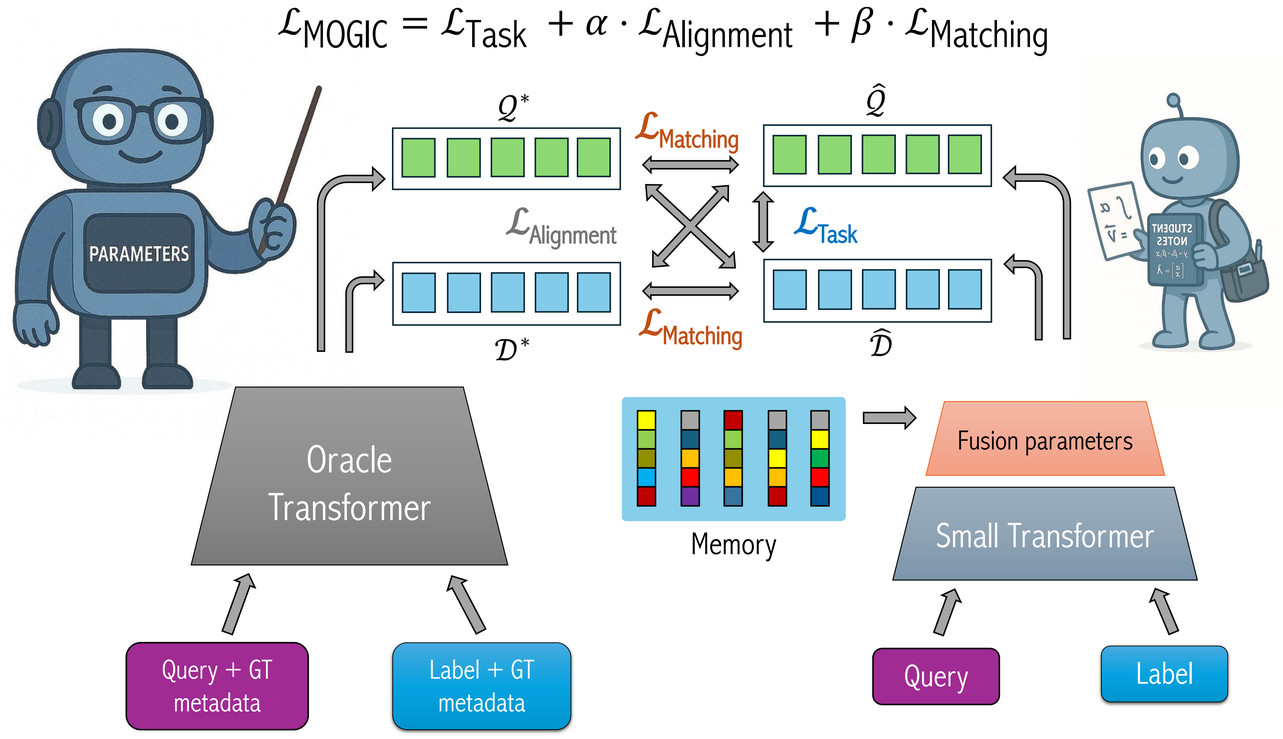
Data Scarcity: labels and memory items have limited training data, hindering accurate prediction.

Memory augmentation: XC & beyond



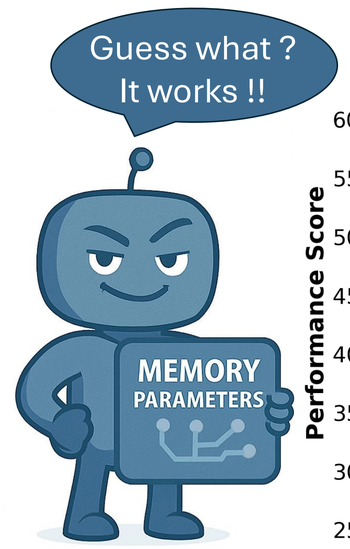
Architecture

$$\mathcal{L}_{\text{MOGIC}} = \mathcal{L}_{\text{Task}} + \alpha \cdot \mathcal{L}_{\text{Alignment}} + \beta \cdot \mathcal{L}_{\text{Matching}}$$

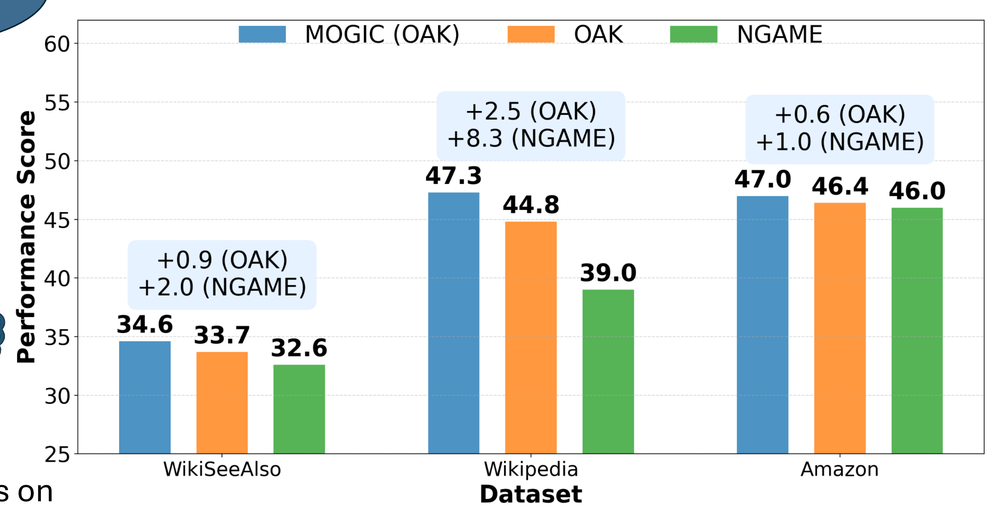


- MOGIC uses parametric memory to inject contextual metadata.
- Oracle models with metadata access guide disciple models.
- Trainable Components: Parametric memory bank, usion parameters, Encoder model

Results



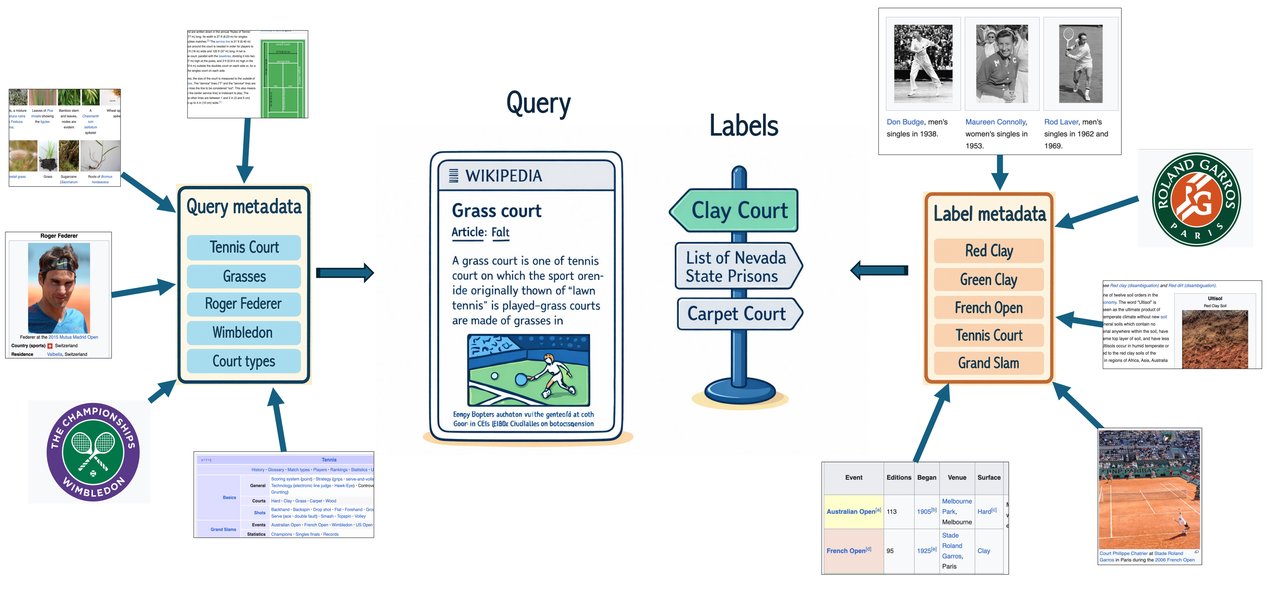
MOGIC vs Baselines on short text datasets



Models	P@1	P@5	N@5	PSP@5
MOGIC(DEXA)	**32.75**	**16.92**	**26.88**	**31.82**
DEXA	31.57	16.14	25.64	29.99
MOGIC(NGAME)	**32.37**	**16.38**	**26.87**	**31.08**
NGAME	30.72	15.42	25.18	28.88

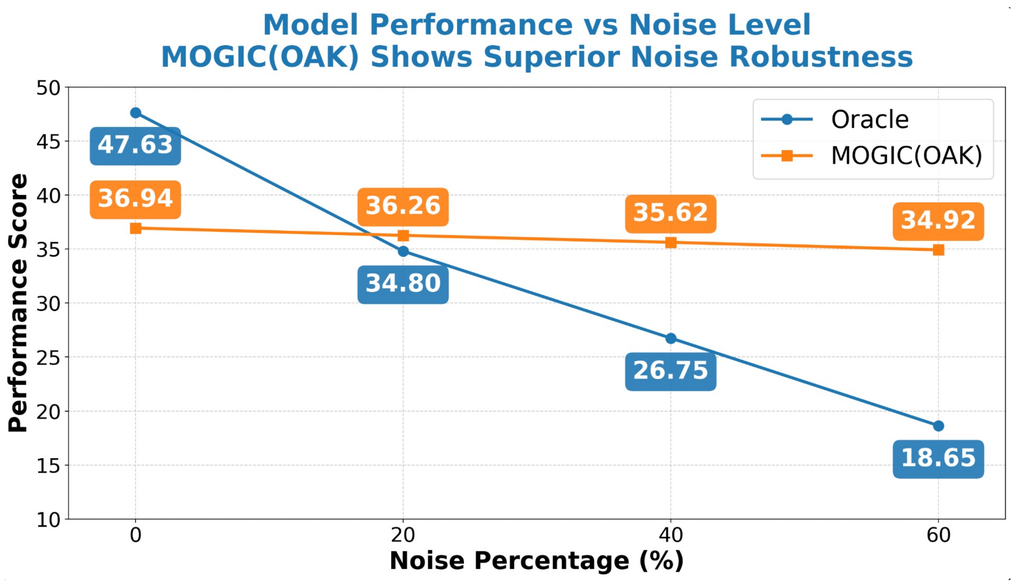
Oracle (#param)	P@1	P@5	N@5	PSP@5
	MOGIC(OAK)			
DistilBERT (65M)	34.62	17.92	**27.44**	**33.18**
LLaMa-2 (7B)	**34.64**	**17.93**	27.28	33.02
Phi-2 (2.7B)	34.34	17.73	27.09	32.07

Leveraging Metadata for Extreme Label Learning



MOGIC with multiple disciples

Various LLMs as oracles (on WikiSeeAlso)



ICML International Conference On Machine Learning