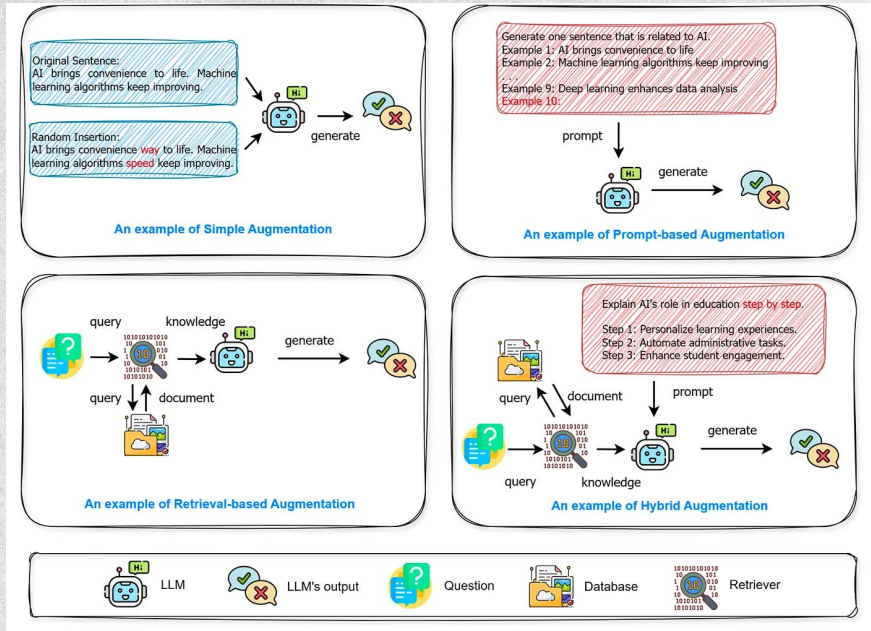
Fixing "Not Enough Data" for Large Language Models

A 2025 Survey Review (Chai et al.)

Mohibkhan Pathan
CMPE 297 - Special Topics
Professor Vijay Eranti

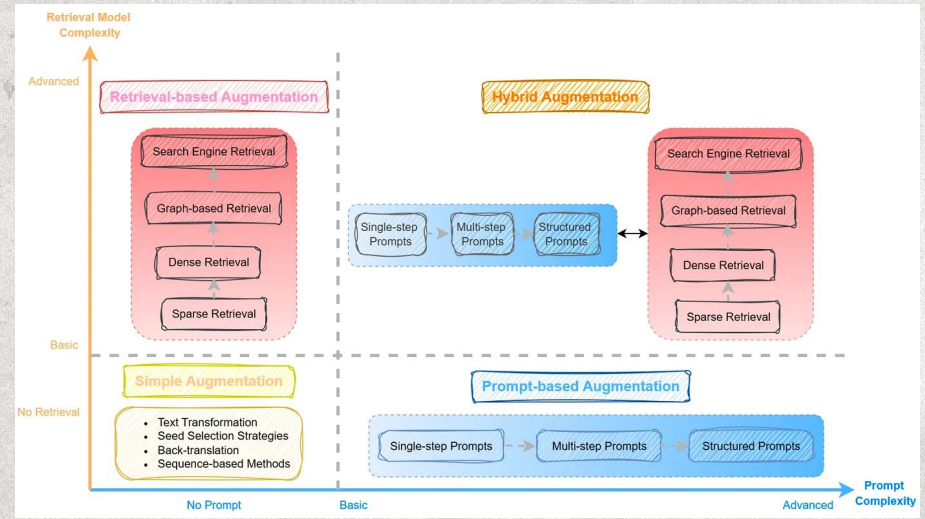
The Four Families of Augmentation

- **Simple:** Edit words (synonyms).
- **Prompt-based:** Ask LLM to generate.
- **Retrieval-based:** Search for facts first.
- **Hybrid:** Mix prompting + search.



Complexity Spectrum

- Simple methods have "No Prompt/Retrieval."
- Hybrid methods are the most complex "Advanced" end of the spectrum.



Design Dials: Granularity

- **Token Level:** Changing small words.
- **Sentence Level:** Generating new sentences (most common).
- **Document Level:** Generating whole docs.

Methods	Token Level	Token-span Level	Sentence Level	Passage Level	Context Level	Document Level
Simple Augmentation						
TransformersDA [10]	✓	✓				
DAGAM [11]	✓					✓
GenAug [12]	✓					
AuGPT [13]			✓			
COCA [14]	✓					✓
Selection-DA [15]			✓			
LAMBADA [16]	✓		✓			
LeCA [17]	✓		✓			
G-DAUGc [18]			✓			
MRC-QA [19]		✓				
Prompt-based Augmentation						
GPT3Mix [20]			✓			
DA-intent [21]			✓			
WANLI [22]			✓			
FlipDA [4]	✓		✓			
AugESC [23]						✓
AugGPT [2]			✓			
Read- Com [24]					✓	
DAIL [25]			✓			
DA-NMT [26]			✓			
EPA [27]			✓			
ZeroShotDataAug [28]			✓			
Dialogue-Convert [29]					✓	
HiPSTG [30]						✓
SUNGEN [31]			✓			
LLM-powered [1]					✓	
LLM-PTM [32]			✓			
Generative-DA [33]					✓	
ICLEF [34]			✓			
LLM-DA [35]		✓	✓			
Synthetic-DA [36]					✓	
LLM–Assisted [3]	✓					✓
LLM2LLM [37]					✓	
PromptMix [38]			✓			
Unnatural-instructions [39]					✓	
GENIUS [5]		✓	✓			
TAPP [40]					✓	
X-GEAR [41]	✓			✓		
InPars [42]			✓			
ConvAug [43]	✓				✓	
Promptagator [44]			✓			
DAPDR [45]			✓			
UDAPDR [46]			✓			
Retrieval-based Augmentation						
AugmentedSBERT [47]			✓			
zicl [48]			✓			
RetGen [7]						✓
Internet-Aug [49]						✓
DialogGen [50]						✓
ChatPLUG [51]						✓
EDGE [52]					✓	
RGQA [53]					✓	
CGRG [54]			✓			
IM-RAG [55]						✓
EAE-RAG [56]						✓
SeeKeR [57]						✓
Efficient-RAG [58]						✓
LAPDOG [59]						✓
Personae-DA [60]						✓
Hybrid Augmentation						
DAICL [61]					✓	
KAPING [62]					✓	
ALCE [63]				✓		
RADA [64]					✓	
UniMS-RAG [65]					✓	
QA-Internet [8]					✓	
ReAct [66]			✓			

Post-Processing

- **Consistency Checks:** Ask model "Are you sure?".
- **Filtering:** Remove bad/unsafe examples.
- **Human Pass:** Manual check for safety.

Evaluation Metrics

- **Auto:** Accuracy, F1, BLEU (Fast).
- **Human:** Coherence, Safety (Slow but necessary)

DA techniques	Evaluation	Metrics	Methods
Simple Augmentation	Automatic Evaluation	Accuracy	DAGAM [11], MRC-QA [19]
		Exact Match score	MRC-QA [19]
		Recall	AuGPT [13], MRC-QA [19]
		F1 Score	AuGPT [13], Selection-DA [15], MRC-QA [19]
		Perplexity	GenAug [12]
		BLEU	GenAug [12], AuGPT [13], LeCA [17]
Prompt-based Augmentation	Automatic Evaluation	Accuracy	WANLI [22],LLM2LLM [37],FlipDA [4],GPT3Mix [20],DA-intent [21] DAIL [25],EPA [27],SUNGEN [31],PromptMix [38]
		ROUGE	Dialogue-Convert [29],Synthetic-DA [36],EPA [27],TAPP [40]
		Exact Match score	Read-Com [24],TAPP [40],Generative-DA [33]
		Recall	DA-intent [21],GENIUS [5]
		F1 Score	Dialogue-Convert [29],LLM-DA [35],X-GEAR [41],Read-Com [24] Generative-DA [33],EPA [27],FlipDA [4]
		Perplexity	GENIUS [5]
		BLEU	DA-NMT [26],EPA [27]
	Human Evaluation	Consistency Coherence Informativeness Safety	AugESC [23]
Retrieval-based Augmentation	Automatic Evaluation	Accuracy	zicl [48]
		ROUGE	LAPDOG [59]
		Exact Match score	IM-RAG[55]
		Recall	DialogGen [50]
		F1 Score	EAE-RAG [56],X-GEAR [41],IM-RAG [55],AugmentedSBERT [47] Internet-Aug [49],DialogGen [50],Efficient-RAG [58],LAPDOG [59]
		Perplexity	Internet-Aug [49],DialogGen [50]
		BLEU	RetGen [7],Efficient-RAG [58],LAPDOG [59]
	Human Evaluation	Consistency	SeeKeR [57]
		Coherence	ChatPLUG [51],RetGen [7]
		Informativeness	ChatPLUG [51],RetGen [7]
		Safety	ChatPLUG [51]
		Hallucination	ChatPLUG [51]
		Knowledgeable	Internet-Aug [49], SeeKeR [57]
		Engaging	Internet-Aug [49]
Hybrid Augmentation	Automatic Evaluation	Accuracy	KAPING [62],ReAct[66],QA-Internet [8],DAICL [61]
		ROUGE	UniMS-RAG[65]
		Exact Match score	RGQA [53],ReAct [66],QA-Internet [8]
		Recall	RGQA [53],ConvAug [43]
		F1 Score	RGQA [53],RADA [64],DAICL [61],UniMS-RAG [65]
		BLEU	CGRG [54],UniMS-RAG [65]
	Human Evaluation	Consistency	SeeKeR [57]
		Coherence	UniMS-RAG [65]
		Fluency	ALCE [63]
		Factually Correct	ALCE [63]

Key Findings

- **Retrieval-based methods** significantly reduce hallucinations compared to pure Prompt-based methods.
- They ground the output in real facts.

Resources

Chai, Y., Xie, H., & Qin, J. S. (2025). *arXiv:2501.18845*
https://arxiv.org/pdf/2501.18845