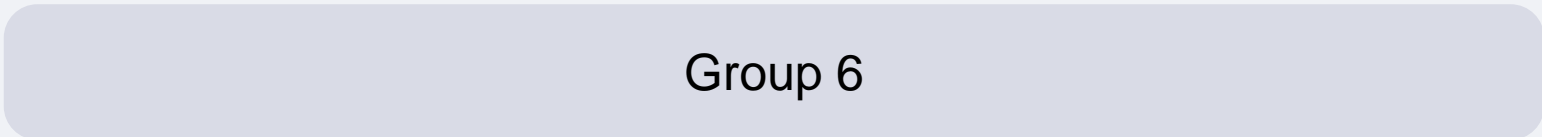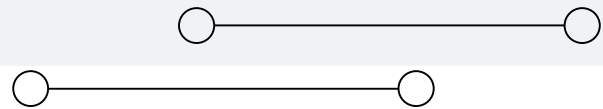# ADL
Final project

Group 6

資工所碩一 R13922022 王凱右
資工系大五 B09501048 方群鈞
資工系大四 B10902011 賴詠晴
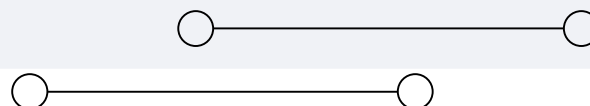
Final result

Medical Diagnosis Classification Task

Task	Accuracy	Leaderboard
Public	0.796	6
Private	0.773	

SQL Generation Task

Task	Accuracy	Leaderboard
Public	0.333	17
Private	0.163	
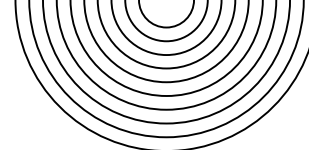
Baseline Performance

**Qwen/
Qwen2.5-7B-Instruct**

Task Type	Classification	SQL Generation
Dataset	DDXPlus	BIRD
Zero-shot	43.48% (~12.5 m)	26.08% (~63.9 m)
Self-StreamICL	51.53% (~49.0 m)	30.64% (~62.0 m)

**meta-llama/
Llama-3.1-8B-Instruct**

Task Type	Classification	SQL Generation
Dataset	DDXPlus	BIRD
Zero-shot	44.95% (~26.5 m)	26.08% (~78.1 m)
Self-StreamICL	49.38% (~57.5 m)	30.77% (~82.5 m)

**prince-canuma/
Ministral-8B-Instruct-2410-HF**

▶ ▶ ▶

Task Type	Classification	SQL Generation
Dataset	DDXPlus	BIRD
Zero-shot	29.48% (~30.6 m)	22.36% (~86.3 m)
Self-StreamICL	39.29% (~58.7 m)	29.27% (~90.5 m)
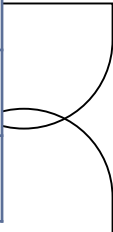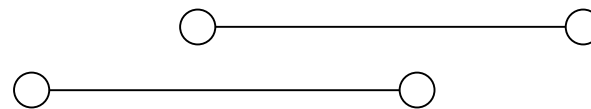
Results

- **Role Prompting**

Model	Gemma-2-9b-it	Gemma-2-9b-it
Accuracy	0.765	0.766
GPU	RTX 3090	A100
4-bits quantization	True	False
Training time	2.5hr	2.3hr
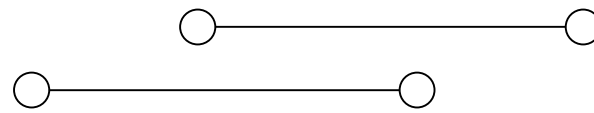
Results

- **Self-Stream Contrastive Learning**

- **Dynamic Self-Stream Contrastive Learning**

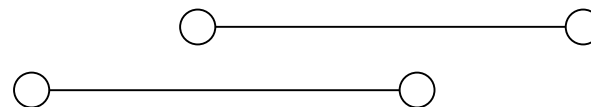- **CoT Role Prompting**

Method	Accuracy
Self-Stream Contrastive Learning	0.702
Dynamic Self-Stream Contrastive Learning	0.764
CoT Role Prompting	0.773
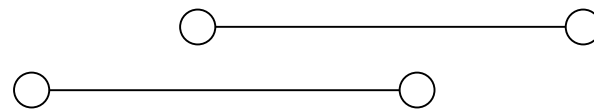
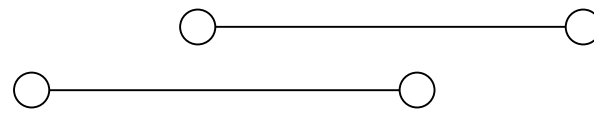# Results

- **Multi-Agentic-Memory StreamICL**
- **MedPrompt**

Multi-Agentic-Memory StreamICL		
Model	**Gemma-2 + Qwen2.5-7B-Instruct**	**Gemma-2 + Llama-3.1-8B-Instruct**
Accuracy	0.724	0.703
MedPrompt		
Model	**Randomized Ordering**	**Randomized Ordering + Voting**
Accuracy	0.7631	0.7423

Results

- **Role Prompting with Format Restrictions**

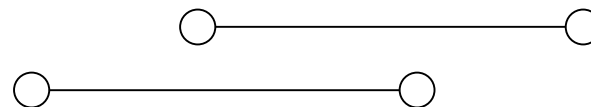| Format	json	yaml	xml
Accuracy	0.762	0.782	0.791

Results

- **Final Medical Diagnosis Classification Result**

Method	Role Prompting with xml Format Restrictions
Accuracy	0.796
Training Time	3.3 hr

Results

- ## Text-to-SQL Generation

Format	NL	yaml	json	xml
Accuracy	0.331	0.315	0.320	0.312

- ## Schema Linking & Self Revision

Method	Self Revision	Schema Linking
Accuracy	0.333	0.320
Inference Time	6.5hr	3.0hr
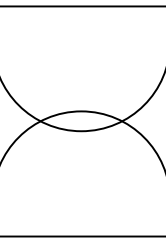
Private Dataset Result

Task Accuracy	Medical Diagnosis Classification	SQL Generation
Public	0.796	0.333
Private	0.773	0.163
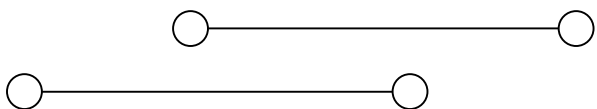
Conclusion

Medical Diagnosis Classification

Medical Diagnosis Classification
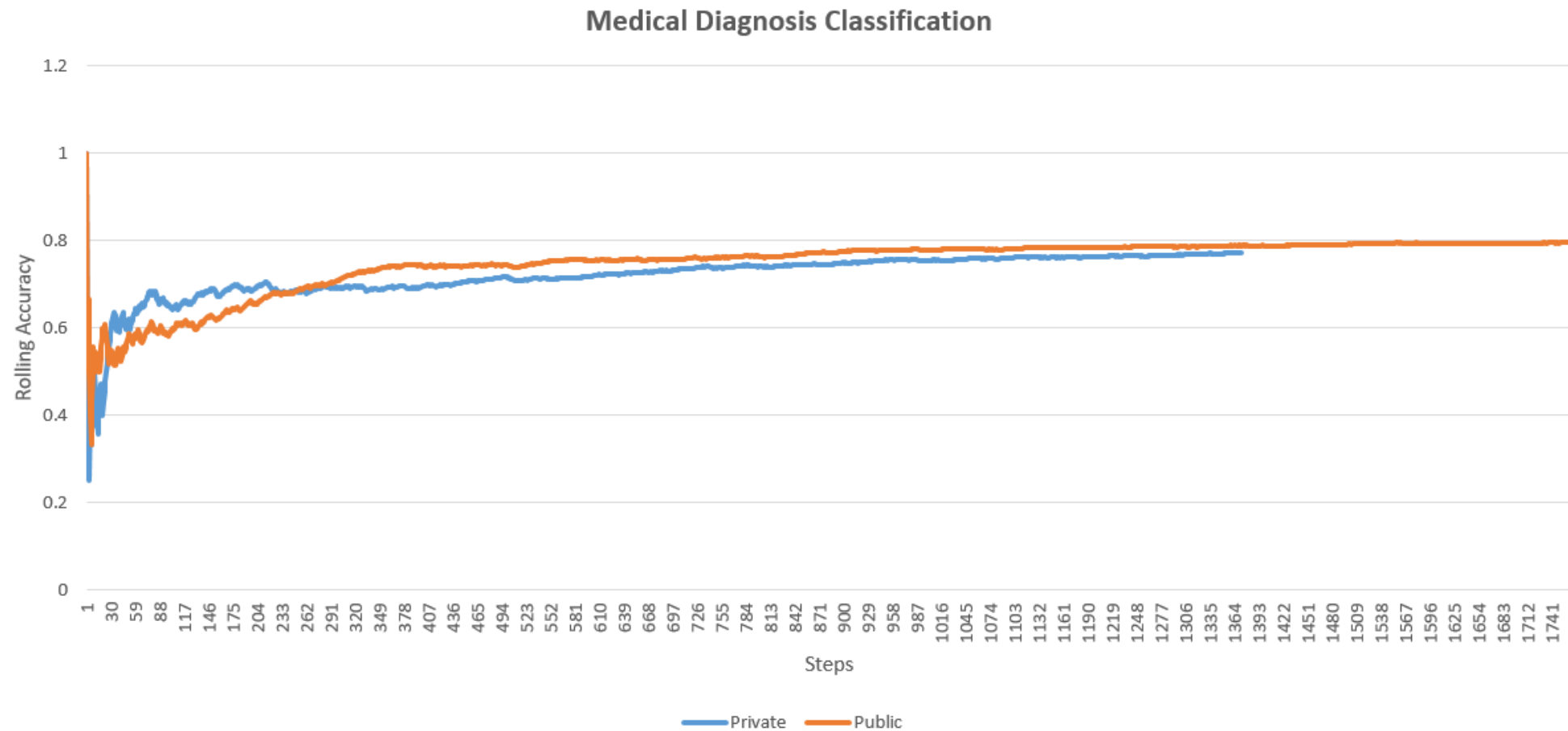
Conclusion

SQL Generation

Conclusion

Thanks