Open LLMs are Necessary for Current Private Adaptations and Outperform their Closed Alternatives

Vincent Hanke, Tom Blanchard, Franziska Boenisch, Iyiola E. Olatunji, Michael Backes, Adam Dziedzic

CISPA Helmholtz Center for Information Security

Summary of Findings

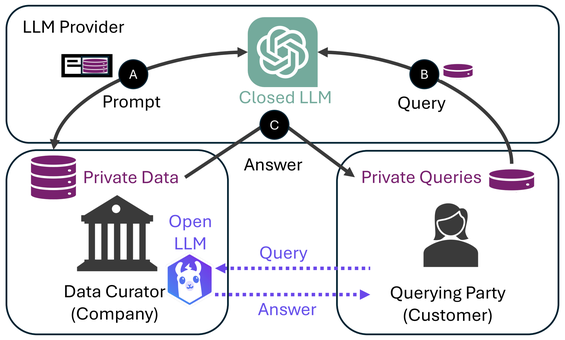
1. **Open LLMs are strictly preferable over closed LLMs** since their adaptations are more private, more performant, and more cost-effective.
2. All private ICL (In-Context Learning) methods **leak query data** (potentially sensitive) to the LLM provider during inference.
3. Methods that **protect private data** from leaking to LLM providers **require a local open LLM**.
4. All private **ICL methods for closed LLMs exhibit lower performance** compared to three private gradient-based adaptation methods (e.g., PEFT - Parameter Efficient Fine Tuning) for local open LLMs.
5. Private adaptation methods for **closed LLMs incur higher monetary** training and query **costs** compared to their open counterparts.

Privacy Risks of Closed vs. Open LLMs

A: The data owner's private data leaks to the LLM provider during the creation of the prompt.

B: The private query of the querying party leaks to the LLM provider.

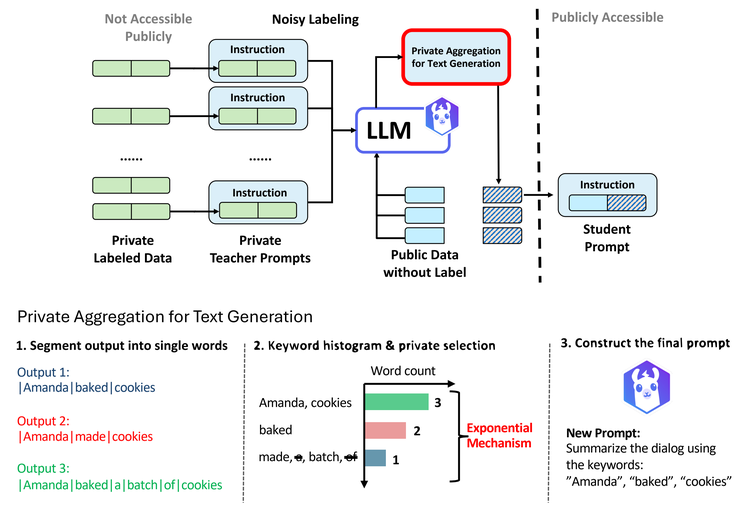
C: Private information from the data owner leaks to the querying party through the returned answers of the prompted LLM.



We advocate that the data owner should privately adapt the open LLM locally and let the querying party interact with this LLM (dashed purple lines), protecting against **A, B, C**.

Method	Ⓐ	Ⓑ	Ⓒ	Open LLM
DP-ICL [Wu et al. ICLR 2024]	✗	✗	✓	Not Needed
PromptPATE [Duan et al. NeurIPS 2023]	✗	✗	✓	Not Needed
DP-FewShotGen(1) [Tang et al. ICLR 2024]	✗	✗	✓	Not Needed
DP-FewShotGen(2) [Tang et al. ICLR 2024]	✓	✗	✓	Needed
DP-OPT [Hong et al. ICLR 2024]	(✓)	✗	✓	Needed
Private Local Adaptation	✓	✓	✓	Needed

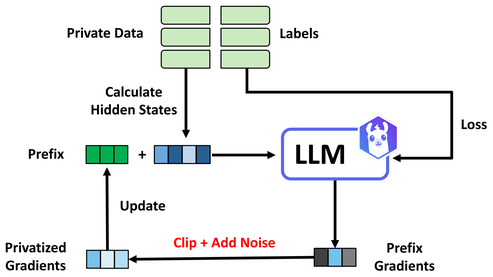
PromptPATEGen



Benefits of **PromptPATEGen**:

- Generated private prompt does **not incur privacy cost per use**
- **Lightweight** prompt

PromptDPSGDGen



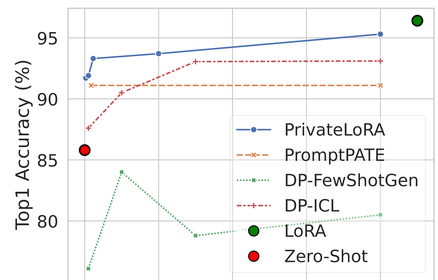
Benefits of **PromptDPSGDGen**:

- Capable of **multitask** batching
- **Lightweight** adaptation method

Find out more!

Text Classification

Method	Model	Acc(Avg)	Cost($)
	Closed LLMs		
DP-OPT	GPT3 Davinci	81.4	8.1
PromptPATE	Claude 2.1	84.5	53.6
DP-FewShotGen	GPT3 Babbage	64.2	2.0
DP-ICL	GPT4 Turbo	68.2	138.0
	Open LLMs		
DP-FullFineTune	RoBERTa Large	89.4	6.15
PrivateLoRA	Vicuna 7B	**90.3**	14.6
PrivateLoRA	Llama3-8B(Instruct)	<u>90.2</u>	28.4
PrivateLoRA	Pythia 160M	78.6	2.1



Text Generation (Dialog Summarization)

Method	Model	Rouge-1	Cost($)
	Closed LLMs		
DP-ICL	GPT3 Davinci	41.2	665.91
DP-ICL	GPT3.5 Turbo	42.6	449.16
DP-ICL	GPT4 Turbo	41.8	3419.42
	Open LLMs		
PromptPateGen	Vicuna 7B	41.3	6.03
PromptPateGen	OpenLlaMA 13B	43.4	19.43
PromptDPSGDGen	Bart-Large	46.4	2.13
PrivateLoRA	Bart-Large	<u>49.1</u>	3.59
PrivateLoRA	Mixtral-8x7B	**52.98**	67.95