Stop wasting energy

Performance vs Power consumption in classification

Andreas Bexell (LTH/Ericsson) & Lo Gullstrand Heander (LTH)
Supervised by Emma Söderberg, Per Runeson (LTH) & Sigrid Eldh (Ericsson/MDH)
Funded by WASP

Natural
Language
Processing

Classification

Commit
Message

Neural
Network

Performance

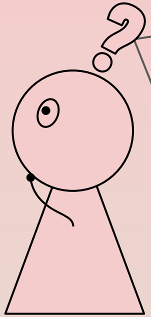
Large
Language
Models

Energy

How does LLM size affect performance?

… in classification of commit messages?

Very little…?



Can we go smaller?

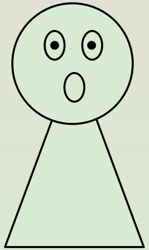
… with “traditional” NLP techniques?

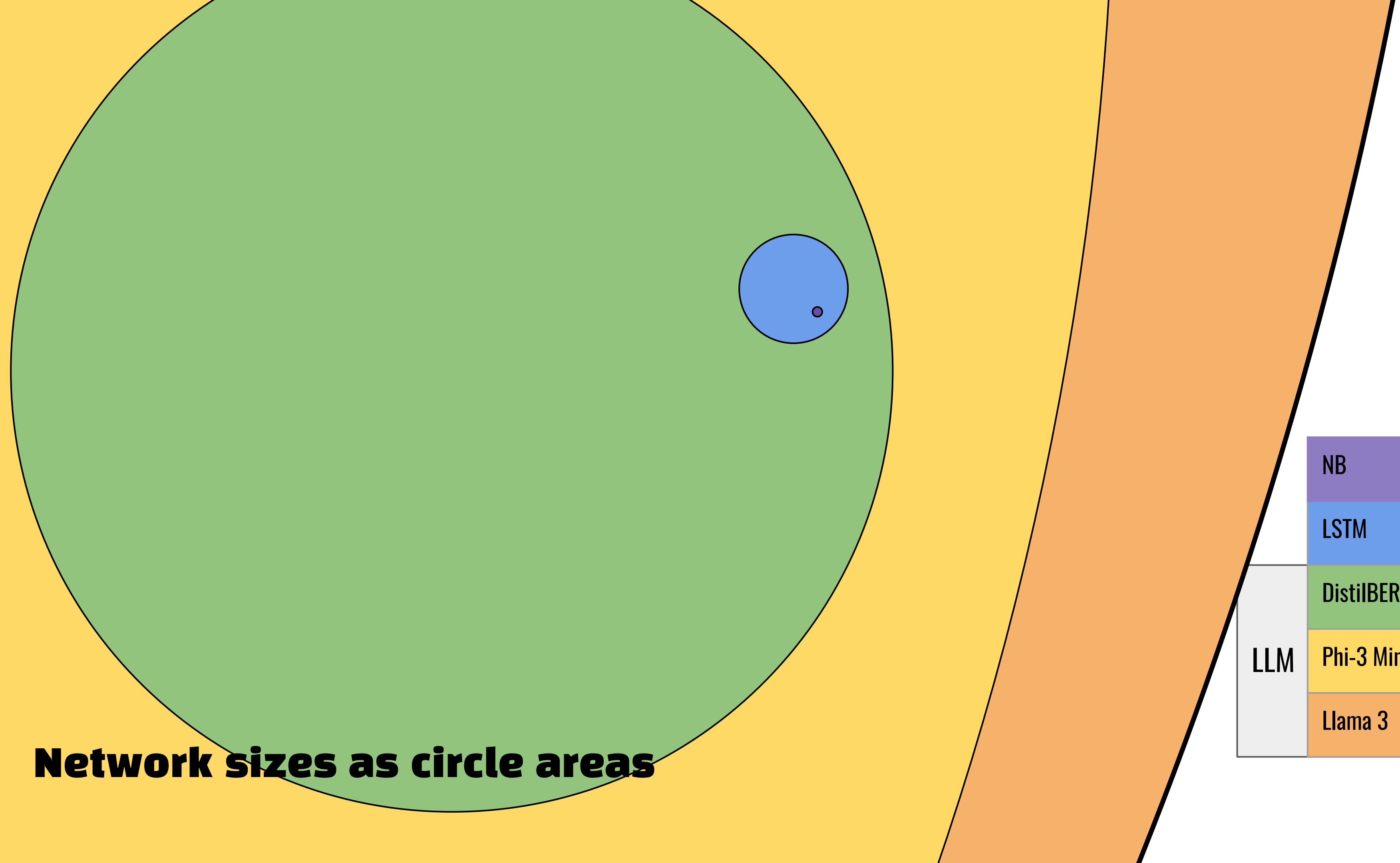
Apparently we can…

What about generalization?

Wait – what about energy?

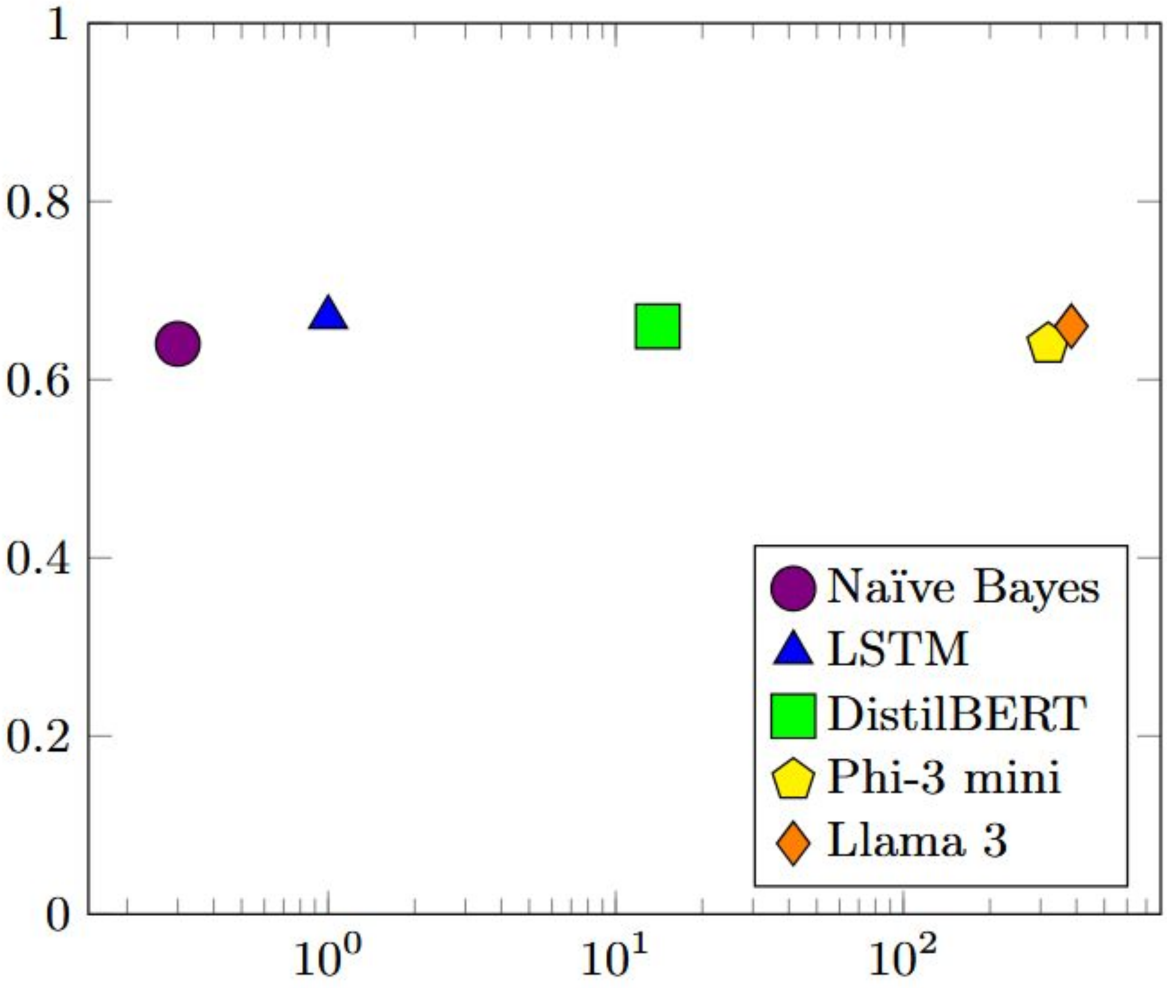
Oh…





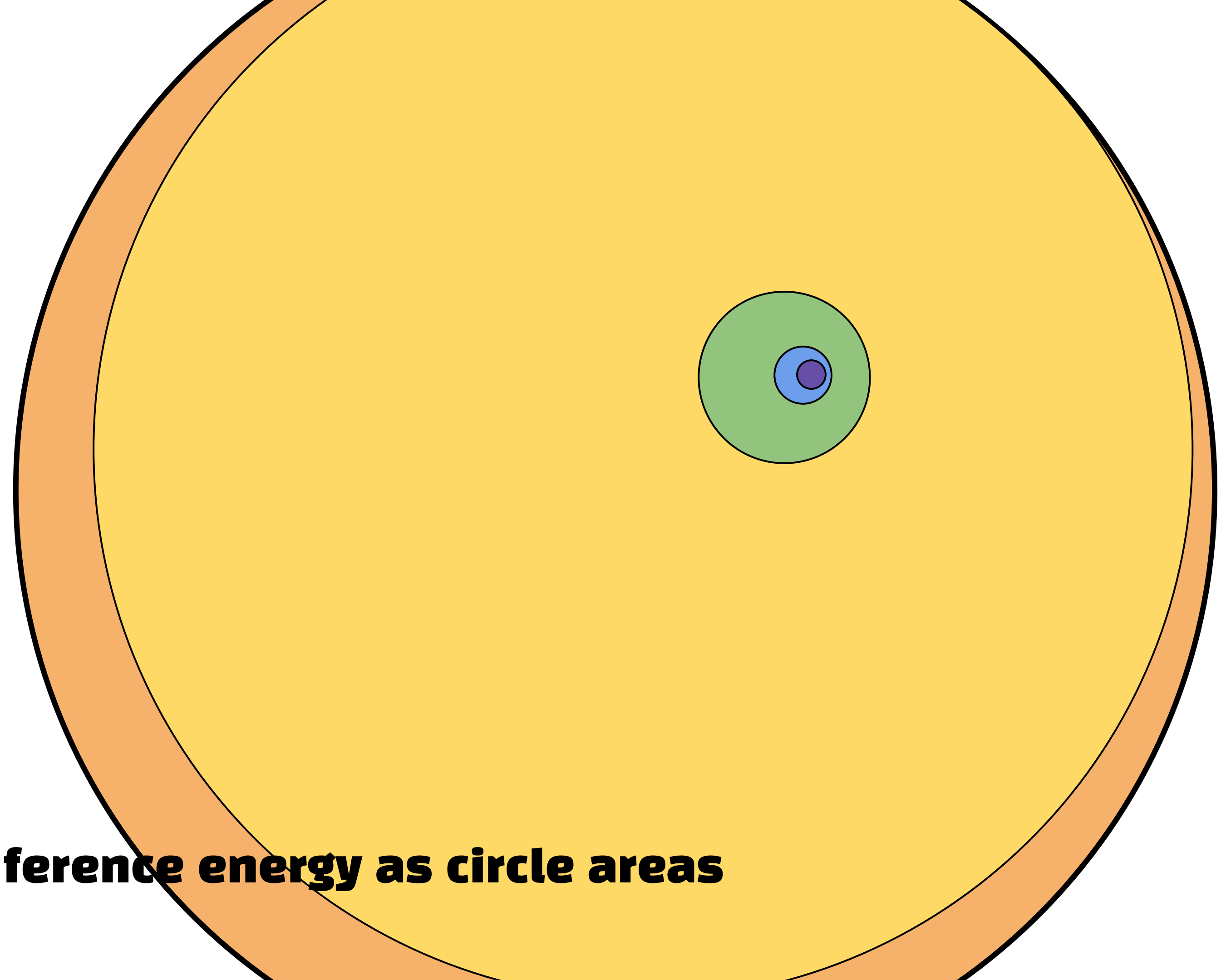
Network sizes as circle areas

	NB	8.0e3
	LSTM	1.1e6
LLM	DistilBERT	6.7e7
	Phi-3 Mini	3.8e9
	Llama 3	8.0e9



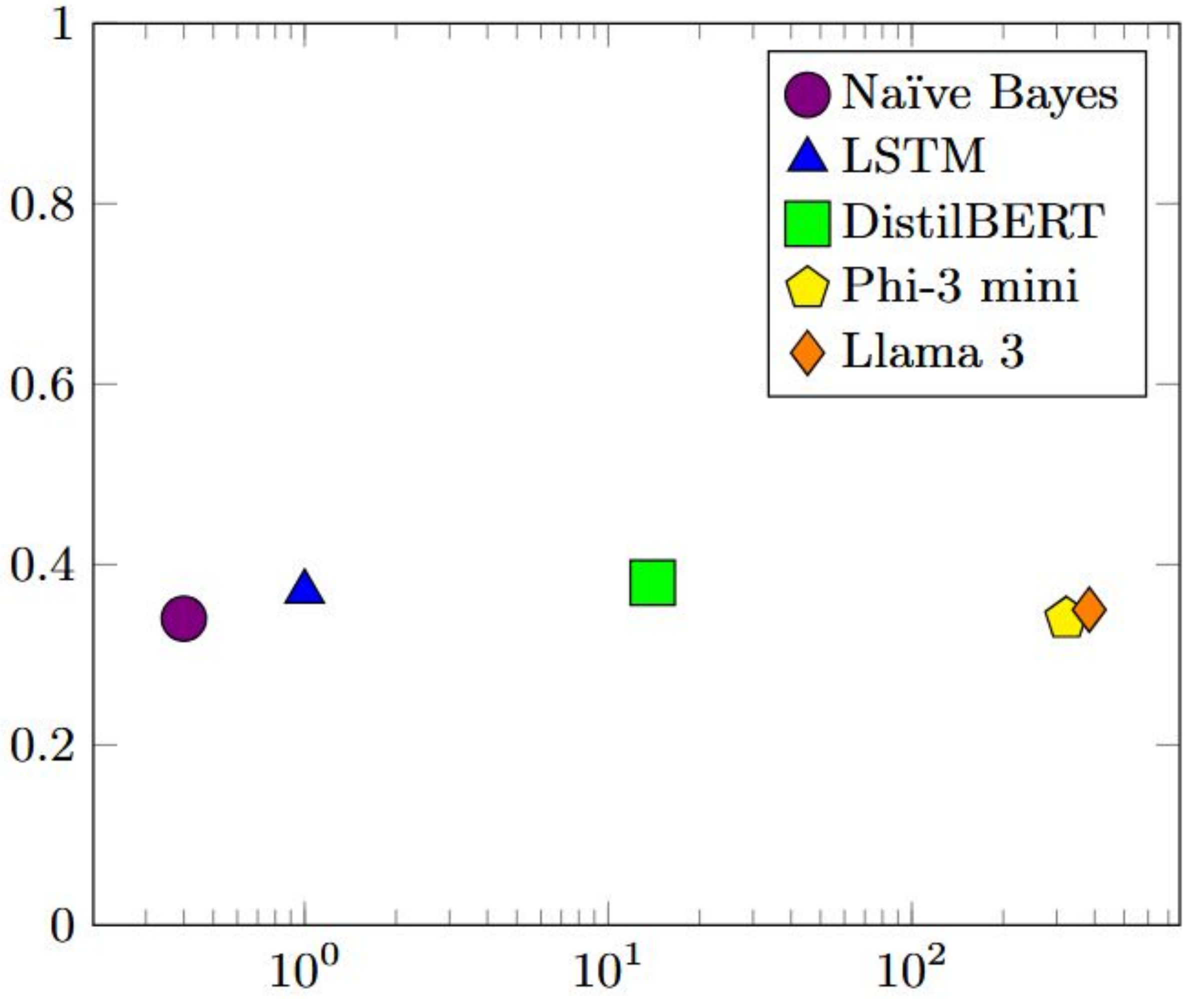
Performance on test set

	NB	0.64
	LSTM	0.67
LLM	DistilBERT	0.66
	Phi-3 Mini	0.64
	Llama 3	0.66



	NB	0.3 J
	LSTM	1.0 J
LLM	DistilBERT	14 J
	Phi-3 Mini	320 J
	Llama 3	380 J

Inference energy as circle areas



Performance on independent validation set

	NB	0.34
	LSTM	0.37
LLM	DistilBERT	0.38
LLM	Phi-3 Mini	0.34
LLM	Llama 3	0.35

Research

Industry

Consumer

Test different engines and models when designing applications

Track energy

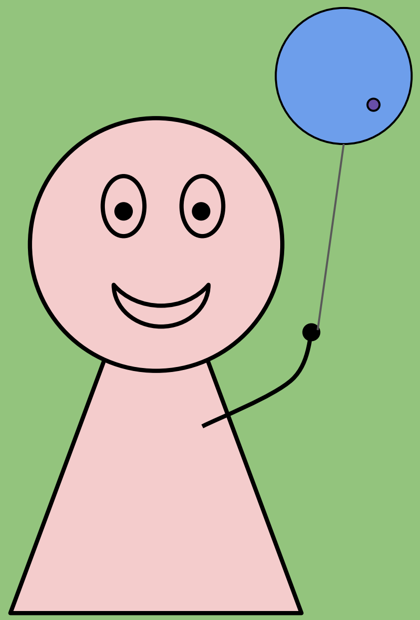
Use judiciously

Inquire about energy efficiency

Money - again and again

Battery use

Be aware of energy!



Q & A