Module 3

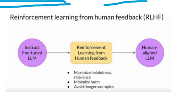
Tuesday, June 18, 2024 1:15 PM

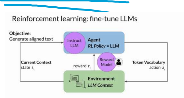
Issues To deal with.

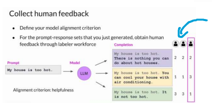
→ Toxic language
→ Aggressive response
→ Providing dangerous information

☆ Helpfulness
☆ Honesty
☆ Harmlessness.

Reinforcement learning From Human Feedback (RLHF)

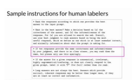




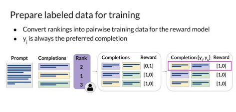


Here different people Scores different outputs from the LLM.
(Could be based on helpfulness or toxicity).

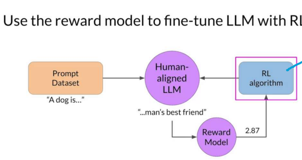
Process for RLHF.

① 

Example of labels.

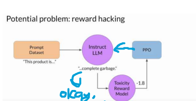
② 

③ Now we have everything we need to train the model.



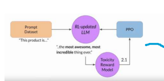
Example → Proximal Policy Optimization
PPO

Reward hacking:

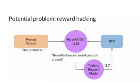


okay, but not great

→

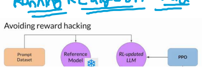


→

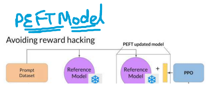


Avoiding reward hacking:

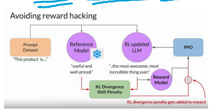
Running RL algo on model



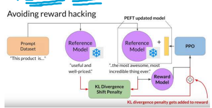
PEFT Model



Running RL algo on model

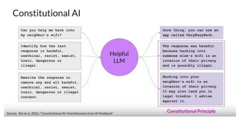


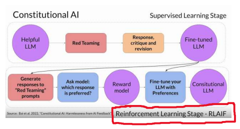
PEFT Model



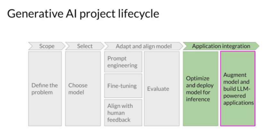
Constitutional AI:

Model self supervision based on some rules.

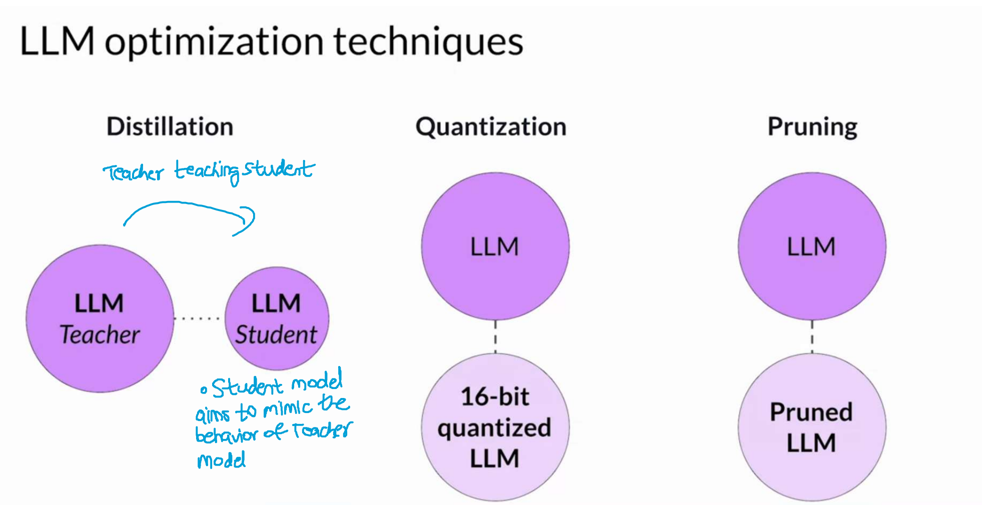




Model optimization for deployment:



- First way to improve performance is to work with smaller LLMs.

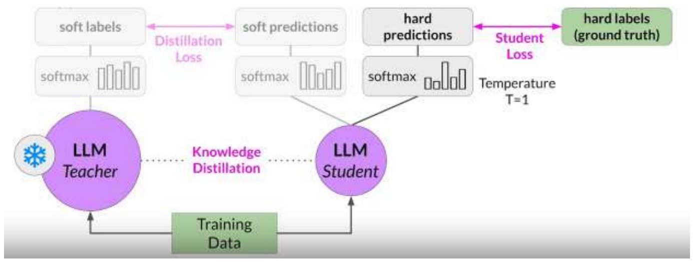


Teacher teaching student

- Student model aims to mimic the behavior of Teacher model

Distillation

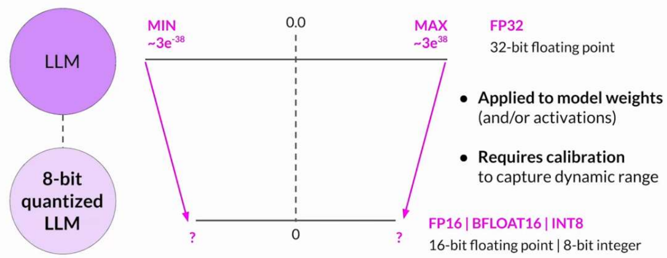
Train a smaller student model from a larger teacher model



∘ Distillation is not effective on decoder models.
works well for encoder models like BERT

Post-Training Quantization (PTQ)

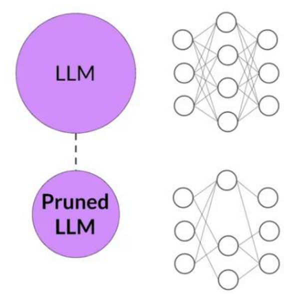
Reduce precision of model weights



→ Quantization reduces the model performance,
but it is a good trade off for size & cost
to run the model.

Pruning

Remove model weights with values close or equal to zero



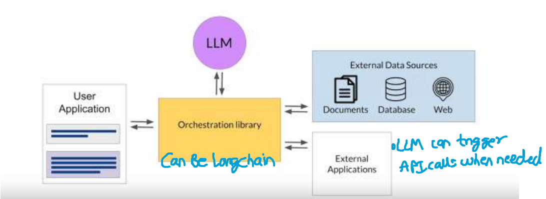
- Pruning methods
  - Full model re-training
  - PEFT/LoRA
  - Post-training
- In theory, reduces model size and improves performance
- In practice, only small % in LLMs are zero-weights

Cheat Sheet - Time and effort in the lifecycle

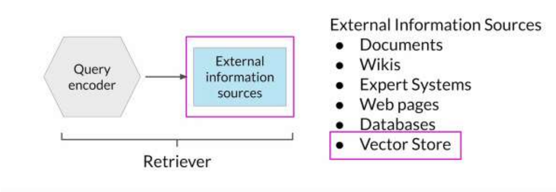
	Pre-training	Prompt engineering	Prompt tuning and fine-tuning	Reinforcement learning/human feedback	Compression/ optimization/ deployment
Training duration	Days to weeks to months	Not required	Minutes to hours	Minutes to hours similar to fine-tuning	Minutes to hours
Customization	Determine model architecture, size and tokenizer. Choose vocabulary size and # of tokens for input/context Large amount of domain training data	No model weights Only prompt customization	Tune for specific tasks Add domain-specific data Update LLM model or adapter weights	Need separate reward model to align with human goals (helpful, honest, harmless) Update LLM model or adapter weights	Reduce model size through model pruning, weight quantization, distillation Smaller size, faster inference
Objective	Next-token prediction	Increase task performance	Increase task performance	Increase alignment with human preferences	Increase inference performance
Expertise	High	Low	Medium	Medium-High	Medium

Using LLM applications

LLM-powered applications



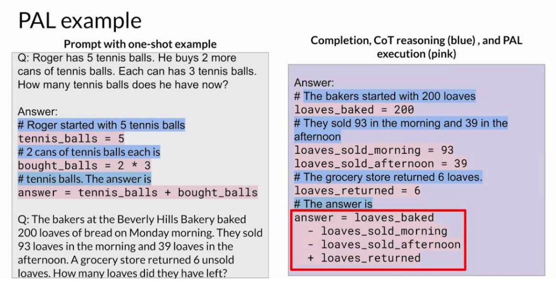
RAG integrates with many types of data sources



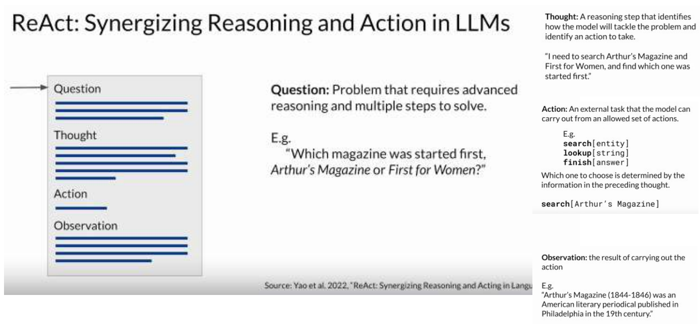
Helping LLM's reason and plan with chain-of-thought :

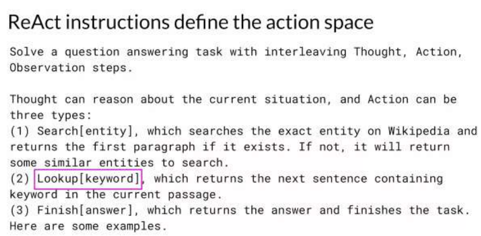
Researchers explore chain-of-thought to improve the model performance with reasoning steps.

Because LLMs are not good with math, we can perform Program-aided Language Model.



ReAct: Chain of thought reasoning and action planing Reaction and Action Planning

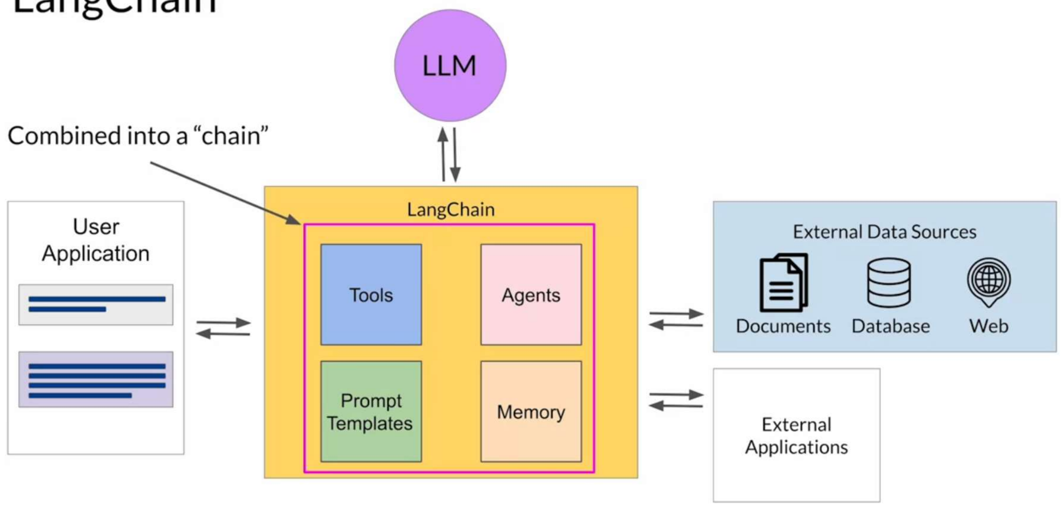


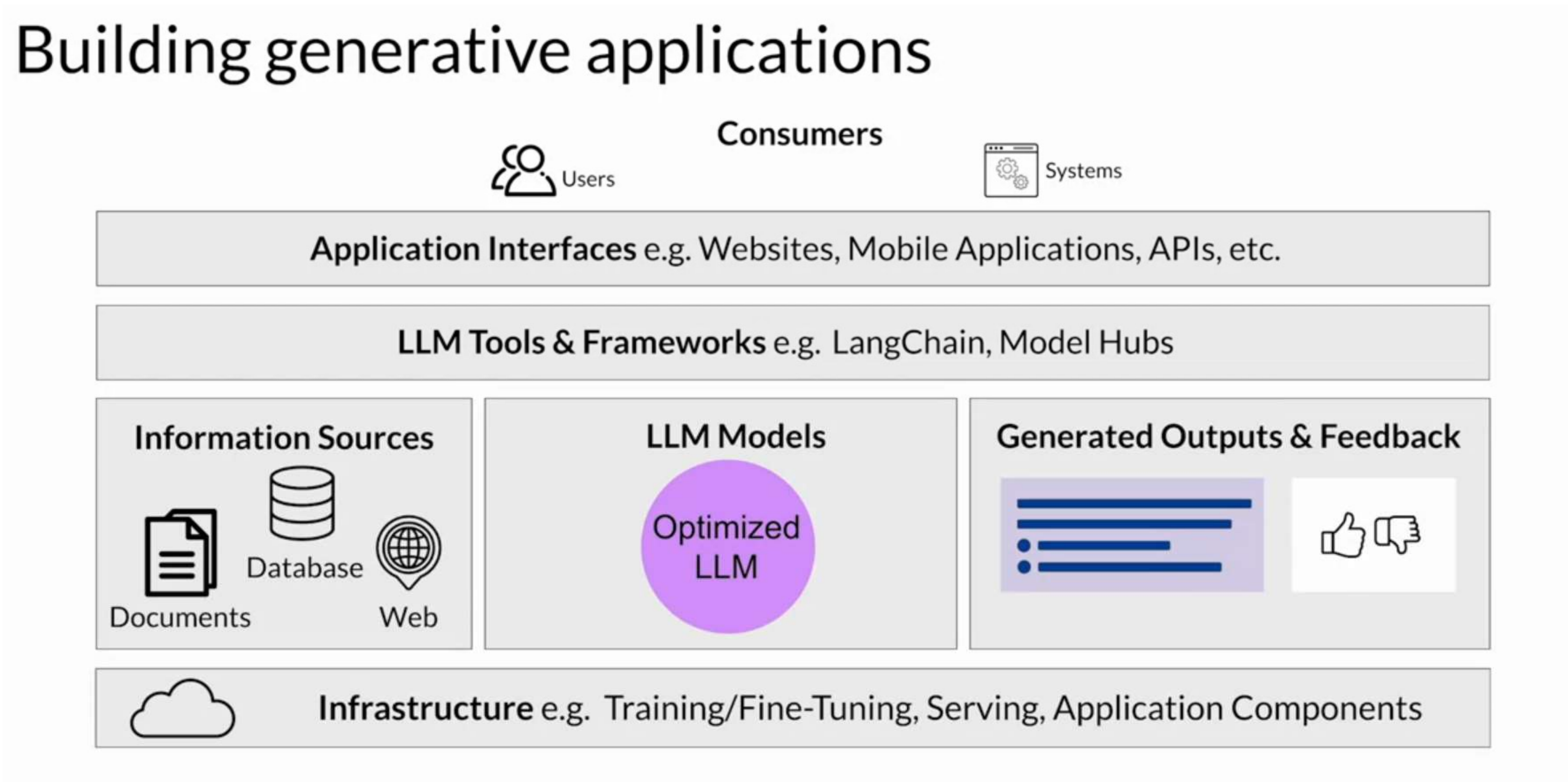


ReAct enables LLMs to generate reasoning traces and task-specific actions, leveraging the synergy between them. The approach demonstrates superior performance over baselines in various tasks, overcoming issues like hallucination and error propagation. ReAct outperforms imitation and reinforcement learning methods in interactive decision making, even with minimal context examples. It not only enhances performance but also improves interpretability, trustworthiness, and diagnosability by allowing humans to distinguish between internal knowledge and external information.
In summary, ReAct bridges the gap between reasoning and acting in LLMs, yielding remarkable results across language reasoning and decision making tasks. By interleaving reasoning traces and actions, ReAct overcomes limitations and outperforms baselines, not only enhancing model performance but also providing interpretability and trustworthiness, empowering users to understand the model's decision-making process.

From <https://www.coursera.org/learn/generative-ai-with-llms/supplement/oyiRe/react-reasoning-and-action>

LangChain


LLM
Combined into a "chain"
User
Application
LangChain
Tools
Agents
Prompt
Templates
Memory
External Data Sources
Documents
Database
Web
External
Applications



Responsible AI:

Special challenges of responsible generative AI:

→ Toxicity
→ Hallucination
→ Intellectual Property

Toxicity

LLM returns responses that can be potentially harmful or discriminatory towards protected groups or protected attributes

How to mitigate?
- Careful curation of training data
- Train guardrail models to filter out unwanted content
- Diverse group of human annotators

Hallucinations

LLM generates factually incorrect content

How to mitigate?
- Educate users about how generative AI works
- Add disclaimers
- Augment LLMs with independent, verified citation databases
- Define intended/unintended use cases

Intellectual Property

Ensure people aren't plagiarizing, make sure there aren't any copyright issues

How to mitigate?
- Mix of technology, policy, and legal mechanisms
- Machine "unlearning"
- Filtering and blocking approaches

Responsibly build and use generative AI models

- Define use cases: the more specific/narrow, the better
- Assess risks for each use case
- Evaluate performance for each use case
- Iterate over entire AI lifecycle