Q* is the currently accepted notation for the Optimal Action Value Function in RL.

Reasoning

The biggest gains on reasoning come from strong reward models, as opposed to more SFT data or tools.

Much of (unpublished) research is now focused on finding a general planning algorithm for LLMs, i.e. some equivalent of the dlPFC. So PLANNING is the name of the game.

https://lnkd.in/d_r9JTBh

Maths

In the literature, we have seen different approaches to teaching math to AI models like Transformers + Beam Search or Large language models, which are capable of solving tasks that require complex multistep reasoning by generating solutions in a step-by-step chain-of-thought format.
One effective method in the second involves training reward models to discriminate between desirable and undesirable outputs.
In the literature we see two distinct methods
for training reward models: outcome supervision & process supervision.

https://lnkd.in/d7Z685NS
https://lnkd.in/dThAb9mS
https://lnkd.in/dZ6F4ZBF

Q-Star (Q*) reinforcement learning is a theoretical concept in the field of artificial intelligence and machine learning, specifically within the realm of reinforcement learning (RL).

Here's a brief overview:

1. Foundational Concepts in Reinforcement Learning: Reinforcement learning is a type of machine learning where an agent learns to make decisions by performing actions in an environment and receiving feedback in the form of rewards. The goal is to learn a policy that maximizes the cumulative reward over time.

2. The Role of the Q-Function: In many RL algorithms, a critical component is the Q-function (or Q-value), which estimates the expected cumulative reward of taking a particular action in a given state, and then following a specific policy thereafter. The Q-function helps the agent evaluate the potential of different actions.

3. Q-Star (Q*): Q* represents the optimal Q-function. It gives the expected cumulative reward for taking the best action in each state under the optimal policy. In other words, Q* embodies the best possible strategy for an agent in a given environment. It tells the agent what to expect in terms of rewards if it always makes the best possible decisions.

4. Learning Q* in Practice: The challenge in reinforcement learning is that Q* is unknown and must be estimated from interactions with the environment. Algorithms like Q-learning, Deep Q

Networks (DQN), and others are designed to iteratively approximate Q* by updating their estimates based on the rewards received for actions taken.

5. Significance: Knowing or accurately estimating Q* is crucial for an agent to make optimal decisions that maximize rewards over time.

AIFI - Artificial Intelligence Finance Institute