YouTube DNN model in recommender system use two-tower architecture, which is like other larger-scale video platforms use.\

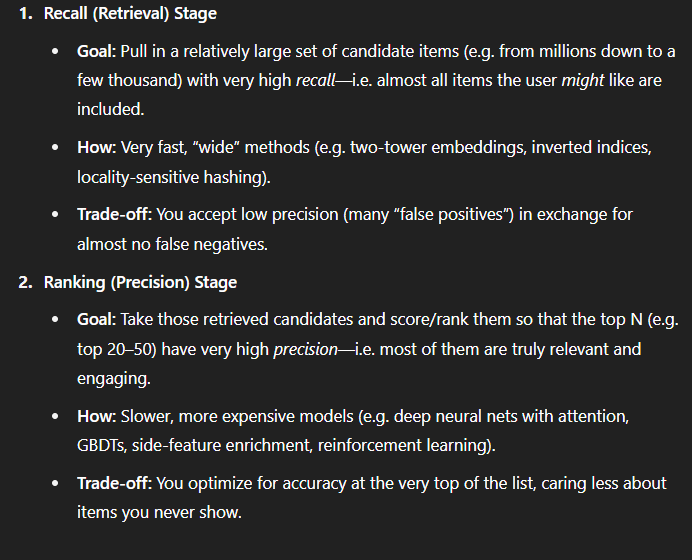
The advantage of two-tower architecture is

This architecture splits recommendation into two main stages

A screenshot of a black screen

Description automatically generated

The recommender system process have three stages, recall ,ranking , re-ranking



A screenshot of a black and white list

Description automatically generated

A screenshot of a black and white screen

Description automatically generated

A screenshot of a computer screen

Description automatically generated

“We will implement an online‐SGD update for each incoming like/dislike to the user‐tower (latency budget: ≤100 ms) and run a 5-minute micro-batch job for item‐tower updates. For our user-study, we’ll replay 10 K real student interactions over 1 hour (preserving original timestamps) to measure both throughput and latency; if real logs are unavailable, we will instead generate synthetic sessions using a probabilistic click model calibrated from pilot data.”

The recommender system architecture