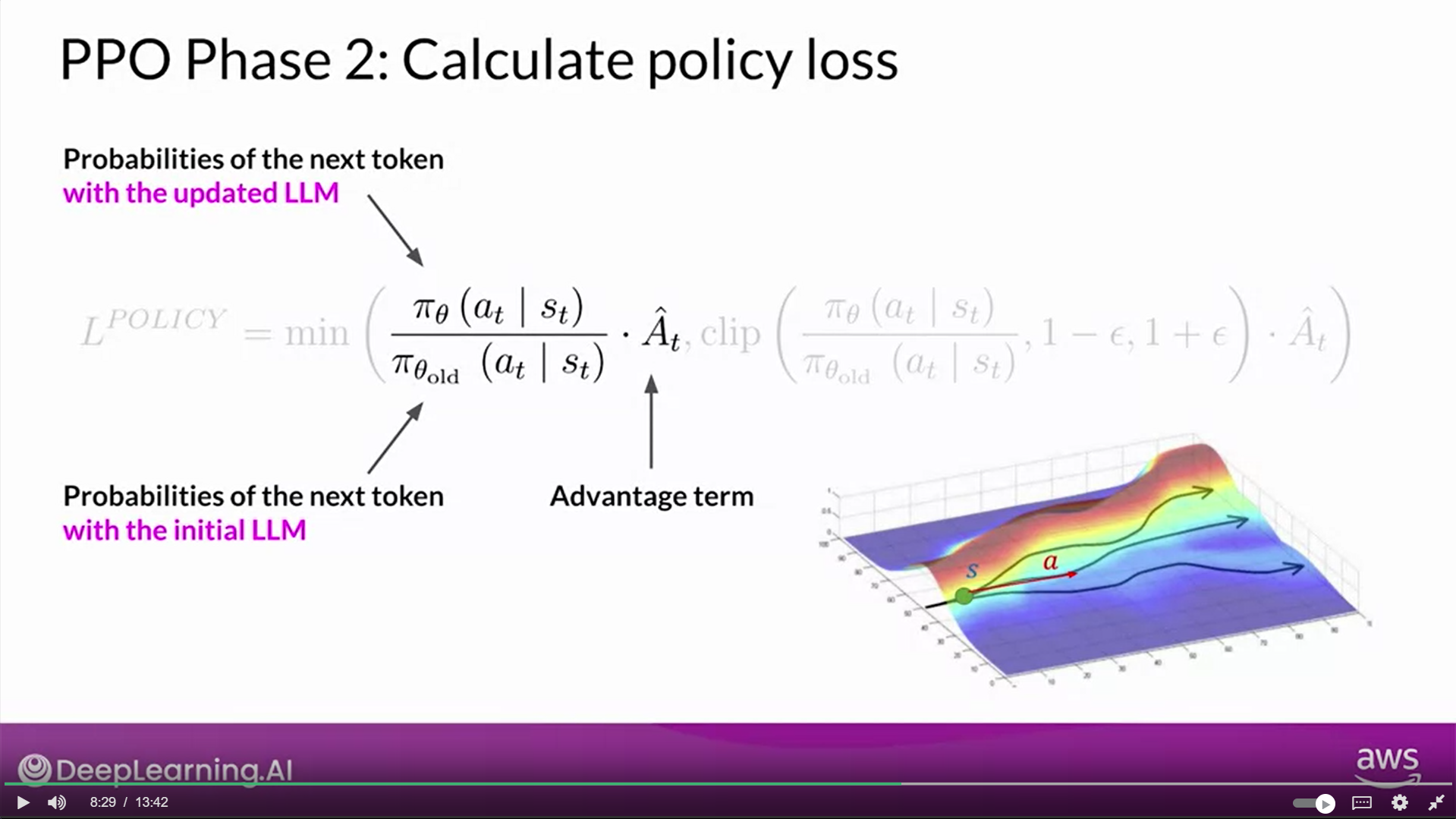
PPO ( Proximal Policy Optimization)

Policy Loss Formula



A diagram of a graph

Description automatically generated

Entropy Loss

A screenshot of a computer

Description automatically generated

Higher the entropy, more creative the answer will be. This is similar to the temperature settings in LLMs.

PPO Objective Function

