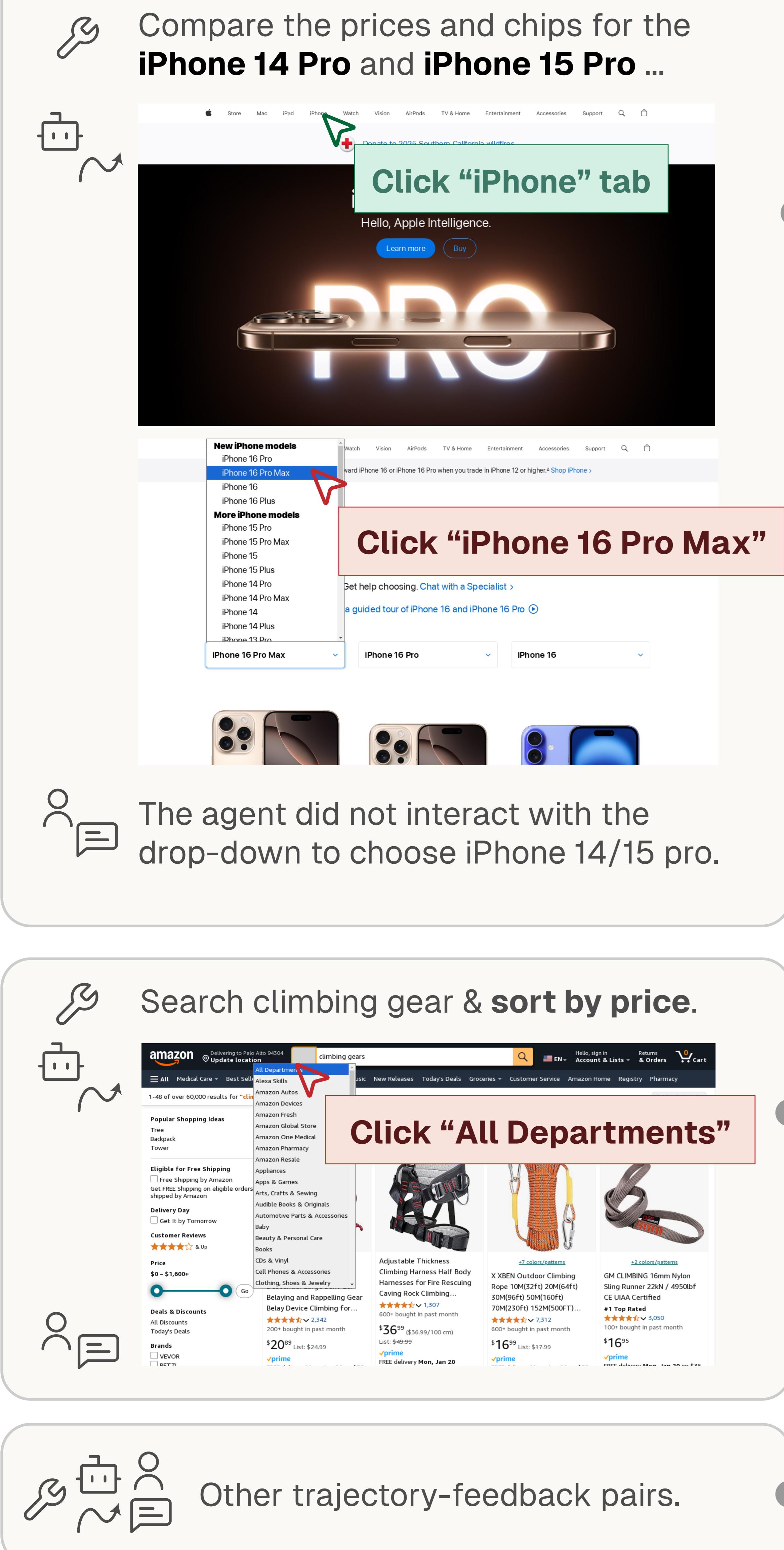
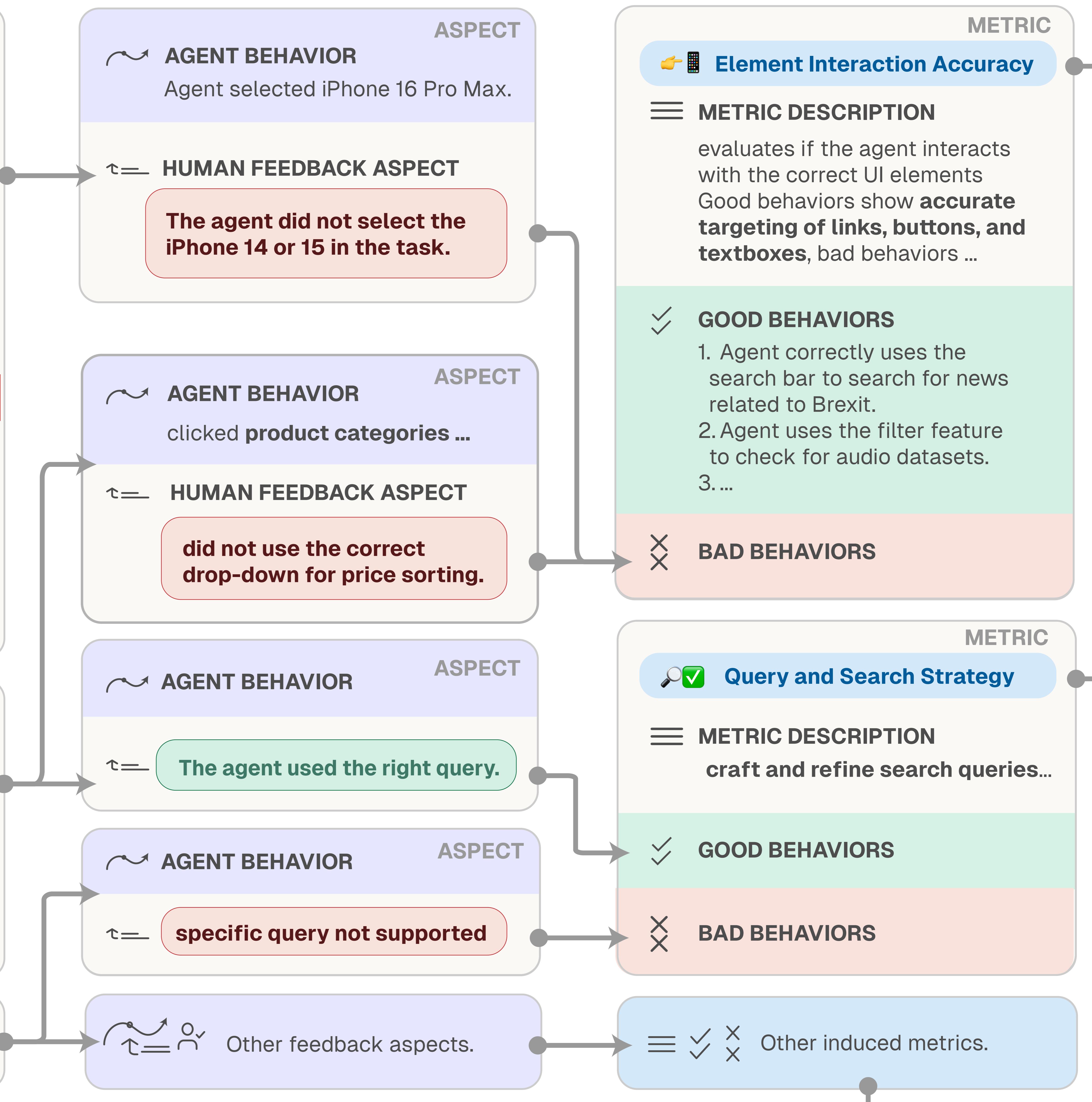


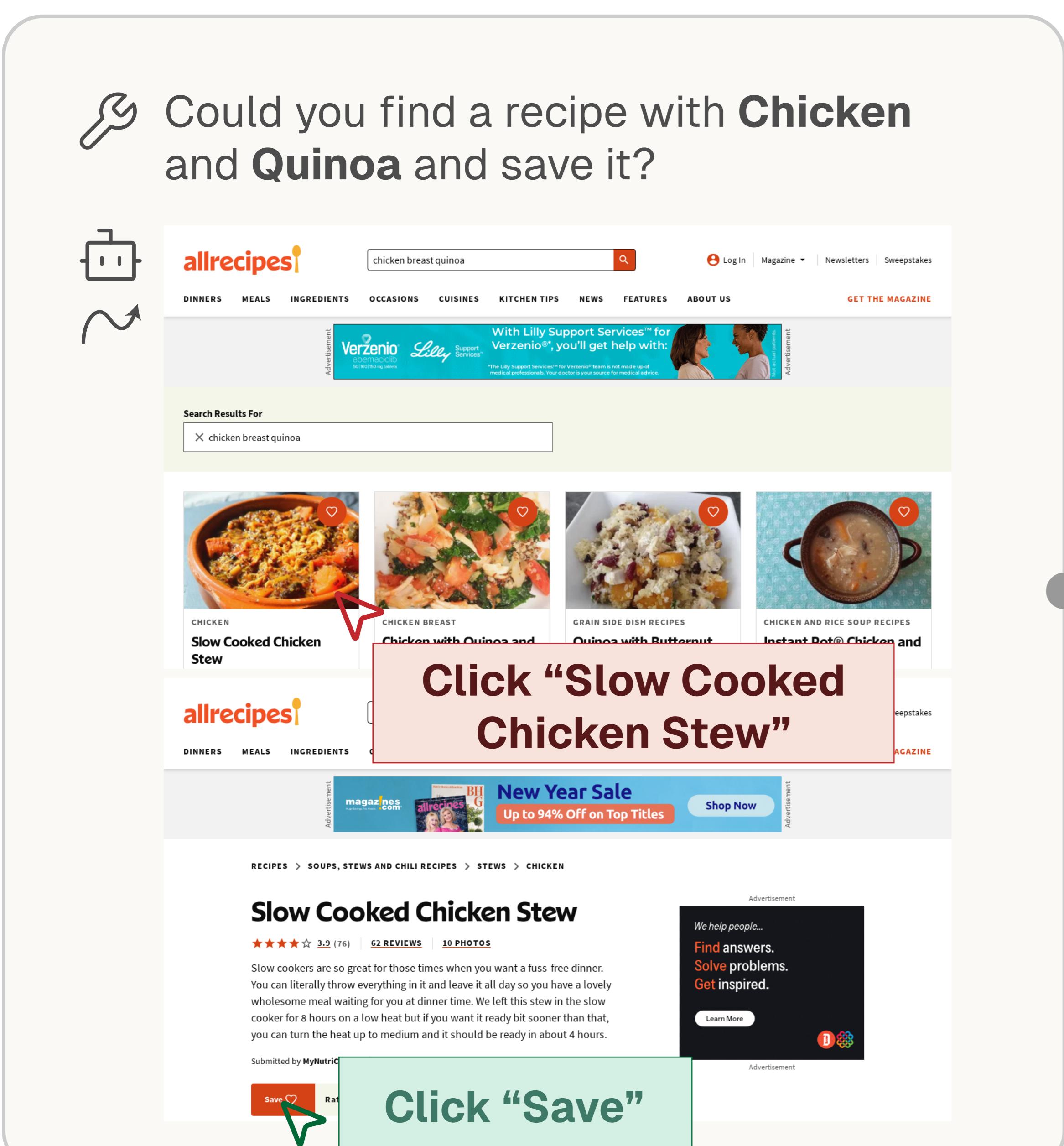
Collect human feedback



Induce concrete and new metrics



Ready to use for agent evaluation



Evaluate quality of metrics

