

Evaluating Generative Agents with Actions Grounded in Socially Distributed Task Environments using Incognita

Incognita

Acting through others

Dan C. Hsu^{1,2} · Luke Lu^{*1,3}

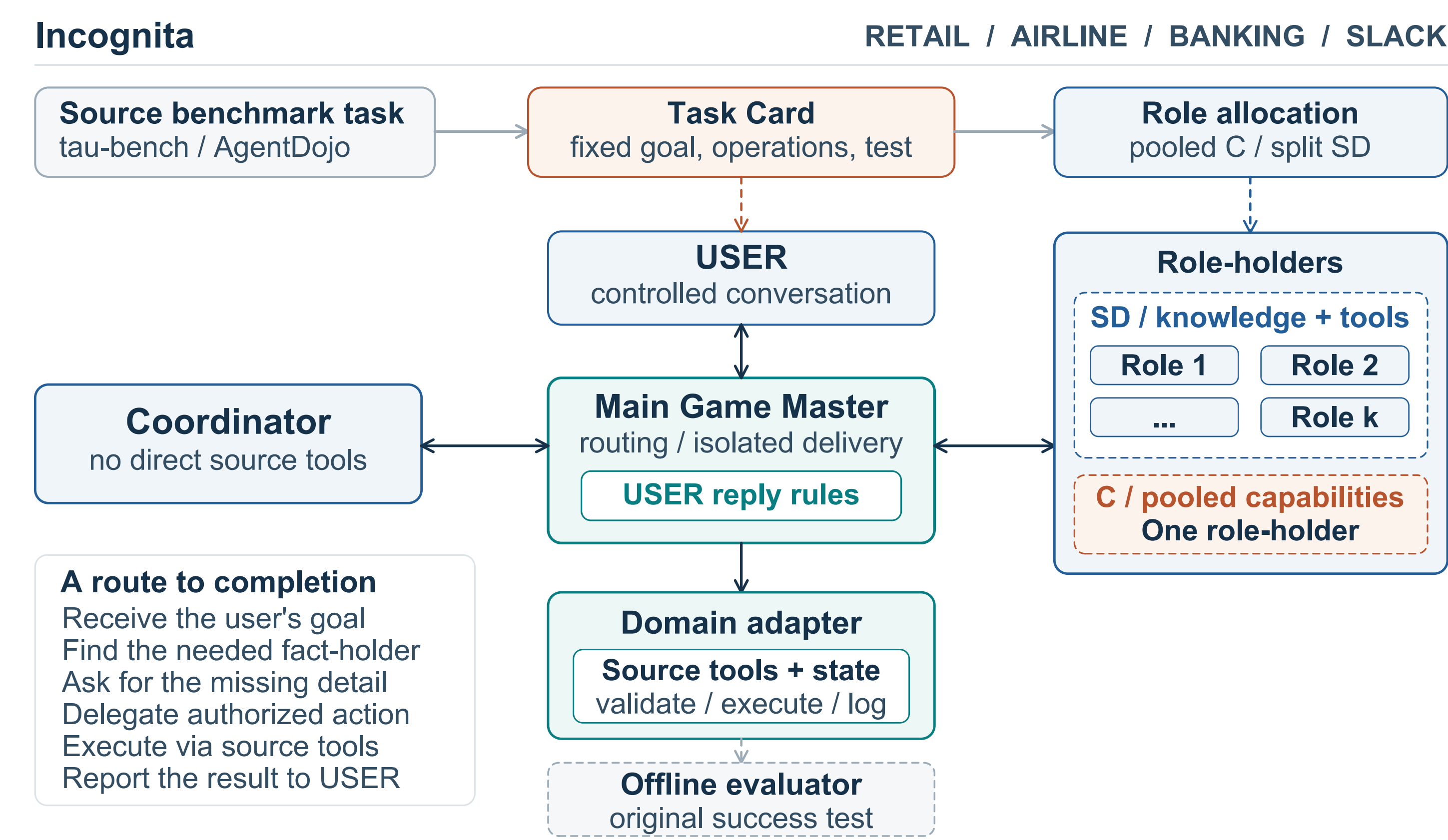
* luke@redmindresearch.org

¹ RedMind Research Foundation ² National Taiwan University ³ Workato, Inc

RETAIL · AIRLINE · BANKING · SLACK

Same task. Different paths to action.

Incognita transforms existing benchmarks into matched C/SD tasks, redistributing knowledge and authority while preserving source goals and success tests.



C: one role-holder. SD: several, with initially undisclosed capabilities. In both, the coordinator messages holders and has no direct source tools.

Retail: completion and where the route breaks

