Dorsal UX vs Codex web search

Chris Benson's adversarial agent swarm experiment - November 18, 2026

Dorsal’s reply answered the specific question. Mine missed Benson’s experiment and substituted broader examples.

Aspect	Sol 6.1 + web search	Dorsal UX reply
Benson’s experiment	Couldn’t find it. Described his general discussions of swarms and cyber defense.	Found him experimenting with opposing agent roles to improve enterprise architecture.
Closest engineering example	Multiagent debate for reasoning and factual accuracy.	AI Jason’s five-agent debugging: competing hypotheses, shared findings, debate, and final reports.
Other use cases	Scientific discovery, vulnerability exploitation, parallel research.	Verified parallel coding, strategic review, vulnerability remediation, product evaluations, independent forecasting.
Practical detail	Explained broad patterns and published results.	Identified concrete roles and workflows: workers, challengers, verifiers, and independent predictions.
Sources	Direct links to relevant primary sources, but none established the requested Benson experiment.	Eight source links, including the decisive Benson transcript record and closely related demonstrations.
Qualifications	Acknowledged that I hadn’t found the experiment.	Distinguished experiments from production results, thematic similarities from recorded connections, and benefits from coordination costs.
Recorded performance	Sol 6.1 at xhigh effort. 3 min 35 s from request to final reply; given a 120-second allowance.	37 seconds, 6 Dorsal calls, 8 full records fetched.

The substantive advantage was discovery: Dorsal surfaced the exact experiment and more applicable engineering examples. My reply provided credible background but failed the central lookup.

An informal comparison of two actual runs, not a controlled benchmark. Both used Sol 6.1; the web-only run used xhigh effort in Codex. The prompts and retrieval capabilities differ. The web-only duration is request-to-reply time; Dorsal’s duration is its server research time. These are different timing boundaries. Dorsal’s effort setting was not recovered.