

AGENT SIGNAL

17 October 2026 | Edition 04

Model updates

Beam 1.0 weights released

A 501B-total/23B-active mixture-of-experts model aimed at coding and long-running agents, originally announced October 5. What's new: On October 16, Reflection published Apache-2.0 weights in BF16, FP8 and NVFP4 plus serving integrations and beta API access. Why: Open weights let engineers inspect and independently evaluate the model; self-hosting such a large checkpoint demands serious compute. [Source](#)

Muse Spark Open checkpoint

Meta's multimodal agent and coding model has 816B total/42B active parameters and a claimed million-token context. What's new: Separate weights were released October 15, according to Meta's model card; not a repeat of October 12 Muse Wonder. Why: Another inspectable multimodal model, but the custom Muse Model License includes attribution and service-provider conditions. [Source](#)

Repositories to try

reflection-oss/mirror | Apache-2.0

A coding agent that reads repositories, edits code and runs commands with approvals and isolated worktrees. Released alongside Beam's October 16 open-weight ecosystem, with a Python interface and Rust agent core. [Source](#)

AGENT SIGNAL

17 October 2026 | Edition 04

Agent ecosystem

Dependency license compliance reaches enterprise GA

Enterprise rules can reject pull requests adding packages with disallowed open-source licenses. Agents choosing dependencies can consult the same policy as reviewers instead of discovering compliance problems at merge time.

[Source](#)

Research watch

FDE-Gym: software delivery with vague requirements | arXiv v1 preprint — not peer-reviewed

Benchmarks agent delivery in customer-like systems with ambiguous requirements and stakeholders. Authors report 95 tasks, 13 model configurations; best mean score 0.56, just 14 fully resolved at a strict success bar. Limit: Scripted customers and simulated systems; not direct evidence of live production results. [Source](#)

LEASH: smarter escalation for computer-use agents | arXiv v1 preprint — not peer-reviewed

Routes a small agent to a stronger model on uncertainty signals, then monitors the stronger agent. In OSWorld, authors report 6.55–11.94-point gains over random escalation; one configuration reached 73.66% success at \$0.4117/task. Limit: Needs access to small-model hidden states and failure trajectories; benchmark-specific results and costs. [Source](#)

Today's experiment

Try reflection-oss/mirror: review a failing-test repair

Create a one-test Python bug fixture; inspect Mirror CLI and its manual approvals; optionally assess its fix. Use test output and Git diffs rather than trusting an agent's success claim. Tested: Sources reviewed for the 17 October edition; fixture, Git baseline and independent repair diff checked in a 19 October correction. Mirror installation and inference not run. [Source](#)