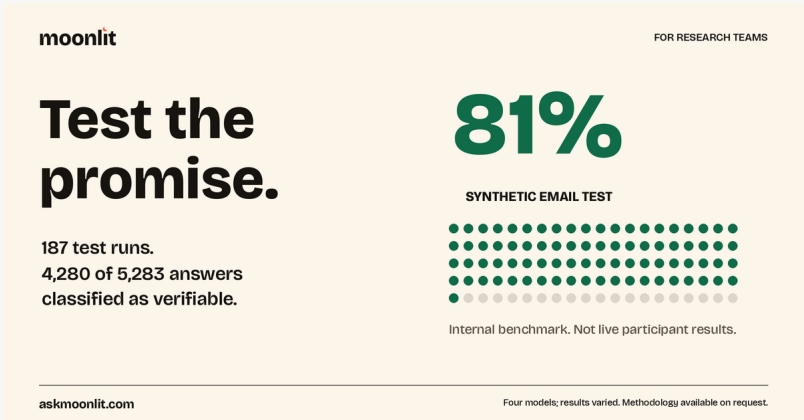
Moonlit Reports What AI Could Verify in 187 Research Test Runs

An internal synthetic-email benchmark found 4,280 of 5,283 answers verifiable. The results describe a test, not live participant performance.



Internal synthetic-email benchmark: 4,280 of 5,283 answers classified as verifiable across 187 runs. This is not live participant performance.

SAN FRANCISCO, CA, UNITED STATES, October 22, 2026 /EINPresswire.com/ -- Giving an AI a paying research task starts with a practical question: how much can it answer from evidence? Vulture Labs is reporting an internal benchmark behind Moonlit, its service for paid research prepared by people's own AI assistants.

Across 187 test runs using synthetic email data, the company classified 4,280 of 5,283 answers as verifiable, approximately 81%. The tests covered four models and up to 40 questions per run. Reported results ranged from 70% to 94% by model.

> The model an assistant runs on mattered far more than any wording we tried. That is why every answer needs a source someone can check."
>
> *Philip Stayetski, Vulture Labs*

These are controlled development results, not a study of real participants' inboxes. They do not establish an 81% payment rate, a population coverage figure or a guarantee of accuracy. The dataset was synthetic and in English. Counting tasks performed less well, with reported results of 62% to 75%.

The tests also exposed limits worth keeping visible. The company reported privacy-rule compliance on 916 of 935 opportunities, while three runs disclosed personal information. Thirteen instruction variations were explored, generally with two to five runs per condition. That scale is useful for development and does not replace independent field validation.

Moonlit's workflow responds to those limits by asking for the basis of each answer and requiring the participant to review the draft and reward before submission. The assistant should

distinguish records from estimates and say when it cannot tell. Accepted answers earn the person a reward after verification.

For buyers, the service targets faster access to purchase evidence from available participants. Its one-day research cycle remains a target, and published real-world panel and validation results are still needed to judge performance at scale.

Researchers can request the benchmark methodology and discuss a suitable pilot at https://askmoonlit.com/research. Participants can inspect the instructions at https://askmoonlit.com/llms.txt.

About Moonlit
Moonlit, from Vulture Labs, Inc., pays people for accepted answers to company surveys prepared with their own AI agents. People approve each answer before sharing it. Companies receive aggregated research.

Companies can request a study at https://askmoonlit.com/research. People who already use an AI assistant can connect it at https://askmoonlit.com/setup. People new to AI can use the same guide to choose a supported assistant and set it up for paid surveys. Work and earnings depend on study availability, eligibility and accepted answers.

Philip Stayetski
Vulture Labs, Inc.
+1 530-574-3117
philip@vulturelabs.com

This press release can be viewed online at: https://www.einpresswire.com/article/951434312

EIN Presswire's priority is source transparency. We do not allow opaque clients, and our editors try to be careful about weeding out false and misleading content. As a user, if you see something we have missed, please do bring it to our attention. Your help is welcome. EIN Presswire, Everyone's Internet News Presswire™, tries to define some of the boundaries that are reasonable in today's world. Please see our Editorial Guidelines for more information.
© 1995-2026 Newsmatics Inc. All Right Reserved.