MOBIUS FORGE

Mobius Forge vs. expert prompting: a controlled Meta ad experiment

Methodology report v1

$1.84	4.3x	84%
cost per lead (engine, blended)	cheaper than off-the-shelf AI	of CBO budget auto-allocated to engine

Document version:	v1
Published:	May 3, 2026
Experiment window:	April 5 to April 28, 2026 (23 days)
Test environment:	FPGATek (FPGA hardware education brand)
Platform:	Meta Ads, lead generation objective
Prepared by:	Mobius Forge (mobiusforge.ai)

Executive summary

Between April 5 and April 28, 2026, Mobius Forge conducted a controlled three-arm Meta ad experiment to test whether the engine produces creative concepts that outperform both off-the-shelf AI generation and refined AI prompting workflows on a live ad account.

Three conditions were tested against the same audience, brief, and product set:

- **Condition A:** off-the-shelf AI (single-pass generation)
- **Condition B:** refined AI workflow (multi-step prompting with critique)
- **Condition C:** Mobius Forge engine

The campaign ran in two phases. Phase 1 used Meta's Campaign Budget Optimization (CBO), letting Meta's algorithm allocate budget across the three ad sets in real time. Phase 2 used Ad Set Budget Optimization (ABO) with equal budgets, isolating creative performance from delivery effects.

Headline results:

- The Mobius Forge engine produced 162 leads at $1.84 per lead.
- The refined AI workflow produced 51 leads at $2.92 per lead, 1.6x more expensive.
- The off-the-shelf AI baseline produced 9 leads at $7.94 per lead, 4.3x more expensive.
- During the CBO phase, Meta's algorithm directed 84% of total spend to the engine ad set within days.

A counterintuitive secondary finding: in the ABO phase, the refined AI workflow produced higher click-through rates than the engine (2.30% vs 1.64%), but at 47% higher cost per lead. Click rates and lead quality moved in opposite directions.

This document contains the full methodology and raw data behind those numbers. The campaign represents N=1 advertiser, approximately $518 in total spend, and 222 total leads. It is directional evidence, not a universal performance guarantee. A direct replication on a consumer DTC brand is in progress.

Background and research question

State of AI ad creative generation in 2026

Most commercially available AI ad tools produce convergent output. Different products wrap the same underlying frontier models, and template-driven prompts generate variations of the same creative idea: different headlines for the same hook, different visuals for the same message, different colors for the same layout. The output is what is called fake variety. It looks different on the surface but tests the same underlying buyer belief.

A skilled human practitioner using a frontier model with iterative prompting can match or beat the output of these tools. This is observable in side-by-side comparisons. The wrapper critique applied to commercial AI ad tools is largely correct.

The question this experiment was designed to answer

The critical question for any AI ad engine that claims to be different is narrower: does it produce concepts that a skilled prompter cannot reach, and does Meta's delivery algorithm reward those concepts when they exist?

Mobius Forge's hypothesis is that structural buyer-barrier diversity, where concepts test different reasons a customer might not buy, outperforms surface-level variation, where concepts test different ways of saying the same thing, on Meta lead-generation campaigns. The engine is designed to generate concepts in buyer-barrier territory that existing creative does not cover, rather than to generate variations of existing winners.

Choice of test environment

FPGATek is an FPGA hardware education brand. We selected this category deliberately. Mobius Forge's claim is about structural buyer-barrier diversity, not category-specific tactics. If the engine outperforms AI baselines on hardware education buyers, the framework's claim of cross-category transfer holds up. Testing within our target customer base (consumer DTC) would have made the result less informative about generalizability, not more.

Methodology

Three conditions

Each condition was implemented as one ad set inside the test campaign. All three conditions received identical creative briefs. The only variable was the generation method.

	Condition A	Condition B	Condition C
Label	**Off-the-shelf AI**	**Refined AI workflow**	**Mobius Forge engine**
Method	Single-pass generation from a strong meta-prompt. No iteration, no critique loop.	Frontier model with multi-step prompting, including critique and refinement steps. Built by an experienced practitioner.	Standard engine output. No special tuning for this experiment.
Represents	The naive baseline most teams encounter when they try a generic AI ad tool.	The smart baseline most in-house growth teams can build internally.	Default Mobius Forge output for any new advertiser.

Two phases

Phase 1 (CBO). All three ad sets ran in a single campaign with shared budget under Campaign Budget Optimization. Meta's algorithm allocated budget across sets in real time based on performance signals it observes (delivery efficiency, conversion rate, audience response). This phase measured how Meta's delivery system perceives each creative set, independent of human judgment. The CBO algorithm's allocation pattern is a third-party signal of which creatives Meta considers more efficient against the campaign objective.

Phase 2 (ABO). After the CBO phase, all three ad sets were migrated to Ad Set Budget Optimization with equal budgets. This isolates pure creative performance from Meta's allocation behavior. The ABO comparison is the cleanest direct comparison of cost per lead across conditions.

Controls

- Identical broad audience definition across all three ad sets
- Same product set (FPGATek's core SKUs)
- Same campaign objective (Meta lead generation)
- Same bidding strategy
- Same time window (April 5 to April 28, 2026, 23 days)
- Same creative formats and aspect ratios across conditions
- Same number of unique creatives per condition

Reported parameters

Total spend	Approximately $518 across all three conditions
Total leads	222 across all three conditions
Duration	23 days (April 5 to April 28, 2026)
Leads (engine)	162
Leads (refined AI)	51
Leads (off-the-shelf AI)	9
Primary metric	Cost per lead (CPL)
Secondary metrics	Click-through rate (CTR), Meta CBO budget allocation

Results

Headline comparison (Phase 2, ABO)

Condition	Leads	CPL	Share of leads	Cost ratio vs. engine
A. Off-the-shelf AI	9	$7.94	4%	4.3x more expensive
B. Refined AI workflow	51	$2.92	23%	1.6x more expensive
C. Mobius Forge engine	162	$1.84	73%	baseline

Phase 1 (CBO): Meta's algorithm voted

Within days of campaign launch, Meta's algorithm shifted 84% of total CBO spend to the Mobius Forge ad set. The off-the-shelf AI variant received minimal spend almost immediately. The refined AI workflow received intermediate allocation. This pattern held throughout the CBO phase.

Why this matters: the budget allocation was not a human decision. Meta's delivery system selected the engine ad set based on real-time performance signals (delivery efficiency, click and conversion behavior, audience response) that the experiment did not control or influence. This is third-party validation that does not depend on Mobius Forge's judgment, the audience's subjective response, or any post-hoc analysis.

Phase 2 (ABO): the counterintuitive finding

In the ABO phase, click-through rate and cost per lead moved in opposite directions for the refined AI workflow and the engine.

Condition	Link CTR	Cost per lead
Refined AI workflow	2.30%	$3.06
Mobius Forge engine	1.64%	$2.09

The refined AI workflow generated higher CTR but cost 47% more per lead. The engine's clicks were pre-qualified by the angle they responded to. Higher click rates on cleverer-sounding copy did not translate to better leads behind those clicks.

This finding has broader implications for performance creative evaluation. Teams that optimize creative based on CTR alone, a common practice given how visible CTR is in Meta's interface, may select for clever creative that performs poorly downstream. The engine's lower CTR but better lead quality suggests a different optimization target: not engagement, but accuracy of buyer self-selection.

Interpretation

The CBO and ABO results agree in direction and at meaningful magnitude. The engine outperformed both AI baselines on the primary metric (CPL) and was independently selected by Meta's delivery algorithm during the budget allocation phase.

The CBO result is the harder result to argue with

We did not pick which ad set Meta favored. Meta's delivery system did, based on signals the experiment had no access to manipulate. The 84% allocation to the engine ad set is a third-party signal of creative quality from Meta's own delivery infrastructure.

The ABO result is the cleaner direct comparison

With equal budgets and identical audiences, the engine produced 1.6x cheaper leads than a refined AI workflow built by an experienced practitioner. This is not a comparison against a poorly-prompted baseline. The refined AI workflow represents the realistic alternative most sophisticated growth teams can build internally.

On cross-category transfer

The engine was tested in a category outside Mobius Forge's target customer base. The underlying mechanism the engine optimizes for, concepts targeting different buyer barriers rather than different surface-level variations, is described in terms that apply to any consumer-facing creative challenge: skepticism, urgency, identification, switching cost, claim fatigue, social proof, risk reversal. These appear in skincare, supplements, software, and FPGA hardware education. The product changes; the buyer psychology does not.

The fact that the engine outperformed AI baselines on FPGA buyers is informative because it suggests the result is not driven by the engine being tuned to a specific category. The next replication, on a consumer DTC brand, will test whether the magnitude of the improvement holds in a category closer to the target ICP.

Limitations and what this study does not establish

This experiment is one campaign on one advertiser in one category over 23 days. It is directional evidence supporting the core hypothesis. It is not universal proof.

Specifically, this experiment does not establish:

- That the engine outperforms expert prompting in every category.
- That the engine outperforms expert prompting on every campaign objective. This experiment was lead-generation only.
- That the magnitude of improvement (1.6x cheaper leads vs. refined AI) generalizes to other advertisers.
- That the result holds at significantly larger spend levels.

The total spend was modest (approximately $518), the duration was short (23 days), and the lead sample (222 across three conditions) is small. A direct replication on a consumer DTC brand is in progress. Subsequent experiments will report whether the result holds across categories and at larger scale.

We are publishing this document at this stage because the FPGATek experiment is one of the few rigorous controlled tests of an AI ad engine against expert prompting on a live Meta account. The public methodology and raw numbers allow sophisticated buyers to evaluate the work themselves.

Appendix A: Condition A method (off-the-shelf AI)

Condition A represents what teams produce when they engage a generic AI ad tool with default settings. The objective was to capture the realistic floor of AI-generated ad creative, not a strawman.

Method

Each concept was generated by a single-pass call to a frontier large language model. The model received product information sourced from FPGATek's public website, the campaign's target audience definition, and a standard request to produce ad creative including hook, primary copy, headline, and visual direction. No iteration, critique, or refinement was applied. The full set produced by the model was submitted to the ad set as-is, in the same image-and-copy format used by the other two conditions.

What this method does not include

- Iterative critique or refinement loops
- Explicit instruction about creative diversity or buyer-barrier coverage
- Curation or filtering of outputs after generation
- Adjustment based on the brand's existing creative
- Examples or stylistic anchors from prior winning ads

Output volume

The same number of unique image-and-copy concepts as Conditions B and C ran in this ad set, ensuring volume parity across all three conditions.

Why describe the method rather than publish the prompt

Conditions A and B are documented at the level of method rather than as literal prompt text. Both conditions are described at the same level of abstraction. The methodology is documented in sufficient detail for a sophisticated reader to understand what each method tests and how the contrast was constructed. The verbatim prompts and the evaluation criteria used in Condition B are part of Mobius Forge's internal research process. Buyers who require deeper methodological verification before contracting can request a detailed walkthrough under a non-disclosure agreement.

Appendix B: Condition B method (refined AI workflow)

Condition B represents what an experienced practitioner produces when they apply a structured prompting workflow to the same frontier model used in Condition A. The objective was to capture the realistic ceiling of AI-generated creative without specialized tooling. This is the comparison that matters for sophisticated growth teams who can build custom prompting workflows internally.

Method

Each concept passed through a three-step workflow. The workflow was implemented and tuned by an experienced direct-response practitioner before the experiment began. The same frontier model and the same product information used in Condition A were used here.

- **Step 1: Initial generation.** A more structured prompt than Condition A's was used, containing the same inputs plus explicit framing about creative diversity, audience research notes drawn from public discussion of FPGATek's product category, and instructions to vary concepts along multiple persuasion dimensions.
- **Step 2: Evaluation.** Each generated concept was evaluated by the model against quality criteria including specificity of the hook, plausibility of the visual direction, and absence of generic copy patterns.
- **Step 3: Refinement.** Concepts flagged in evaluation were regenerated with the feedback from Step 2 folded into the prompt.

What this method adds beyond Condition A

- Explicit framing about creative diversity in the initial generation
- Audience research inputs from public discussion of the product category
- A self-evaluation step against quality criteria
- A refinement loop that regenerates flagged concepts

Output volume

The same number of unique image-and-copy concepts as Conditions A and C ran in this ad set. The refinement loop replaced flagged concepts; it did not add to the output count.

Appendix C: Raw data summary

Combined totals across both phases

Phase 1 (CBO) and Phase 2 (ABO) totals across the 23-day window. Detailed per-day breakdowns are available on request.

Metric	A. Off-the-shelf AI	B. Refined AI	C. Engine	Total
Leads	9	51	162	222
Cost per lead (blended)	$7.94	$2.92	$1.84	-
Share of total leads	4%	23%	73%	100%

Phase 2 (ABO) direct comparison: engine vs. refined AI

The ABO phase is the cleanest direct comparison between the engine and the strongest AI baseline. Equal budgets across both ad sets, identical audience, same time window. The off-the-shelf AI condition received negligible delivery in this phase and is excluded from the comparison.

Metric	B. Refined AI	C. Engine	Delta
Link CTR	2.30%	1.64%	Engine 29% lower
Cost per lead	$3.06	$2.09	Engine 32% cheaper

Phase 1 (CBO) budget allocation

Meta's algorithm directed 84% of total CBO spend to the Mobius Forge ad set within days of campaign launch. The off-the-shelf AI variant received minimal spend almost immediately, with Meta's delivery system identifying it as inefficient against the campaign objective. The refined AI workflow received the remaining intermediate share.

The CBO allocation pattern was stable across the phase. The algorithm did not rebalance toward the alternative ad sets at any point in the 23-day window.

About this study

N=1 advertiser, two campaigns (ABO and CBO), 23 days, 222 leads, approximately $518 total spend. Meta lead generation objective. All variants ran on the same Meta account against the same audience definition. This is directional evidence supporting our hypothesis, not a universal performance guarantee.

About Mobius Forge

Mobius Forge generates structurally different ad concepts for performance brands. Each concept tests a different buyer barrier, belief, or motivation, so growth teams can find winners instead of producing more surface-level variants. Mobius Forge is a product of Quiet Leap.

Web: mobiusforge.ai
Contact: hello@mobiusforge.ai
Document version: v1, published May 3, 2026