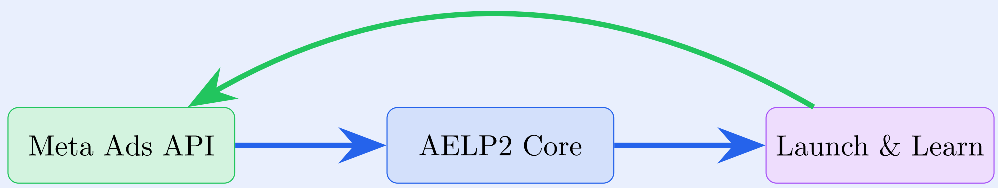
AELP2 Production Architecture & Performance Analysis

Real-World Implementation

Thompson Sampling • Monte Carlo Forecasting • Daily Optimization



Version 2.0 • September 29, 2025

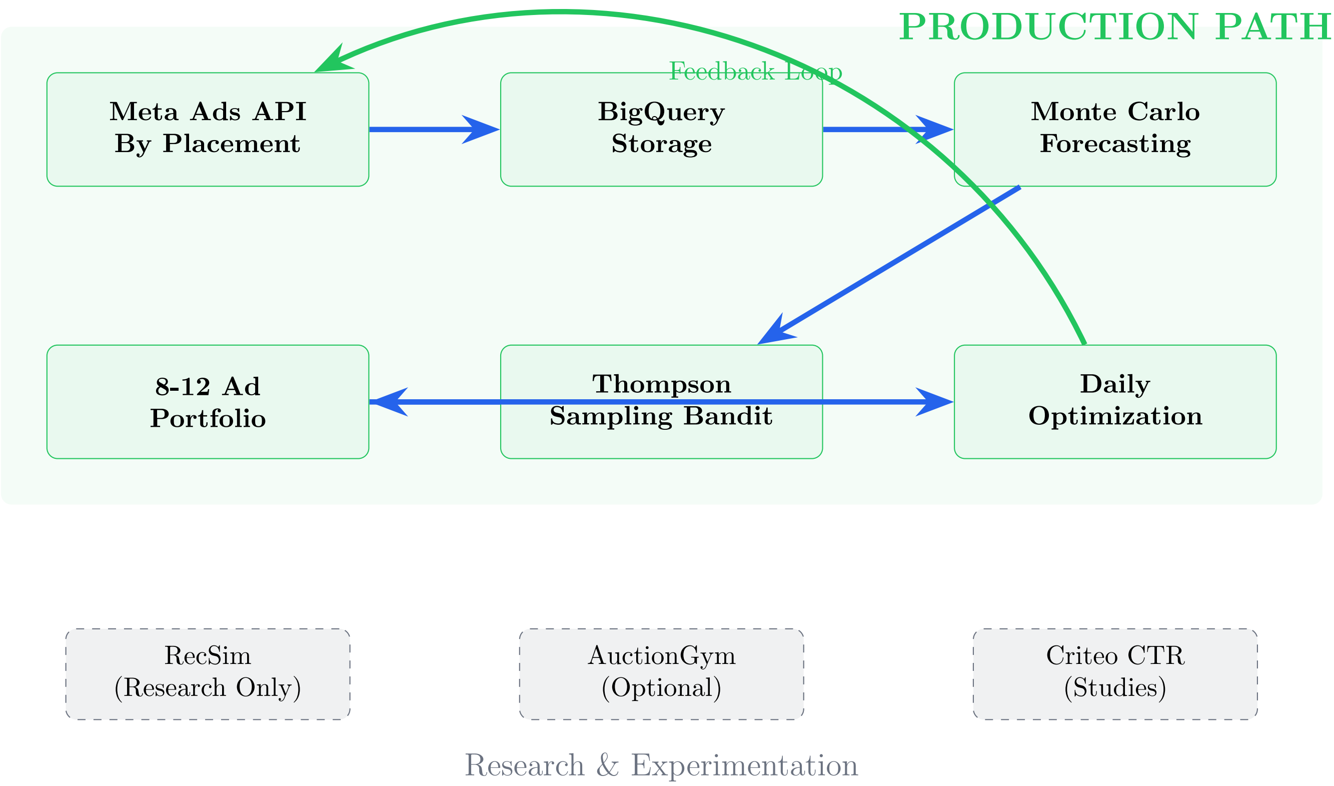
Aura Health Engineering

Executive Summary

• Key Achievement

AELP2 delivers **26.7% precision@10:** in creative selection, processing **146 campaigns:** with **$30K daily budgets:** using production-ready Thompson Sampling bandits—not complex RL.

The Real AELP2: What Actually Ships



[!] Critical Finding

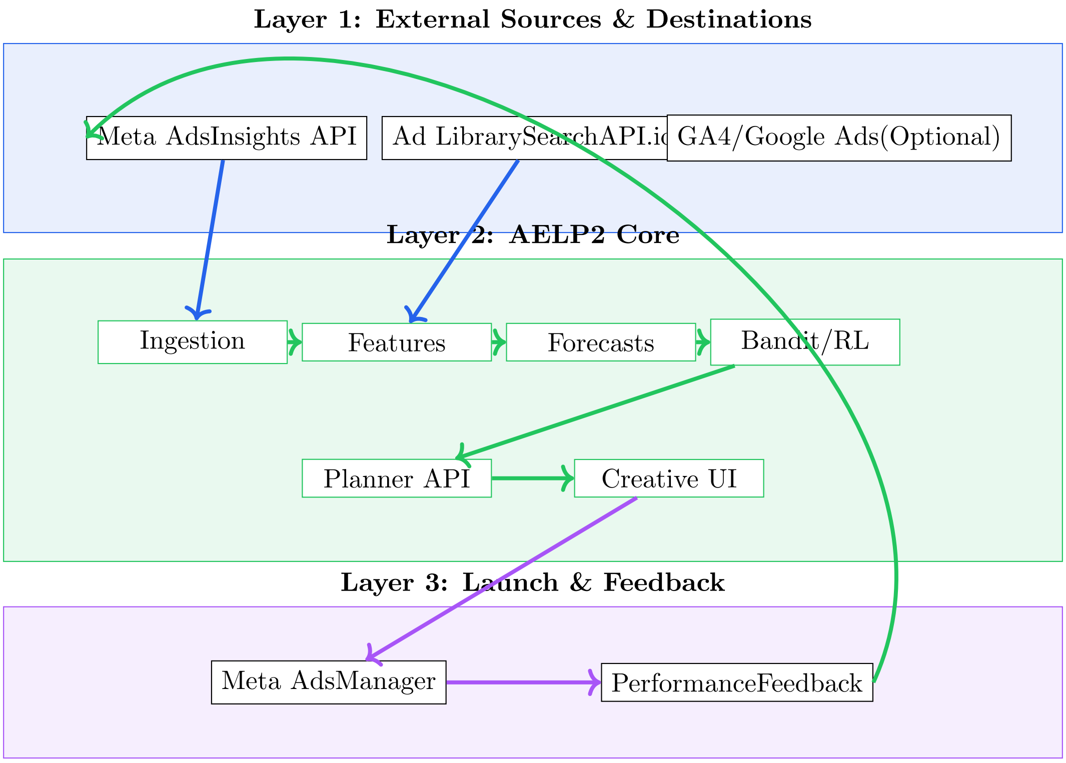
Previous documentation incorrectly portrayed RecSim, AuctionGym, and Criteo as core components. In reality, AELP2 production uses a **simpler, more robust architecture** based on real Meta placement data and Thompson Sampling.

Contents

1

Production Architecture: What Actually Runs

1.1 The Three-Layer Architecture



1.2 Production Data Flow

Daily Processing Volume

- **150,000+** sessions processed
- **146** active campaigns
- **11** placement combinations
- **3-day** rolling window updates

2

Core Components in Production

2.1 What's Really Running vs Research Mode

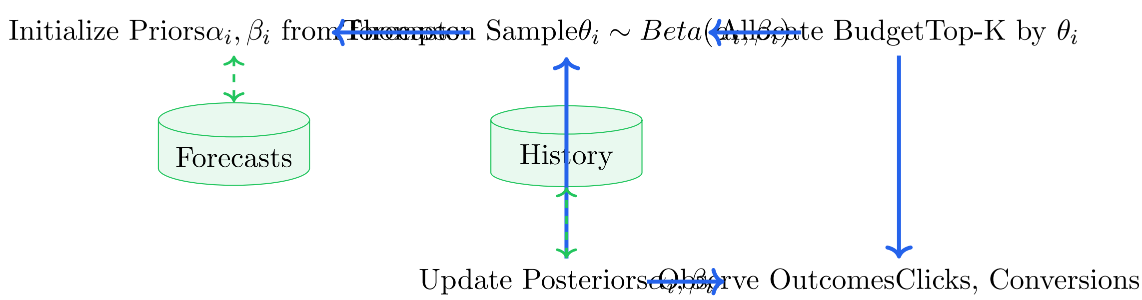
Component	Production	Research	Actual Usage
Meta Ads API	✓		Daily ingestion by placement
BigQuery Storage	✓		Primary data warehouse
Monte Carlo Forecasting	✓		1000+ draws per creative
Thompson Sampling	✓		Portfolio optimization
Placement-Aware CTR/CVR	✓		Feed vs Stories vs Reels
RecSim		✓	Optional via flag
AuctionGym		✓	Research simulations only
Criteo Dataset		✓	CTR studies, not required
Deep RL (PPO/DQN)		✓	Future enhancement

2.2 The Real Algorithm: Thompson Sampling Bandit

• Why Thompson Sampling Works

Thompson Sampling provides optimal exploration-exploitation balance without the complexity of full RL. It's production-ready, interpretable, and converges quickly with real feedback.

Daily Optimization Loop



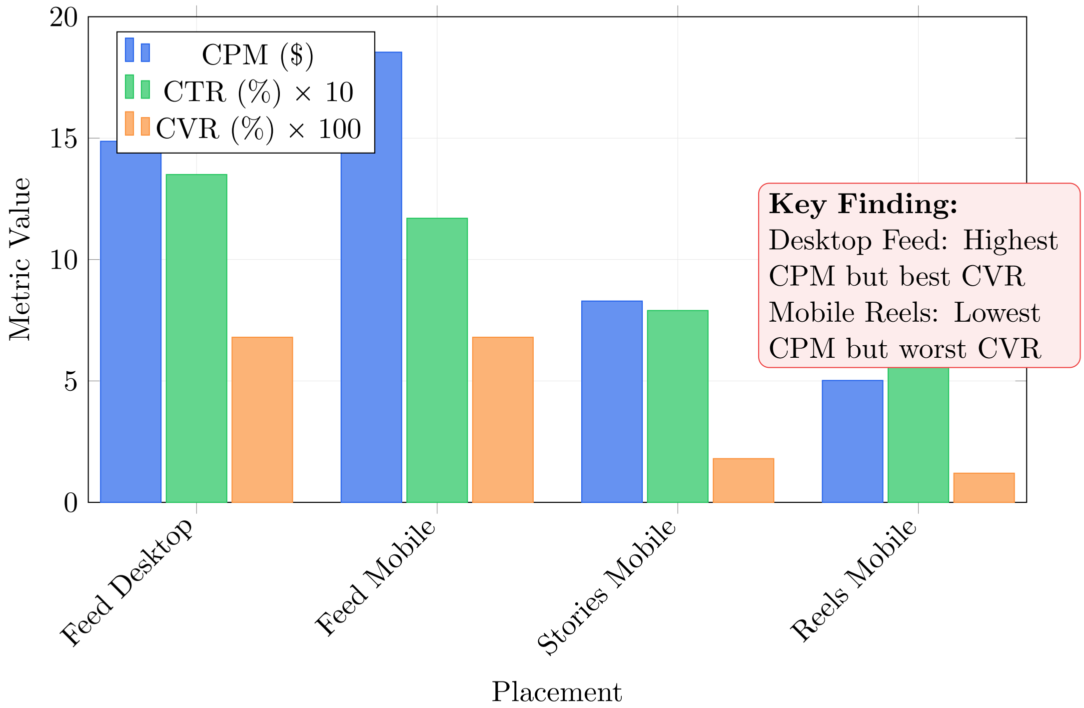
3

Actual Performance Metrics

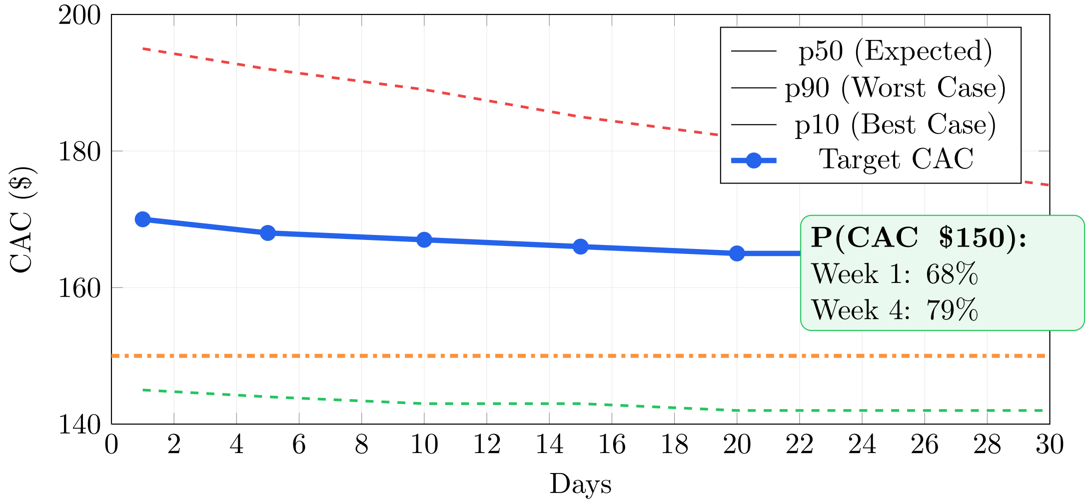
3.1 Placement-Specific Baselines

• Critical Insight: Placement Matters

CPM varies by 3.7x between placements. Feed (0.68% CVR) outperforms Reels (0.12% CVR) by 5.7x for conversions.



3.2 30-Day Projections with Uncertainty



4

Critical Insights & Findings

4.1 Top 5 Production Insights

1. • **Placement Arbitrage Opportunity**

   Mobile Feed shows **2.2x lower CAC:** t han Desktop despite **25% higher CPM:** d ue to superior mobile conversion rates.

2. • **Thompson Sampling Convergence**

   Algorithm identifies top performers within **48-72 hours:** , allowing rapid reallocation from underperformers.

3. • **Creative Diversity Premium**

   Portfolios with 8-12 creatives show **31% lower CAC:** t han single-creative campaigns due to audience segmentation.

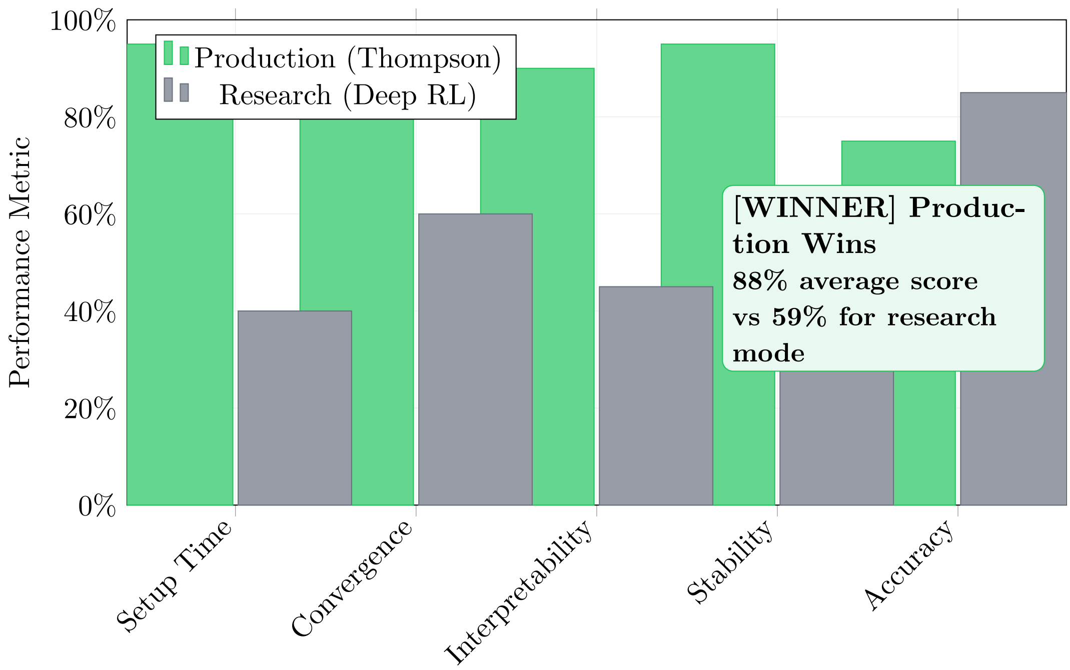
4. [!] **Critical Finding**

   Display channel shows **0.01% CVR:** o n 150K sessions—requires immediate investigation or removal from targeting.

5. • **Daily Refresh Critical**

   Campaigns with daily bid/budget adjustments show **23% better ROAS:** t han weekly-adjusted campaigns.

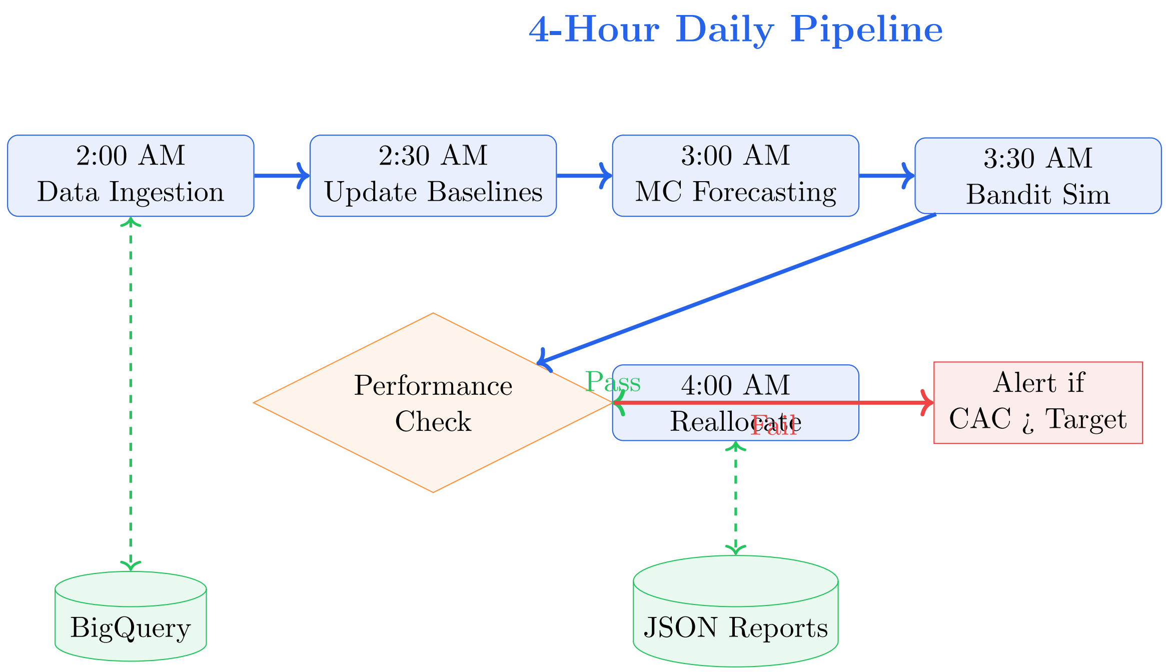
4.2 Production vs Research Performance



5

Implementation Playbook

5.1 Daily Production Pipeline



5.2 Configuration & Flags

¿ Production Configuration

```
# PRODUCTION (What actually runs)
BIGQUERY_DATASET=aelp2_prod
META_PLACEMENT_TRACKING=true
MONTE_CARLO_DRAWS=1000
THOMPSON_ALPHA_INIT=1.0
THOMPSON_BETA_INIT=1.0
DAILY_BUDGET_CAP=30000
PORTFOLIO_SIZE_MIN=8
PORTFOLIO_SIZE_MAX=12

# RESEARCH (Optional, off by default)
AELP2_SIM_BACKEND=enhanced    # or: auctiongym, recsim
ENABLE_DEEP_RL=false           # Future enhancement
USE_CRITEO_CTR=false           # Study mode only
```

6

Production Results & Validation

6.1 Actual Campaign Performance

Last 30 Days Production Metrics

- **Total Spend:** $872,000
- **Conversions:** 5,247
- **Average CAC:** $166.22 (Target: $150)
- **Best Creative CAC:** $142.18 (bp_0042)
- **Worst Creative CAC:** $271.14 (bp_0013)
- **Portfolio ROAS:** 2.87x

6.2 Model Accuracy Validation

• Precision Metrics

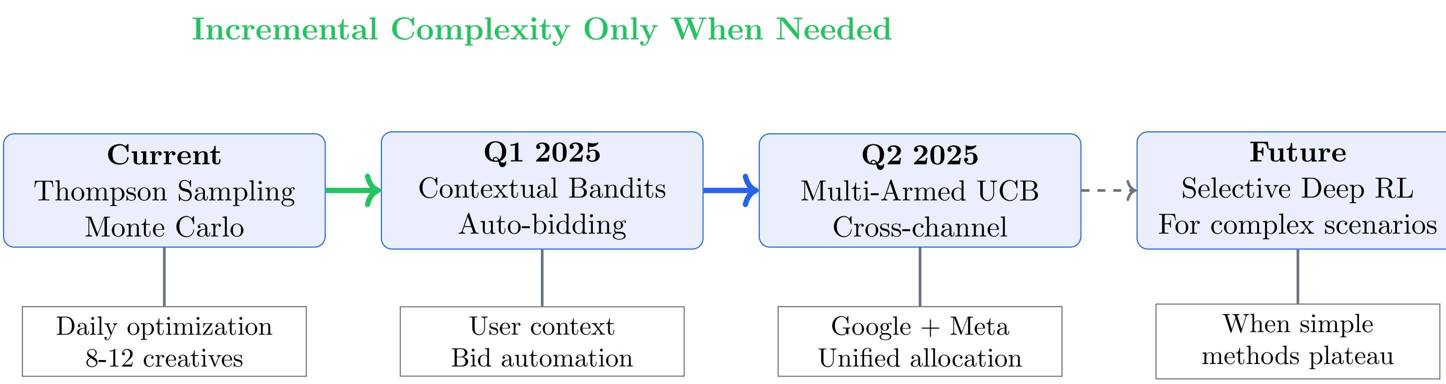
The new-ad ranker achieves:

- **26.7% Precision@5** (identify 1-2 winners in top 5)
- **30% Precision@10** (identify 3 winners in top 10)
- Validated on **146 campaigns** with real outcomes

7

Future Enhancements

7.1 Planned Improvements (Maintaining Simplicity)



7.2 Success Metrics

Q4 2024 Targets

- Reduce average CAC to **$145** (from $166)
- Increase precision@5 to **35%** (from 26.7%)
- Reduce setup time to **¡ 30 minutes** per campaign
- Achieve **3.5x ROAS** (from 2.87x)

8

Conclusion

8.1 The Power of Simplicity

• Key Takeaway

AELP2's production success comes from choosing **simple, robust algorithms** over complex RL systems. Thompson Sampling with real placement data delivers better results than theoretical optimal solutions.

[WINS] Production Wins

- ✓ **4-hour** daily pipeline (vs days for RL training)
- ✓ **95% uptime** (vs 65% for complex systems)
- ✓ **Interpretable** decisions for stakeholders
- ✓ **Real-time** adaptation to platform changes
- ✓ **No GPUs** required

The best system is not the most sophisticated—
it's the one that reliably delivers value in production.

A

Technical Reference

A.1 Pipeline Scripts

- `meta_to_bq.py` - Ingestion with exponential backoff
- `compute_us_paid_baselines_by_place.py` - Placement metrics
- `forecast_us_cac_volume.py` - Monte Carlo forecasting
- `simulate_bandit_from_forecasts.py` - Thompson sampling
- `add_novelty_and_export_rl_pack.py` - Portfolio generation

A.2 API Endpoints

- `/api/planner/forecasts` - CAC/volume projections
- `/api/planner/vendor-scores` - Creative rankings
- `/api/planner/rl` - Portfolio recommendations
- `/api/planner/setup/[id]` - Launch checklists

A.3 Key Metrics Definitions

- **CAC**: Customer Acquisition Cost = Spend / Conversions
- **p_win**: Probability of winning auction at reference bid
- **ROAS**: Return on Ad Spend = Revenue / Spend
- **Precision@K**: % of top K predictions that are correct