Building Digital Twins for Marketing Concept Evaluation

Overview

This memo outlines an end-to-end approach for creating digital twins that can evaluate marketing concepts with believable "yes/maybe/no" responses. After researching the subject extensively using Claude's deep search function (accumulating knowledge from hundreds of websites), I found the subject extremely wide with many different approaches. For this assignment, I decided to focus specifically on social media based ads (Facebook, Instagram, TikTok).

Data Collection: The Simulated Feed Approach

First, I wanted to tackle the interview process, as it is the source of our data and we can then base our model and predictions on it. A standard interview would be too direct, surveys and interviewing people on things they know would have significant bias. I drew this parallel from startup customer interview guides where most sources recommend interviewing people about products in person and trying to make it feel the least possible like an interview, asking about the product indirectly for best results.

My first thought was to track phones through an app monitoring all social media behavior, but due to user data being a sensitive subject, I decided to opt for a simpler solution: make a replica of feed-based social media platforms. The app should look like any other social media app where you scroll down and go through posts and videos as you would on Instagram. The contents of these feeds should be as close as possible to real-life feeds. We add test ads to this feed that should be as diverse as possible—from the format and fonts to the products marketed themselves, appearing naturally as they would on a regular platform.

The key feature here is that we can freely track user interaction with the feed: time spent on each item, whether they've gone through multi-page posts, and engagement patterns. This lets us build up the user's persona by their interaction with the feed, but also have direct marketing insights as we track direct impact of marketing material on that persona, learning about their affinities with every aspect of the ad (the more diverse the ad content in terms of variance, the better the understanding).

Interview Process: The user is initially only aware they're taking part in a study about human interaction with social media (ensuring the user doesn't have a goal and doesn't fake the test). After a 10-minute traditional survey covering demographics, the user is presented with a phone containing our feed. They have a fixed 30 minutes to scroll through (fixing time rather than feed length ensures users don't sprint through to finish faster).

Data Structure: We create a comprehensive category system for posts (e.g., an "Elon Musk meme" = [politics, funny, US]; "cooking video" = [cooking, family, food]). We can add format categories like [image, reel, video]. This gives us a matrix where each row represents a post and each column is a binary category indicator. We then track each user's interaction time, scroll patterns, and engagement with each post, creating an intersection graph of metrics for each user and content type.

From Interaction Data to Personas

Based on state-of-the-art research, K-means clustering is the most popular method (19.5% of data-driven persona studies), followed by non-negative matrix factorization (16.9%). Our approach combines multiple techniques:

1. **Initial Clustering:** K-means on user-feature matrices (user × category × avg_engagement_time)
2. **Dimensionality Reduction:** PCA or t-SNE to visualize high-dimensional behavioral patterns

3. **Hybrid Enhancement:** Exit interviews asking users to explain 3-5 of their longest interactions
4. **Dynamic Personas:** Create ML-powered twins that update with new data

Twin Logic & Decision Making

When evaluating new concepts, digital twins extract features across multiple dimensions:

- Content features: categories, emotional tone, visual style
- Technical features: format (image/video/reel), duration, complexity
- Marketing features: CTA type, urgency indicators, brand positioning

The ML pipeline uses an ensemble approach:

1. **Similarity scoring:** Compare new ad features to historical engagement patterns
2. **Predictive model:** XGBoost for engagement probability prediction
3. **Sentiment generation:** LLM (GPT-4) generates natural language explanations

Scaling Strategy

From 10 to 1000 users: Deploy as downloadable Progressive Web App. Fix main categories upfront; subcategories auto-prune based on statistical significance. Update personas through batch processing (weekly retraining).

Validation: Use 80/20 split—train on 80% of user-ad interactions, validate on remaining 20%.

Future Improvements & Alternative Approaches

Several enhancements could strengthen this system. First, we could personalize the feed for each user based on their initial questionnaire responses, ensuring we capture their full behavioral range rather than showing identical content to everyone. This would help reveal preferences that might not surface in a generic feed.

The category matrix could evolve from binary (0/1) to continuous values between 0 and 1, capturing the intensity of category membership (a post that's "somewhat funny" vs "extremely funny"). This nuanced approach would better reflect the multi-dimensional nature of content.

We could also investigate collaborative filtering algorithms similar to Netflix's recommendation system. By predicting user ratings for posts they haven't seen, we can fill gaps in our behavioral data and create more robust personas even with limited interaction history. This would be particularly valuable for understanding responses to novel ad formats or emerging content categories.

Additional paths worth exploring include incorporating eye-tracking for attention measurement beyond time spent and A/B testing different feed algorithms to understand how platform mechanics influence behavior.