Thesis Proposal: AutoMMM - An AI-Powered System for Market Mix Modeling

Prateek Nigam

June 2025

Abstract

This thesis will develop AutoMMM, an AI-powered system to analyze marketing data, like ad spend and sales, automatically. Using Python, LangChain, and LangGraph, AutoMMM will use multiple AI agents to process data, run models, and suggest better marketing strategies. It will generate synthetic data to test the system and validate its performance. This report explains the problem, goals, methods, expected results, and a two-month plan.

1 Introduction

Market Mix Modeling (MMM) helps businesses see how marketing efforts, like ads and promotions, drive sales. It guides smart budget decisions to maximize return on investment (ROI) [13]. Traditional MMM is slow and needs a lot of manual work, which struggles in today's fast digital marketing world [1].

Artificial Intelligence (AI) can improve MMM by automating tasks and finding data patterns [6]. However, there are few systems that fully automate MMM using multiple AI agents. AutoMMM will solve this by using Python, LangChain, and LangGraph to analyze ad spend, sales, and external factors, delivering clear business advice.

The problem is the lack of fully automated MMM systems with multi-agent collaboration. The goals are to:

1. Build an automated MMM system with AI agents.

2. Test it using synthetic data and validate performance.

3. Add new ideas to AI in marketing analytics.

2 Literature Review

2.1 Understanding Market Mix Modeling (MMM)

Market Mix Modeling (MMM) is a method that uses statistics to understand how different marketing actions—like advertising, promotions, and pricing—affect product sales. It often uses regression models and includes features like:

- **Adstock:** The effect of ads continues over time.

- **Saturation:** Spending more gives less extra benefit after a point.

- **Seasonality:** Certain times of year affect sales more than others.

Modern MMM tools like Adobe Mix Modeler and NextBrain AI use machine learning to make this process faster. But they still need people to clean the data, choose features, and understand the results [19, 1, 7].

2.2 What Are Multi-Agent Systems (MAS)?

A Multi-Agent System (MAS) is a group of AI agents that work together. Each agent can do its own task but also talk to other agents. MAS systems are good for handling complex problems where many things happen at once [20, 9].

2.2.1 MAS Structures

There are two common ways to build MAS:

- **Hierarchical:** One supervisor agent manages others.

- **Decentralized:** All agents work independently but still coordinate.

These designs help divide work and manage big projects, which fits well with a system like AutoMMM.

2.2.2 MAS in Marketing and Similar Areas

MAS has not been used directly in MMM yet, but it has been helpful in related areas:

- **Marketing simulations:** MAS is used to model how customers behave and respond to marketing [16, 5].

- **Advertising tools:** Platforms like AdSim use MAS to test how ad spend affects ROI by simulating customer behavior [3].

- **Social campaigns:** MAS helps understand how ideas or products spread in networks [14].

2.2.3 MAS in Other Fields

MAS has also been used in:

- **Supply chains:** Simulating suppliers, factories, and deliveries.

- **City systems:** Modeling how people move and interact in cities [11, 18].

These examples show how MAS can help in large, complex systems like MMM.

2.2.4 Tools for Building MAS

There are several frameworks and tools that help build MAS, such as:

- langchain

- openAI

- CrewAI

- Agno

These tools support agent messaging, workflows, and system design [15, 10].

2.3 Key Insights

From this review, we can see:

- MMM still relies heavily on regression and needs human support.

- MAS is used in marketing and other fields but not yet in MMM.

- MAS features—like automation and teamwork—match well with AutoMMM's design.

This shows that AutoMMM is filling a real gap by combining MMM and MAS in a new, automated way.

MMM and MAS are both strong on their own, but they have not yet been combined. AutoMMM will be one of the first systems to do this, creating a smart and automated way to analyze marketing performance.

3 Methodology

3.1 System Design

AutoMMM is designed as a multi-agent system. It uses LangChain and LangGraph to let agents talk to each other and work together. The system is managed by a special agent called the **Supervisor Agent**. This agent controls the overall workflow and makes sure each step is done in the correct order. Each agent has a clear job. They pass results to the next agent in a smart way, depending on the outcome. The Supervisor Agent watches the whole process and can tell an agent to repeat a task or fix a problem if needed.

- **Data Team**:

  - **Data Processing Agent**: Will run Python scripts for exploratory data analysis (EDA) on ad spend, sales, and external factors, creating Excel reports with charts and notes on data issues like missing values or outliers.
  - **Insight Agent**: Will summarize key performance indicators (KPIs) and data stories using LangChain to guide modeling.

- **Modeling Team**:

3

Rough Workflow for MMM agents

— Prateek Nigam
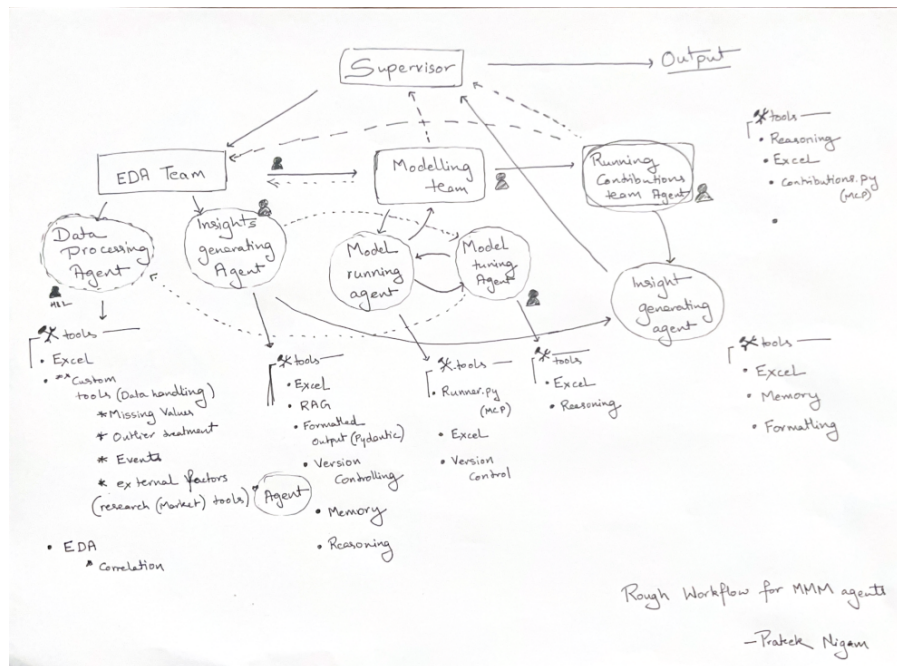
   – **Model Running Agent**: Will run a simple regression model (MCP) to predict sales, saving model summaries and graphs (PNG files).

   – **Model Tuning Agent**: Will adjust hyperparameters using insights from the Data Team to improve accuracy.

- **Contribution Agent**: Will run `run_contributions.py` to calculate marketing contributions using the Shapley method, producing an Excel file.

- **Final Suggestions**: Will turn results into clear business advice, like optimizing ad spend for better ROI.

How Agents Interact

The system works in steps:

1. The Supervisor gives data to the Data Processing Agent.

2. After cleaning, results go to the Insight Agent.

3. Then the model is run. If it is not good, the Supervisor calls the Tuning Agent.

4. Once the model is good, the Contribution Agent finds which ads worked best.

5. Finally, the Recommendation Agent gives business advice.

If any step fails or needs improvement, the Supervisor can go back and repeat it. This makes the system flexible and reliable.

Parallel and Reusable Steps

Some tasks can run at the same time, like testing different models. The system can also reuse previous results, making it fast and efficient.

This setup follows a popular design called "orchestrator-worker," where one main agent manages many workers [12, 2].

3.2 Synthetic Data Generation

Synthetic data helps test our model in a safe and controlled way. It acts as a "gold standard" because we know the correct answers ahead of time. This is important in marketing mix modeling (MMM), where collecting real-world causal data (like from randomized experiments) is expensive and slow [4].

Instead of running real tests, we simulate realistic sales data using known values. This lets us check if the model gives back the correct results—just like how tests work in science [8].

To test AutoMMM, I will generate synthetic data that mimics real marketing data. The process will include:

- **Data Structure**: Create 104 weeks (2 years) of weekly data for ad spend (e.g., TV, digital), sales, and external factors (e.g., holidays, seasonality).

- **Realistic Patterns**: Add positive correlations between ad spend and sales, diminishing returns (saturation), and carryover effects (adstock) using decaying functions and logistic transformations.

- **Noise and Outliers**: Include Gaussian noise and occasional outliers to simulate real-world data challenges.

- **Ground Truth**: Set known coefficients (e.g., 0.3 for TV, 0.4 for digital) to compare with model outputs.

- **Config File**: Use a YAML config to define parameters like adstock decay, saturation, and noise levels for reproducibility.

The Data Processing Agent will handle this synthetic data, ensuring it's ready for modeling.

This synthetic setup helps in three ways:

- **Model Testing:** We can check if the model correctly recovers the true input values (coefficients).

- **Fast Feedback:** We avoid the delay and cost of real-world experiments.

- **Safe and Scalable:** No real user data means no privacy issues [17].

3.3 Model Validation

To check AutoMMM's performance, I will:

- **Compare Coefficients**: Compare model-estimated coefficients to ground truth coefficients from the synthetic data to measure accuracy.

- **Evaluate Metrics**: Use metrics like Mean Absolute Error (MAE) and R-squared to assess prediction quality.

- **Test Robustness**: Validate the model on noisy data and outliers to ensure it handles real-world challenges.

- **ROAS Analysis**: Check if the Contribution Agent's outputs align with expected return on ad spend (ROAS) based on ground truth.

The Supervisor Agent will ensure all agents work together to validate results.

3.4   Implementation Details

- **Tools**: Python for coding, LangChain for agent logic, LangGraph for workflows.

- **Modeling**: Use a simple regression model (MCP) for sales prediction.

- **Contributions**: Calculate using `run_contributions.py` with the Shapley method.

- **Data Generation**: Create synthetic data with realistic MMM patterns.

Table 1: AutoMMM Agent Jobs

Agent	Job
Supervisor Agent	Manages agents using LangGraph
Data Processing Agent	Analyzes data, generates Excel reports
Insight Agent	Summarizes KPIs using LangChain
Model Running Agent	Runs MCP model, saves results
Model Tuning Agent	Improves model settings
Contribution Agent	Calculates contributions with `run_contributions.py`
Final Suggestions	Gives business advice

4   Expected Results

4.1   Academic Results

- Develop a new way to automate MMM with AI agents.

- Create a framework for multi-agent systems in marketing analytics.

4.2   Practical Results

- Build a tool that automates MMM, saving time.

- Provide clear advice to improve marketing and boost profits.

5 Timeline

The project will take two months:

- **Week 1**: Study past work and design system.

- **Weeks 2-3**: Build Data Team and synthetic data generator.

- **Weeks 4-5**: Develop Modeling Team.

- **Week 6**: Create Contribution Agent.

- **Week 7**: Test system and validate performance.

- **Week 8**: Finalize testing and write report.

Table 2: Project Timeline

Task	Timeline
Study and System Design	Week 1
Data Team and Synthetic Data	Weeks 2-3
Modeling Team	Weeks 4-5
Contribution Agent	Week 6
Testing and Validation	Week 7
Final Testing and Report	Week 8

References

[1] Ruler Analytics. What is marketing mix modeling?, 2024. Accessed: 2025-06-16.

[2] Anthropic. How to build a multi-agent system with a supervisor pattern. Online, 2023. https://www.anthropic.com/index/multi-agent-orchestration.

[3] AnyLogic. Marketing analytics with adsim simulation, 2023. Accessed: 2025-06-16.

[4] Ron Berman and Christine Moorman. Incrementality testing in marketing: The gold standard of measurement. *Journal of Marketing*, 86(4):1–20, 2022.

[5] Sebastiano A. Delre, Wander Jager, Marco A. Janssen, and Tammo H. A. Bijmolt. Will it spread or not? the effects of social influences and network topology on innovation diffusion, 2007.

[6] DiGGrowth. The role of ai in marketing mix modeling. https://diggrowth.com/blogs/marketing-mix-modeling/the-role-of-ai-in-marketing-mix-modeling/, 2024.

[7] DiGGrowth. The role of ai in marketing mix modeling, 2024. Accessed: 2025-06-16.

[8] Mengnan Du, Ninghao Liu, Xia Hu, et al. Synthetic data generation for machine learning in healthcare. *Nature Machine Intelligence*, 2(10):442–450, 2020.

[9] Stan Franklin and Art Graesser. Is it an agent, or just a program?, 1997. Available online, often cited for defining MAS properties.

[10] Arnaud Grignard, Patrick Taillandier, Benoît Gaudou, Duy Cuong Vo, Nhan Huynh, and Alexis Drogoul. Gama 1.6: Advancing the art of complex agent-based modeling and simulation, 2017.

[11] Nitin Julka, Ramesh Srinivasan, and Iftekhar A. Karimi. Agent-based supply chain management—1: Framework, 2002.

[12] LangChain Team. Langgraph documentation. Online, 2024. https://docs.langchain.com/langgraph.

[13] LatentView. A complete guide to marketing mix modeling. https://www.latentview.com/marketing-mix-modeling/, 2024.

[14] Charles M. Macal and Michael J. North. Agent-based modeling and simulation: Desktop abms, 2007.

[15] Steven F. Railsback and Volker Grimm. Agent-based and individual-based modeling: A practical introduction, 2011.

[16] William Rand and Roland T. Rust. Agent-based modeling in marketing: Guidelines for rigor, 2011.

[17] Tanya Sharma. Synthetic data and privacy: How artificial data helps protect users. https://getrecast.com/synthetic-data-and-privacy. Accessed: 2025-06-16.

[18] Cheng Tao and Yisong Yue. Urban simulation using multi-agent systems: A case study, 2022. Available in urban systems and smart city modeling literature.

[19] Wikipedia contributors. Marketing mix modeling, 2024. Accessed: 2025-06-16.

[20] Wikipedia contributors. Multi-agent system, 2024. Accessed: 2025-06-16.