

Exploring LLM Biases in Hotel Booking

Kevin Nathanael Hamonangan, Yugam Surana, Olivia Li

Under the Supervision of Professor Ali Hirsa, Miao Wang, Jim Strugger and Bas Jasper from Morgan Stanley

Abstract and Objective

This project investigates whether large language models (LLMs) exhibit bias toward **aggregator booking platforms** (e.g., Booking.com, Expedia) versus **direct hotel websites**. We design a structured **six-factor prompts framework** (City, Travel Type, Budget, Amenities, Role preface, and Prompt Style) yielding 810 distinct booking scenarios.

Prompts are queried to **Perplexity** and **Gemini**, and responses are parsed to extract and categorize URLs. The framework allows us to observe how different prompt formulations influence link type selection, revealing model tendencies in real-world booking contexts.

LLM Prompts Generation

Design: Deterministic factorial grid across **6 factors** to create controlled booking scenarios: **City** (New York, Boston, Portland) × **Travel Type** (Solo, Business, Family) × **Budget** (Low, Medium, High) × **Amenities** (None, Pool+Gym, Wi-Fi+Breakfast) × **Role** (system preface on/off) × **Prompt Style** (5 different wording of the prompts)

Sample prompt:

“You are an expert travel booking assistant. I’m planning a **Business** trip to **Boston**. My budget is **Medium** relative to average hotel prices in Boston. I also prefer amenities like **free Wi-Fi and breakfast**. Please recommend 5 hotels and include a link to more info or the site you used, if possible.

Extraction Pipeline

Stage 1 - Input Preparation:

Structured prompts generated from the six-factor framework.

Stage 2 - Model Querying:

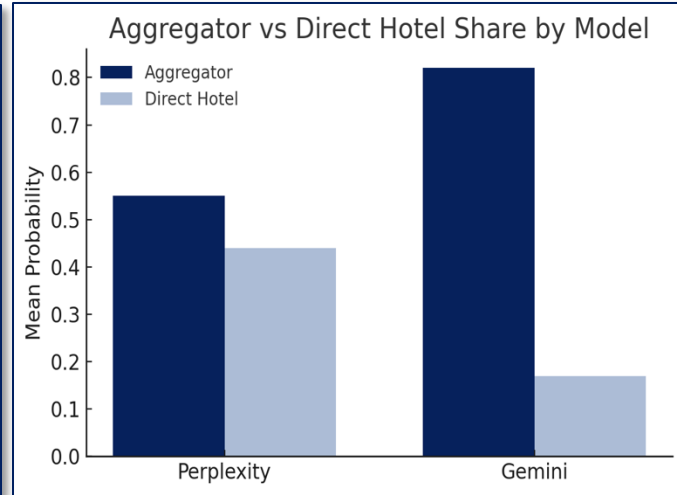
Prompts queried to LLMs (Perplexity, Gemini) using their respective APIs.

Stage 3 - URL Extraction:

Responses parsed to identify and normalize booking links.

Stage 4 - Output Storage:

Links categorized as **direct hotel sites** vs **aggregators**, with results logged for analysis.



Output Analysis

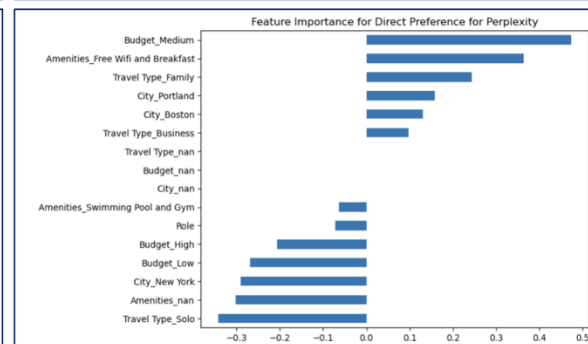
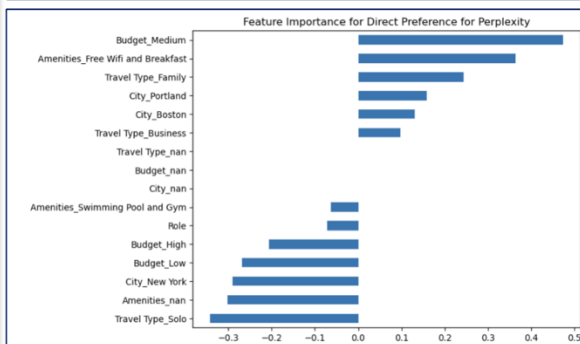
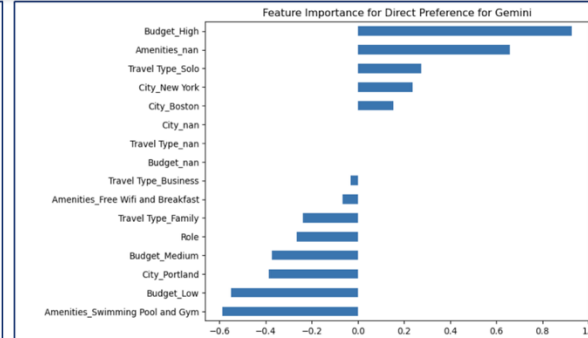
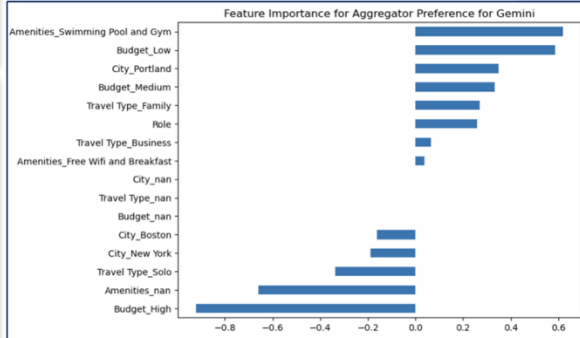
Direction of Bias:

- **Gemini** skewed toward **aggregator platforms**.
- **Perplexity** also skews **aggregator platforms** but less strongly compared to Gemini.

Underlying reasons:

- **Perplexity** supports **active online search**, enabling it to retrieve the most current official hotel pages. This real-time access allows it to surface more direct hotel websites, reducing (but not eliminating) its aggregator bias.
- **Gemini**, by contrast, relies more on its **comparative training data** (heavily populated with aggregator domains such as Booking and Expedia). This explains its stronger **aggregator bias**.

Certain conditions (e.g., high-budget prompts or leisure-oriented city types) increased the odds of aggregator lean, but overall Perplexity remained more balanced, with relatively stronger weights for direct links than Gemini. Across most categories, Gemini shows **positive coefficients for aggregator leaning**. Direct-leaning coefficients are weaker, indicating a systematic bias toward aggregator citations across conditions.



Recommendation and Next Steps

Expand models: Include ChatGPT, Claude, and better reasoning models of Perplexity/Gemini.

Broaden Prompts: Test more factors like trip duration, traveler profiles, more cities,

Vary Prompting styles: Test directive vs casual, minimalist vs detailed, and strict vs flexible phrasings.

Richer evaluation: Assess trustworthiness, recency, diversity, and redundancy of links.

Practical impact: These extensions will further study how prompt design drive LLM output in real-world scenarios.