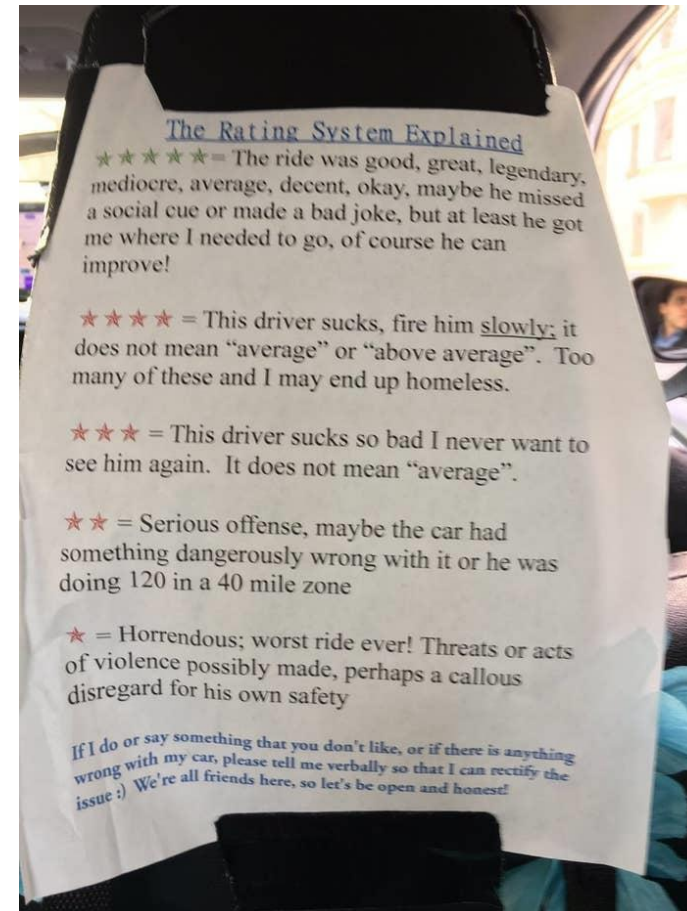


# Large Language Models in Marketing

Instructor: Davide Proserpio

# A few things

- Course evaluations open November 30



# A few things

- Presentations (45% of the project grade): Dec 1 and 3
  - Due at midnight of November 30
  - 16456 (12:30 pm session): 8 groups
  - 16457 (2 pm section): 10 groups
- Presentation Time: 12 mins + 3 Q&A
- Final project doc (notebook + pdf) due Dec 3 (40% of the project grade)
- Peer evaluations (multiplier between 0.9 and 1 so your project grade can decrease by as as much as 10%)
  - Due in class (Zoom) on Dec 12
  - [Form](#)

# What is a Large Language Models (LLMs)?

- A **neural network** trained to predict the next token in a sequence
- Learns patterns from massive amounts of text
- Can now:
  - Understand natural language
  - Generate new content
  - Extract insights
  - Reason (imperfectly)
- Examples: GPT-4/5, Claude, Gemini, Llama

# Large Language Models (LLMs)

- AI is shifting from **prediction** to **generation**
- LLMs power applications across:
  - Customer service
  - Creative development (e.g., ads)
  - Segmentation & personalization
  - Consumer insights
  - Ad performance & measurement
- Marketing teams increasingly need **AI fluency**, not engineering skills

# What Marketers can do with LLMs?

- **Customer Insights**

- Summarizing thousands of reviews
- Extracting feature sentiment (e.g., “battery life complaints”)
- Topic detection in open-text surveys (NPS)
- Simulate customers

- **Customer Experience & Support Automation**

- Chatbots (e.g., customer service)
- Assistants

- **Advertising**

- Generating ad copy variants
- Dynamic creative optimization
- Understanding search queries at scale

What else am I missing?

# Four use cases about LLMs

1. Are LLMs that useful? How much human prompting matter?
2. Brand optimization
3. Market research (surveys, conjoint analysis)
4. Advertising



# Are LLMs really that useful?

Study: Generative AI results depend on user prompts as much as models

## **Experiment:**

- 1900 participants assigned to use DALL-E1, 2, or 3
- Participants were shown a reference image and asked to re-create it by typing instructions into the AI
  - They had 25 minutes to submit at least 10 prompts
  - They were told that the top 20% of performers would receive a bonus payment, which motivated them to test and improve their instructions.

# Are LLMs really that useful?

- Upgrading to a more advanced generative-AI model (in the study, moving from DALL-E 2 to DALL-E 3) **only explains about half** of the performance uplift.
- The **other half** comes from improved user prompting: better prompt length, clarity, and iteration.
- Interestingly: when prompts were automatically rewritten by an AI (without user's full control), performance actually **fell by ~58%** compared to manual prompt-writing

# Are LLMs really that useful?

- Investment in model upgrades alone isn't enough: firms must invest in user training, interface design, and iterative learning for prompting.
- Prompting is less about technical coding ability and more about clear communication. Even non-tech users improved performance substantially.
- Caution:
  - Automation of prompt rewriting (to help users) may backfire if it misaligns user intent or adds unwanted detail.
- For marketing teams:
  - embed prompting best-practices into operations (creative generation, ad copy, segmentation tasks) and treat prompt-refinement as an analytics process in its own right.

# Brand optimization

## Forget What You Know About Search. Optimize Your Brand for LLMs

- Consumers are shifting away from traditional search engines toward generative AI platforms
- In a survey of 12,000 consumers, **58%** reported using Gen AI tools for product/service recommendations (vs. 25% in 2023).
- Key takeaway: The digital consumer journey is changing.
  - it's no longer about keyword search → website visit → purchase
  - t's moving to AI-mediated dialogue and recommendation.

# Brand optimization

- Brands must shift from optimizing for clicks/keywords (traditional SEO) to optimizing for **resolution** (i.e., solving user tasks/questions with clarity and authority) rather than just attention.
  - Brands want to be cited by LLMs

# Brand optimization

- Strategic guidelines:
  - Create content that addresses use-cases (e.g., “best EV for winter driving”) rather than generic branding.
  - Provide structured, expert-backed information (trust signals, clear feature/use-case focus) for LLMs.
  - Recognize each LLM has its own “lens” of what it values (e.g., unique features, flexibility, local options) and tailor strategy accordingly.

# LLMs for Market Research

- [Using LLMs for Market Research” by Brand, Israeli, and Ngwe](#)
- **Can Large Language Models (LLMs) replace or augment traditional market research?**
- The authors explore whether LLMs can:
  - Generate **realistic consumer preference data**
  - Produce **Willingness to Pay (WTP) estimates** similar to real consumers
  - Reflect **differences across customer segments**
  - Improve with **fine-tuning using past survey data from real consumers**

# LLMs for Market Research

- **LLMs can mimic human average preferences**

- Across several categories (toothpaste, deodorant, laptops, tablets), GPT produced **WTP estimates close in sign and magnitude** to human surveys.
- In some cases, LLMs' WTP was remarkably similar to real-world benchmark studies.

- **But performance is uneven**

- GPT frequently **mis-estimated new or unfamiliar attributes** (e.g., “pancake flavor” toothpaste, laptop projectors).
- Different models (GPT-3.5, GPT-4o, Claude, LLaMA) produced **different preference curves**, showing **model instability**.



# LLMs for Market Research

- **LLMs struggle with heterogeneity**
  - GPT was **poor at capturing segment-level differences** (income, gender, race, politics).
  - It often reproduced the overall population's average preferences rather than group-specific patterns.

# LLMs for Market Research

- **How marketers should use LLMs**
  - **Not a substitute for human surveys**, but a powerful **early-stage simulator**:
    - Rapidly test new features
    - Screen ideas before expensive human studies
    - Explore preference ranges and sensitivity
- Consider LLMs as “synthetic consumers” — useful for ideation, not decision-finalization.

# LLMs for advertising

[Applying Large Language Models to Sponsored Search Advertising](#)  
by Martin Reisenbichler, Thomas Reutterer & David A. Schweidel

- Can large language models (LLMs) be applied to generate ad copy for sponsored search and improve performance compared with human-only content?
  - Context: Search advertising is huge (~\$100B in U.S. ad spend in 2023) and content (ad text) is under-studied compared to bidding.

# LLMs for advertising

- Authors build a human-in-the-loop framework
  - LLM + info about target keyword, the landing page, and top organic results for those key words
  - Use an LLM to generate ad text that integrates keywords, semantic fit to landing page and organic results
  - Compute a “quality score” of each generated ad copy (based on semantic similarity to webpages + keyword integration) to pick best pieces.
- Empirical test: Two field experiments: one with a B2C higher-ed campaign (education sector) and one B2B (IT & SaaS) campaign.

# LLMs for advertising

- **Productivity:** LLMs reduced human time by ~60% for generating 208 ads for 208 keywords ( $\approx 18.56$  hours saved) in their experiment
- **Performance:**
  - More impressions, clicks, ad quality,
  - Cost advantage only in the low budget scenario

# LLMs for advertising

- Implication for marketers:
  - LLMs can be a **scalable tool** to generate keyword-specific ad copy quickly and cost-efficiently
  - Valuable for firms with **limited budget**.
  - But they need a **business application layer**: keyword data + landing page context + quality scoring
    - simply using “vanilla” LLM outputs may not suffice.
  - Importance of **holistic optimization**: LLMs can help if given the right context.