

Using GPT for Market Research

Ayelet Israeli

Harvard Business School

May 2024

with James Brand and Donald Ngwe

Office of the Chief Economist, Microsoft

GPT can simulate humans

- Replicates results of classic experiments in social psychology and economics (Aher et al., 2022)

Research shows how inputs impact GPT outputs:

- Partisan text generation and voting behavior, based on ideology inputs (Argyle et al., 2022)
- Results of classic behavioral economic experiments, with heterogeneity based on “beliefs” (Horton 2023)



GPT may be useful for social science research

Research Questions

Can GPT be used as a tool for market research?

- A. Are GPT responses aligned with fundamental properties of consumer demand?
- B. Are GPT's responses (about demand) realistic and meaningful?

How can we learn about customer preferences?

- To conduct market research, we typically utilize data from many (randomly sampled) consumers
 - Focus groups, surveys, observational studies
 - Costly, time-consuming, labor-intensive, ideal data may not exist
- First question: can we use GPT to emulate this type of data collection?
- Answer: we can query GPT hundreds of times for each bundle of goods we consider, generating responses from a “random sample of consumers”
 - We maximize variation using GPT’s randomness input (“temperature”)
 - Training data likely include product websites and UGC such as blogs, reviews

Our prompts aim to emulate a customer survey

- Ask GPT to fill in the response to a survey question as if it were a customer that was shopping in the category of interest and was randomly selected to participate in a survey
- Describe relevant features of the customer (e.g., annual income)
- Describe one or two products, as well as the option to opt-out
- Ask GPT to either make a choice or to provide WTP

We repeat each prompt hundreds of times to generate our sample

We design six studies to answer our research questions

A. Are GPT responses aligned with fundamental properties of consumer demand?

- 1) Downward-sloping demand curve
- 2) Impact of income on demand
- 3) State dependence / impact of past choice
- 4) Diminishing marginal utility of consumption

B. Are GPT responses (about demand) realistic and meaningful?

- 1) WTP for products
- 2) WTP for product attributes

Study A1: Does GPT generate a downward-sloping demand curve?

- Binary choice: laptop or no purchase
- Choice among multiple options: two goods versus no purchase

Study A1: We examine GPT's responses

“A customer is randomly selected while shopping for laptops. Their annual income is \$70,000.

While shopping, the customer has three options:

- a Surface Laptop 3, Price: **PRICE**, Processor: Intel Core i5, RAM: 8GB, Screen Size: 13.5in, SD: 128GB
- A Macbook Air (2019) Laptop, Price: \$999, Processor: Intel Core i5, RAM: 8GB, Screen Size: 13.3in, SD: 128GB

They also have the option not to purchase a laptop. The customer is asked, after they finish shopping: Which laptop, if any, did you purchase?

Customer:”

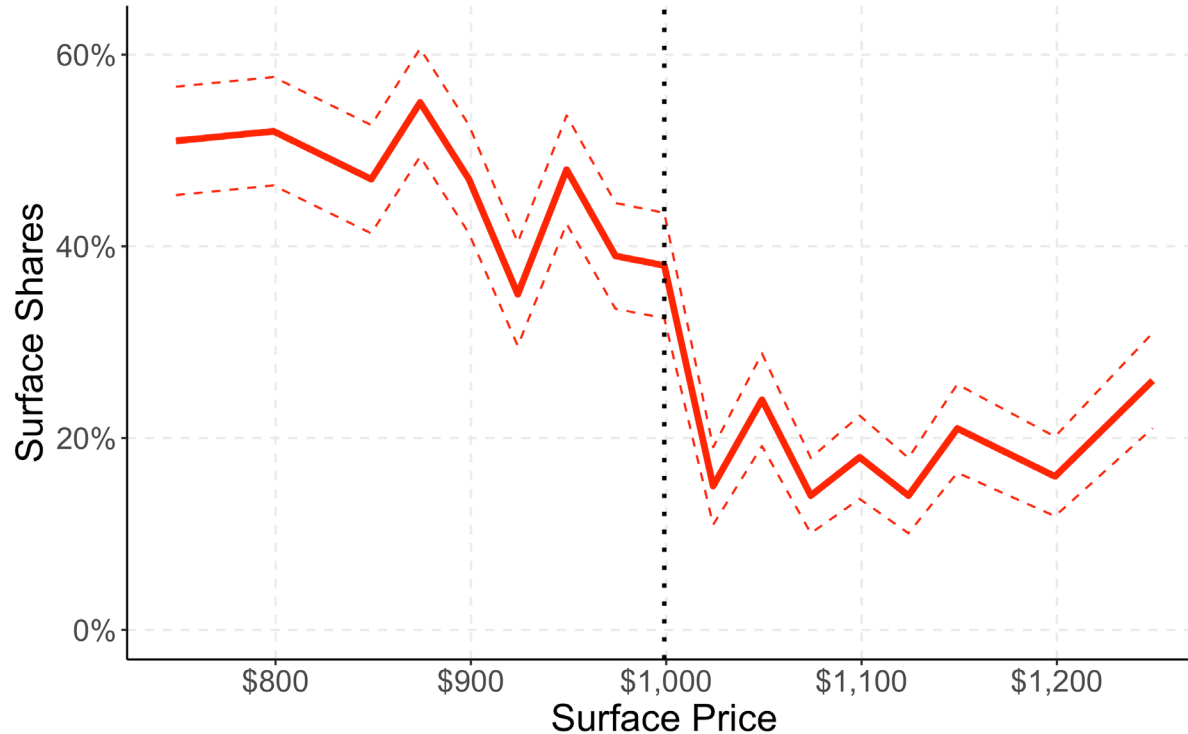
Where: **PRICE** ∈ [\$750...\$1,250]

Prompt for OpenAI's GPT API with the “text-davinci-003” model, also replicated with GPT 3.5 Turbo models

GPT generates hundreds of responses

- I purchased the Macbook Air (2019).
- I decided to purchase the Surface Laptop 3. It meets my needs and the price point was right.
- I decided to go with the Macbook Air (2019). I decided to go with this model since it had the same processor and RAM at a lower cost than the Surface Laptop 3.
- None. I decided not to purchase a laptop at this time. My current laptop works for me.
- None. I decided to wait until I can find a better deal.
- ...
- ...
- I purchased neither laptop. I decided to wait until I can save up more money to purchase a laptop with a better processor and more RAM.

We aggregate the responses and plot choice shares



INSIGHTS:

- GPT is price sensitive
- Context matters to GPT
- Reference price matters to GPT
- It has brand preferences
- GPT gravitates toward the first option presented

Study A1: We also examine supermarket goods

“A customer is randomly selected while shopping at the supermarket. Their annual income is \$70,000.

While shopping, the customer passes by the toothpaste aisle and sees two options:

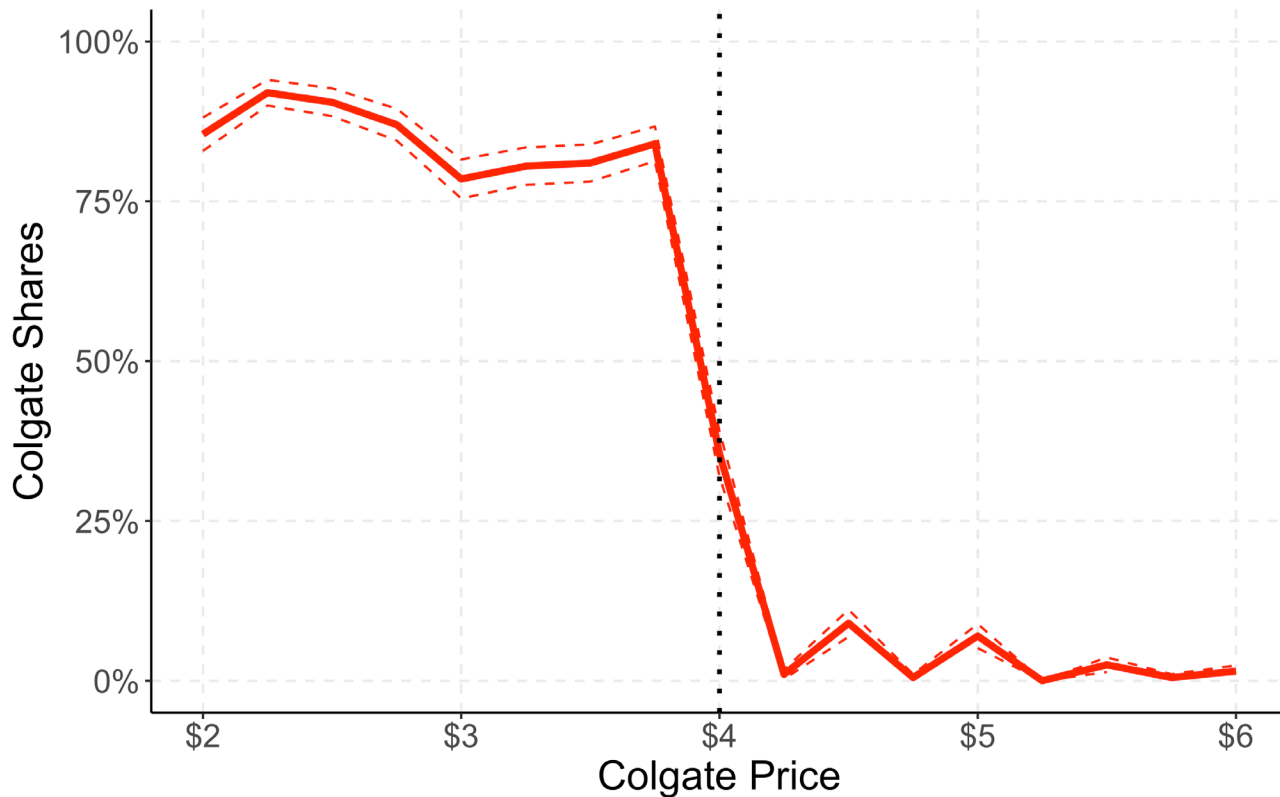
- Colgate whitening toothpaste with fluoride, price **PRICE**.
- Crest whitening toothpaste with fluoride, price \$4.

They also have the option not to purchase toothpaste. The customer is asked, after they finish shopping: Which toothpaste, if any, did you purchase?

Customer:”

Where: **PRICE** \in [\$2...\$6]

Multiple options results – Colgate (vs. \$4 Crest)



INSIGHTS:

- Toothpaste is not as differentiated as laptops for GPT
- GPT has diminishing sensitivity

Overall, GPT responses correspond with fundamental properties of consumer demand

- 1) Downward-sloping demand curve
- 2) Impact of income on demand
- 3) State dependence
- 4) Minimal evidence of diminishing marginal utility of consumption

Are GPT's responses realistic and meaningful?

- Study B1: Recovering realistic Willingness to Pay (WTP) for **products**
 - “...what is the maximum price you would be willing to pay for...?”
 - Shown in the paper: GPT provides reasonable distributions of WTP for multiple categories/products
- Study B2: Recovering realistic Willingness to Pay (WTP) for **attributes**

Study B2: Recovering realistic WTP for attributes

- Can we recover WTP for attributes of products?
 - Not likely to appear directly in GPT's corpus
 - Three approaches:
 - Direct solicitation (WTP prompts)
 - Indirect solicitation (via demand curves)
 - Conjoint analysis
- Comparing to Fong et al (JMR, 2024)'s results
 - In a different context, study WTP for fluoride in toothpaste, aluminum in deodorant
 - Conduct conjoint studies and thoroughly validate outcomes (prices, shares) with real-world data

WTP for attributes: recovering preferences via conjoint

- We test whether the effect of price and non-price attributes on choice probabilities is consistent with economic predictions
- We use the responses to estimate a multinomial logit to evaluate the realism of the model-based WTP
- Instead of having to derive a few choice sets which are orthogonal and balanced across configurations, we use the full set of options

Study B2: Recovering preferences for attributes via conjoint

“A customer is randomly selected while shopping at the supermarket. Their annual income is \$70,000.

While shopping, the customer passes by the toothpaste aisle and sees two options:

- Colgate whitening toothpaste **COLGATE_WITH** fluoride, price **COLGATE_PRICE**.
- Crest whitening toothpaste **CREST_WITH** fluoride, price **CREST_PRICE**.

They also have the option not to purchase toothpaste. The customer is asked, after they finish shopping: Did you purchase any toothpaste, if so, which one?

Customer:”

Where: **_PRICE** $\in \{0.99, 1.99, 2.99\}$; **_WITH** $\in \{\text{with}, \text{without}\}$;

We use a total of $6 \times 6 = 36$ combinations

Our results are consistent with human studies

- First, we corroborate that choices are consistent with economic theory and substitution patterns:
 - When Colgate price is low / Crest price is high: more likely to choose Colgate
 - When Colgate contains fluoride / Crest doesn't: more likely to choose Colgate
- We use the responses to estimate a multinomial logit to evaluate the realism of the model-based WTP
 - Our implied measure is \$3.4, consistent with Fong et al (2023) \$3.27.
 - We also confirm our approach for other products.

Comparing our approach to real-world conjoint studies

GPT-based conjoint	Real-world conjoint
Run all possible configurations	Run sample configurations, simulate additional configurations
1 GPT	100s+ humans
\$3 to collect all 10,800 responses 35 minutes total	\geq \$1 per participant \geq 5 minutes per participant
The script that parses responses and generates csv files takes a couple of more minutes to run and verify	Data collected immediately for each participant

Summary

- We used GPT as a random customer sampling mechanism and examined the distribution of its responses
- Our first set of results highlights that when prompted as a random customer, GPT exhibits behaviors consistent with economic theory
- Our second set of results demonstrates that GPT-based estimates are realistic and consistent with values obtained from existing research



GPT can serve as a powerful tool for understanding customer preferences

What does this mean?

- At a minimum, GPT can serve as a realistic simulator of consumer choice
- GPT can be provided with additional “knowledge” (product attributes, customer personas, representative demographics, proprietary data), or fine-tuning to generate results
- As LLMs advance, we expect them to become more useful

Limitations and words of caution

- We are still working on understanding how useful and generalizable this is
- GPT responses need to be examined critically in different contexts
- Limitations of the model and training data: Pre-trained; bias;
- Use of GPT should be disclosed for ethical and external validity reasons

Next steps: generalizability and usefulness

- Examining the boundaries of our approach
 - Evidence of boundaries for GPT (e.g., Goli and Singh 2024, Gui and Toubia 2023, Manning et al. 2024)
 - B2B vs B2C settings
- Validating our approach for new products / features
- Examining heterogeneity in responses due to different customer personas and demographics

Thank You!

aisraeli@hbs.edu

Questions / Comments?