

Language Models for Automated Market Research: A New Way to Generate Perceptual Maps

MSI 2023 Immersion: How to Use ChatGPT for Research and Marketing



Peiyao Li, Noah Castelo, Zsolt Katona, Miklos Sarvary



Haas School of Business University of California Berkeley





### Using GPT to create perceptual maps



# **Typical Data Collection – Human Survey**

1. How similar are the car brands BMW and Audi? (0-10)

	Audi	BMW	Chev	Dodg	Ferr	Ford	Hond	Hyun	Jeep	Lamb	Lexus	Mazd	Merc	Niss	Pors	Rena	Suba	Tesla	Тоуо	vw	Volvo
Audi		7.4	3.3	3.5	5.1	3.7	3.9	3.7	3.2	5.3	6.8	4.4	7.3	4.2	6.5	5.0	4.5	5.1	4.3	6.6	5.5
BMW	7.4		2.8	2.8	4.6	2.6	3.7	4.3	3.0	5.7	6.3	3.9	8.2	3.9	7.0	4.2	4.1	5.1	3.3	5.7	5.1
Chevrolet	3.3	2.8		7.8	2.0	7.4	5.5	5.6	6.0	2.0	3.8	4.9	2.8	5.8	2.9	3.1	5.4	2.5	5.8	4.2	3.9
Dodge	3.5	2.8	7.8		3.0	7.4	4.7	4.7	6.1	2.0	3.3	4.6	2.8	4.3	2.3	4.1	4.4	1.5	5.2	4.1	3.7
Ferrari	5.1	4.6	2.0	3.0		2.1	1.5	1.4	1.9	8.4	4.8	3.5	6.2	1.7	7.3	4.5	2.3	3.6	2.2	2.6	3.3
Ford	3.7	2.6	7.4	7.4	2.1		6.3	5.0	5.7	1.4	3.7	5.0	3.6	6.2	2.3	4.2	4.7	1.4	6.4	5.4	5.8
Honda	3.9	3.7	5.5	4.7	1.5	6.3		7.2	4.4	1.8	4.6	6.8	3.0	7.1	1.9	3.8	6.3	2.5	8.7	5.1	5.5
Hyundai	3.7	4.3	5.6	4.7	1.4	5.0	7.2		4.5	0.8	3.5	6.2	3.0	6.2	2.8	4.5	6.3	2.3	7.9	4.7	5.0
leep	3.2	3.0	6.0	6.1	1.9	5.7	4.4	4.5		1.3	3.3	2.9	2.1	3.5	2.2	3.6	5.5	1.2	5.5	4.6	4.4
Lamborghini	5.3	5.7	2.0	2.0	8.4	1.4	1.8	0.8	1.3		4.5	2.3	5.7	1.7	7.5	5.1	2.0	3.1	1.6	3.2	2.1
Lexus	6.8	6.3	3.8	3.3	4.8	3.7	4.6	3.5	3.3	4.5		5.5	6.2	5.2	5.2	4.5	5.3	4.9	5.4	4.8	5.3
Mazda	4.4	3.9	4.9	4.6	3.5	5.0	6.8	6.2	2.9	2.3	5.5		3.8	7.2	3.0	3.8	6.2	3.0	7.6	5.8	5.2
Mercedes	7.3	8.2	2.8	2.8	6.2	3.6	3.0	3.0	2.1	5.7	6.2	3.8		3.4	7.0	4.6	3.6	5.6	3.9	4.0	6.4
Nissan	4.2	3.9	5.8	4.3	1.7	6.2	7.1	6.2	3.5	1.7	5.2	7.2	3.4		2.2	4.2	6.3	3.2	8.3	5.8	5.5
Porsche	6.5	7.0	2.9	2.3	7.3	2.3	1.9	2.8	2.2	7.5	5.2	3.0	7.0	2.2		4.1	2.6	5.7	2.7	3.6	2.9
Renault	5.0	4.2	3.1	4.1	4.5	4.2	3.8	4.5	3.6	5.1	4.5	3.8	4.6	4.2	4.1		4.2	4.0	3.3	4.7	5.3
Subaru	4.5	4.1	5.4	4.4	2.3	4.7	6.3	6.3	5.5	2.0	5.3	6.2	3.6	6.3	2.6	4.2		2.8	6.4	5.6	5.0
Tesla	5.1	5.1	2.5	1.5	3.6	1.4	2.5	2.3	1.2	3.1	4.9	3.0	5.6	3.2	5.7	4.0	2.8		2.8	2.2	3.9
Toyota	4.3	3.3	5.8	5.2	2.2	6.4	8.7	7.9	5.5	1.6	5.4	7.6	3.9	8.3	2.7	3.3	6.4	2.8		5.1	6.6
VW	6.6	5.7	4.2	4.1	2.6	5.4	5.1	4.7	4.6	3.2	4.8	5.8	4.0	5.8	3.6	4.7	5.6	2.2	5.1		6.7
Volvo	5.5	5.1	3.9	3.7	3.3	5.8	5.5	5.0	4.4	2.1	5.3	5.2	6.4	5.5	2.9	5.3	5.0	3.9	6.6	6.7	

# **Problems with Human Data**

- Expensive
- Question Formulation
- Inconsistent
- Respondent Heterogeneity

# Natural Language Processing

# NLP Natural Language Processing NLU Natural Language NLG Understanding Natural Language Generation

Large Language Models (GPT) are very good at both

Generate information based on public/private data

Extract/interpret any text (including voiceto-text) data within and outside your company

# Language models main idea

- Main idea
  - GPT-2, GPT-3, GPT-3.5: predict next word / token

they went to the store to buy a gallon of [HIDE]

predict: milk

- BERT: predict a word / token that was removed

they went to the [HIDE] to buy a [HIDE] of milk

predict: store predict: gallon

GPT: Generative Pretrained Transformer

BERT: Bidirectional Encoder Representations from Transformers

# General Language Understanding Evaluation (GLUE) Tasks

Dataset	Description	Data example	Metric
CoLA	Is the sentence grammatical or ungrammatical?	"This building is than that one." = <b>Ungrammatical</b>	Matthews
SST-2	Is the movie review positive, negative, or neutral?	"The movie is funny , smart , visually inventive , and most of all , alive ." = <b>.93056 (Very Positive)</b>	Accuracy
MRPC	Is the sentence B a paraphrase of sentence A?	<ul> <li>A) "Yesterday, Taiwan reported 35 new infections, bringing the total number of cases to 418."</li> <li>B) "The island reported another 35 probable cases yesterday, taking its total to 418."</li> <li>= A Paraphrase</li> </ul>	Accuracy / F1
STS-B	How similar are sentences A and B?	<ul> <li>A) "Elephants are walking down a trail."</li> <li>B) "A herd of elephants are walking along a trail."</li> <li>= 4.6 (Very Similar)</li> </ul>	Pearson / Spearman
QQP	Are the two questions similar?	<ul> <li>A) "How can I increase the speed of my internet connection while using a VPN?"</li> <li>B) "How can Internet speed be increased by hacking through DNS?"</li> <li>= Not Similar</li> </ul>	Accuracy / F1
MNLI-mm	Does sentence A entail or contradict sentence B?	<ul> <li>A) "Tourist Information offices can be very helpful."</li> <li>B) "Tourist Information offices are never of any help."</li> <li>= Contradiction</li> </ul>	Accuracy
QNLI	Does sentence B contain the answer to the question in sentence A?	<ul> <li>A) "What is essential for the mating of the elements that create radio waves?"</li> <li>B) "Antennas are required by any radio receiver or transmitter to couple its electrical connection to the electromagnetic field."</li> <li>= Answerable</li> </ul>	Accuracy
RTE	Does sentence A entail sentence B?	<ul> <li>A) "In 2003, Yunus brought the microcredit revolution to the streets of Bangladesh to support more than 50,000 beggars, whom the Grameen Bank respectfully calls Struggling Members."</li> <li>B) "Yunus supported more than 50,000 Struggling Members."</li> <li>= Entailed</li> </ul>	Accuracy
WNLI	Sentence B replaces sentence A's ambiguous pronoun with one of the nouns - is this the correct noun?	<ul> <li>A) "Lily spoke to Donna, breaking her concentration."</li> <li>B) "Lily spoke to Donna, breaking Lily's concentration."</li> <li>= Incorrect Referent</li> </ul>	Accuracy

# **Information Extraction**



# What is the source of information in Large Language Models?

GTP3 trained on Tbytes of text data including:

- Web crawls
- Reddit links
- Books
- Academic journal
- Wikipedia

	Wikipedia	Books	Journals	Reddit links	СС	Other	Total
GPT-1		4.6					4.6
GPT-2				40			40
GPT-3	11.4	21	101	50	570		753
The Pile v1	6	118	244	63	227	167	825
Megatron-11B	11.4	4.6		38	107		161
MT-NLG	6.4	118	77	63	983	127	1374
Gopher	12.5	2100	164.4		3450	4823	10550

**Table 1. Summary of Major Dataset Sizes.** Shown in GB. Disclosed in **bold**. Determined in *italics*. Raw training dataset sizes only.

# Where is the information "stored"?



# ChatGPT: Reinforcement Learning from Human Feedback (RLHF)

Step 1

Collect demonstration data and train a supervised policy.

A prompt is sampled from our prompt dataset.

A labeler demonstrates the desired output behavior.

This data is used to fine-tune GPT-3.5 with supervised learning.



We give treats and punishments to teach ...



Step 2

Collect comparison data and train a reward model.

A prompt and several model outputs are sampled.

0 Explain reinforcement learning to a 6 year old. A In reinforcement learning, the

agent is C D In machine learning.

B

Explain rewards

A labeler ranks the outputs from best

This data is used to train our reward model.

to worst.





D > C > A > B

#### Step 3

Optimize a policy against the reward model using the PPO reinforcement learning algorithm.



### Back to our objective



**Data Generation Using GPT2 and GPT3** 

Algorithm: GPTNeo

The car brand Audi is similar to \_\_\_\_\_

**The car brand Audi is similar to** its main competitor Mercedes. In its brand statement, the automaker explains: ...

How often does BMW show up in the continuation? How many times does Mercedes?...

# **Resulting Frequency Table**

Prompt

			Chevro	Dod	For		Hyund		Lex	Maz	Merced	Nissa	Subar		Toyot
	Audi	BMW	let	ge	d	Honda	ai	Jeep	us	da	es	n	u	Tesla	а
Audi	183	143	10	8	46	35	74	63	97	43	204	33	10	27	42
BMW	390	165	12	14	110		Base	elin	e	5	185	90	32	47	41
Chevrolet	15	4	93	24	67		Prot	blen	n	5	5	31	1	17	11
Dodge	0	0	6	186	10	2	2	18	6	10	4	4	2	7	3
Ford	230	852	1530	1140	251	400	389	1157	328	977	639	543	817	923	908
Honda	46	35	17	37	176	191	91	32	147	76	15	154	163	50	279
Hyundai	7	0	2	3	17	22	320	7	7	3	6	24	8	3	8
Јеер	10	1	2	97	18	4	3	209	2	6	0	9	4	3	8
Lexus	14	10	1	10	12	15	18	2	94	3	6	15	24	4	17
Mazda	0	0	1	0	1	6	0	0	1	74	0	4	7	0	3
Mercedes	246	205	E	24	22	11	ГЭ	21	74	24	61	25	0	E 2	11
Niccon	240	205	<u> </u>	24	21		55	10	74	54 1	101	23	0	32	41
NISSan	12	14	9	35	31	/3	55	18	90	41	15	218	/8	33	81
Subaru	1	2	1	6	10	23	21	5	15	13	1	11	111	0	2
Tesla	87	68	3	8	214	83	55	28	42	48	103	107	13	332	29
Toyota	88	42	32	114	361	693	251	81	243	185	122	439	274	68	145

Mentioned

### **Data Generation Using ChatGPT**

## Algorithm: GPT3.5 and GTP4

I want you to act as a person filling out a survey. I will ask you a question and you must answer using only an integer, no words. You will reply with an integer between 0 and 10. My first question is in your opinion, how similar are the car brands a and b on a scale of 0 to 10 where 10 means very similar?

# **Data Collection Summary**

	AI	Human
Numerical data	GPT4 numerical responses	Direct rating on a
Open-ended data	GPTNeo sentence completion	Open-ended responses

# **Data Collection Summary – Example: Jeep and Ford**

	AI	Human
Numerical data	7	7
	The car brand	
	Jeep is similar to	
	the car brand	The car brand
Open ended data	Ford. They	Jeep is similar to
Open-ended data	combine high	the car brand
	quality, comfort	Ford.
	and affordability	
	in	

# The Maps (Open-Ended)



# The Maps (Numerical)



### **Comparison using our "Triplet Method"**



### Which brands are problematic – cause disagreement?

- Idea: Using only AI data, remove brands one by one and see if selfconsistency increases
- Largest increases identify problematic brands
- Removing problematic brands increases agreement with human data
- Bottom-line: we can use AI data only to assess how well the method works for given brands

# Which brands are problematic – cause disagreement?



# **Prompt selection**

- Simple: On a scale of 0 to 10, how similar are the car brands A and B?
- RTF (role, task, format): *I want you to act as a person filling out a survey. I will ask you a question and you must answer using only an integer, no words. You will reply with an integer between 0 and 10. My first question is in your opinion, how similar are the car brands a and b on a scale of 0 to 10 where 10 means very similar?*
- Few-shot:

*Question: On a scale of 0-10, how similar are the car brands X and Y on a scale of 0 to 10 where 10 means very similar? Answer: 5* 

# **Prompt selection**

Type of prompts	Self-consistency rate	Agreement rate	Adjusted agreement rate
Simple	.980	.800	.848
	(.971,.988)	(.795,.804)	(.843,.853)
Few-shot	.986	.821	.868
	(.971,.988)	(.820,.830)	(.867,.877)
RTF	.964	.815	.872
	(.947,.978)	(.810,.823)	(.866,.880)
Combined	.977	.800	.850
	(.967,.989)	(.794,.808)	(.844,.859)

# Comparison with Trade-in Data

- We use JD Powers data on car trade ins
- Which brand of car was traded in for which?
- Annual data between 1999 and 2008
- Can be converted to similar format as our human surveys

# Comparison with Trade-in Data

year	vs human open-ended	vs human direct rating	vs AI open-ended	vs AI direct rating
1999	0.739	0.713	0.641	0.689
2000	0.731	0.717	0.638	0.706
2001	0.733	0.714	0.651	0.716
2002	0.750	0.737	0.667	0.717
2003	0.759	0.751	0.661	0.738
2004	0.760	0.755	0.653	0.740
2005	0.755	0.756	0.653	0.742
2006	0.760	0.771	0.667	0.739
2007	0.756	0.766	0.680	0.736
2008	0.754	0.761	0.681	0.750

# **Year-Specific Prompt**



# **Attribute-based Analysis**

- Collect a set of attributes potentially describing cars:
  - sportiness, power, stylish-ness, technology, spaciousness, eco-friendly-ness, fuel efficiency, popularity, reliability, safety, comfort, durability
- $\circ$  Human-based data
  - ``For the following words, please use the slides to indicate how well each word describes the car brand A" (Likert 0-10)
  - Use ratings directly
- Al-based data
  - Use prompt: The most X car brand is \_\_\_\_\_
  - Count mentions of car brands in the responses, generating an attribute-brand frequency matrix

# **Mapping Attributes**

#### Machine

Human



# **Comparing Segments**

- Prompt idea: "A young and wealthy male's favorite car brand is..." or "A young and wealthy female's favorite car brand is..."
- Varying age, income and gender in these types of prompts allows us to extract preferences of different segments
- We choose these three variables, because it's relatively easy to obtain corresponding human data for validation
- Opportunity: collect data on segments or in contexts that are not feasible with human respondents.

### **Comparing Segments – Effect of Demographics on Preferences**



# Conclusion

Takeaways

- AI-assisted or even pure AI-based market research can yield good results
- Can generate perceptual maps based on similarity, attributes (factors)
- Able to incorporate context: demographics, time, etc

Limitations

- Requires a good amount of data
- Need to be careful with comparisons (analysis)
- Prompts matter
- Important to identify brands that don't work well

#### **Group Activity – Market Research Using LLMs for Information Extraction** 15 + 15 minutes

- Think about a market research application for your firm, where you extract information from a language model (optional add your proprietary text data)
- Go through some of the steps outlined on next slide can be similar/same to what we just saw
- Feel free to do some quick experimentation with ChatGPT to see what prompts would work.
- Discuss the main limitations/challenges you foresee

### **Group Activity – Market Research Using LLMs for Information Extraction**

Objective	Create Perceptual Maps	Your Alternative Method
Prompt Design	The car brand X is similar to	
Add Optional Proprietary Data	Focus group interviews, forum data, customer support chats, etc	
Data Collection	Choose set of brands, determine number of data points	
Quantitative Analysis	Control for baseline, use "triplet" method	
Validation	Compare with human data	
Refinements	Time, Demographics, Context	