Creating Synthetic Experts with Generative Artificial Intelligence

Daniel M. Ringel

dmr@unc.edu

Wharton Business & Generative AI Conference
September 8, 2023
Wharton San Francisco Campus

AI Research, classified in four categories

Today's Talk

Image by MidJourney

Classification is paramount in today's data rich environment

Organizations increasingly depend on machine learning to distill ***intelligence from*** vast amounts of ***unstructured data***

- News articles
- Reports and policies
- Social media
- Internal communications

Classification models can swiftly ***identify constructs of interest*** in data

- Sentiment
- Opinions and arguments
- Type of rhetoric
- Product categories

Classification models ***unlock hidden insights and potentials*** that managers can leverage to drive operational efficiency and facilitate business growth

Image by MidJourney

Versatility *of classification models* ***extends*** *their* ***utility across sectors*** *and* ***functions***

Efficacy of classification models relies heavily on their training

Training a classifier typically requires ***many labeled examples***

easy for ***simple constructs***

- crowdsourcing (e.g., Amazon mTurk)
- fast and relatively low cost

problematic for more ***complex constructs***

- requires experts
- scarce and expensive resource

Complex constructs are often ***more insightful,*** *e.g.,*

General sentiment of consumers towards a brand

vs.

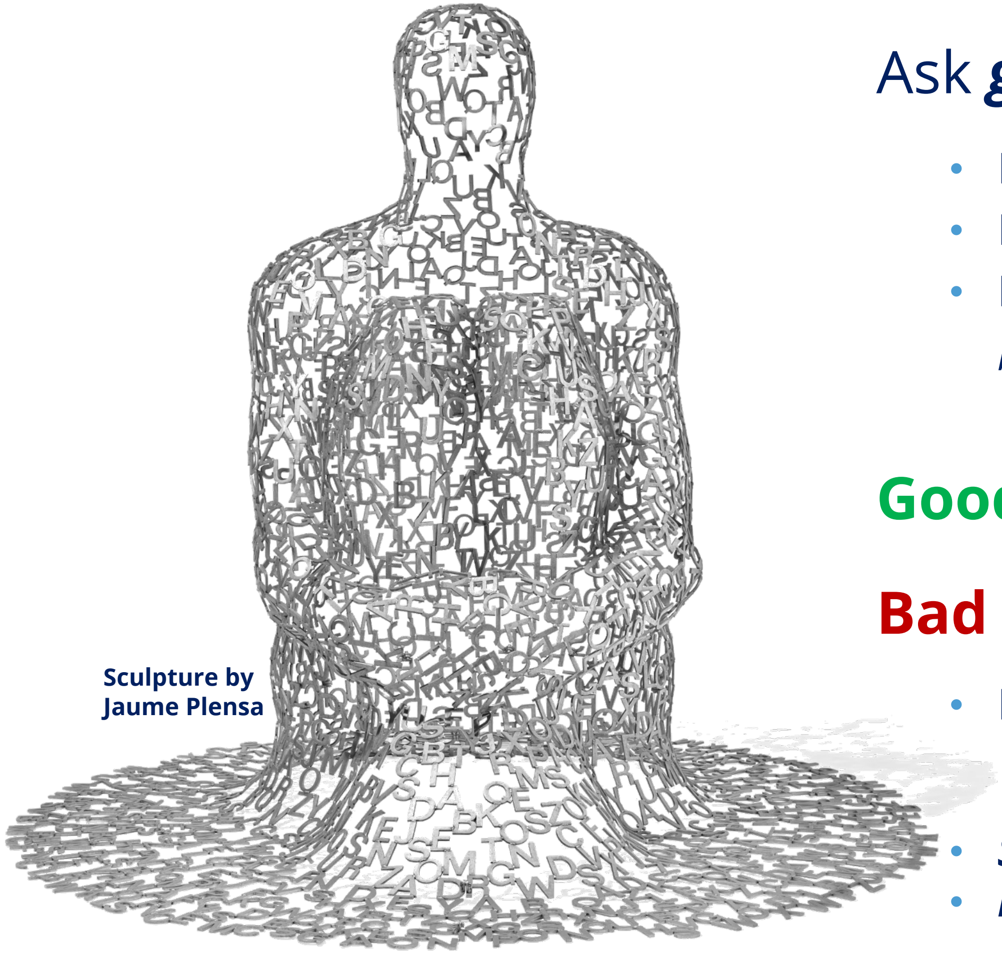
What exactly consumers perceive as positive/negative
(e.g., Marketing Mix: Product, Place, Price, Promotion)

Inform many downstream tasks

Can be used as DV or as IV

Useful to answer research questions, test hypotheses, and explain mechanisms

Idea: Let generative AI label data



Sculpture by
Jaume Plensa

Ask ***generative AI*** to identify complex construct of interest

- Easily accessible | *e.g.,* ***OpenAI's ChatGPT***
- Represents ***vast body of knowledge***
- Many ***theoretically founded constructs*** in training data: *books, reports, news articles, websites, etc.*

Good News: Works well!

Bad News: Limitations

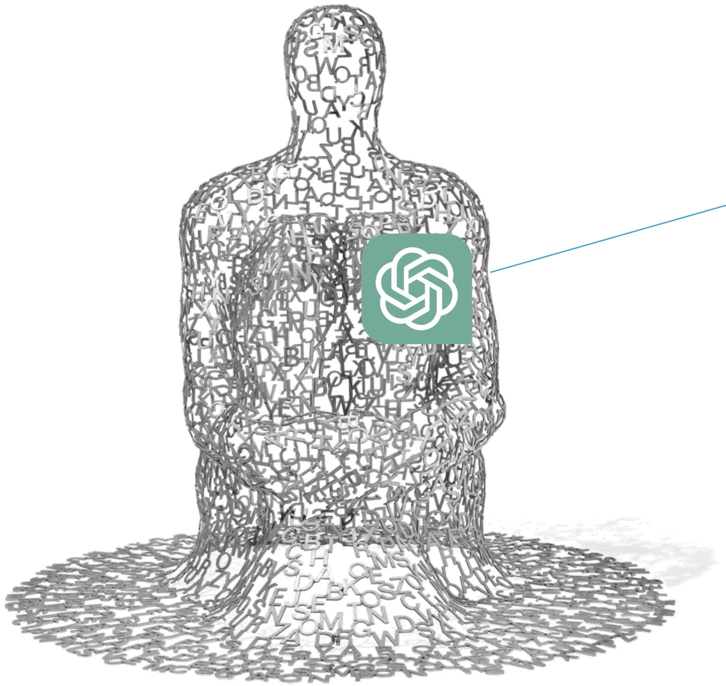
- Largely ***proprietary***
  - Full dependence on provider: access, pricing, capabilities
  - Privacy and confidentiality risks
- ***Slow***
- ***Expensive***

Not appropriate for many research and production environments

Introducing: Synthetic Experts

What we don’t need

- Vast body of knowledge
- Ability to carry out many different tasks



Sculpture by Jaume Plensa

What we need

- Faithful identification of a ***specific*** (complex) construct
- Free of third-party constraints
- Efficiency

just a tiny piece

Synthetic Experts

Approximate powerful Artificial Intelligence
with an open-source Large Language Model (LLM)

- Fine-tune pretrained LLM for classification task of interest
- Use powerful generative AI to label training data

Creation Process: Synthetic Expert

Steps	Define Construct of Interest	Collect Examples	Query AI for Labels	Approximate AI	Classify Data
This Study	Marketing Mix Variables (MMX) • Product • Place • Price • Promotion	Tweets that mention Major Brands (N = 699) Twitter Research API	ChatGPT4 via RTF prompting (role, task, format) OpenAI's API	Fine-Tune pre-trained LLM "RoBERTa" for multi-label classification Tune Hyperparameters Out-of-Sample Validation	Rapidly classify thousands of Tweets at low cost: *1 million Tweets in less than 3 hours on my MacBook* Share classifier on Hugging Face (open-source, open-access platform and community)

Synthetic Experts *are* ***task specific approximations*** *of powerful Artificial Intelligence (AI) models*

- *Highly scalable*
- *Fully independent*
- *Free from third-party constraints*

Empirical Application: The Marketing Mix

Construct of Interest: The Marketing Mix (MMX)
- Marketing mix is at the heart of marketing strategy
- Theoretically founded
- 4 Ps: Product, Place, Price Promotion
- Multi-label classification task

Data: Consumers' posts on Twitter mentioning brands
- Vast and unstructured
- Valuable information source to marketers
- Brevity and informal language challenging for text analysis

Twitter posts from 2019 to 2021
Brands: 699 mentioned in posts

Validation Sample: 1,000 Tweets; brand stratified
- ***Experts***: 4 human experts x 6 workshops x 2 hours = 48 hours; ***priceless***
- ***Amateurs***: 3 per Tweet, 52 mTurk Masters; 13.3% HIT rejection; 1 day; Krippendorff's-α = .391; **$410**
- ***ChatGPT4*** : 3 per Tweet; 30min; Krippendorff's-α = .700; **$6**

Training Sample: 30,000 Tweets; brand stratified
- ***ChatGPT4***: 3 per Tweet; 15 hours; Krippendorff's-α = .716; **$180**

Mouse-over Twitter Bird for real-time MMX labels

@Sony's XM3's ain't as sweet as my bro's airpod pros but got a real steal 🤑 the other day #deal #headphonez

Best cushioning ever!!! 🤗🤗🤗 my zoom vomeros are the bomb🏃‍♀️💨!!! @nike #run #training

It's inspiring to see religious leaders speaking up for workers' rights and fair wages. Every voice matters in the #FightFor15! 💪🏽✊🏼 #Solidarity #WorkersRights

To celebrate this New Year, @Nordstrom is DOUBLING all donations up to $25,000! 🎉 Your donation will help us answer 2X the calls, texts, and chats that come in, and allow us to train 2X more volunteers!

Identifying marketing mix variables in consumers' posts on Twitter

Human Experts vs. Amateurs vs. ChatGPT4

Table 1. Bootstrapped Label Agreement among Human Experts, Amateurs, and Generative AI

	Mean Krippendorff's α			Mean Classification Metrics		
	Expert	Amateur	ChatGPT4	Precision	Recall	F1-score
Expert	1 *(.000)*			1 *(.000)*	1 *(.000)*	1 *(.000)*
Amateur	.512 *(.000)*	1 *(.000)*		.835 *(.000)*	.546 *(.000)*	.660 *(.000)*
ChatGPT4	**.786** *(.000)*	.470 *(.000)*	1 *(.000)*	**.893** *(.000)*	**.833** *(.000)*	**.862** *(.000)*

Notes: $N = 4{,}000$ labels. Analysis based on 1,000 bootstrapped Tweets. Standard errors are in parentheses. Human expert labels are the ground truth for classification metrics.

Findings

1. Crowdsourced labels insufficient for complex constructs
2. ChatGPT4 is a viable surrogate for human expertise

Synthetic Expert: Training and Performance

1. Get MMX labels for ***pool of 30,000 Tweets*** from ChatGPT4
2. Randomly ***sample N labeled Tweets*** from Tweet pool
3. For each N, use ***5 different seeds*** for sampling and model initialization
4. Evaluate ***label agreement with validation sample*** for each *N* and *seed*: *(40 fine-tuned models)*

   Area und ROC Curve: accounts for ***different probability thresholds*** for binary classification

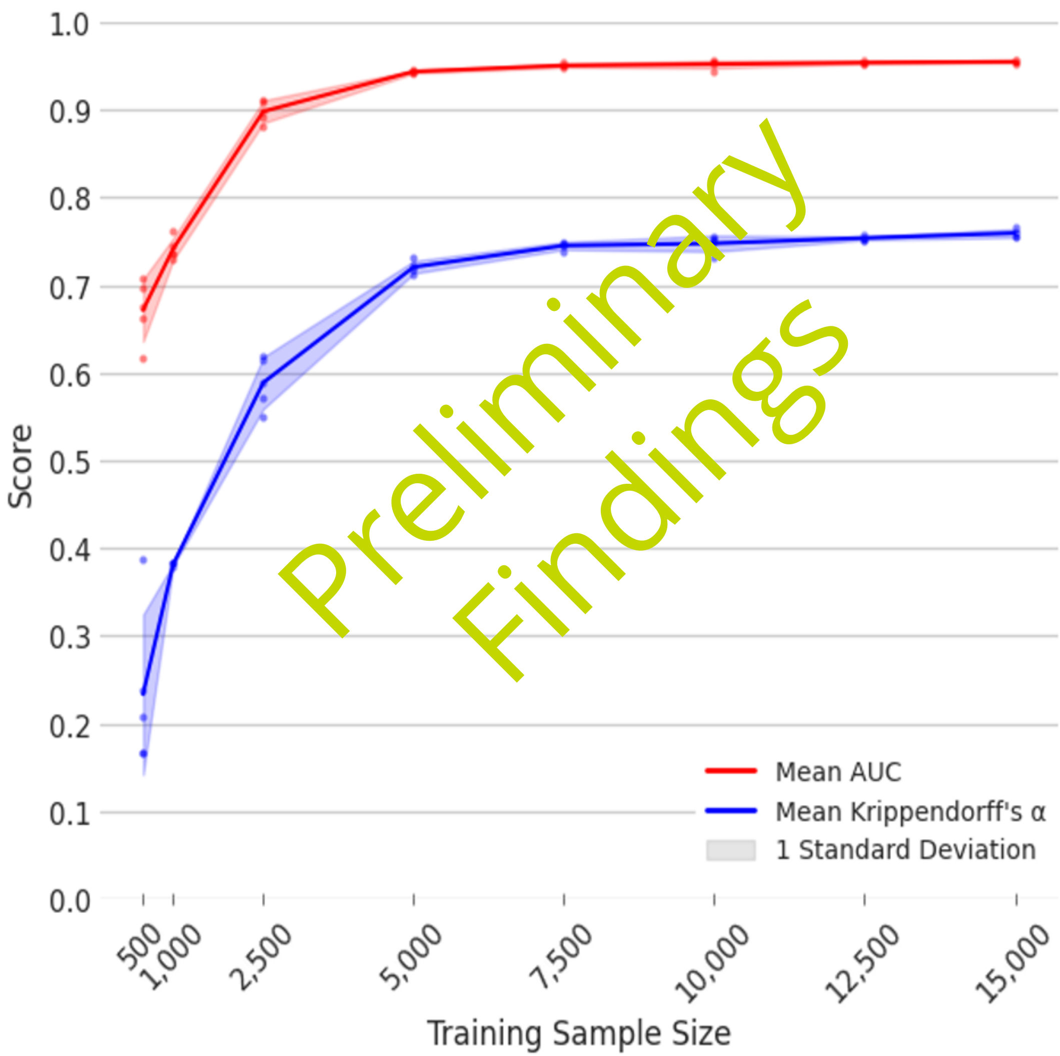
   Krippendorff's Alpha: accounts for ***chance agreement***

Findings

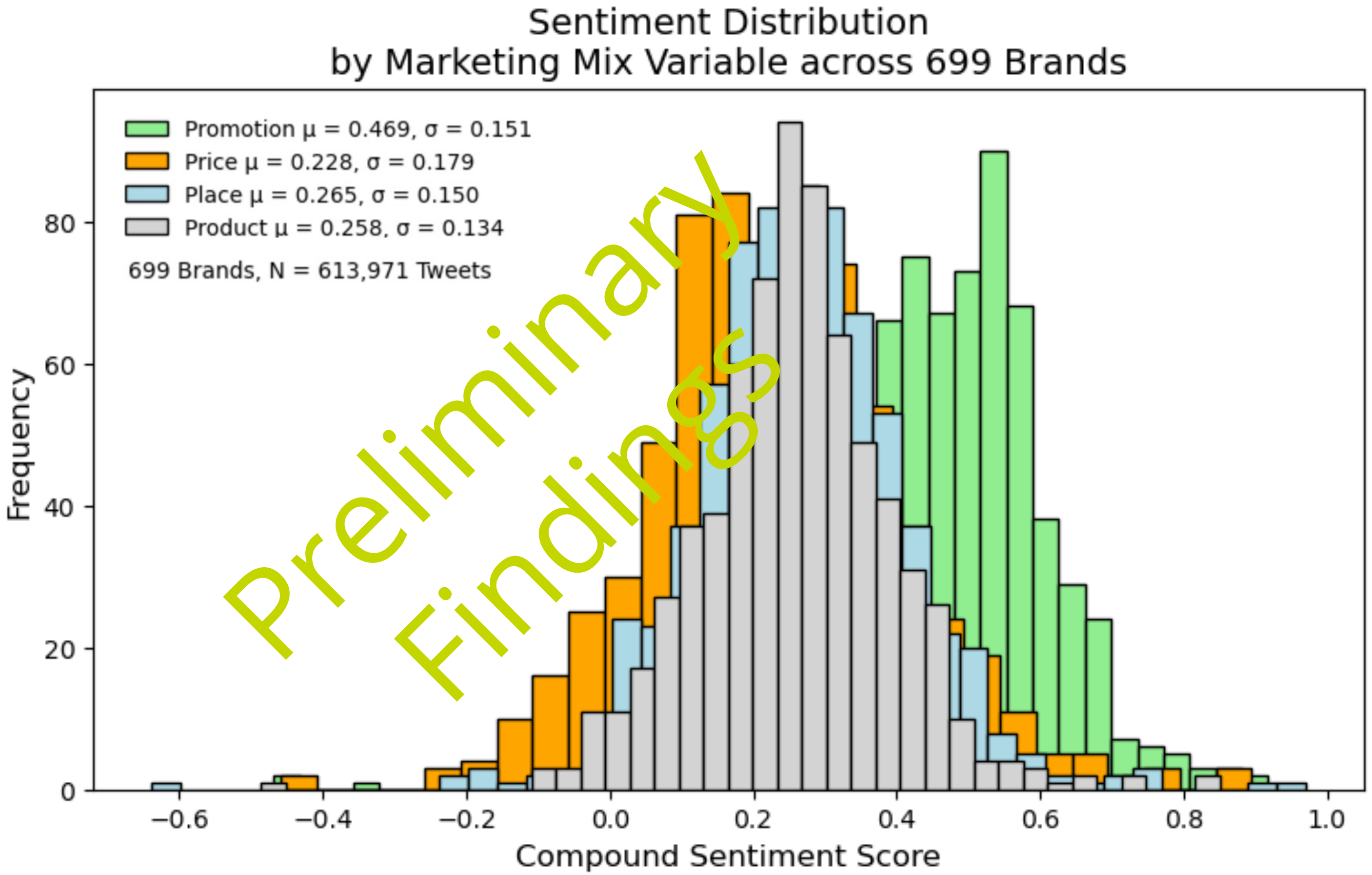
1. Synthetic Expert performs only **3.4%** below ChatGPT4
2. Performance converges after 7,500 training samples
3. Sample selection and model initialization less relevant for larger training samples

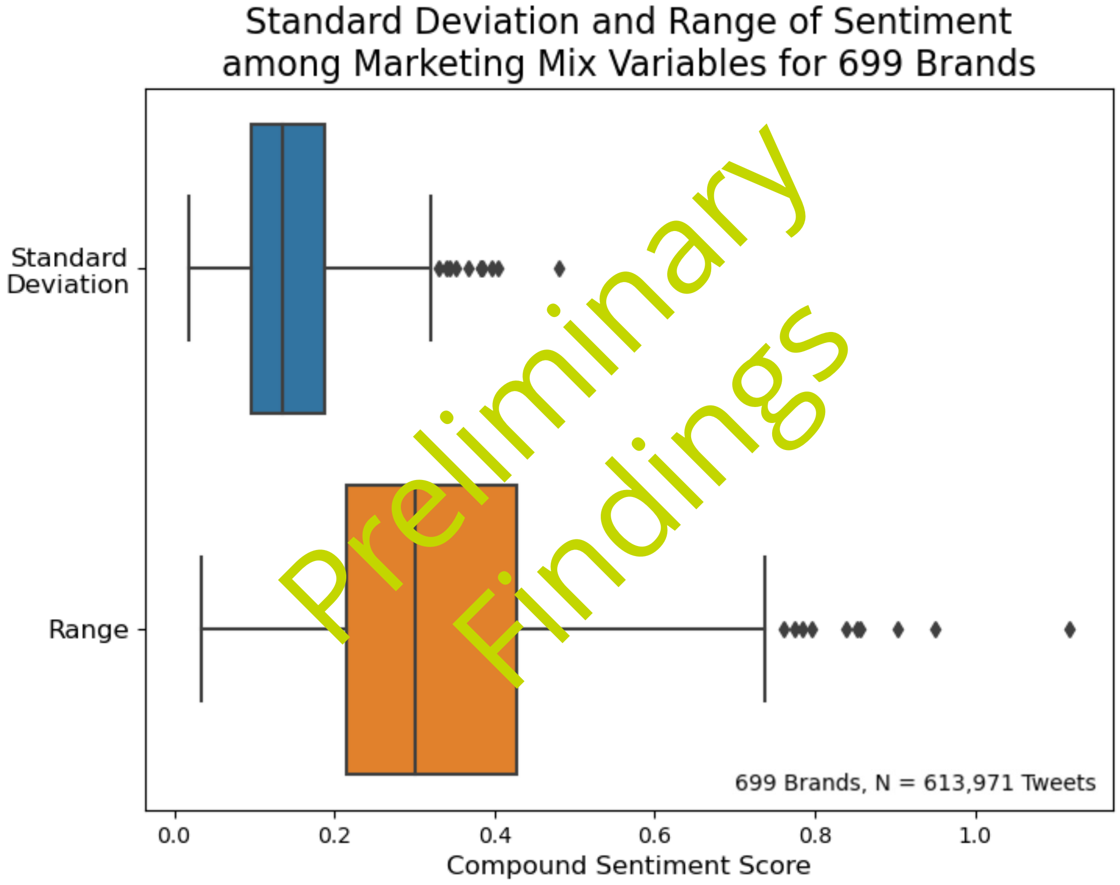
Note: We observe some variance in performance of the synthetic expert across the four marketing mix variables (F1-Score σ^2 = .001; Krippendorff's alpha σ^2 = .005), with the lowest relative performance for promotion and the highest for price.

Figure 1. **Label Agreement with Validation Sample**



Sentiment Distributions differ by MMX Variable





Notes: *VADER Sentiment Analysis (Hutto and Gilbert 2014), 699 Brands, N = 613,971 Tweets, Overall Sentiment: μ = 0.258, σ = 0.133, Range = $Sent_{mam}$ - $Sent_{min}$*

1. Considerable differences in brand sentiment among marketing mix variables → Average can mislead managers
2. Differentiated evaluation of marketing mix variables → Discover strengths and weaknesses in MMX → Informs MMX adjustments

Brands' Strengths and Weakness differ across MMX

	Brand	Sentiment Overall	Sentiment by Marketing Mix Variable			
			Product	Place	Price	Promotion
Apparel	Calvin Klein	.256	.258	.339	.419	.330
	Abercrombie & Fitch	.286	.302	.209	.347	.494
	Polo Ralph Lauren	.291	.295	.336	.192	.408
Snacks	Tostitos	.245	.243	.303	.135	.410
	Doritos	.261	.258	.148	.317	.438
	SunChips	.188	.189	.295	.486	.275
Airlines	Spirit Airlines	.024	.002	.025	.072	.532
	JetBlue	.237	.264	.193	.119	.541
	Southwest Airlines	.280	.301	.279	.200	.571

Legend lowest **Color indicates Relative Sentiment *** highest

* column-wise for overall sentiment; row-wise across marketing mix variables

***Notes**: VADER Sentiment Analysis (Hutto and Gilbert 2014), 9 Brands, N = 9,000 randomly sampled Tweets from 2020; stratified by brand*

Preliminary Findings

Zooming-in for deeper Insights

Topics in “Place” Tweets that mention *Abercrombie & Fitch*

LOL anyone remember when @abercrombie used to have those shirtless greeters outside their store?😂😂

I wish @abercrombie would stop using #usps to deliver their goods on this occasion, they give an email and text stating delivery between a 4 hour period. This is the 3rd delivery recently where I've been in all day waiting and nothing has arrived 😡 infuriating

Why ? Why are you closing the crossgates mall store? Now no store from Albany to buffalo or NYC. 21 years I shopped there. And now I will have to pay for shipping if you close that down. I'm striking against you now. This is BS. @abercrombie

There’s talk in one of my Facebook groups about racism @abercrombie. Apparently u get stuck up the back of the shop if you r African American. I wonder if they have cameras in the change rooms like Victoria’s Secret?

***Notes**: Four topics identified by BERTopic (Grootendorst 2022) using sentence embedding (SBERT), dimensionality reduction (UMAP), density-based clustering (HDBSCAN), bag-of-words with n-grams (SciKit-learn), and cluster-based TF-IDF.*

Contribution and Outlook

Hugging Face Model Hub
> 2,000 sentiment classifiers

***NEW**: 1 MMX Classifier*

Research bridges the gap between conventional text classification methods and generative AI

Synthetic experts are a scalable, independent solution for complex classification tasks: You can train and deploy them on a laptop computer!

Facilitate extraction of theoretically-founded variables from unstructured data

Synthetic experts inform downstream tasks as DV or IV to answer research questions, test hypotheses, and explain mechanisms. They guarantee reproducibility.

Rapidly generate deeper managerial insights from unstructured data

Application extends to various sectors such as law, public policy, medicine, and management

Future innovations in generative AI promise to broaden the application of synthetic experts in classification tasks across different media types

Image by MidJourney

Dissemination

Website **Code** **Working Paper** **MMX Synthetic Expert**

synthetic-experts.ai 50%

How to Create your own Synthetic Expert using Generative AI

This website provides supporting materials, code, data, and tutorials for the paper ***Creating Synthetic Experts with Generative AI*** **by Daniel M. Ringel (2023).**

News

[September 8, 2023] Presentation of ***Synthetic Experts*** at the **Wharton Business & Generative AI Workshop** in San Fransico.

[September 6, 2023] How to create ***Synthetic Twins*** of text: Python notebook now available on GitHub

[August 26, 2023] Create your own ***Synthetic Expert*** with new Python notebooks available on GitHub. Label text with generative AI and fine-tune LLMs for your purpose.

[August 23, 2023] Work with several demo **Datasets** to explore the use and creation of Synthetic Experts.

[August 19, 2023] Easily apply the ***MMX Synthetic Expert*** of this research to YOUR data. New Python notebook available on GitHub.

[August 16, 2023] *This research is currently being revised and extendet. I will make materials available as they are completed.*

Paper

- Ringel, Daniel, *Creating Synthetic Experts with Generative AI* (July 15, 2023). Available at SSRN: https://ssrn.com/abstract=4542949.
- Appendix: Details, Notes, Parameters, IRB information

Code

- Python Notebooks: GitHub Repo
- Functions: UseSynExp.py

Data

- List of Brands
- List of Tweets
- Synthetic Twins of Data

Supporting Documents

- How to set-up GPU computing on Apple Silicone [ipynb]
- How to set-up your own Deep Learning Machine: Install Ubuntu with CUDA, CuDNN and PyTorch on a computer [pdf]
- How to run the code on Google Colaboratory (for free) [pdf]
- **BONUS**: Synthetic Twins of real-world textual data [ipynb]

Tutorials

- Get ready to fine-tune LLMs [Setup_Python_GPU.ipynb]
- Query the OpenAI API [Query_OpenAI_API.ipynb]
- Parse OpenAI API responses [Parsing_API_Responses.ipynb]
- Fine-tune a pre-trained LLM from Huggingface [Fine-tuning_LLMs.ipynb]
- Hyperparametertuning with Optuna [Hyperparametertuning.ipynb]
- Deploy your Synthetic Expert and use it for Inference [ipynb]

Classroom Materials

- Slides [pptx]
- Notebook [ipynb]
- Data [zip]
- Instructor Notes [pdf]
- Video [YouTube]

Repository

Code will be maintained on GitHub

Fine-Tuned Model

Marketing Mix Classifier on Huggingface's Model Hub

Questions, Comments?

Get in touch at dmr@unc.edu

contact **dmr@unc.edu**

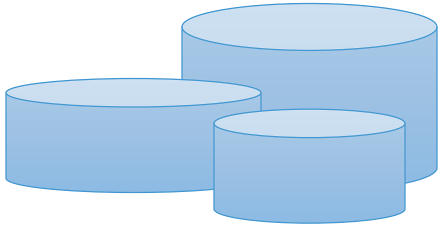
Appendix: Data

*Anonymizing texts with **Synthetic Twins***

Basic Idea: Use generative AI to create replicas of texts that reflect their idea and meaning but obfuscate identifying information

Synthetic Twins

- Correspond semantically in idea and meaning to original texts
- Wording, people, places, firms, brands, and products changed by AI
- Mitigate, to some extent, possible privacy, and copyright concerns
- Can be useful to create variations of existing texts



Multiple Demo Datasets available [here]()

Create your own Synthetic Twins [here]()

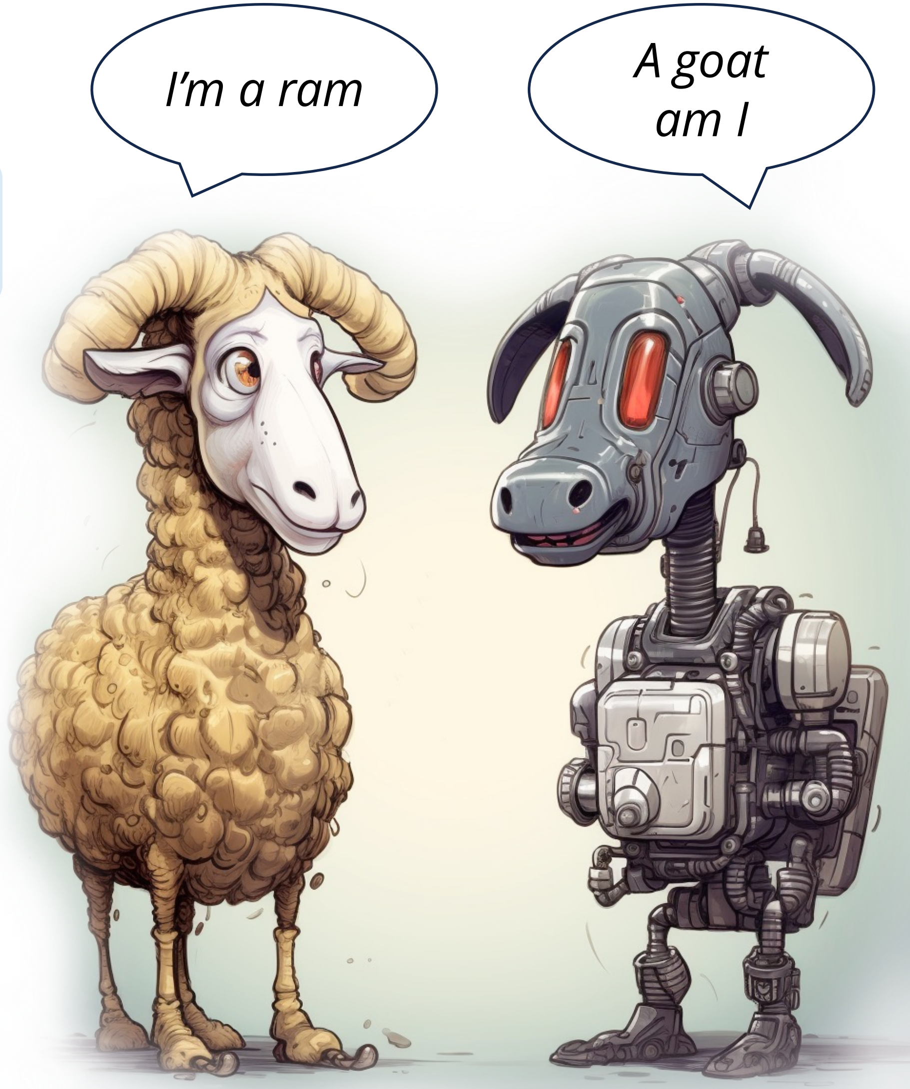


Image by MidJourney