Data Science in the Business World

Daniel M. Ringel

https://github.com/dringel
https://www.linkedin.com/in/dmr76/
https://www.ringel.ai

Class 21

Smarter, cheaper, greener: *Vertical AI for business analytics*

2026

Pre-Reading

Creating Synthetic Experts with Generative Artificial Intelligence

Kenan Institute of Private Enterprise Research Paper No. 4542949

25 Pages • Posted: 20 Aug 2023 • Last revised: 7 Dec 2023

Daniel Ringel
University of North Carolina at Chapel Hill

Date Written: December 5, 2023

Abstract

Classification is paramount in today's data-rich environment as marketers increasingly depend on machine learning to distill intelligence from vast amounts of unstructured text such as news, reports, and social media. Modern classification models swiftly identify constructs of interest, such as sentiment or product categorizations to inform research and managerial decision-making. Training an effective classification model requires many correctly labeled examples. While simple constructs can be labeled via crowdsourcing, more abstract and multifaceted constructs necessitate expert labelers—a scarce resource. We study whether generative AI, specifically ChatGPT4, can replace domain experts for identifying a central marketing construct in microblogs: brands' marketing mix. We find that, unlike crowdsourced labels, those generated by ChatGPT4

New opportunities with generative AI

Arora et al. (2024 Journal of Marketing) propose an **AI-Human hybrid approach** to qualitative and quantitative marketing research using GPT-4.

Goli and Singh (2024 Marketing Science) explore the viability of using GPT-4 to **emulating human survey respondents**, focusing on intertemporal choices.

Cao et al. (2024 Journal of Financial Economics) show that **combining AI computational power with human institutional knowledge** produces **superior stock return predictions** compared to either alone.

Boussioux et al. (2024 Organization Science) find that **individuals using GPT-4 generate creative solutions** comparable to human crowds at **95% lower cost,** though with reduced collective diversity.

Shalpegin et al. (2025 Journal of Operations Management) demonstrate how **generative AI accelerates** empirical operations research tasks like **literature classification** and **qualitative coding** from months to days.

de Kok (2025 Management Science) **establishes generative LLMs as valid research instruments for textual analysis in accounting**, solving both traditional tasks and previously human-only coding challenges.

Reisenbichler et al. (2025 Marketing Science) tailors **open-source LLM with human-in-the-loop layer** to generate better performing sponsored-search ad copy; show **when GenAI-driven SEA gains hold.**

Cost and peril of generative AI

Forbes

Agentic AI's Token Paradox: When Cheaper Means More Expensive

November 3, 2025

THE WALL STREET JOURNAL.

OpenAI Made a $12 Billion Loss Last Quarter, Microsoft Results Indicate

October 31, 2025

MIT report: 95% of generative AI pilots at companies are failing

July 2025

Environmental toll of generative AI

One query to ChatGPT uses approximately ***as much electricity*** as one ***light bulb*** for about ***20 minutes*** *(Allen Institute for AI 2024)*

Google's greenhouse gas emissions in 2023 were ***48% higher*** than in 2019, according to its latest environmental report, ***due to increasing energy demands from the greater intensity of AI compute.***
2024 Environmental Report

Starting in 2023, **Google** wrote in its sustainability report that it was ***no longer maintaining operational carbon neutrality.***

In its sustainability report, **Microsoft** said its ***emissions grew by 29%*** since 2020 due to the construction of more ***datacenters*** that are designed and optimized to ***support AI workloads.***

Challenges of generative AI

The **COST** of training large AI models has grown exponentially, making it challenging for firms to sustain their investments.
Stanford University (2023). AI Index Report 2023

The aggregation of datasets for AI poses inherent **PRIVACY RISKS**, necessitating governance to protect sensitive information.
McKinsey & Company (2020). Derisking AI by design: How to build risk management into AI development

Less than half of organizations mitigate the risk they consider most relevant: **INACCURACY**.
McKinsey & Company (2023) The State of AI in 2023: Generative AI's Breakout Year

The AI industry could soon be one of the largest contributors to **CARBON EMISSIONS**, if current trends continue. *Sundberg, N. (2023). MIT Sloan Management Review. Tackling AI's Climate Change Problem*

Artificial intelligence models are getting bigger. But sometimes **SMALLER CAN BE SMARTER**.
Leffer, L. (2023). Scientific American. When It Comes to AI Models, Bigger Isn't Always Better

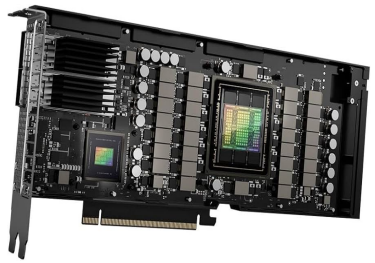
Creating and operating generative AI

Training

Supplier	Meta	**Training H100 GPUs**	16,000	**Training electricity (MW)**	**52.428**
Model	llama-v3p1-405b-instruct	**Training GPU Hours**	30.840M	**Training CO_2 emissions (tons)**	**19,346**
Parameters	405 billion	**Training Days**	80	*corresponds to electricity for*	*14,979 households*

H100 80GB

NVIDIA



Nvidia H100 NVL Tensor Core GPU
$39,000.00
Pre-owned
itcreations.com
Free delivery

Nvidia Tesla H100 80GB HBM2e 5120-Bit PCI-Express 5.0 x16
$28,283.17
Usually $31,345
Renewtech & more

NVIDIA H100 Tensor Core GPU
$24,500.00
Uvation
Get it by Jun 27 (Free)

NVIDIA Redstone-Next GPU Baseboard 4 H100 SXM5 320GB...
$149,271.18
anafrashop.com

HPE P54868-001 NVIDIA H100 80GB PCIe Accelerator
$32,195.50 ~~$33,890~~
ServerSupply.com
Get it by Tue (Free)

genAI is not your typical software

Inference

Quantized: INT8

	VRAM (GB)	GPU (count)	Watts /h	CO_2 kg/h	CapEx GPU invest	CapEx /h	Power /h	Own h	Rent /h
H100 SXM 80GB	486	7	4,075	**1.50**	**$ 210,000**	$ 5.99	$ 0.92	**$ 7.01**	**$ 26.23**

Generative AI - bigger, better, overkill?

Horizontal AI

*today, predominately **genAI** models*

→ Applicable to a broad range of tasks
→ Promises to revolutionize knowledge work

→ Unwieldy
→ Resource-hungry
→ Difficult to build and run
→ Largely closed and proprietary
→ Centralized and under the control of few

Source: Scientific American (2023)

Example: OpenAI's GPT Models

Vertical AI

*such as **specialized classifiers***

→ Build for individual tasks
→ Known for efficiency and accuracy
→ Deterministic

→ Requires domain expertise for training

Example: OpenAI's Toxicity Classifier*

Why buy the Candy Store, when you just need a Lollypop?

credit: anonymous industry manager, 2023

Many applications for *Vertical AI:* Today, *Brand Relationships* in Tweets

Vertical AI by example of classifiers

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

- News and social media
- Customer interactions
- Reports and policies
- Internal communications

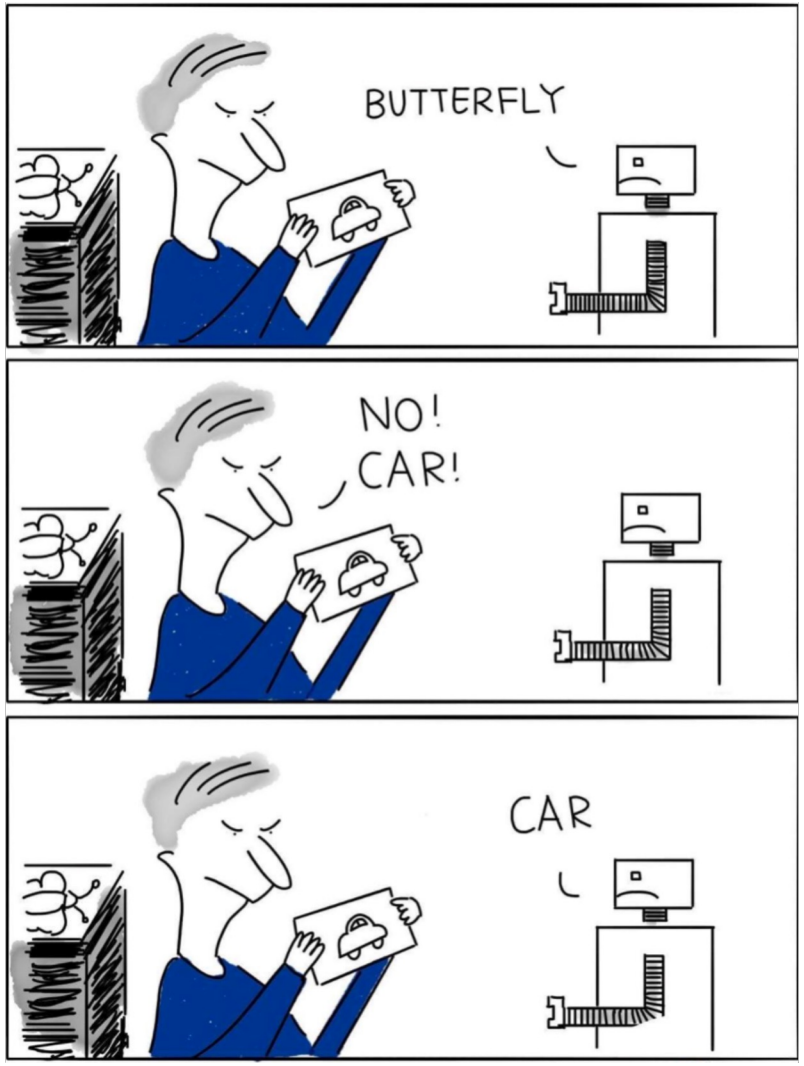
Classifiers can swiftly ***identify constructs of interest*** in data

- Specific topics
- Bias and sentiment
- Compliance
- Emotions
- Customer Experience
- Dimensions of ESG

Image by Midjourney

Versatility *of classifiers* ***extends*** *their* ***utility*** *across* ***marketing functions*** *and* ***problems***

Training a classifier



Training a good classifier requires *labeled examples*

EASY for ***simple constructs***

Straightforward; can be easily defined and measured

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

DIFFICULT for more ***complex constructs***

Multifaceted with higher levels of abstraction and ambiguity

- requires domain experts
- scarce and expensive resource

Complex constructs allow for richer insights

Generative AI labeling

Ask generative AI to identify a complex construct of interest

- Easily accessible: dozes of models available
  e.g., *GPT, Claude, Grok, Gemini, Llama, Mistral, Deepseek, Nova*
- Represents vast body of (latent) information
- 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 concerns
- ***Slow and costly***
- ***Limited reproducibility***

→ ***Not appropriate for many research & production environments***

Synthetic Specialists

What we get from genAI

The Generalist *(horizontal AI)*

- Trained on vast body of information
- Ability to carry out many different tasks
- Excels at standard benchmarks

What we need for analytics

The Specialist *(vertical AI)*

- Identify a ***specific*** construct
- No third-party constraints
- Accurate, efficient, reproducible

just a tiny piece = vertical AI

Synthetic Specialists …

… approximate powerful AI with *much smaller, open-source Models*

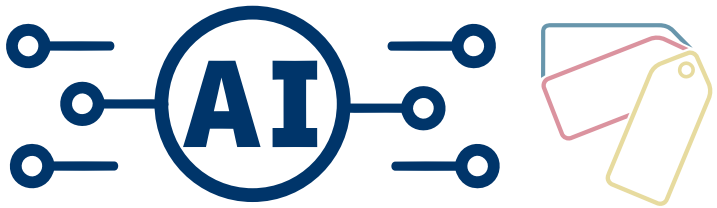
- Fine-tune a pretrained model on specific task …
- … using training data labeled once by powerful genAI

Pull the Lollypop out of the Candy Store

Creating a Synthetic Specialist

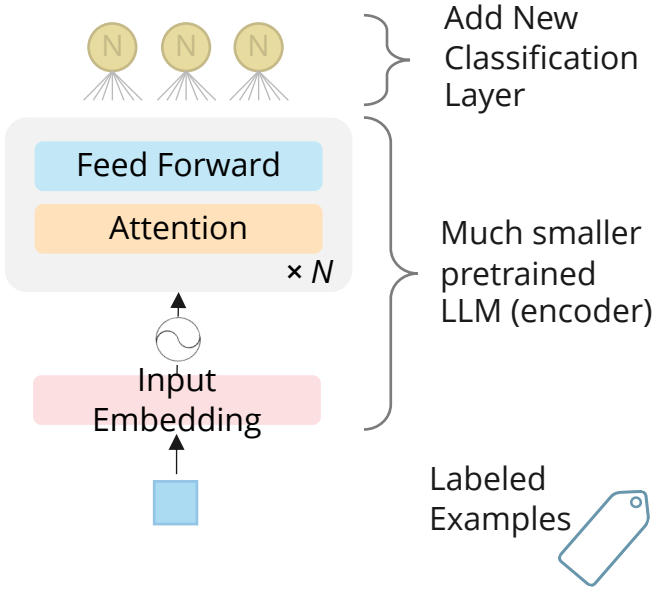
1. **Collect Data** and preprocess

2. Label Data with ***Horizontal genAI Model***



Generative **P**retrained **T**ransformer (decoder) *2 Billion to 1.7 Trillion Parameters*

3. Fine-tune pretrained model to create ***Synthetic Specialist***



Synthetic Specialists *are* ***task specific approximations*** *of Generalist AI Models*

- *Highly scalable and efficient*
- *Run on own infrastructure (e.g., laptop)*
- *Reproducible results*
- *Free from third-party constraints*

Empirical Study

Analysts derive ***insight from*** the ***co-occurrence*** of

firms, ***brands***, and ***products*** in:

- *news*
- *financial reports*
- *online search*
- *social media*
- *…*

What ***relationships*** *might* co-occurrence ***capture***?

- *Competition (substitution)*
- *Collaboration (complementarity)*
- *Neither*

Trying to decide between @Lufthansa and @United for my Munich trip. Who's got the better legroom these days? ✈️🤔

Loving the new codeshare: booked on Lufthansa, connecting on United: seamless check-in and only one bag fee. #Teamwork in the skies! 🚀💼

Just landed in IAD with @United ✈️! Now watching an @Lufthansa Airbus A350 heading down the runway. #sustainability #aviation

Some examples of co-occurrence research

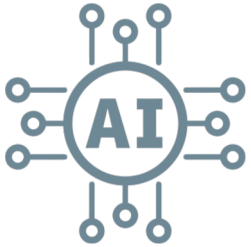
Research	Object	Data	Relationship	Qualified	Volume
Netzer et al. (2012), Marketing Science	Brands, Products	Blogs, forums, chat rooms	Competitive	***No***	~350K user posts
Tirunillai & Tellis (2014), Journal of Marketing Research	Brands	Product reviews, social media chatter	Competitive positioning	***No***	~860K reviews/posts
Araujo, Neijens & Vliegenthart (2017), International Journal of Advertising	Brands	Tweets	Brand endorsement /association	***No***	5,300 tweets
Lee & Bradlow (2011), Journal Marketing Research	Brands, Products	Customer reviews	Substitute / competitive proximity	***No***	450K reviews
He, Zha & Li (2013), Intl J. of Information Management	Brands	Social media posts	Competitive	***No***	~60K social posts
Hassan et al. (2019), Quarterly Journal of Economics	Firms	Earnings conference call transcripts	Political risk association	***No***	~170K call transcripts
Tang, Zhou & Hong (2019), J. of Risk and Financial Management	Firms	Financial news articles	Economic linkage	***No***	~2M news articles

Problem

Co-occurrence carries meaning: rich source of information

Context matters: unqualified co-occurrence can be misleading

Qualification expensive: data volume and complexity of construct



I propose a ***fast, inexpensive***, and ***reliable approach*** to qualify co-occurrence in vast unstructured data: ***Train a specialized, vertical AI model with genAI***

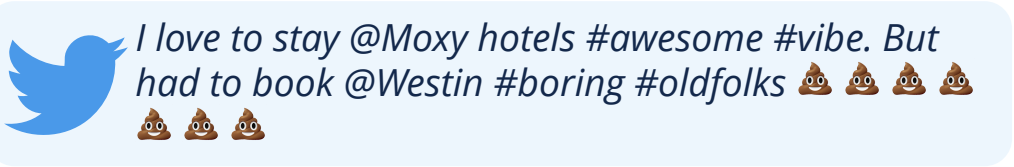
Brand relationships on Twitter

Construct of Interest: Brand Relationships
- **Competitive Intelligence**
- **Market Structure**
- **Opportunities for Collaboration**

Data: Consumers' posts on Twitter co-mentioning brands
- Vast and unstructured information source to marketers
- **Lens on hearts and minds of consumers**

Task: Identify relationship from brand co-mentions

substitutive, complementary, neither



Data collection

Brands	499	**Time**	2019 to 2021
Microblogs	Tweets co-mentioning brands	**Source**	Twitter research API

Validation Sample

N = 1,000

Human Experts

4 human experts
x 10 workshops x 1 hour = 40 hours

Generative AI

34 genAI Models
3 sets of labels per Tweet

Brand Relationships

Substitutes	33%
Complements	27%
Neither	40%

Training Sample

N = 15,000

Generative AI
Sample

3 top performers on Validation

3 sets of labels per Tweet

Brand Relationships

Substitutes	32%
Complements	27%
Neither	41%

Labeling Tweets with genAI

Preprocess

Replace links with “URL”

Parse HTML (`&` → &)

Remove excessive spaces and line breaks

(Anonymize)

User

System (Role, Task, Format)

You are an expert in social media content analysis and marketing scholar [...]
When given a Tweet, you examine it carefully. Determine the relationship between co-occurring brands [definitions ...]
Return a single word: substitutes, complements, or neither. Do not include explanations, disclaimers, or any other text.

Parameters

Model = Various
Temperature = 0
Max length = 30
Stop sequences = None
Top P = 1
Frequency penalty = 0
Presence penalty = 0

Prompt

API call

ANTHROP\C OpenAI
MISTRAL AI_ Fireworks AI

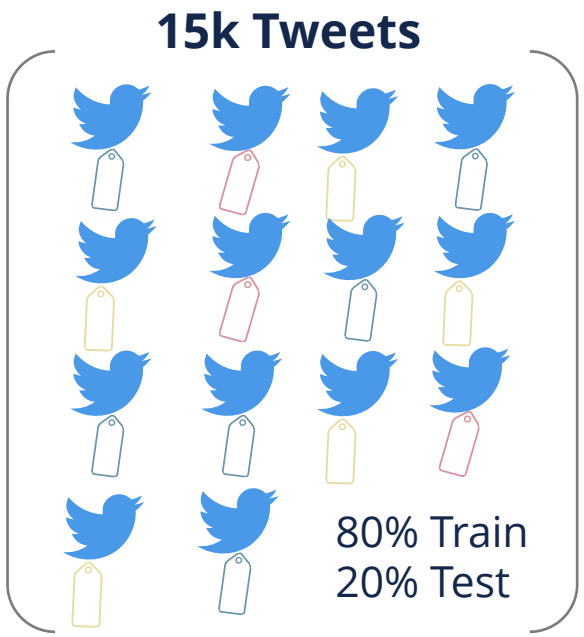
Postprocess Model Response

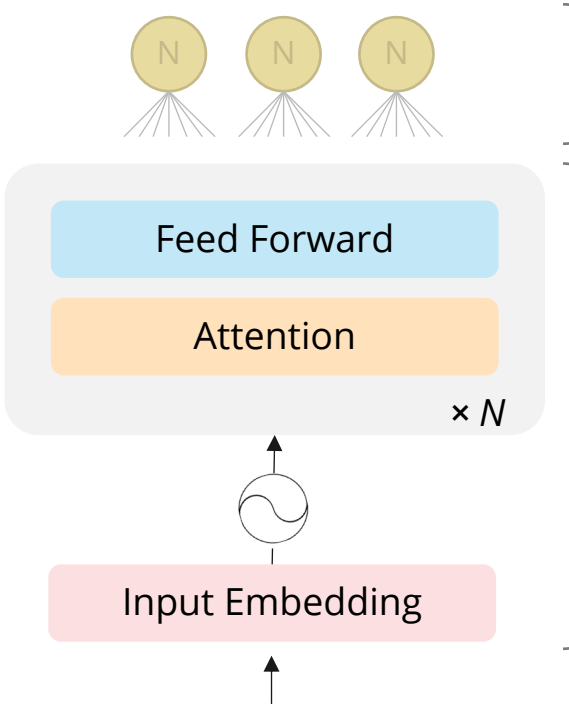
[Substitutes]

Procedure

- Label each Tweet 3 times
- Majority label

Fine-tuning a pretrained LLM





Labeled Examples

New Classification Layer

Pretrained Transformer (encoder)

Fine-tuning
- *Cross-entropy loss*
- *Additional parameters: 1,024 × 3 + 3 = 3,075*
- *Low learning rate (2e-5)*
- *Small batches (16)*
- *4 Epochs with early stopping*

BERTweet Large
Nguyen, Vu, Nguyen (2020)
- *Encoder model*
- *RoBERTa training procedure*
- *873 million English Tweets*

Training Data
- *GenAI Pooling: Majority vote of 3 models for more accurate labels:*
  - *Grok 3-mini (Reason)*
  - *Gemini 2.5 Flash (Reason)*
  - *o4-mini (Reason)*

Generative **P**retrained **T**ransformer (GPT-4o) ***~ 350 Billion*** *Parameters*

vs.

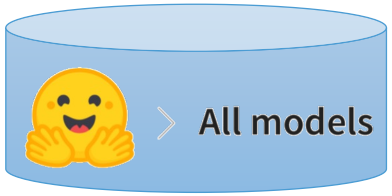
~ 354 Million *Parameters* BERTweet Large Bi-directional encoder

1,000 x smaller model

Using the brand relationship Synthetic Specialist

Oh man, the latte @Starbucks is so horrible and boring! They taste bland. Their special flavors are outdated and suck. 🤮 I wish they had something novel like a @Oreo latte!

```python
# Imports
import pandas as pd, numpy as np, warnings, torch, re
from transformers import AutoModelForSequenceClassification, AutoTokenizer
from bs4 import BeautifulSoup
warnings.filterwarnings("ignore", category=UserWarning, module='bs4')
# Helper Functions
def clean_and_parse_tweet(tweet):
    tweet = re.sub(r"https?://\S+|www\.\S+", " URL ", tweet)
    parsed = BeautifulSoup(tweet, "html.parser").get_text() if "filename" not
    return re.sub(r" +", " ", re.sub(r'^[.:]+', '', re.sub(r"\\n+|\n+", " ", p
def predict_tweet(tweet, model, tokenizer, device, threshold=0.5):
    inputs = tokenizer(tweet, return_tensors="pt", padding=True, truncation=Tr
    probs = torch.sigmoid(model(**inputs).logits).detach().cpu().numpy()[0]
    return probs, [id2label[i] for i, p in enumerate(probs) if id2label[i] in
# Setup
device = "mps" if torch.backends.mps.is_built() and torch.backends.mps.is_ava
synxp = "dmr76/mmx_classifier_microblog_ENv02"
model = AutoModelForSequenceClassification.from_pretrained(synxp).to(device)
tokenizer = AutoTokenizer.from_pretrained(synxp)
id2label = model.config.id2label
# ---->>> Define your Tweet  <<<----
tweet = "Best cushioning ever!!! 🤗🤗🤗  my zoom vomeros are the bomb🏃💨!!!  \
# Clean and Predict
cleaned_tweet = clean_and_parse_tweet(tweet)
probs, labels = predict_tweet(cleaned_tweet, model, tokenizer, device)
# Print Labels and Probabilities
print("Please don't forget to cite the paper: https://ssrn.com/abstract=45429
print(labels, probs)
```



huggingface.co/**models**

Substitutes	0.11
Complements	0.84
Neither	0.05

	Speed (Tweets/s)	Cost ($/Tweet)
OpenAI o3 via API*	~ 0.05	~ $ .009
Synthetic Specialist *on MacBook Pro***	**~ 43**	**~ 9.20e-7**

*** EUR 2.700 depreciated over 3 years, 50-watt energy consumption @ USD 0.17per kWh during inference*
** 20 watt less laptop energy consumption than local inference*

Evaluation metric

Krippendorff's α captures the degree of agreement beyond what would be expected by chance.
Here, it measures the agreement between model labels and expert labels.

$$\alpha = 1 - \frac{O}{E}$$

$\alpha = 1$	perfect agreement
$\alpha = 0$	agreement expected by chance
$\alpha < 0$	suggests systematic disagreement

$\alpha > .800$	considered ***reliable*** *
$.670 < \alpha < .800$	considered ***acceptable*** *
$\alpha < .670$	considered ***insufficient*** *

O is the observed disagreement between expert and model
E is the expected disagreement by chance.

* *Krippendorff, K. (2004). Content Analysis: An Introduction to Its Methodology. 2nd ed., Sage Publications.*

$$O = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{2(2-1)}\sum_{l=1}^{L}\sum_{m=1}^{L} o_{il}\cdot o_{im}\cdot d_{lm}$$

vs.

$$E = \frac{1}{N}\sum_{i=1}^{N}\sum_{l=1}^{L}\sum_{m=1}^{L} p_l \cdot p_m \cdot d_{lm}$$

$where$:

o_{il} is 1 if the rater assigned label l to Tweet i, otherwise 0

for nominal data (like the 4Ps): $d_{lm} = 1 \;\; if \;\; l \neq m$
$d_{lm} = 0 \;\; if \;\; l = m$

$where$:

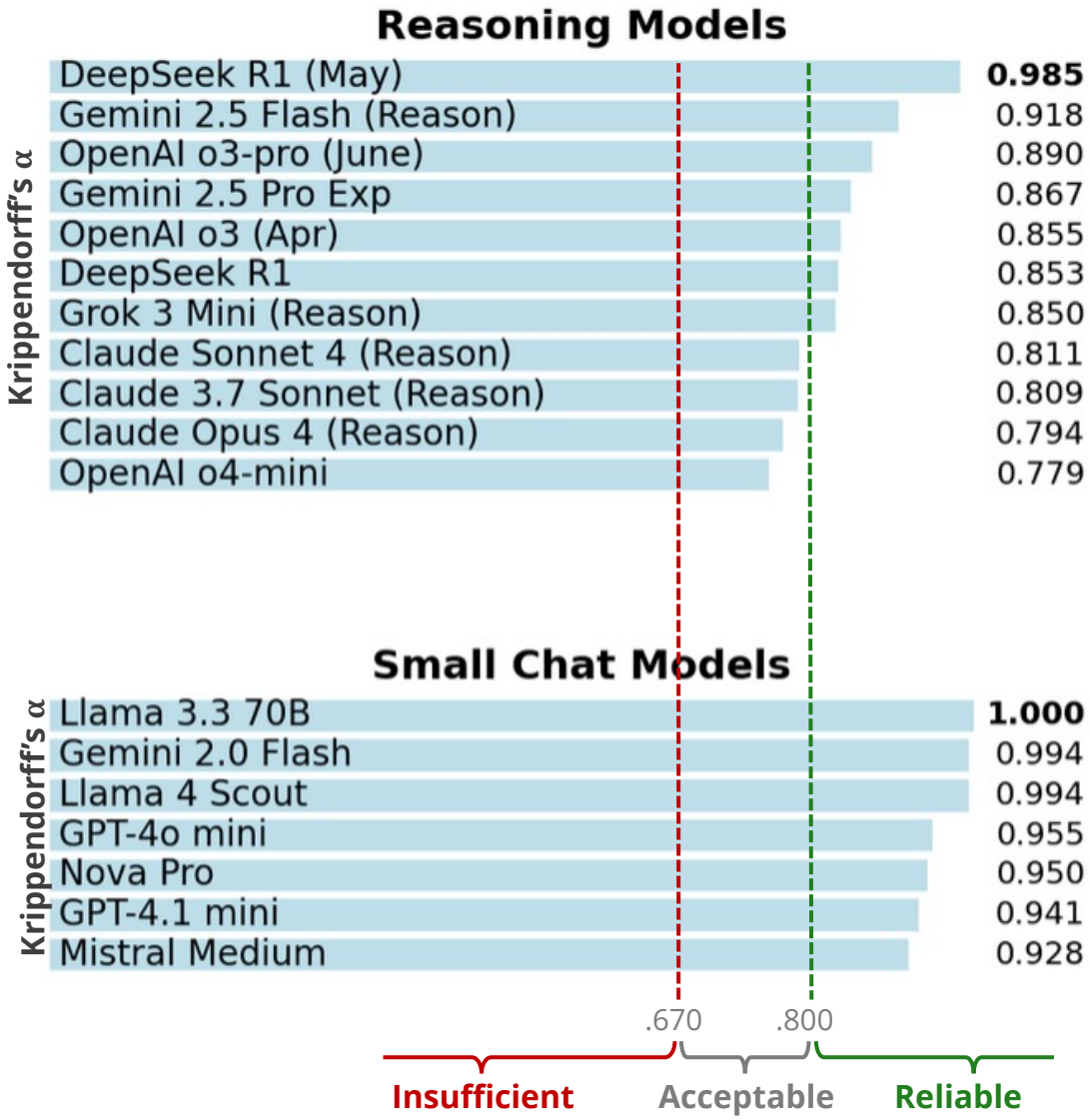
p_l is the proportion of Tweets labeled l, calculated as:

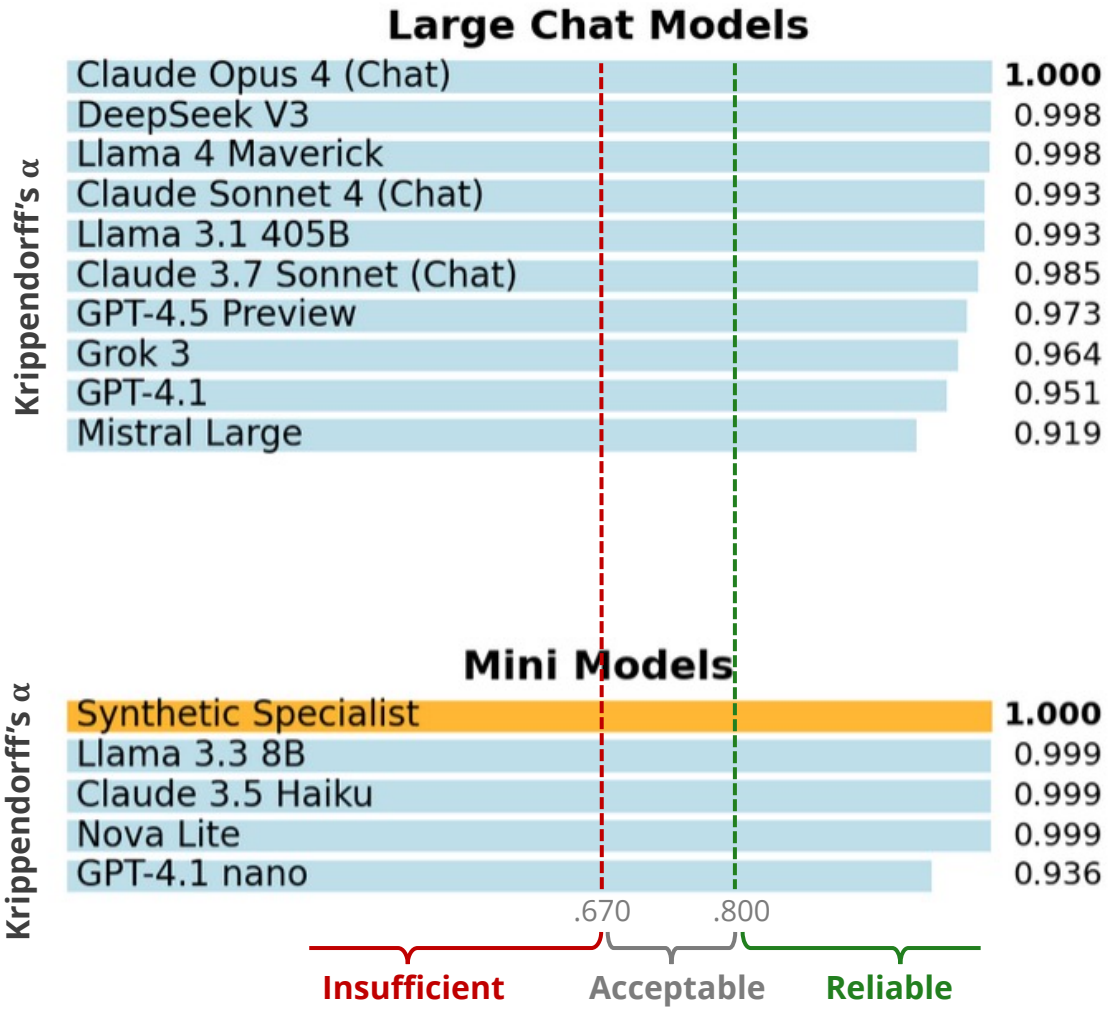
$$p_l = \frac{\sum_{i=1}^{N} o_{il}}{N \cdot 2}$$

N	total number of Tweets (examples) being labeled.
$m = 2$	number of raters: Experts and Model(s).
$L = 4$	number of labels corresponding to the 4Ps of Marketing (Product, Price, Place, Promotion).

o_{il}	observed frequency of raters assigning label l to Tweet i.
p_l	proportion of labels assigned to category l across all Tweets.
d_{lm}	disagreement function for categories l and m.

Reproducibility: Label agreement across 3 runs

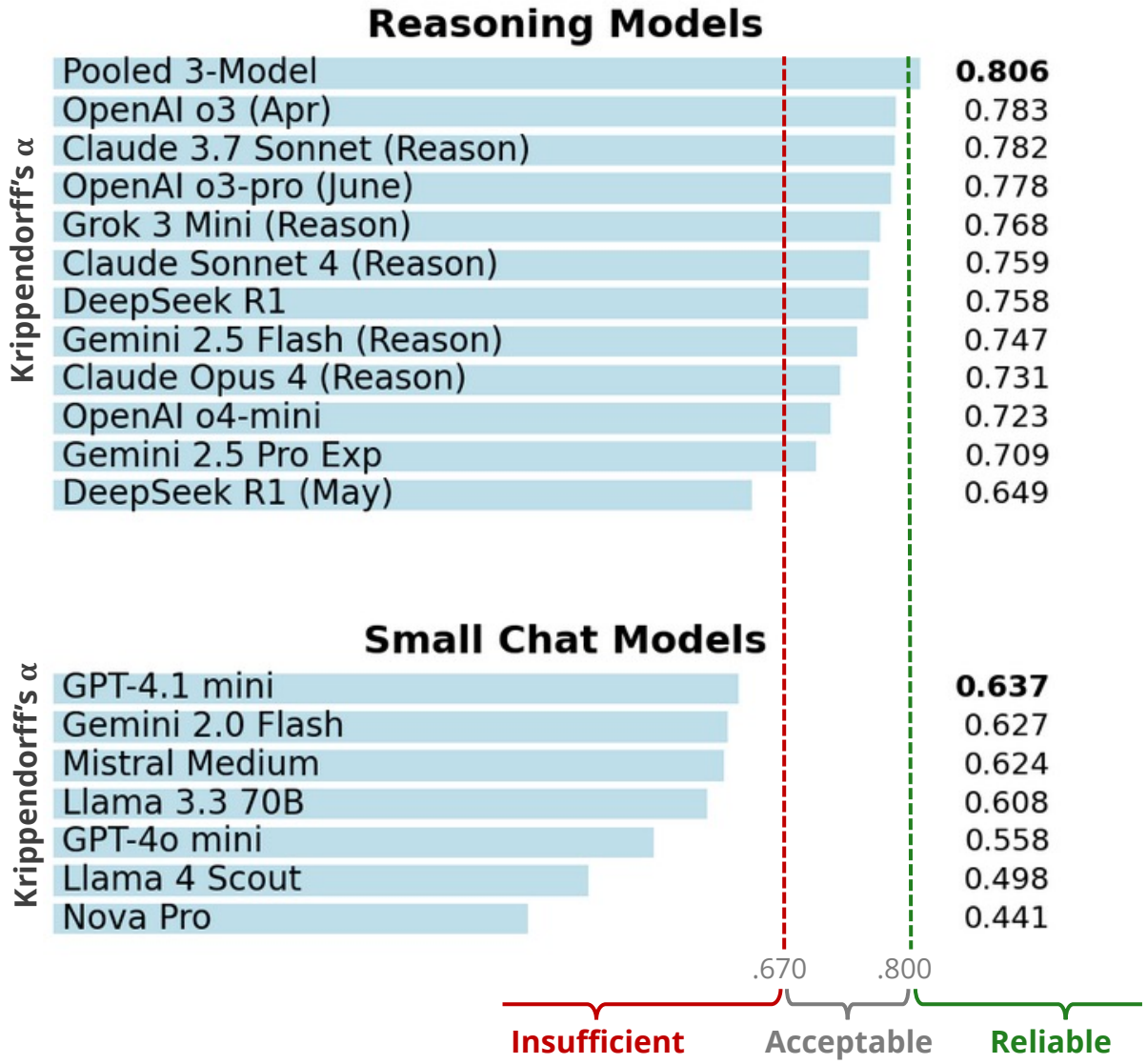


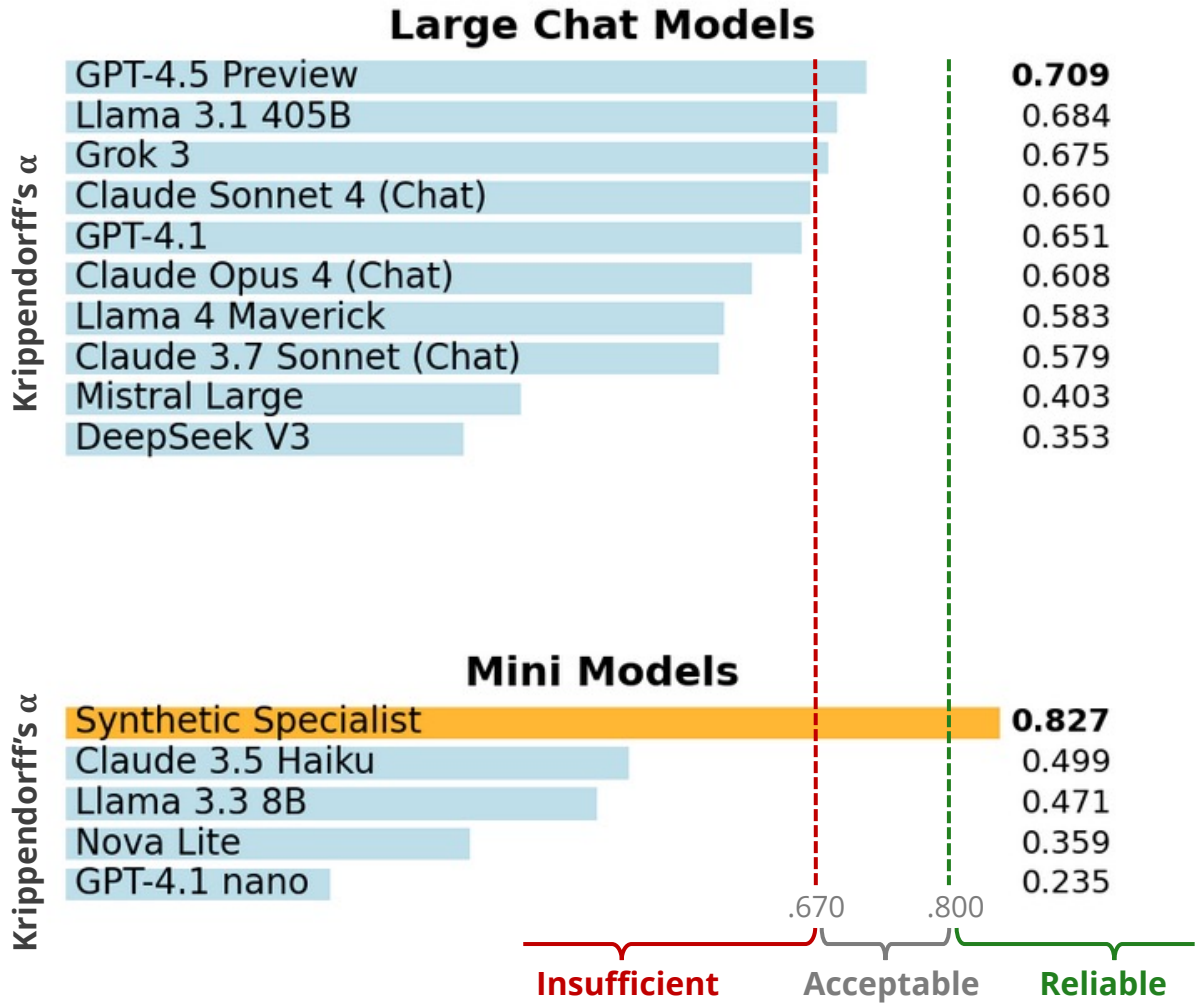


Krippendorff's α captures the degree of agreement beyond what would be expected by chance.

GenAI Pooling: *Grok 3-mini (Reason), Gemini 2.5 Flash (Reason), o4-mini (Reason)*

Agreement with human expert labels

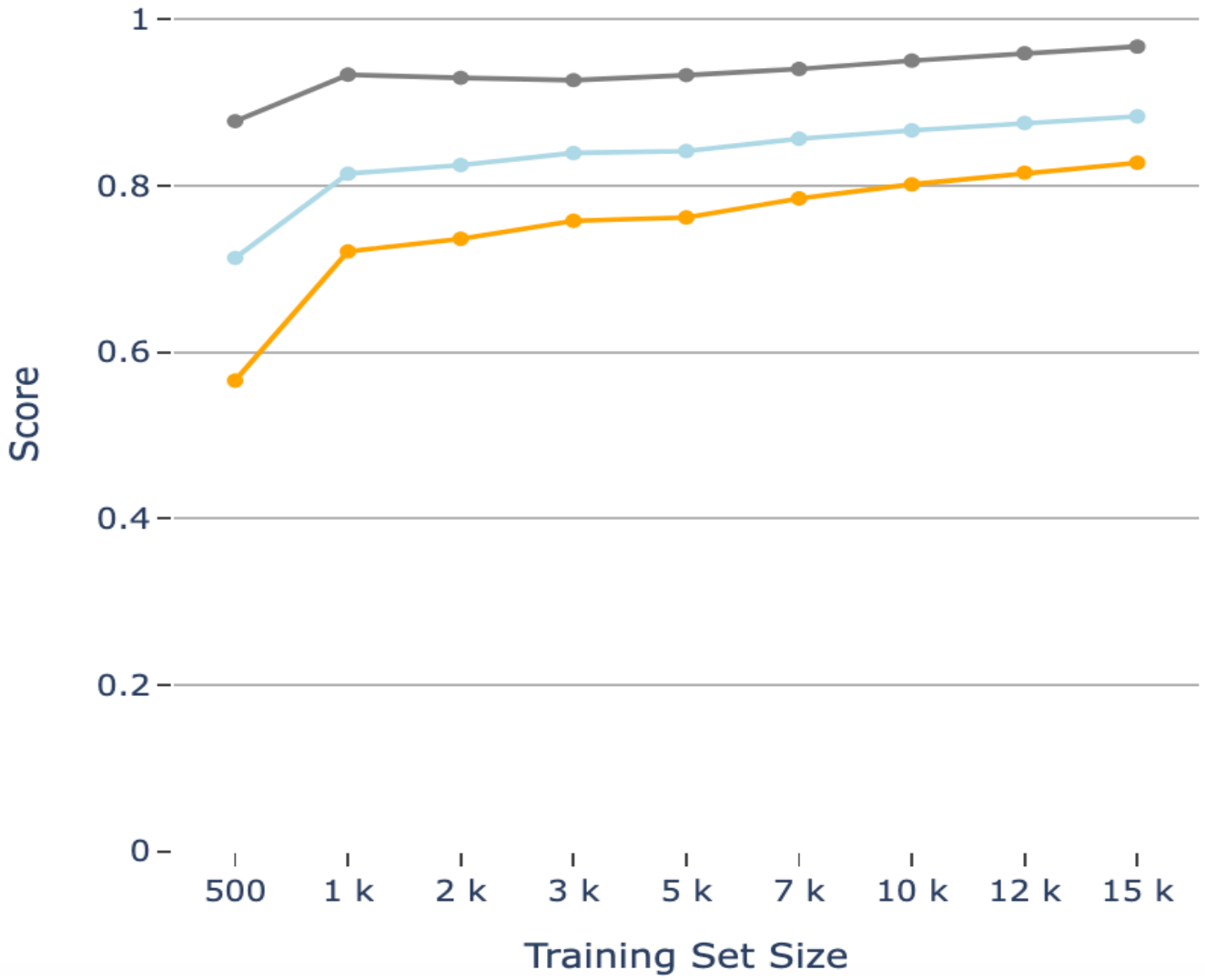




Krippendorff's α captures the degree of agreement beyond what would be expected by chance.

Pooled 3-Model: *Grok 3-mini (Reason), Gemini 2.5 Flash (Reason), o4-mini (Reason)*

Ablation: Training sample size



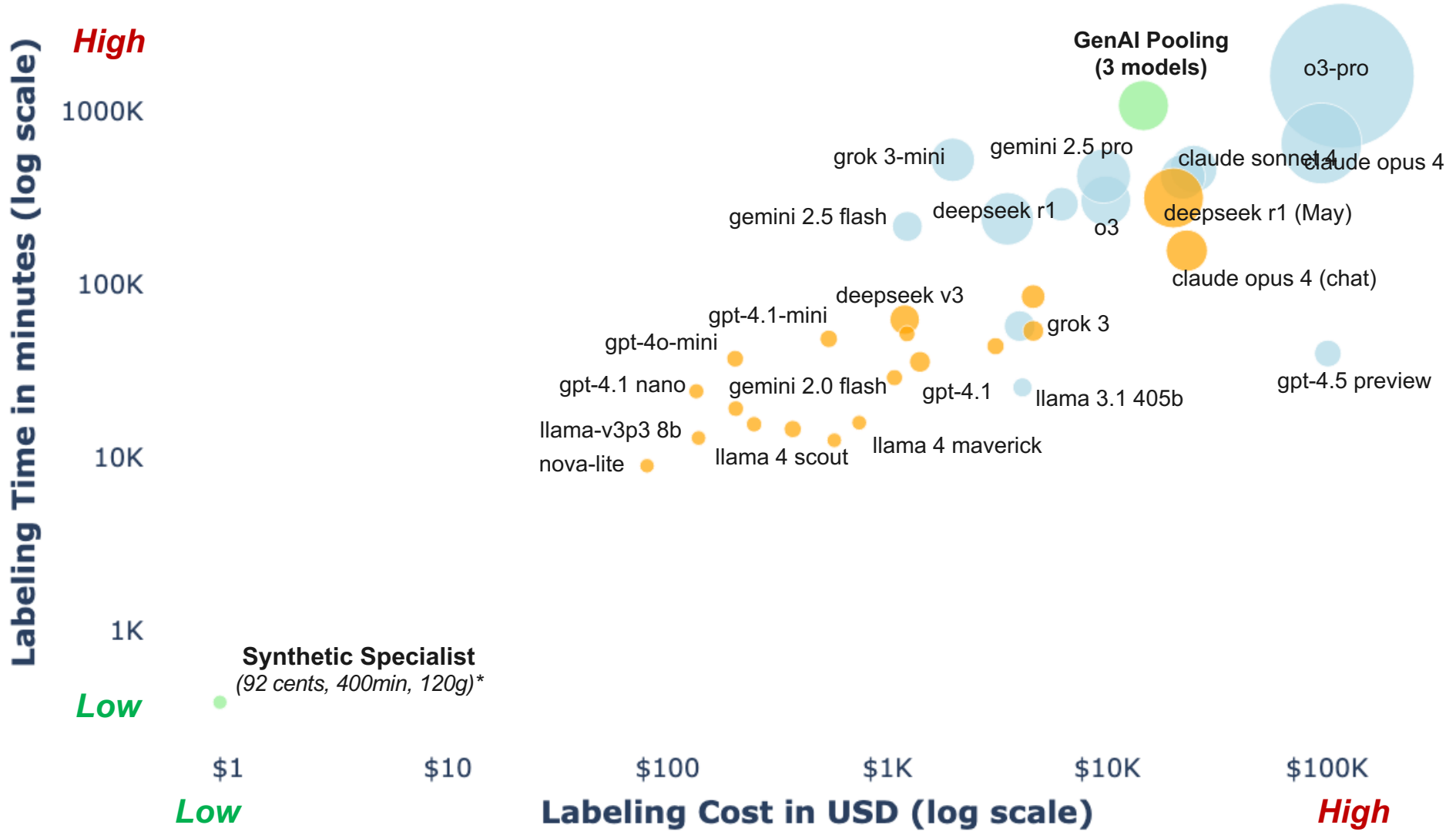
Area under Curve	accounts for ***different probability thresholds***
F1 Score	Harmonic ***mean of precision*** and ***recall***
Krippendorff's α	accounts for ***chance agreement***

Sampling Procedure

1. Label ***pool of 15,000 tweets*** with ***three GenAI Models*** and take majority vote.
   Gemini 2.5 Flash, Grok 3-mini, o4-mini
2. Randomly ***sample N labeled Tweets*** from labeled tweet pool for ***different N***
3. Evaluate ***label agreement with validation sample*** for each *N.*

Impact on efficiency, performance, and CO_2 emission

for labeling 1 million Tweets



Expert Label Agreement
Krippendorff's α
- Reliable (α ≥ .800)
- Acceptable (.670 ≤ α < .800)
- Insufficient (α < .670)

CO_2 Emissions

< 25 kg ~ 2,000 kg > 10,000 kg

Notes

GenAI Pooling 3 runs each Gemini 2.5 Flash, Grok 3-mini, o4-mini

Calculations assume INT8 quantization for improved model efficiency.

CO_2 emissions based on 369g per kWh electricity of U.S. energy mix.

Labeling time does not consider potential parallelization.

* one time cost for 15K training labels: $141, 260 minutes, 31 Kg CO_2

Implications

Organizations

Powerful, scalable, and accessible solution for complex classification tasks

- free of third-party constraints
- mitigate privacy and confidentiality concerns
- easily updated or replaced when world and/or task changes

Research

Inform downstream tasks as DV or IV

- access to complex (marketing) constructs
- answer research questions / test hypotheses
- replicability of research that leverages AI

Society

Greater Accessibility at lower Cost

- minimal hardware requirements
- local (edge) AI
- tiny CO_2 footprint

Brand Relationship Synthetic Specialist

6% more accurate
850 times faster*
10,000 times lower cost*
99.99% less CO_2 emissions*

than best genAI

* after one time cost for 15K training labels: $141, 260 minutes, 31 Kg CO_2

The simple math behind it all

Reasoning Models

More powerful, but more expensive

GPT-4.1 $8.0 / 1 M output tokens
o4-mini $4.4 / 1 M output tokens,
but ~20× more tokens per response

→ ***a 45 % price cut turns into a 1000% bigger bill***
when verbosity is factored in

Agentic AI

More complex tasks, automated

More models generating more output per task

→ *Even mode **$$$** and **CO_2***

The Bottom Line

Users ⇧ × Applications ⇧ × Reasoning × Agentic AI = ⇧ **Cost & CO_2**

source: coachapproachministries.org

Looking Ahead

After watching this async lecture:

1. Read Capstone Assignment on Canvas
2. Form your Capstone Team *TODAY* on Canvas
3. Study construct of interests for your Capstone: must pick one from here

Next Class: Tuesday, March 31

Fine-tuning pretrained LLMs

Always check Schedule and Prep on: https://www.ringel.ai/UNC/2026/BUSI488/Overview/