ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Quant Studies in **Social Media Marketing**: Data, Measurements, Models and Apps

Dr. Francisco Villarroel Ordenes

Professor of Marketing

University of Bologna

Most of my work relates to modeling social media (content) data…

JOURNAL ARTICLE

Cutting through Content Clutter: How Speech and Image Acts Drive Consumer Sharing of Social Media Brand Messages

Francisco Villarroel Ordenes ✉, Dhruv Grewal, Stephan Ludwig, Ko De Ruyter, Dominik Mahr, Martin Wetzels   Author Notes

Journal of Consumer Research, Volume 45, Issue 5, February 2019, Pages 988–1012, https://doi.org/10.1093/jcr/ucy032

JOURNAL ARTICLE

Unveiling What Is Written in the Stars: Analyzing Explicit, Implicit, and Discourse Patterns of Sentiment in Social Media

Francisco Villarroel Ordenes ✉, Stephan Ludwig, Ko de Ruyter, Dhruv Grewal, Martin Wetzels   Author Notes

Journal of Consumer Research, Volume 43, Issue 6, April 2017, Pages 875–894, https://doi.org/10.1093/jcr/ucw070

Article

AMERICAN MARKETING ASSOCIATION

Complaint De-Escalation Strategies on Social Media

Journal of Marketing
1-22
© American Marketing Association 2022
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/00222429221119977
journals.sagepub.com/home/jmx
SAGE

Dennis Herhausen, Lauren Grewal, Krista Hill Cummings, Anne L. Roggeveen, Francisco Villarroel Ordenes, and Dhruv Grewal

Article

AMERICAN MARKETING ASSOCIATION

How High-Arousal Language Shapes Micro- Versus Macro-Influencers' Impact

Journal of Marketing
1-22
© American Marketing Association 2024
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/00222429231207636
journals.sagepub.com/home/jmx
Sage

Giovanni Luca Cascio Rizzo, Francisco Villarroel Ordenes, Rumen Pozharliev, Matteo De Angelis, and Michele Costabile

Article

AMERICAN MARKETING ASSOCIATION

Standing Out While Fitting In: Visual Design of Text Overlays in Social Media Communication

Journal of Marketing
1-20
© American Marketing Association 2025
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/00222429251322773
journals.sagepub.com/home/jmx
Sage

Stefania Farace, Francisco Villarroel Ordenes, Dennis Herhausen, Dhruv Grewal, and Ko de Ruyter

Current Work

Virtual Influencers

Customer experience and Brand Impact

Relational communication in Service Chats

Marketing Insights from Unstructured Data (UD)

Keeping the retail mall alive

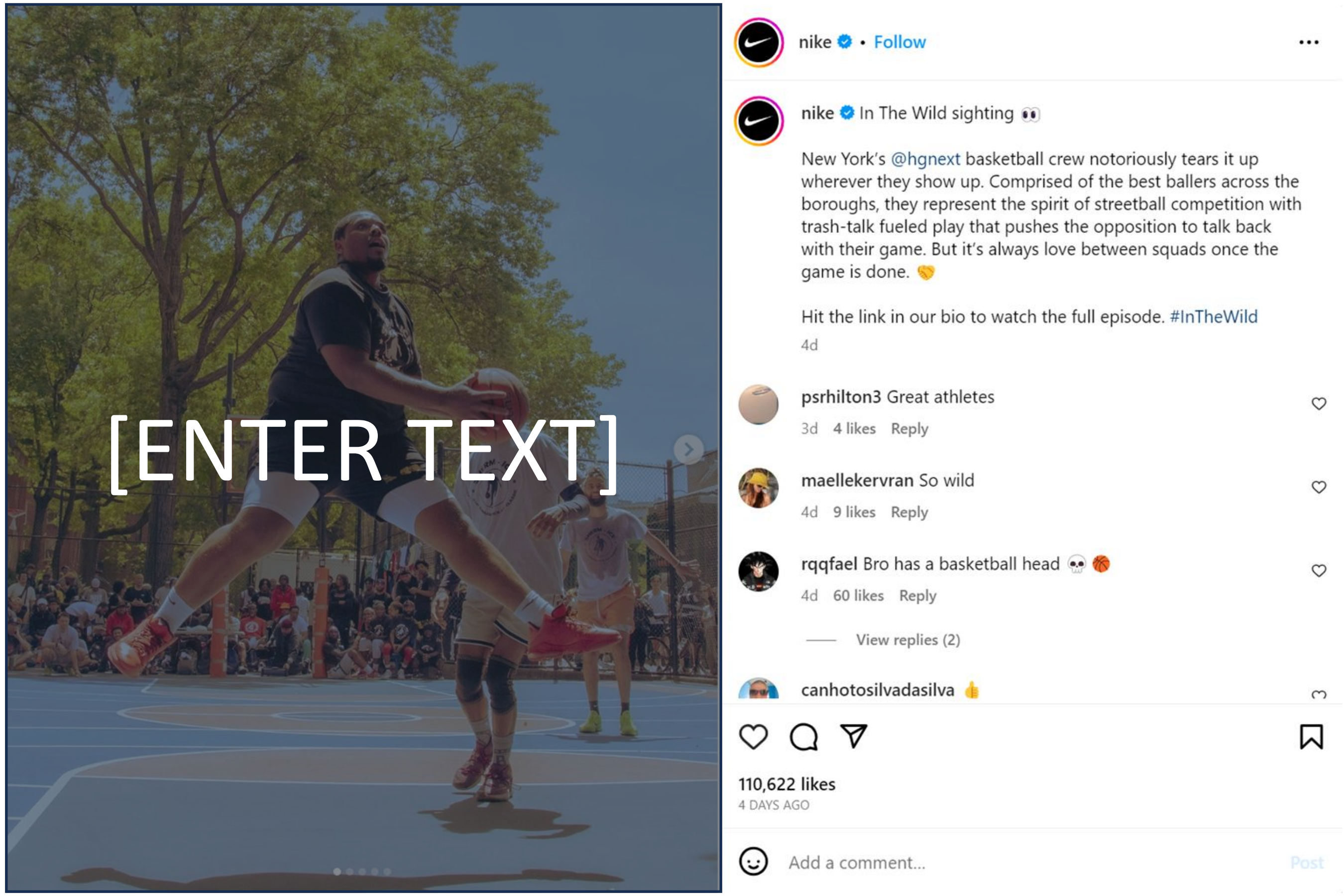
Crypto greedy e-WOM and adoption

Podcast Conversations

Standing out while fitting in

(With S. Farace, D. Herhause, D. Grewal and K. de Ruyter)

[LINK]



Data Scraping

Level of difficulty

python

R

Open for Innovation KNIME

APIFY

upwork

Data scraping from Social Media

Alex Smith

Overview Messages Contract details

Description

This job consists in collecting a data-set of 50 Instagram accounts including all their posts (image or videos) and the engagement they generate (likes, views, comments). I would need an excel file with a column for the ID of the post, the account (e.g., Nike), the URL of the post, the text of the post, number of likes, number of views, and text of the comments (from followers). Having the text of the comments will make the table longer as each brand post will have many comments. Because of comments I might need a separate table for these with the ID of the brand post they refer to. Also I will need a folder with the images and videos. The image/video name should be the same as the ID of the post. So I can identify it. less

View original proposals

Summary

Contract type	Fixed-price
Total spent	$120.00
Start date	Feb 23, 2020

Constructs and Measurement

Focus on two TO salient properties. Overlay **SIZE AND CENTRALITY** (relevant stylistic features) (Pieters and Wedel 2004; Sample, Hagtvedt, and Brasel 2020)

A visual (salient) property that has been substantially studied in marketing with positive behavioral outcomes is **DYNAMISM** (Cian, Elder and Krishna 2014)

TO Size

TO Centrality

High Dynamism (Dynamic)

Low Dynamism (Static)

How to Measure Dynamism in Images?

- Handled empirically in the second field study using Instagram data
- Used annotated social media images from *X* (Study 1) from 8 brands to created "type of image" classifier. N= 6,181
- Augmented this sample to fit the context with 2,075 annotated Instagram images
- Split on Static-Dynamic Scale to perform binary classification task
  - 1-3 = Static (N=6,104)
  - 4-7 = Dynamic (N=2,152)

Steps to Develop the Image Classifier

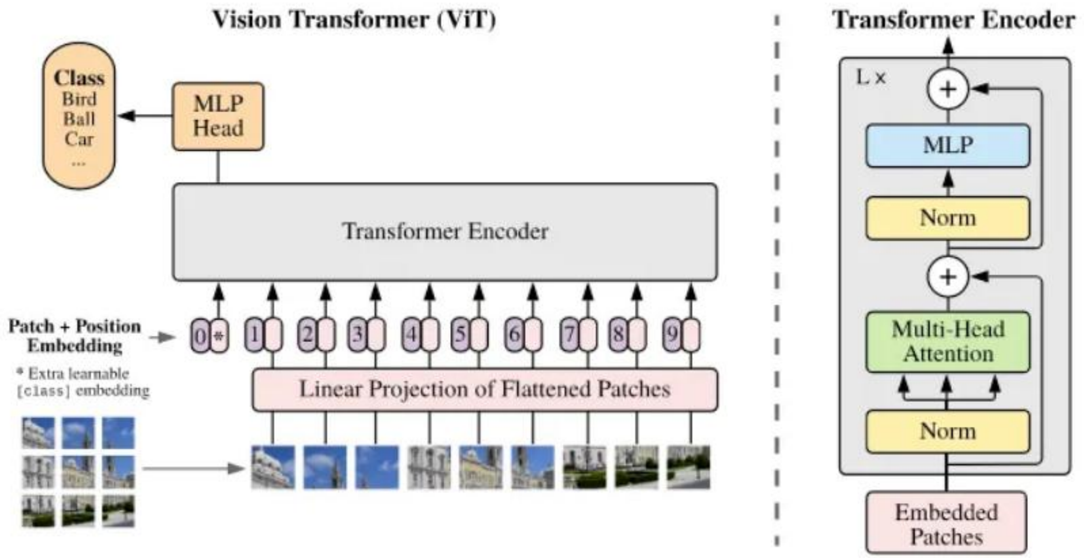
1. Resize (150*150) & Rescale (1/255)

2. Training, Validation and Testing (80/10/10)

3. Vision Transformer (ViT) a pre-trained DL model. Converts image into an embedding, representation includes global/local dependencies, that can be used for classification tasks

4. Grid search for hyper-parameter optimization (batch size, learning rate)

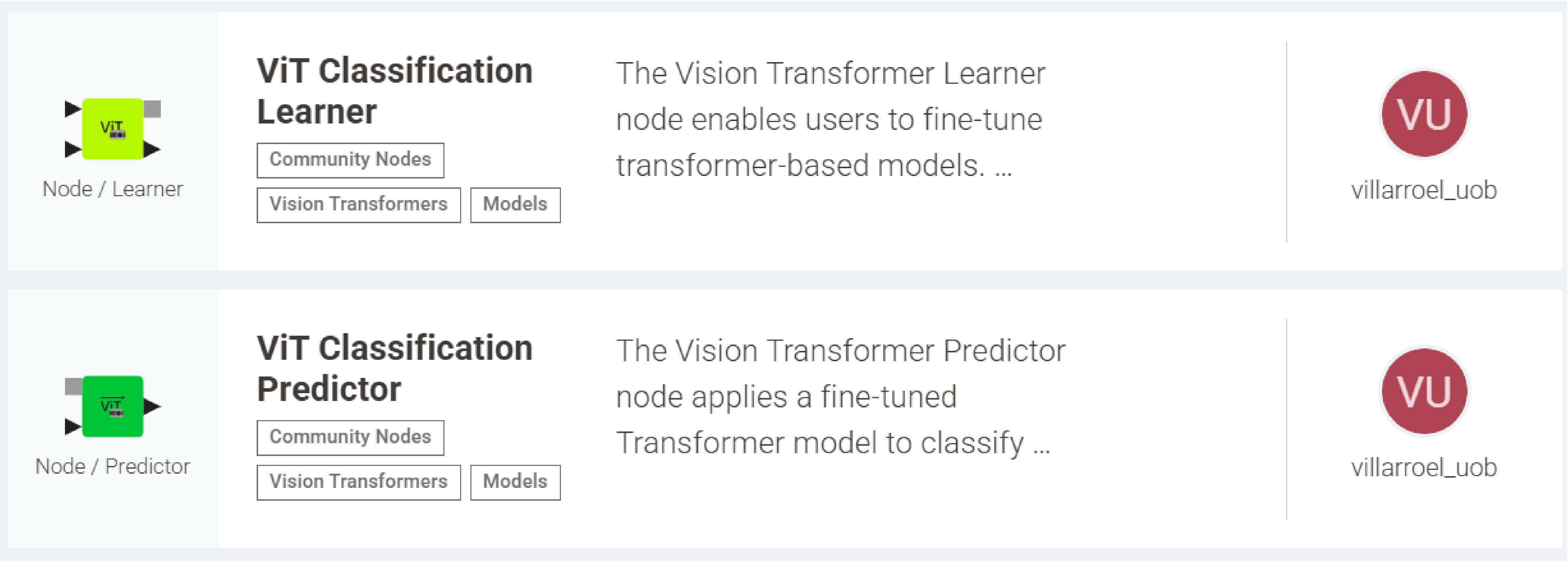
5. Use of dropout regularization to hidden and attention layers avoid overfitting



ViT Package with Low Code

Open for Innovation
KNIME

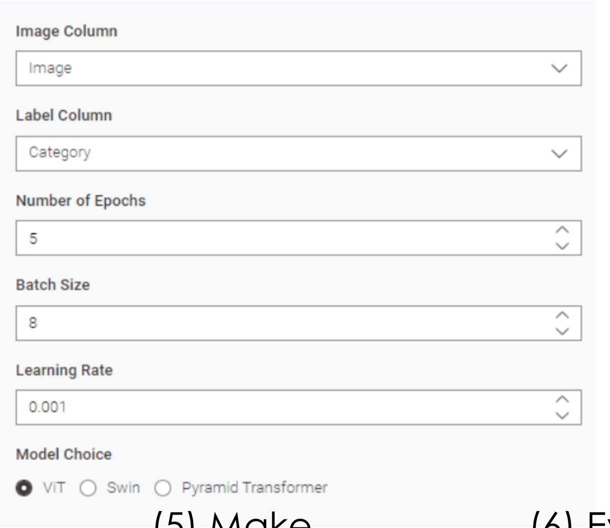
We created two configurable nodes in visual coding language (no scripting required to implement vision transformers. See extension **here**



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Example workflow for Vision Transformer Extension

See workflow implementation **here**



(1) Read "labeled" Images from a folder

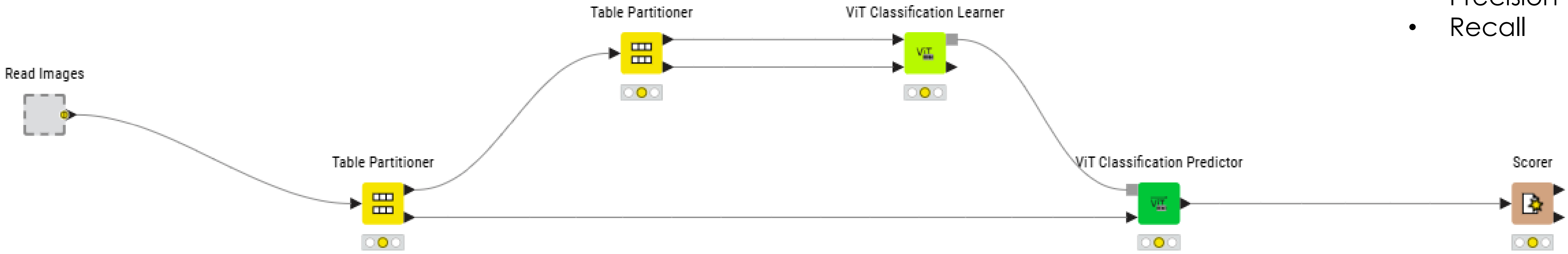
(2) Training vs. Testing

(3) Training and Validation

(4) Configure a ViT algorithm

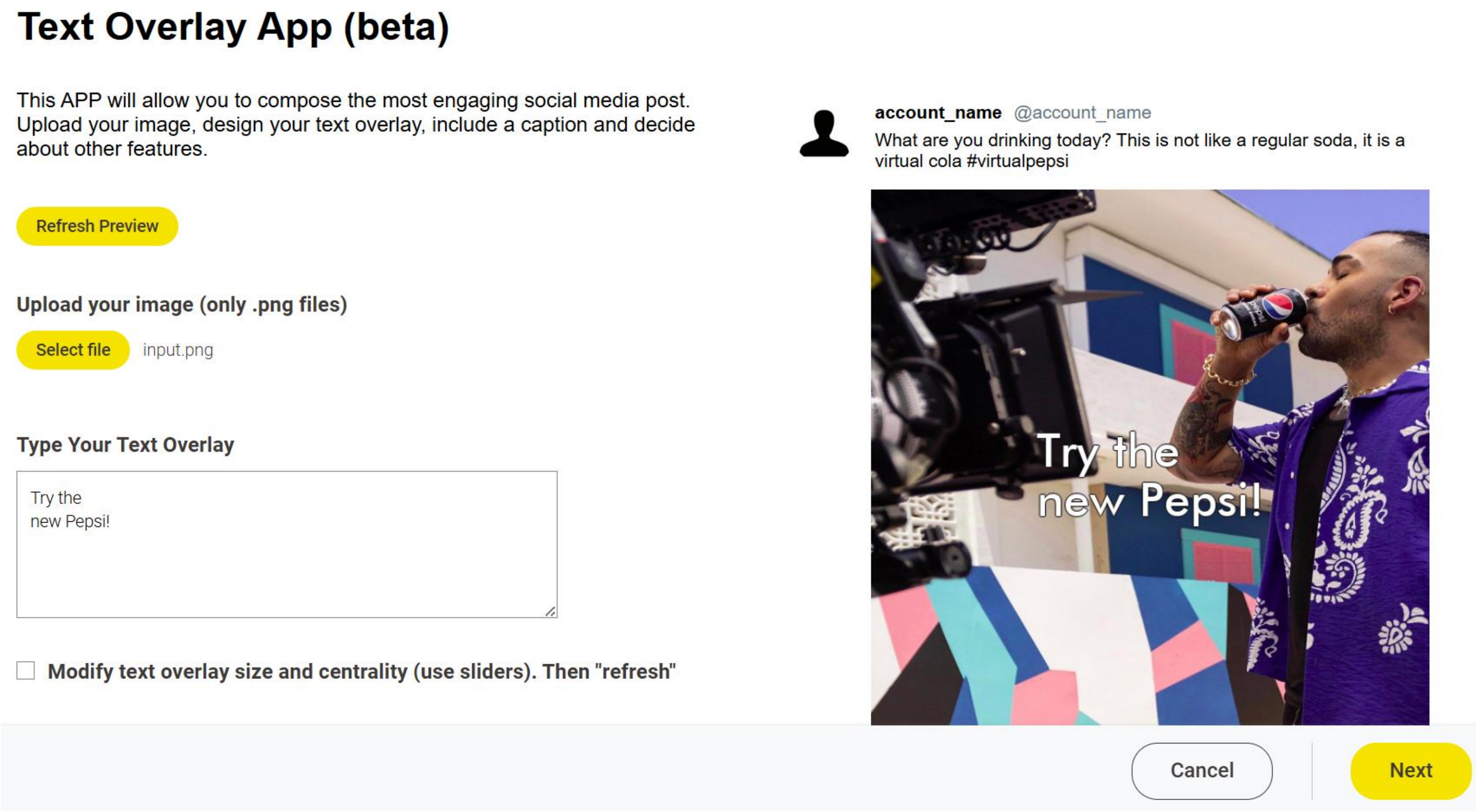
(5) Make Predictions

(6) Evaluation
- Accuracy
- Precision
- Recall



Apps can enhance the value of a practical contribution. Example from Farace et al. (2025) here

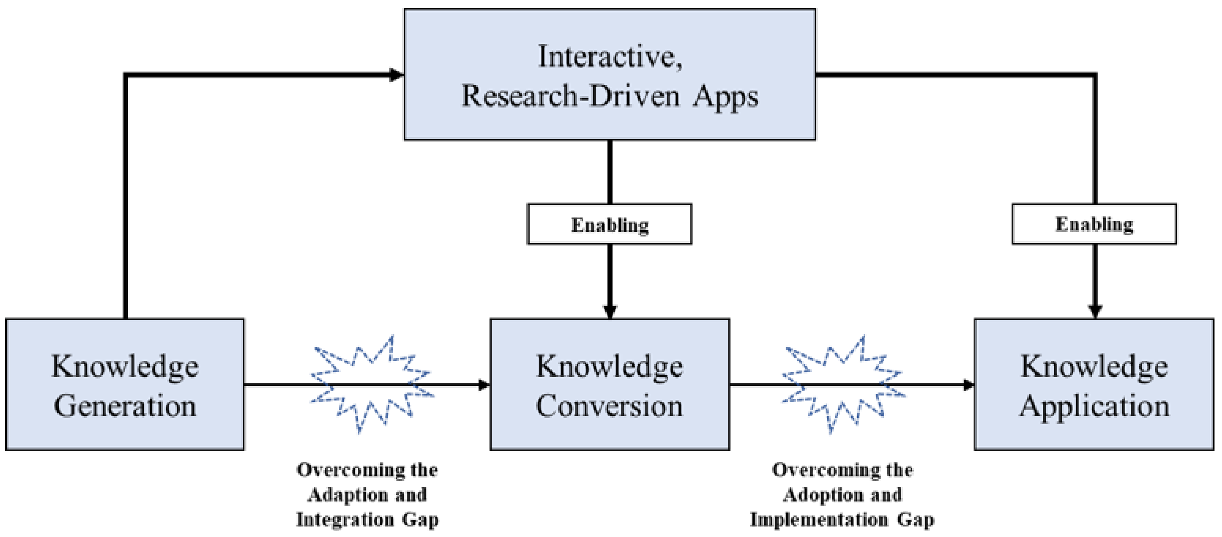
Text Overlays Data App



Why are Interactive Research Driven (IRD) Apps Important?
(with K. Pikal and D. Herhausen 2025)

A **research-driven app** is “an online interactive tool that provides a deeper understanding of the *usability of the research contribution*.”(Chintagunta et al. 2022)

1B: The Marketing Science Value Chain with Interactive, Research-Driven Apps



Note. The underlying diffusion process of the Marketing Science Value Chain is based on Roberts et al. (2014).

Authors
- Associated with more citations
- Outsourced or in-house
- Costs between 5K and 20K
- Target is manager and students

Managers:
- Increases perceptions of interest and relevance of research

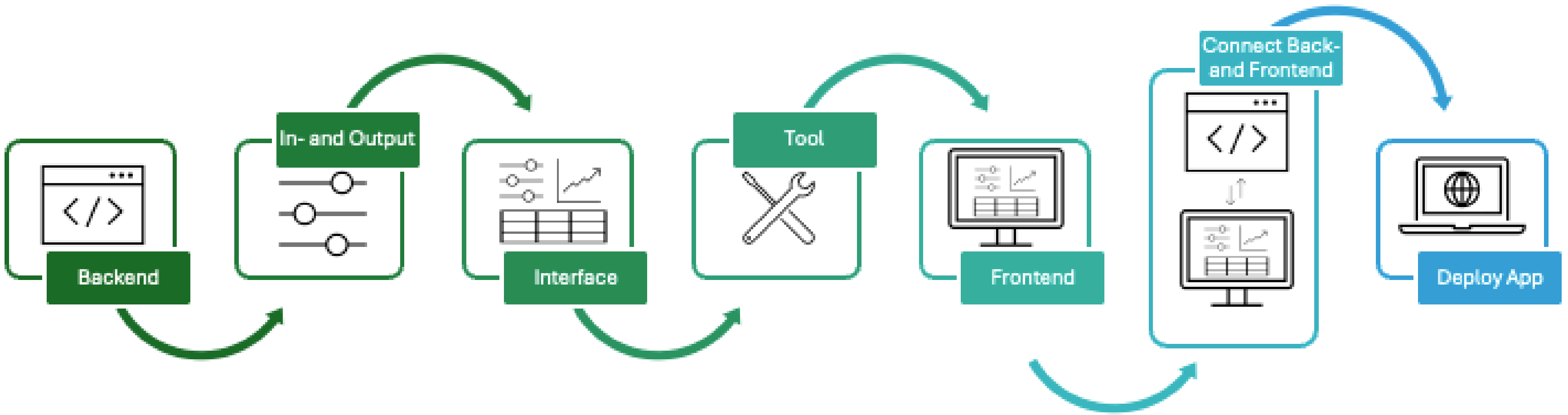
Students:
- Increases perceptions of interesting, useful and relevant

What is the Potential of IRD Apps?

We estimate that **about 25% of marketing articles** have the potential to add an IRD app; 8% predictors, 5% optimizers and recommenders, 9% explorers, and 3% converters.

- E.g., Nguyen, Johnson, and Tsiros (2023) who use large-language models to optimize e-mail headlines have potential for an optimizer and recommender app

How to develop IRD Apps?



(with K. Pikal and D. Herhausen 2025)

Consumer Engagement with **Virtual Influencers:** The Power of Social Tie Presence in Visual Content

(With L. Cascio Rizzo and J. Berger)

LINK

Conceptual Model

- **H1A: Social tie presence** increases engagement with virtual influencers' posts and likelihood to choose the product advertised.

- **H1B:** The effects are serially mediated by **anthropomorphism and perceived trust**.

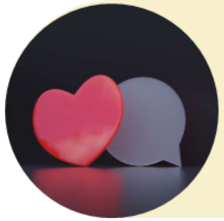
FIGURE 2: Conceptual Model

Social Tie Presence → Anthropomorphism → Perceived Trust → Impact

Field Study and Measurement

Sample

10k virtual influencer posts, 20k photos; 28 virtual influencers; multiple industries

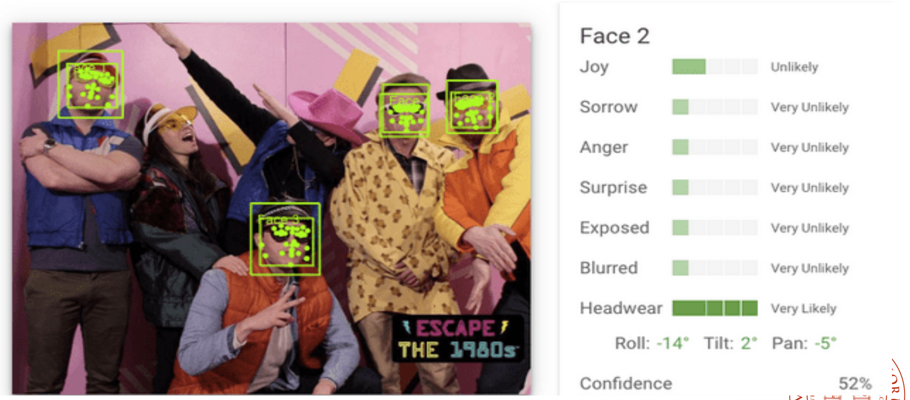


DV

Instagram Engagement (likes + shares)

Focal Variables

Social Tie Presence (Google Cloud Vision; Li and Xie 2020)



	(1) Main Model
Social Tie Presence	**1.207*** (.046)**

Controls:

- Virtual Influencer (e.g., follower count, post count, if verified)
- Image (e.g., visual complexity, emotions, color dominance, saturation)
- Text (e.g., topics, wordcount, questions, emojis, arousal, readability)
- Other: other's followers, time FE

Robustness Checks:

- Selection: is the inclusion random? propensity score matching (2,990 posts, half with/without companion)
- Influencer Heterogeneity: influencer FE
- Model Specification: OLS with log DV, carry-over effects
- Alternative Measures: likes and comments (separately), engagement rate
- Other Sources of Endogeneity: active product consumption, objects (2,005!), social proof , intimacy

TABLE 5. Robustness Tests

	What We Test	How We Test
Influencer heterogeneity	Do the effects depend on influencer heterogeneity?	Influencer fixed effects with cluster SEs (IRR = 1.086, SE = .041, z = 2.20, p = .028; Table WA3, Model 1).
Selection bias	Are the effects driven by the particular sample of influencers used?	Propensity score matching (IRR = 1.090, SE = .048, z = 1.96, p = .050; Table WA3, Model 2).
Modeling approach	Do data ranges make the use of count distributions less appropriate?	OLS with log transformed DV (b = .094, SE = .041, z = 2.30, p = .021; Table WA3, Model 3).
Alternative measures	Do the results hold for likes only?	Likes as DV (IRR = 1.213, SE = .048, z = 4.88, p < .001; Table WA3, Model 4).
	Do the result hold for engagement weighted by follower count?	Engagement rate as DV (b = .958, SE = .310, t = 3.09, p = .002; Table WA3, Model 5).
	Do the results hold for positivity in followers' comments?	Valence as DV (b = .017, SE = .005, t = 3.41, p = .001)
Other sources of endogeneity	Are posts with social ties more likely to depict active product consumption?	Manual annotation of a random sample of 500 posts for product use. Social tie presence and product use were not related (r = .001).
	Are the effects driven by the particular objects depicted in photos?	Controlling for 2,005 objects, the effect of social ties persists (IRR = 1.132, SE = .048, z = 3.31, p < .001).
	Are the effects driven by posts with no one in the image (including the influencer) being detrimental?	Posts that include a social tie receive more engagement than posts containing no one (IRR = 1.439, SE = .113, z = 4.63, p < .001) or just the virtual influencer, IRR = 1.188, SE = .047, z = 4.36, p < .001).
	Are the effects robust to the disclosure of the virtual influencer's nature?	Controlling for disclosure presence, the effects hold (approach 1: IRR = 1.228, SE = .047, z = 5.32; p < .001; approach 2: IRR = 1.210, SE = .047, z = 4.91; p < .001).
Alternative explanations	Might the effects be driven by social proof or intimacy?	No increase or decrease in engagement when more others are present (IRR = .982, SE = .012, z = –1.48; p = .139).

Exploring Mediation?

- Traditionally we use field data to hypothesize main effects and moderation.
- But, how to explore mediation?
- THEORY!

We are hypothesizing that social ties increase engagement **because they trigger humanness perceptions**

-> then **the effect might be weakened or strengthened** in certain field scenarios

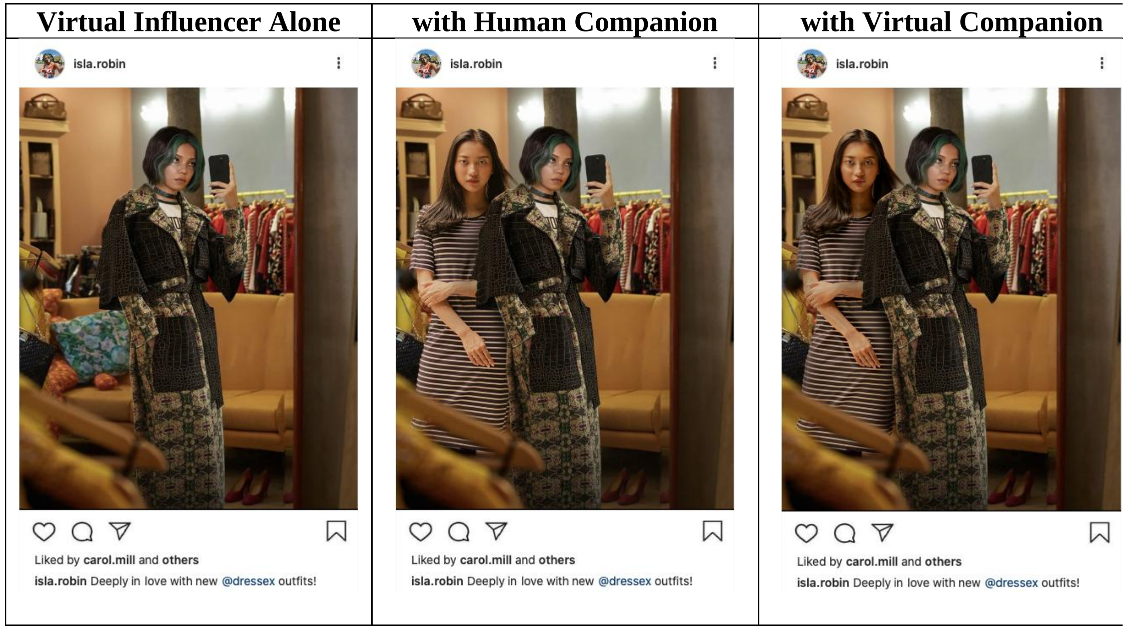
Example of Human (left) and Virtual (right) Ties in the Field

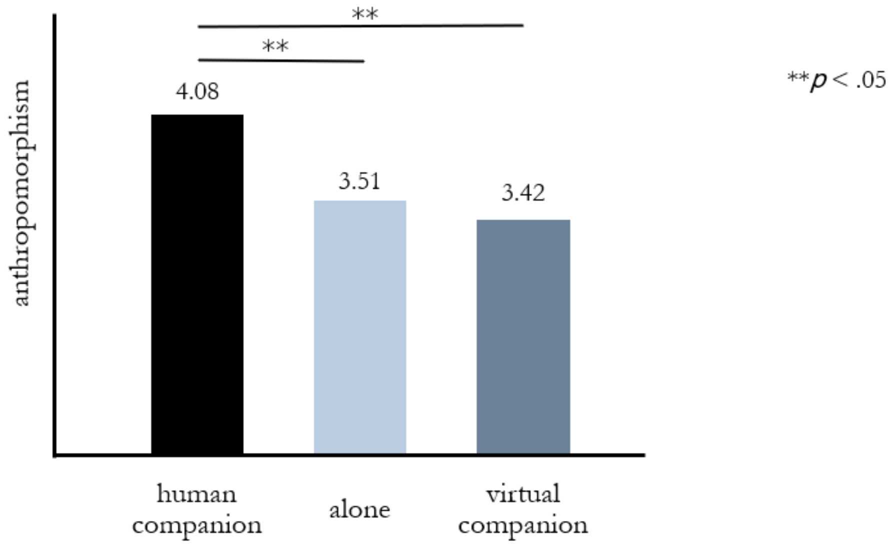
Note: These posts also feature textual mentions of these others, which facilitates verification based on their profiles

- If the companion is non human, then Social Ties effect should dissapear, because there won't be antropomorphism (real interactions)
- The companion was a human (in 1,383 posts) or a virtual character (in 173 posts)
- Social tie's effects go away when the other is a virtual (IRR = 1.131; SE = .115; t = 1.22; p = .223).

Causality? – Only with Experiments

- N = 227; Prolific
- 3 (VI alone vs. human companion vs. virtual companion)
- Measures: Anthropomorphism, Trust, Engagement





How Discourse Concentration and Content Valence Drive Podcast Engagement

(With Miceli et al. 2025)

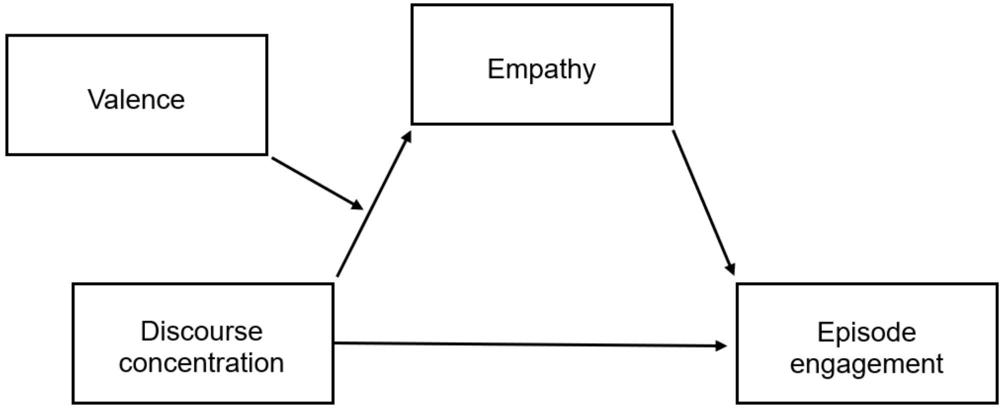
Motivation and Conceptual Model

Growing interest in **cultural items** (Movies, Songs, Books, News, Ted Talks), but no empirical studies focusing on podcasts

Berger and Milkman 2012; Papies and Ban Heerde 2017; Toubia, Berger and Eliashberg 2021; Pyo, Lee, and Park 2022; Cascio Rizzo, Berger and Zhou 2025)

Conversational podcasts are the most popular ones we focus on an unstudied language property called **"discourse concentration."**

- We argue that it's effect on engagement is conditional with **"episode valence"**



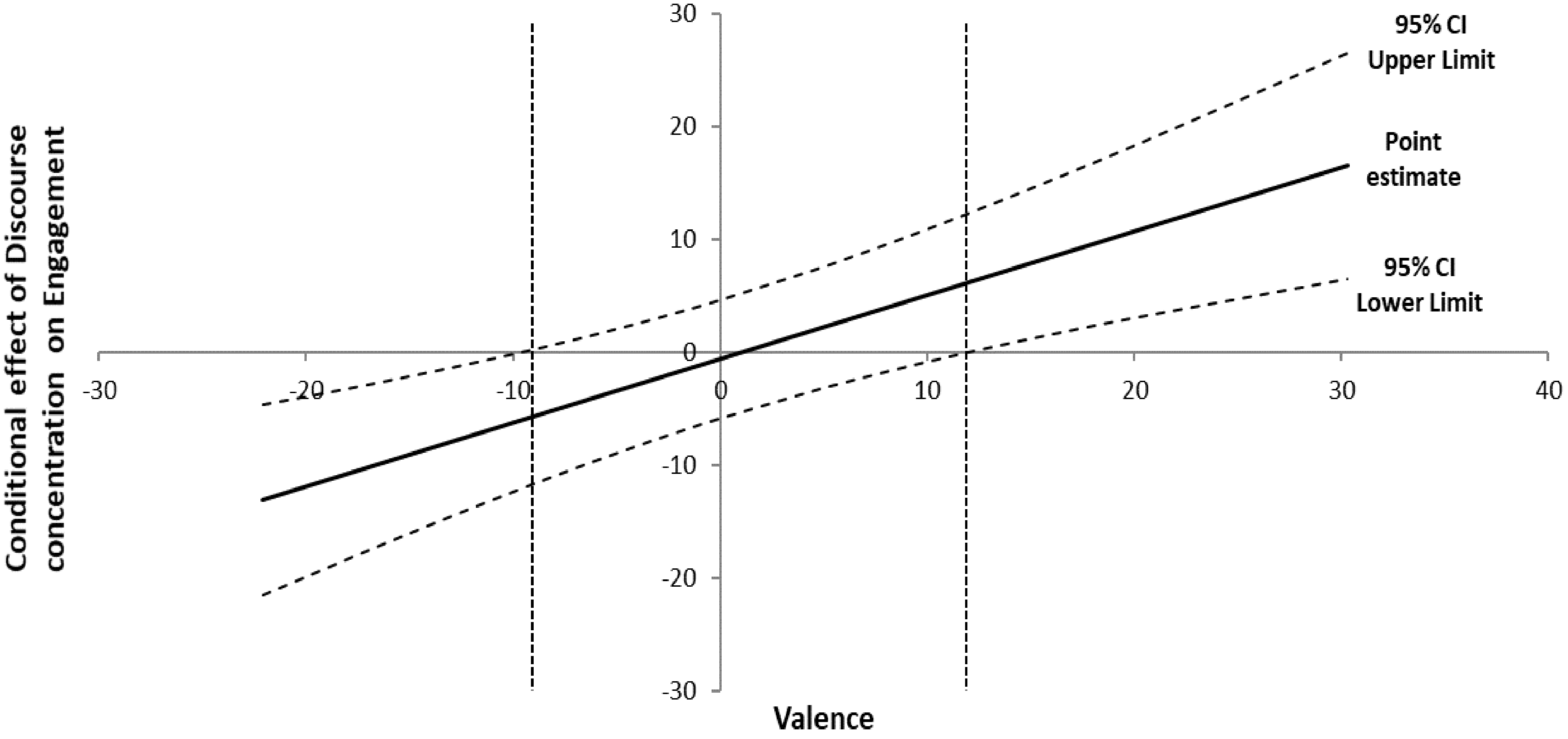
Field Study on 401 Transcripts and Audio files from the New York Times' Podcast, "The Daily"

MEASUREMENTS

- **Discourse concentration:** measured through the Hirschman-Herfindahl Index (HHI), computed on the percentage length of the speakers' interventions (in number of characters).
- **Episode valence:** measured through the LIWC 2022 Tone variable, normalized from 0 to 100 (Robustness with: SieBERT, VAD and EL)
- **Engagement:** measured as the sum of likes and comments associated with each single episode of "The Daily" podcast on Castbox

Table 3: Results of regression analyses

	Linear model, only main variables		Interaction model, only main variables		Linear model, all variables		Interaction model, all variables	
	b	*s.e.*	*b*	*s.e.*	*b*	*s.e.*	*b*	*s.e.*
Intercept	9.15 **	.25	9.28 **	.25	9.15 **	.24	9.29 **	.24
Main Variables								
Discourse concentration	- 1.41	2.11	- 2.02	2.09	-0.14	2.71	-.60	2.67
Valence	-.04 *	.02	-.03	.02	-0.04	.03	-.02	.03
Disc. concentration x Valence			.51 **	.15			.56 **	.15
Control variables								



<u>JN analysis.</u>

- For **negatively-valenced** episodes (up to 23rd perc) the effect of discourse concentration on engagement is negative,
- For **positively-valenced** episodes (from 84th perc.) the effect of discourse concentration on engagement is positive.

Experiment 1– "Broadway shows during pandemic: Six - the Musical" Transcripts examples

Thomas Paulson

The team was overall embarrassed. During rehearsals everyone was checking in on the mental state of everyone, and be like, OK, how are you feeling? And even that it was like "how am I feeling"? They felt crummy and pessimistic most of the time. Ultimately, this constant check on each other made them more stressed, and just slowed them down. And that's been heartbreaking as they've moved forward, just because they felt more distant and tuned off within each other, which is bad news for the show. And of course, for life, most importantly. And everyone involved with "Six" told us this time away from set had really changed them. For the worst. Like, getting that cab from JFK to Midtown, and like, seeing New York in the flesh, in the cement... And I was like, gosh, what a daunting step. After the lockdown, "Six" was going to have its precarious debut on Broadway. They all expressed this new sense of sk rehearsals and being like so fearful that I'm here. And just b disappointed and upset. The word they used over and over, was uncertainty. I think it's safe to say that everyone is conc sell out theatres, and it's not going to sell out theatres for v will enough people show up? Will it be safe? And will this ris middle of a pandemic? I just felt dubious on the last point. Y And the truth is, answers are not obvious at all. I guess it's j uncertainty and even disappointment, they're all kind of ba

Michael Barbaro

What do you mean with this kind of provocative last statem

Thomas Paulson

Well, most shows are just a triumph. Most shows open and sometimes flattery. And you know, just think about all the i King" and his heartening years in the wilderness. Think abo about "A Chorus Line," which is like directly about the kind of people fiercely yearning to make their way in the theatre about challenge and peace of mind and attachment and am stage will build toward a happy ending. Sometimes, ending potential and light-hearting outcome when you first settle i

Michael Barbaro

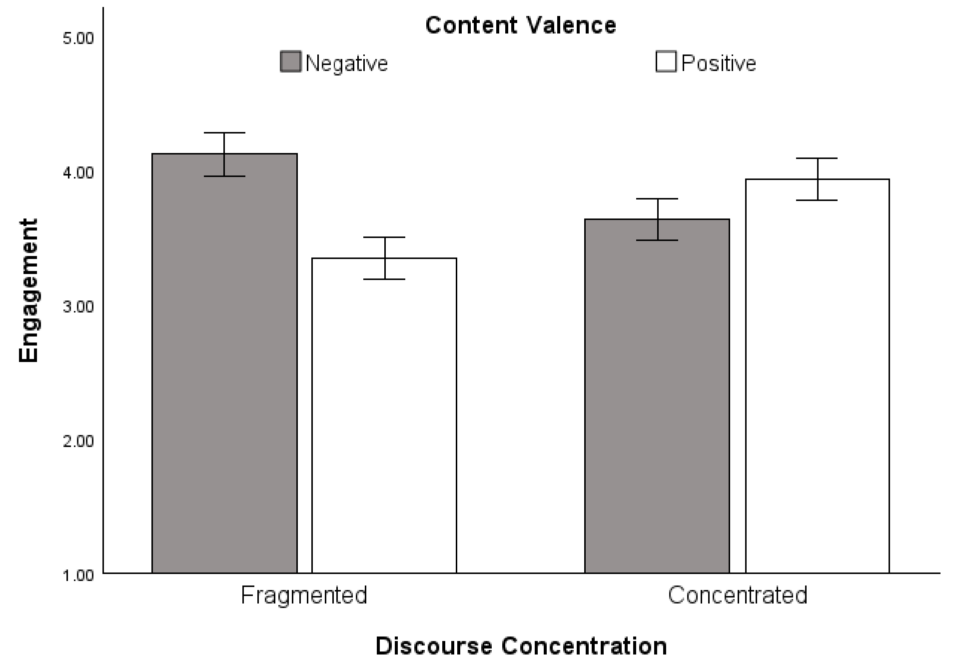
I guess you just get used to it. In the end, how many theatre Thomas, thank you very much. This was The Daily's episode consequences of a serious pandemic on a lucrative and cre and don't forget to follow us on the main listening platforms!

Samantha Pauly

It's been emotionally distressing.

Abby Mueller

I walked into the last session like, guys, I am purposeless, I'm so scared.



to meet the expectations of the audience...

age topic you expect to be watching on stage.

ch a long time away from each other and from theatre, was gratitude.

joyed after Broadway's reopening.

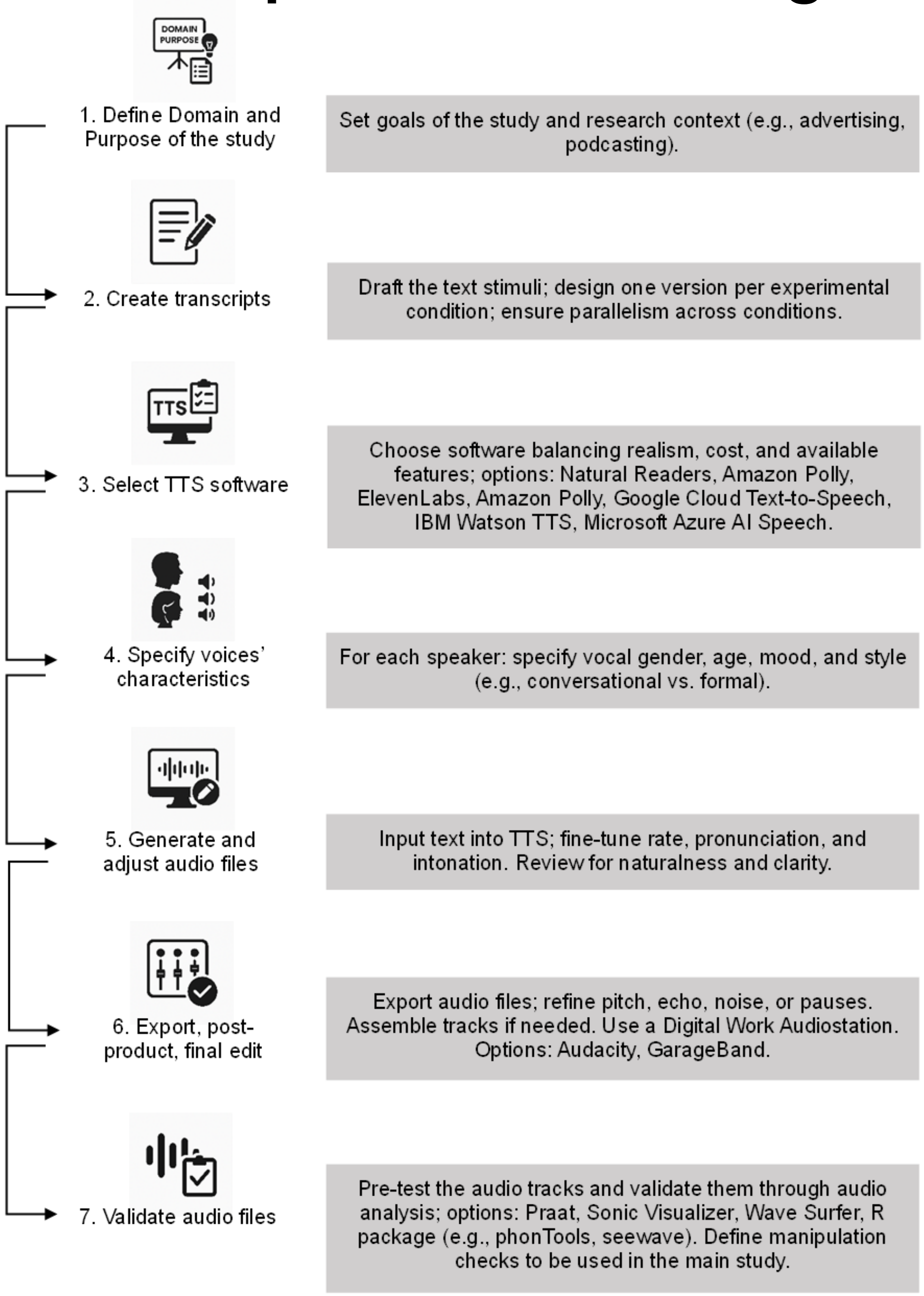
very long.

uestion is, will enough people show up? Will it be safe?

Samantha Pauly

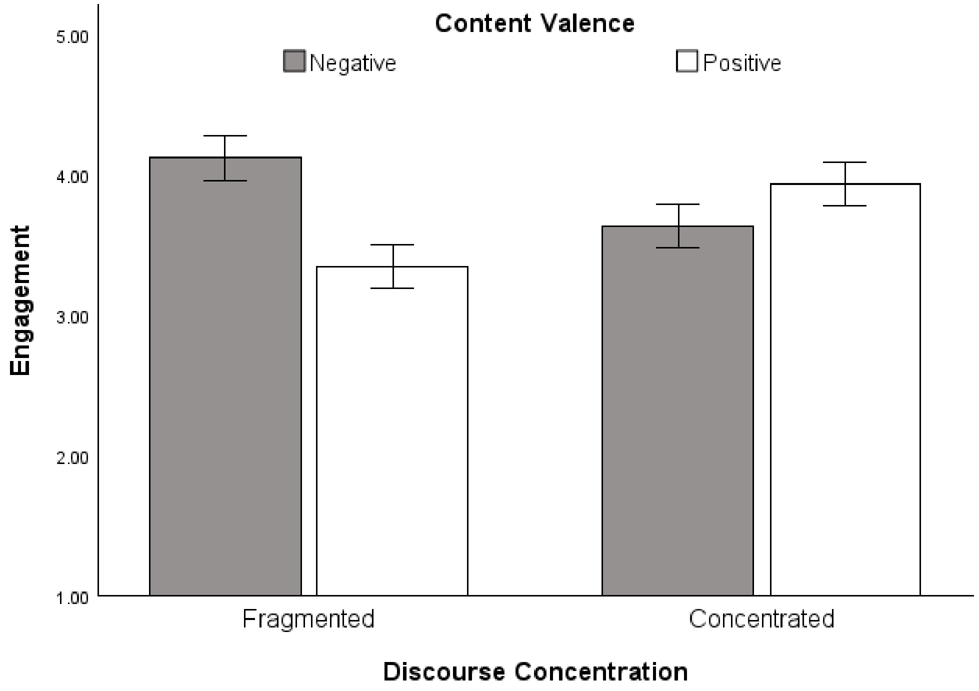
And will this benign experiment work? Will a show like "Six" make it as the pandemic ended?

Experiment 2: Replication but using AI generated voices



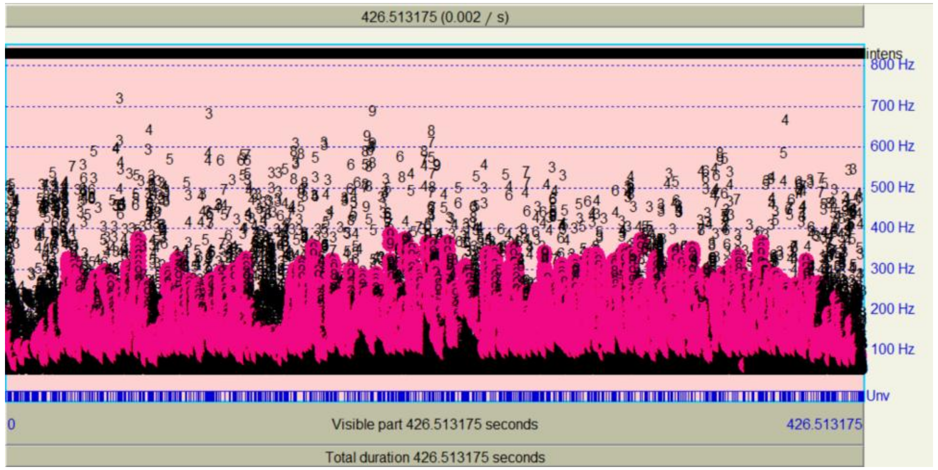
Pipeline for Using AI to create Voice Stimuli

<u>Concentrated x positive valence</u>



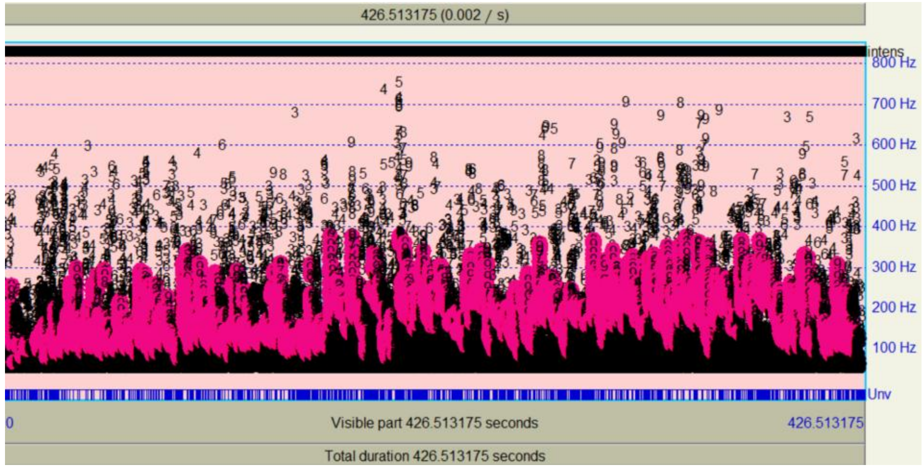
Concentration x positive valence

Medium pitch:162.66953867750433 Hz



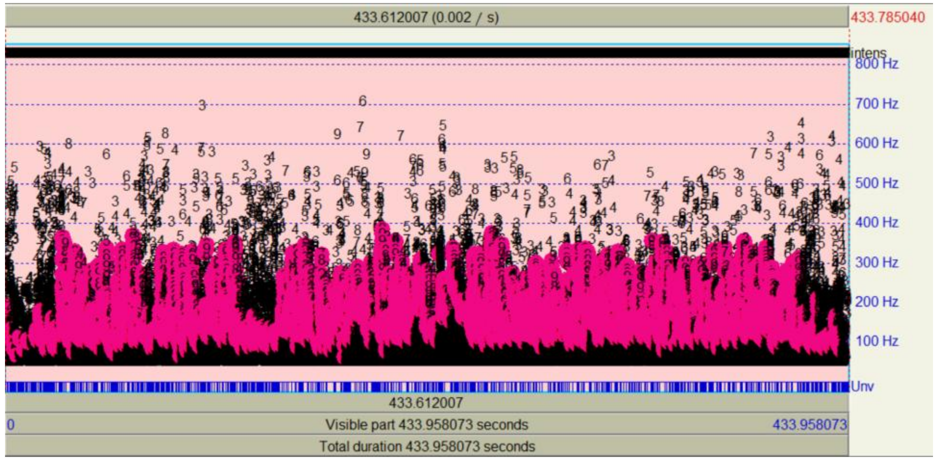
Fragmentation x positive valence

Medium pitch:173.7036728864732 Hz



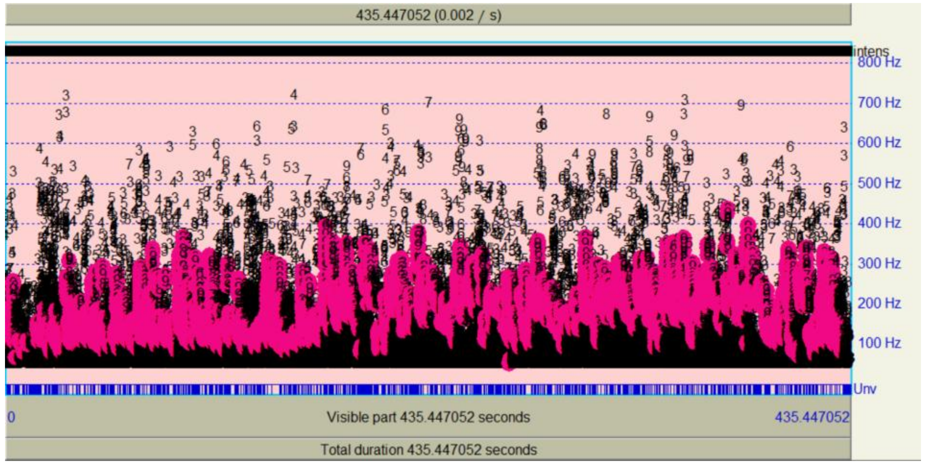
Concentration x negative valence

Medium pitch: 163.68624883309553 Hz



Fragmentation x negative valence

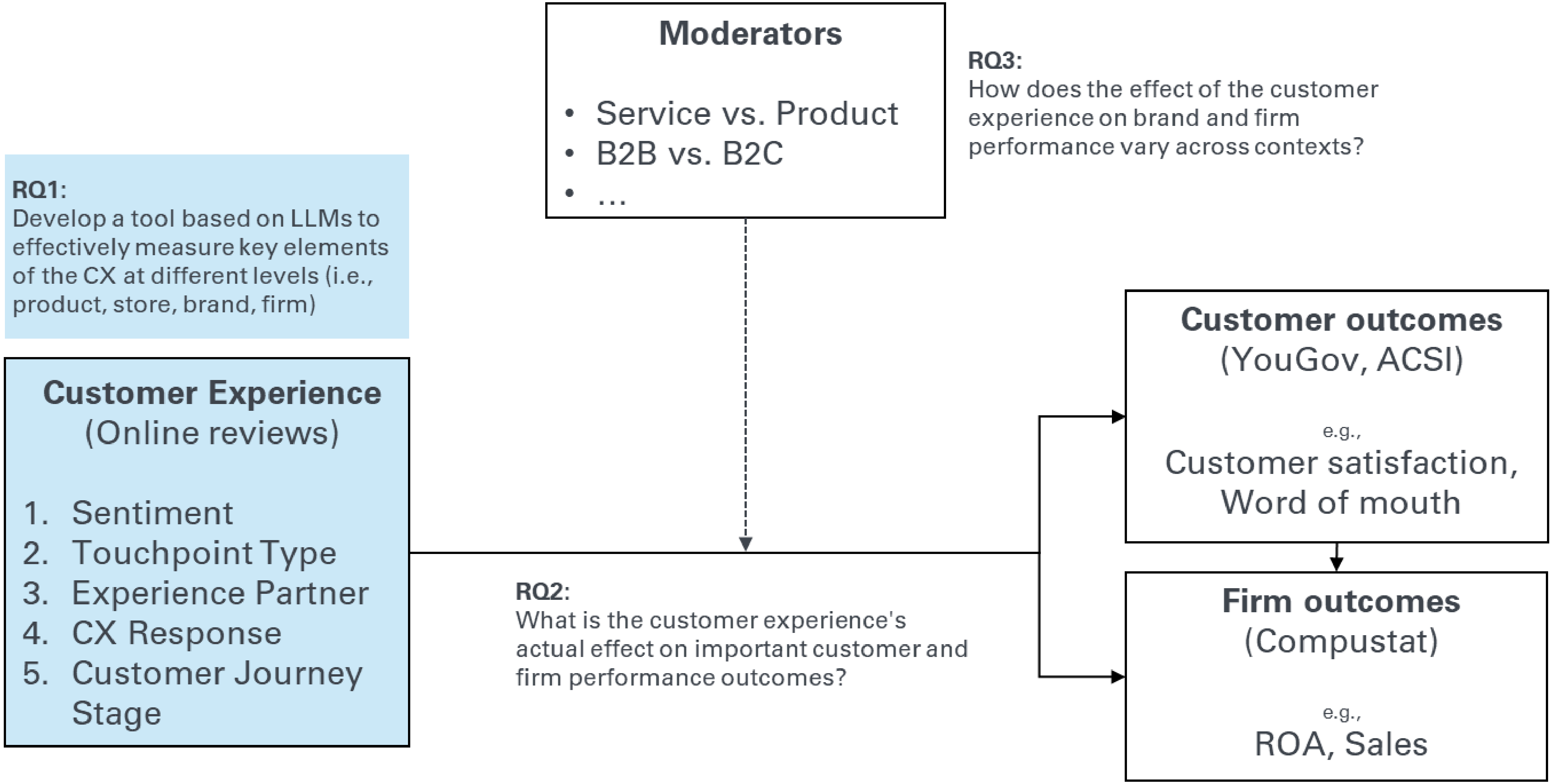
Medium pitch: 172.10675471558412 Hz



Measuring the Customer Experience with Large Language Models: Implications for Brand and Firm Performance

(With A. Wagner, K. Kuehnl, and Dennis Herhausen)

Conceptual Model



"Latent" Measurement of CX with LLMs

Step 1: Data Collection

Total: **6,764,042** online reviews from 79 S&P 500 companies

Online review platforms:

- Yelp
- G2
- App Store
- Trustpilot
- Co-operating B2B firm

Step 2: LLM Classification Task

1. Aspect-based sentiment analysis
2. Classification of CX elements into each identified aspect
   - Sentiment
   - Touchpoint type
   - Experience partner
   - CX response
   - Customer journey stage

Step 3: Model Choice

Comparison of different open-source LLMs:

- Llama 3.3 → **accuracy 89.96%**
- Microsoft Phi 4
- Microsoft Phi 3.5 mini
- DeepSeek-R1

Step 4: CX Score

Determination of CX score based on the identified aspects and the respective sentiment

Prompt: We asked the model to **(1)** perform **aspect-based sentiment analysis**, classify each aspect into the **(2) touchpoint type** and **(3) experience partner**, **(4)** classify the **associated response** into the five CX responses and **(5)** classify the **Customer Journey Stage** of the aspect.

Exemplary online review (Yelp):

"I came into this store after numerous attempts to fix that should've been detected and handled by customer service over the phone. Nick was very patient and helpful, he is knowledgeable, professional, and provides excellent customer service."

Classification:

1. [*A*: **store**; *S*: **Neutral**; *TP*: **offline**; *EP*: **brand**; *R*: **Cognitive**; *CJ*: **postpurchase**]
2. [*A*: **customer service**; *S*: **Negative**; *TP*: **online**; *EP*: **brand**; *R*: **Cognitive**; *CJ*: **prepurchase**]
3. [*A*: **Nick**; *S*: **Positive**; *TP*: **offline**; *EP*: **personnel**; *R*: **Emotional/Social**; *CJ*: **postpurchase**]
4. [*A*: **customer service**; *S*: **Positive**; *TP*: **offline**; *EP*: **personnel**; *R*: **Cognitive/Behavioral**; *CJ*: **postpurchase**]

Validation of Measurement

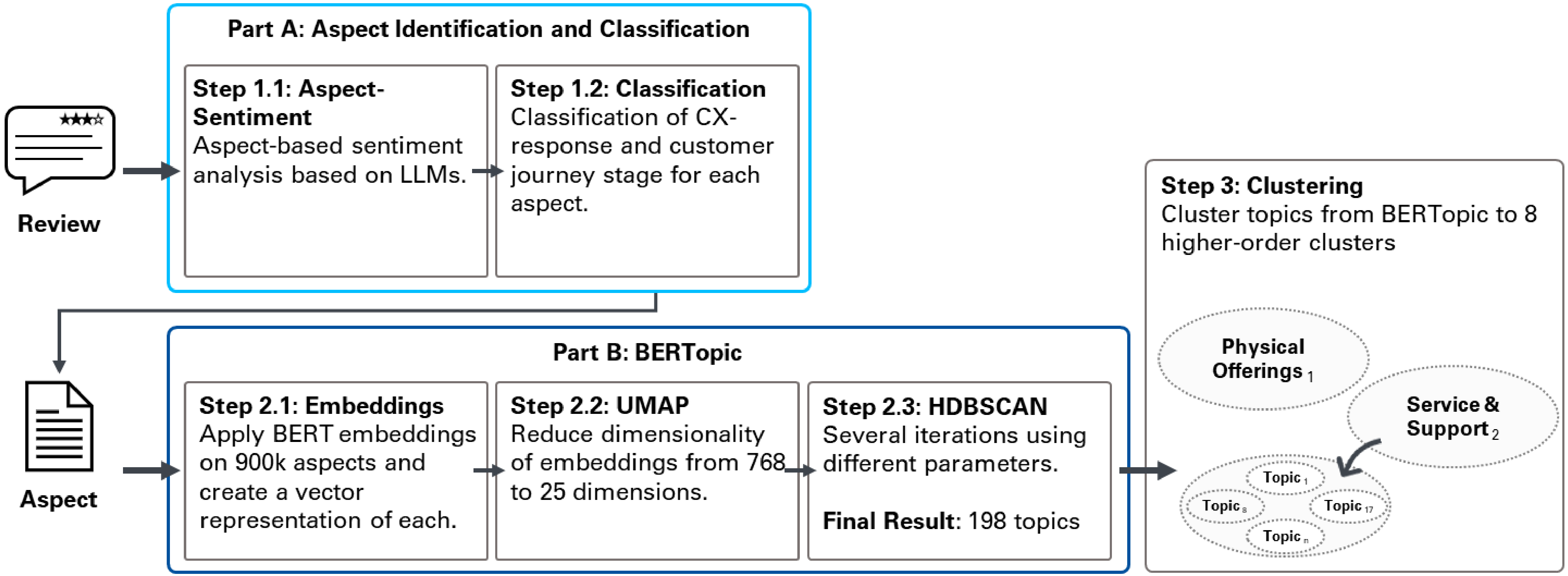
Annotation Study

Comparison of model labels with human labels, which serve as the "ground truth" to showcase the appropriateness of Llama3.3 for the research context

Accuracy of Llama3.3 for the CX measurement context

CX element	Annotator 1	Annotator 2	Combined
Sentiment	95.68%	95.23%	**95.45%**
Touchpoint type	80.21%	83.20%	**81.71%**
Experience partner	97.17%	90.88%	**94.03%**
CX response	82.23%	89.83%	**86.03%**
Customer journey stage	90.48%	94.31%	**92.4%**
Overall	89.11%	90.81%	**89.96%**

BERTopic for CX Aspects/elements



Other Resources

Recent Tutorials

Scholarly Article

Using Traditional Text Analysis and Large Language Models in Service Failure and Recovery

Journal of Service Research
1–7

DOI: 10.1177/10946705241307678
journals.sagepub.com/home/jsr
Sage

Francisco Villarroel Ordenes[1], Grant Packard[2], Jochen Hartmann[3], and Davide Proserpio[4]

Abstract
Service failure and recovery (SFR) typically involves one or more people (or machines) talking or writing to each other in a goal-directed conversation. While SFR represents a prime context to understand how language reflects and shapes the service experience, this subfield has only begun to apply text analysis methods and language theories to this context. This tutorial offers a methodological guide for traditional text analysis methods and large language models and suggests some future research paths in SFR. We also provide user-friendly workflow repositories, in Python and KNIME Analytics, that researchers with (and without) coding experience can use. In doing so, we hope to encourage the next wave of text analysis in SFR research.

Keywords
service failure and recovery, text mining, Natural Language Processing (NLP), language theory, large language models, machine learning

LINK

From Insight to Impact: Closing the Marketing Science Value Chain with Interactive, Research-Driven Apps

(Pikal, K., Villarroel O., F., Herhausen, D., Tamagnini P., 2025)

Abstract

Every year, several thousands of marketing articles are published in academic journals, often with the aim of disseminating new insights not only to the academic community but also to managerial practice. However, there is wide acknowledgment of gaps in the marketing science value chain, hindering the flow of marketing knowledge to other researchers and managers. We posit that interactive, research-driven (IRD) apps that provide a deeper understanding of the usability of the research contribution are a viable solution to improve the diffusion of marketing knowledge. We shed light on the motivations and barriers to develop IRD apps as well as the market potential and impact of IRD apps through a multi-method examination, which includes interviews, secondary data analyses, and experimental studies. We find an untapped potential of IRD apps among published articles and complementary evidence of their value to researchers and managers. We further provide a tutorial to guide the development of IRD apps that complement static research papers, and close with a forward-looking section about the future of IRD apps.

Keywords: Interactive research-driven apps, marketing science value chain, research-practice gap

LINK

Both links lead to workflows to implement: dictionaries, machine learning, topic models, embeddings and LLM-based methods for text mining

Python Workflows

KNIME Workflows

In [33]:
```python
# import model
get_sentiment = pipeline("sentiment-analysis", model="siebert/sentiment-roberta-large-english", device = 0)
```

In [34]:
```python
# predict sentiment of a single sentence
get_sentiment("My flight was cancelled. Where can I request a refund?")
```

Out[34]: `[{'label': 'NEGATIVE', 'score': 0.9993242025375366}]`

In [35]:
```python
# predict sentiment of multiple documents

# load data from GitHub
!wget https://raw.githubusercontent.com/shivang16/movieReview/master/movie_review.csv
d = pd.read_csv("/content/movie_review.csv")

# sample and preprocess data
d = d.iloc[0:100]  # sample to first 100 reviews
text_column = 'text'  # select text column
texts = d[text_column].astype('str').tolist()
```

In [37]:
```python
# print predictions
p = pd.DataFrame(get_sentiment(texts))['label']
p.value_counts()  # check sentiment distribution
```

Out[37]:
```
label
POSITIVE    61
NEGATIVE    39
Name: count, dtype: int64
```

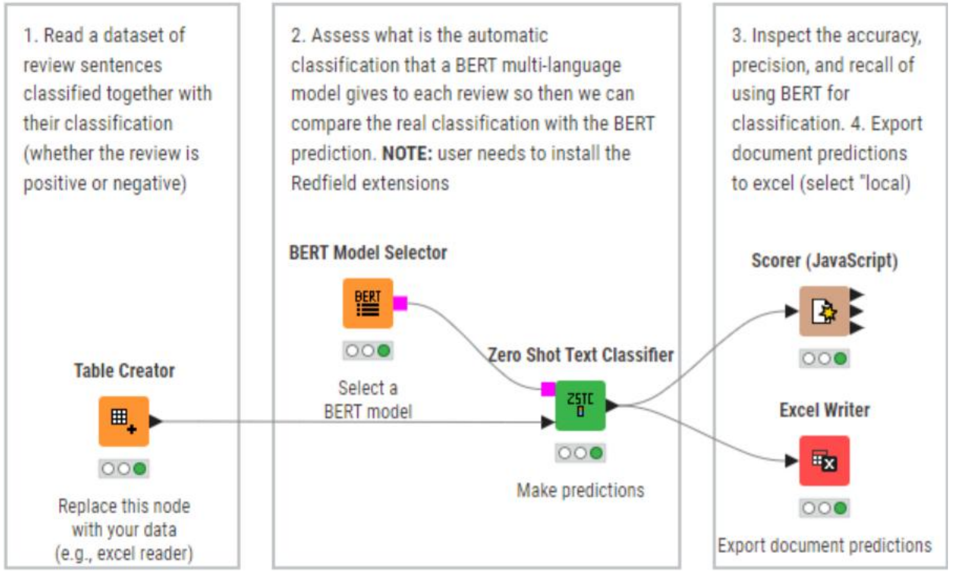
In [38]:
```python
# add predictions to original dataframe
d['sentiment'] = p
d[[text_column, 'sentiment']].head()
```

Out[38]:

	text	sentiment
0	films adapted from comic books have had plenty...	NEGATIVE
1	for starters , it was created by alan moore (...	POSITIVE
2	to say moore and campbell thoroughly researche...	POSITIVE
3	the book (or " graphic novel , " if you will ...	NEGATIVE
4	in other words , don't dismiss this film becau...	POSITIVE

In [39]:
```python
# save results to csv
output_filename = 'Sentiment_Predictions.csv'  # name your output file
d.to_csv(output_filename, index=False)
```

In [42]:
```python
# download output file
files.download(output_filename)
```



Scorer View

Results

Confusion Matrix

	negative (Predicted)	positive (Predicted)
negative (Actual)	9	1
positive (Actual)	0	10
	100.00%	90.91%

Overall Statistics

Overall Accuracy	Overall Error	Cohen's kappa (κ)	Correctly Classified	Incorrectly Classified
95.00%	5.00%	0.900	19	1

CONCLUSION

My Perfect
RECIPE

Is there a best recipe for a Social Media Quant paper?

NO!, but there we can distinguish three clear type of papers

- *<u>The Classic Strategy</u> (theory or empirics first)*. Three strong contributions (ideally not just main effects, but moderations)
- *<u>The Why? paper:</u>* Main effect or moderation hypothesis (very novel), and explanation through a series of experiments
- *<u>The tool:</u>* Developing a tool/app that other researchers or practitioners can use to convert or optimize

Any questions?

Francisco Villarroel Ordenes

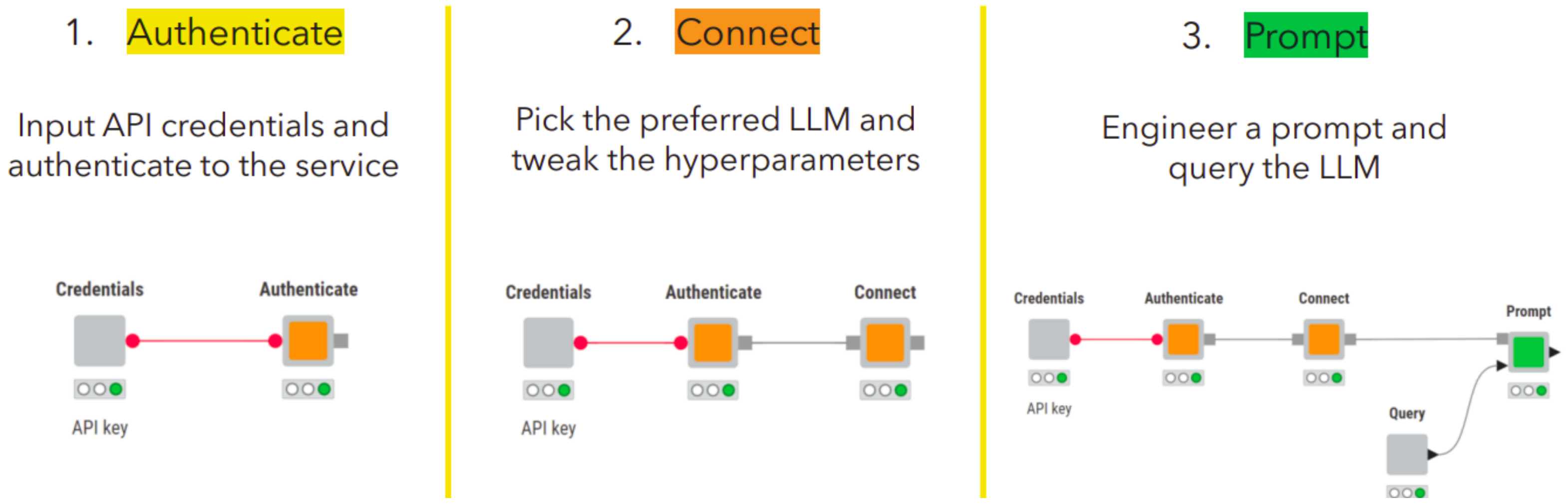
University of Bologna
francisco.villarroel@unibo.it

To connect: LinkedIn

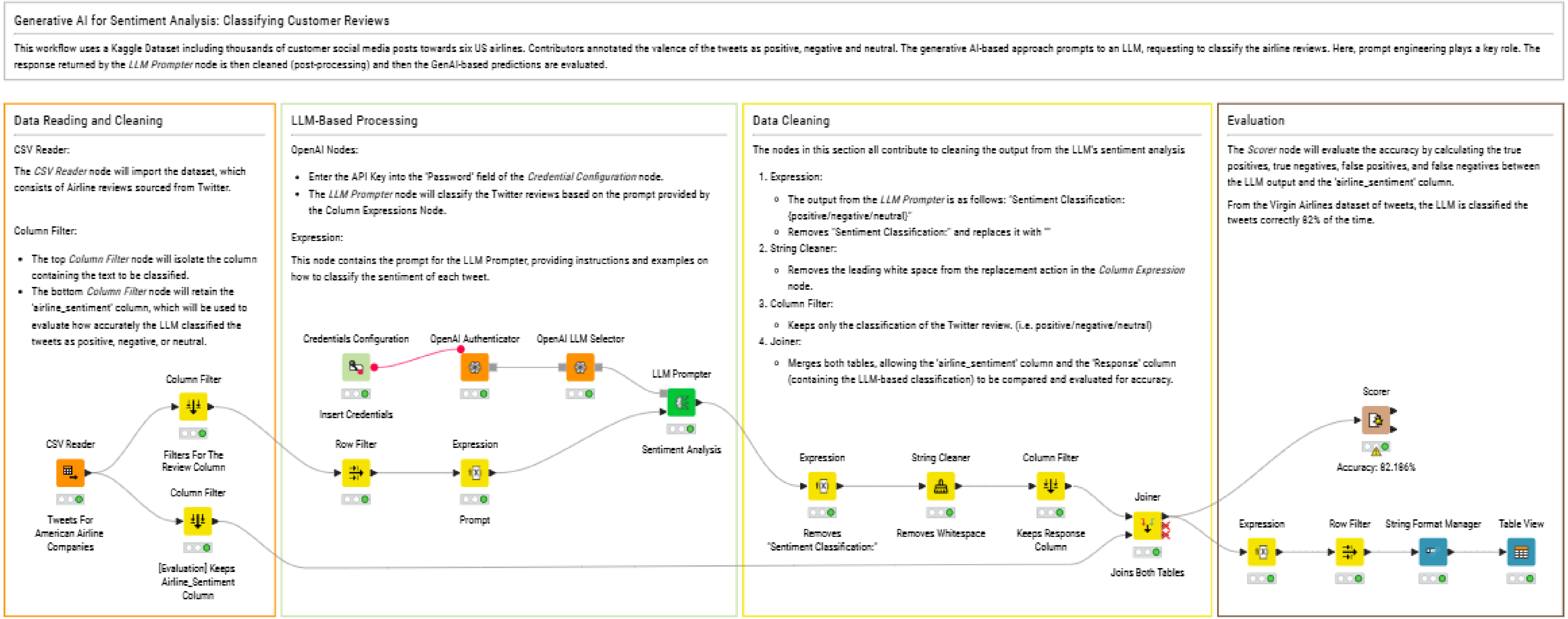
Workflow - Language Models and Generative AI (Blanchard et al. 2025)

Alternatives:

- Sequential API-based approach, in which a computer script sequentially submits each participant's response (along with standardized instructions) to th API and captures the value returned.
- A file-upload approach with code execution, where a structured dataset (e.g., CSV with one participant response to code per row) is uploaded to a chat-based GenAI system, along with instructions. The GenAI generates coding rules, which can be saved.



Workflow for sentiment classification using LLM API: https://hub.knime.com/s/5zGU8SvYdqd_hJAt



UD, Research Objectives, and Methods

Objectives / Type of UD	"Defined" Construct Operationalization E.g., arousal	"Latent" Construct Identification E.g., content type	"Prediction" of variable outside UD E.g., engagement	"Generation" of new data E.g., brand (s)logo
TEXT represented as "Bag of Words" or "Word Embeddings"	Dictionaries, Word co-occurrence, Distances, Machine Learning, Language Models	Topic Models such as LDA, CTM, STM, BERTopic	Machine Learning process and algorithms (Neural Networks, X-Boost, Random Forest, SVM), Causal Machine Learning	Generative large language models and tools (e.g., GPT) Image generation (DALL-E) Audio and music generation (JukeBox)
IMAGES represented as an arrangement of "pixels in RGB"	Pixel based features (colors), Objects, or actions from ML classification (e.g., cloud vision)	Topic Models performed on predicted image objects and actions		
AUDIO represented as "Hertz (time) Frequencies" in audio waves	Algorithms based frequency properties of audio (e.g., pitch), or ML based features (e.g., emotions)	Topic Models performed in time frequency data (bag of frequencies)		

Validity

Measurement error. Whatever you are measuring, you should demonstrate that the measurement you are using is valid (i.e., it doesn't have much measurement error; cause of endogeneity)

- **Construct validity:** "Consistent construct operationalization". Human coders to show high correlation between text mining measurement and human ratings
- **Convergent validity:** "The degree to which measures of the construct correlate to each other". By measuring the construct using different linguistic aspects (also outside from the text)
- **Concurrent validity:** "Ability to draw inferences over many studies". Using dictionaries that have been used in previous studies
- **Discriminant validity:** "To observe consistent patters of difference using opposite dictionaries" (e.g., tentative vs. certain)
- **Predictive validity:** "Using a hold out sample, cross validation and predictions across studies"
- **Model Fit:** Main used in topic models by computing log-likelihood, perplexity and coherence
- **Accuracy:** Mainly used in machine learning, which involves cross-validation, and providing accuracy, precision and recall as main metrics
- **Simulation Extrapolation Method (SIMEX):** a data-driven approach to correcting measurement errors and requires relatively fewer assumptions and information than alternative methods (Peng et al. 2020 JM)

Example Cascio Rizzo et al. (2024)

Table 1. Dictionary Validation for Language Arousal and Informative Goal.

Type of Validity	Validation Procedure
Construct validity: Does the text represent the theoretical concept?	Following Humphreys and Wang (2018), for arousal, we used a top-down approach, combining Mohammad's (2018) VAD dictionary with paralanguage. For informative goal, we used a bottom-up approach, empirically guided by the most frequent words in our data.
Concurrent validity: Does the measurement of the constructs relate to other measurements?	Our measurement indicates concurrence with human ratings for both arousal ($r_{intercoder}$ = .64, r = .67) and informative/commercial goal (average r_{coders} = .79, average r = .61; classification accuracy = .82).
Convergent validity: Do multiple measurements of the construct all converge to the same concept?	The VAD arousal score correlates with the DAL arousal score (r = .69). The commercial word count at a post level correlates with Jalali and Papatla's (2019) list of sale promotional words (r = .33).
Discriminant validity: Does the measurement differentiate from measures of other constructs?	Our arousal measure does not relate to valence (r = .02, n.s.). Our informative goal measure does not relate to arousal (r = .02, n.s.). The informative and commercial goals measures are weakly related (r = .14)
Causal validity: Is the construct in the data set causally related to other constructs?	We include several controls in the model to rule out alternative explanations (e.g., influencer, text, image, other).
Predictive validity: Does the construct have the expected effects of a meaningful variable?	Across different measurement approaches, we confirm the theoretically derived relationship between arousal and informative goal with engagement for macro- and micro-influencers in the field.
Face validity: Does the construct measure what it claims to measure?	High arousal: "Sharing TONS of essentials to help you! 🌸 this sweater is SO soft! 😍 @liketoknow.it #liketkit #ad @nordstrom." Low arousal: "Just out here making my own pizza dreams come true with this Spicy Broccoli Pizza with a quick homemade pizza crust using @fleischmannsyeast. It's really a wonder #ad." Informative goal: "Have you heard? 🤫 Hi-C is BACK at @mcdonalds🥤 Wops World Out Now 🎶🌎🚀 #Ad #McDonalds." Commercial goal: "Holiday Gift from @manscaped 🎁. Use code "TREVOR20" to save 20% #sponsored #ad."
Robustness: Is more than one method used?	We replicate the focal relationships in controlled experimental settings, in which we manipulate arousal (Studies 2–4) and informative goal (Study 4).
Generalizability: Are results based on multiple data sets?	The relationship of arousal, influencer type, and engagement is replicated with two independent samples (i.e., Instagram posts and TikTok videos).

Modelling Text Data

- **Identification Strategies**
  - **Panel Data (fixed effects).** Remove time invariant factors affecting the outcome (e.g., employee characteristics) (Marinova, Singh and Singh 2018)
  - **Difference-in-Differences and Synthetic Control**. Mimic an experimental research design using observational study data and a natural experiment (e.g., a platform allowing vs. another one not allowing service recovery) (Proserpio, Troncoso and Valsesia 2021)
  - **Instrumental Variables or Control Functions**. Require identifying a variable that randomizes the predictor variable but does not directly affect the service recovery outcome (i.e., exclusion condition) (Villarroel Ordenes 2019)
  - **Two Stage Heckman Correction** (selection Bias). When the sample from which data are drawn is not representative of the population being studied (e.g., companies do not provide social media responses to all customers) (Herhausen et al. 2023)
  - **Propensity Score Matching.** When the focus is on a specific characteristic of the data (e.g., tweets containing images; images containing overlays) (Li and Xie 2020)

- **Experiments**
  - They are the **best way to claim causality**, but they might be complicated to design in certain projects (e.g., drivers of no pay in loans) (Packard and Berger 2021)

Modelling Text Data

- **Explainable AI**
  - Increase interpretability of ML models, by relying on solutions such as SHapley Additive exPlanations (**SHAP**) or Local Interpretable Model-Agnostic Explanations (**LIME**) that can help identify the most predictive words (Hartmann, Bergner, and Hildebrand 2023)

- **Causal inference machine learning**
  - **Deep Instrumental Variables** estimates both the first and second stages of the IV framework through deep neural networks, thereby allowing both heterogeneous and nonlinear estimation of causal effects (Tian, Dew, and Iyengar 2024)
  - **Double Machine Learning (DML)** allows researchers to account for a large set of covariates, which is often the case when working with text
  - **Causal Random Forests,** by handling high-dimensional data, Causal Random Forests automatically identify subgroups that exhibit different treatment effects and estimate conditional (on these groups) effects