Beyond Words: A Systematic Multimodal Framework for Text, Images, and Extreme Helpfulness in Online Reviews

Ph.D. Candidate: Alvaro J. Aguado Marin

Program: PhD in Business Data Science – Martin Tuchman School of Management

Advisors: Dr. Jorge E. Fresneda Fernandez, Dr. Chelsey Hill

Committee Members: Dr. Shanthi Gopalakrishnan, Dr. Ellen Thomas, Dr. Plavini Punyatoya, Dr. Javier Cabrera

Date: August 25th, 2025

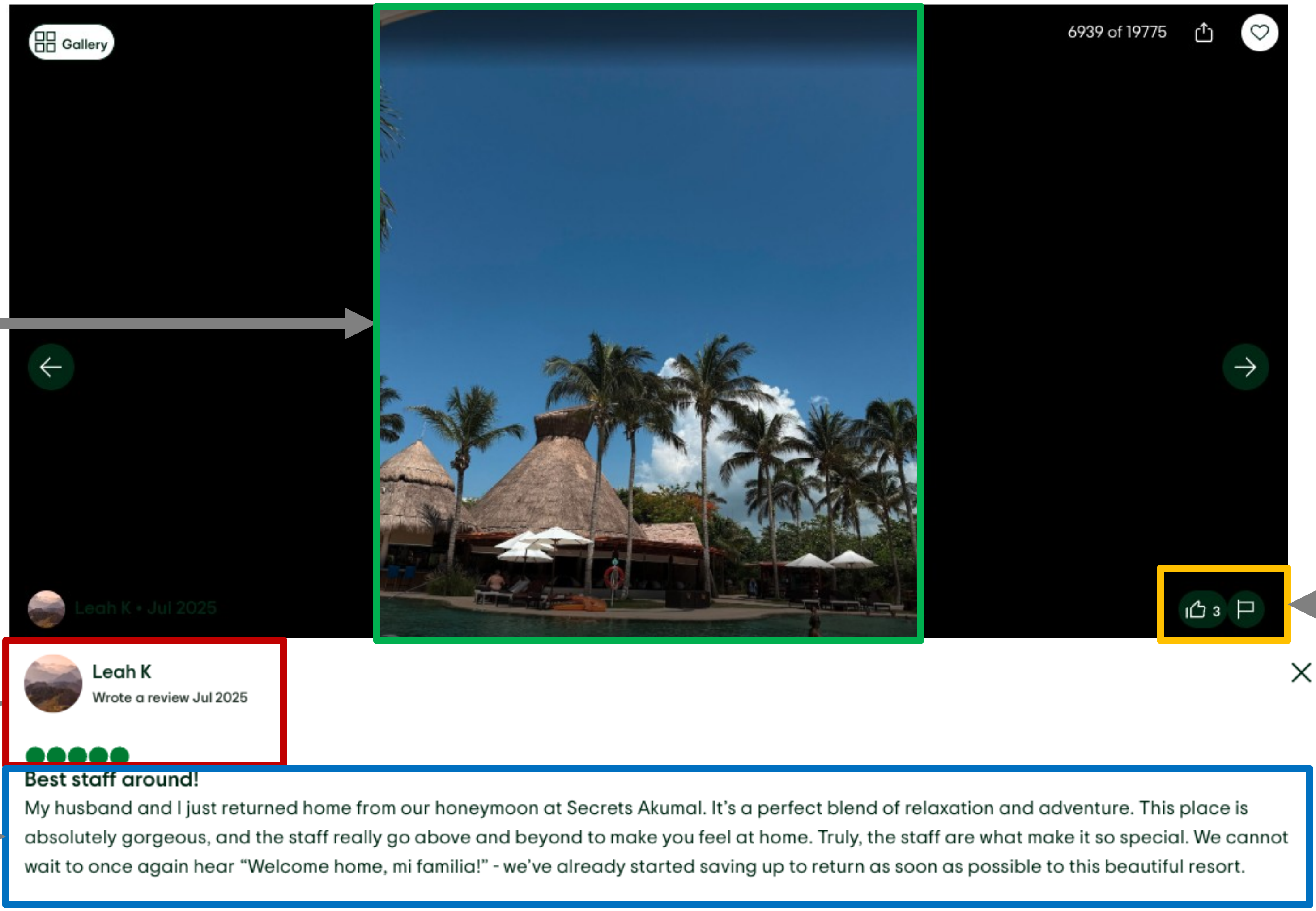
Anatomy of Online Reviews and Helpfulness

Elements of Online Reviews

Visual Elements

Metadata

Textual Elements



Helpfulness

- **Online reviews** are a form of *electronic word-of-mouth (eWOM).*
- A **helpful** review reduces uncertainty and aids consumers in making informed purchase decisions

Anderson, 1988 (WOM); Hennig-Thurau et al., 2004 (eWOM); Kim et al., 2006; Mudambi & Schuff, 2010 (Helpfulness)

Why Focus on Experience Products

Product Type*	Definition	Examples	Characteristics	Helpfulness Challenge
Search Goods	Attributes evaluated before purchase	Clothing, electronics, Furniture	Lower ambiguity: Decision based on price and product characteristics	Consistent Evaluation: Reviews focus on quality and relevance
Experience Goods	Evaluated only after consumption	Restaurants, Hotels, Healthcare Products	Context-dependent and multimodal** Harder to interpret consistently	Interpretation Varies: Harder to judge what is "helpful"

Search Goods – Image Sample

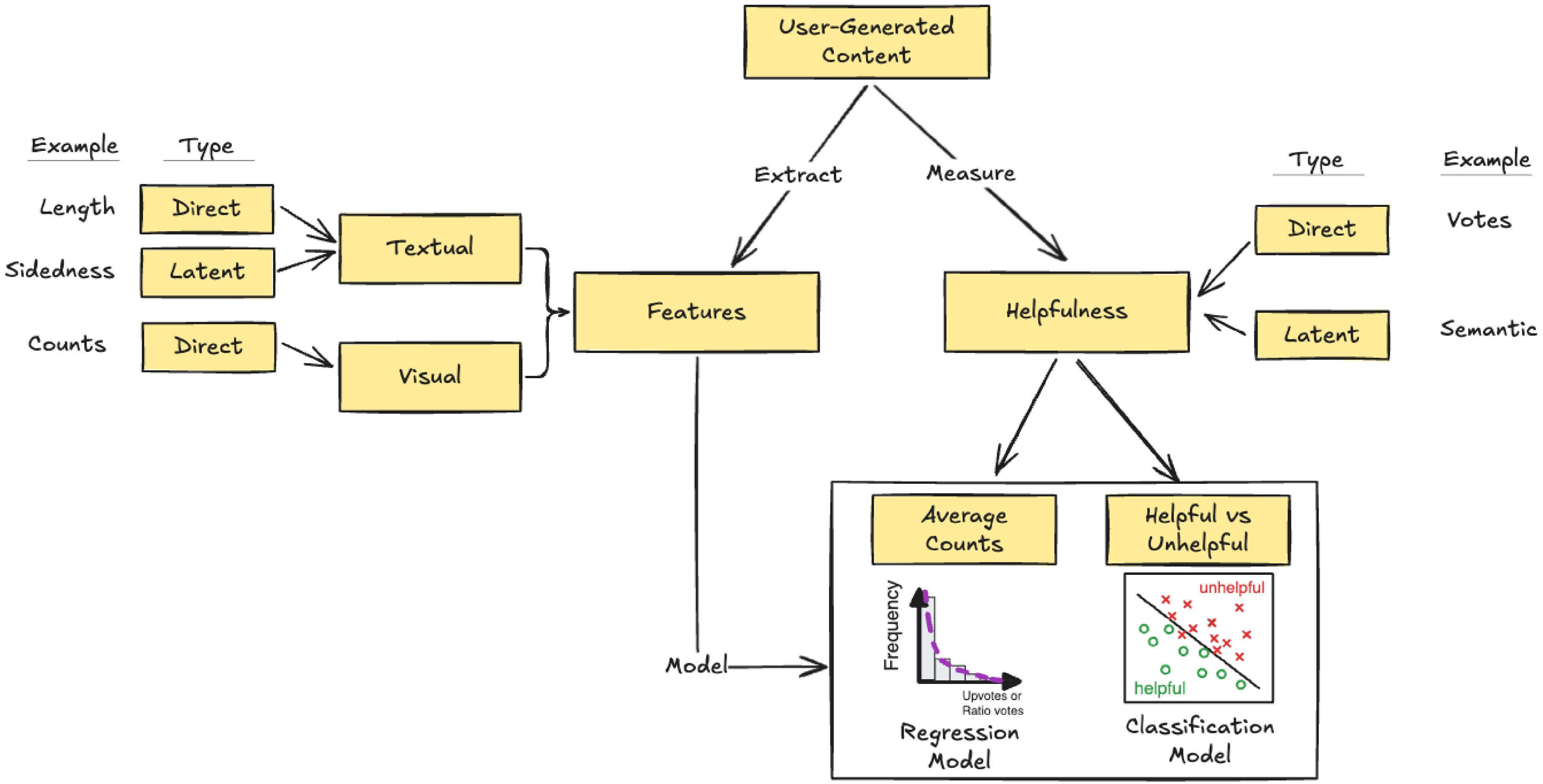
Clear & consistent product attributes

Experience Goods – Image Sample

Multimodal & context-dependent

* Darby and Karni, 1973; Nelson, 1970 (SEC Search-Experience-Credence Framework);
Kotler and Keller, 2012; Kübler et al., 2023; Mudambi and Schuff, 2010; (Characteristics of Search vs Experience Products in Online Review Helpfulness)
** Multimodality = presence of two or more types of information format (e.g., text + images)

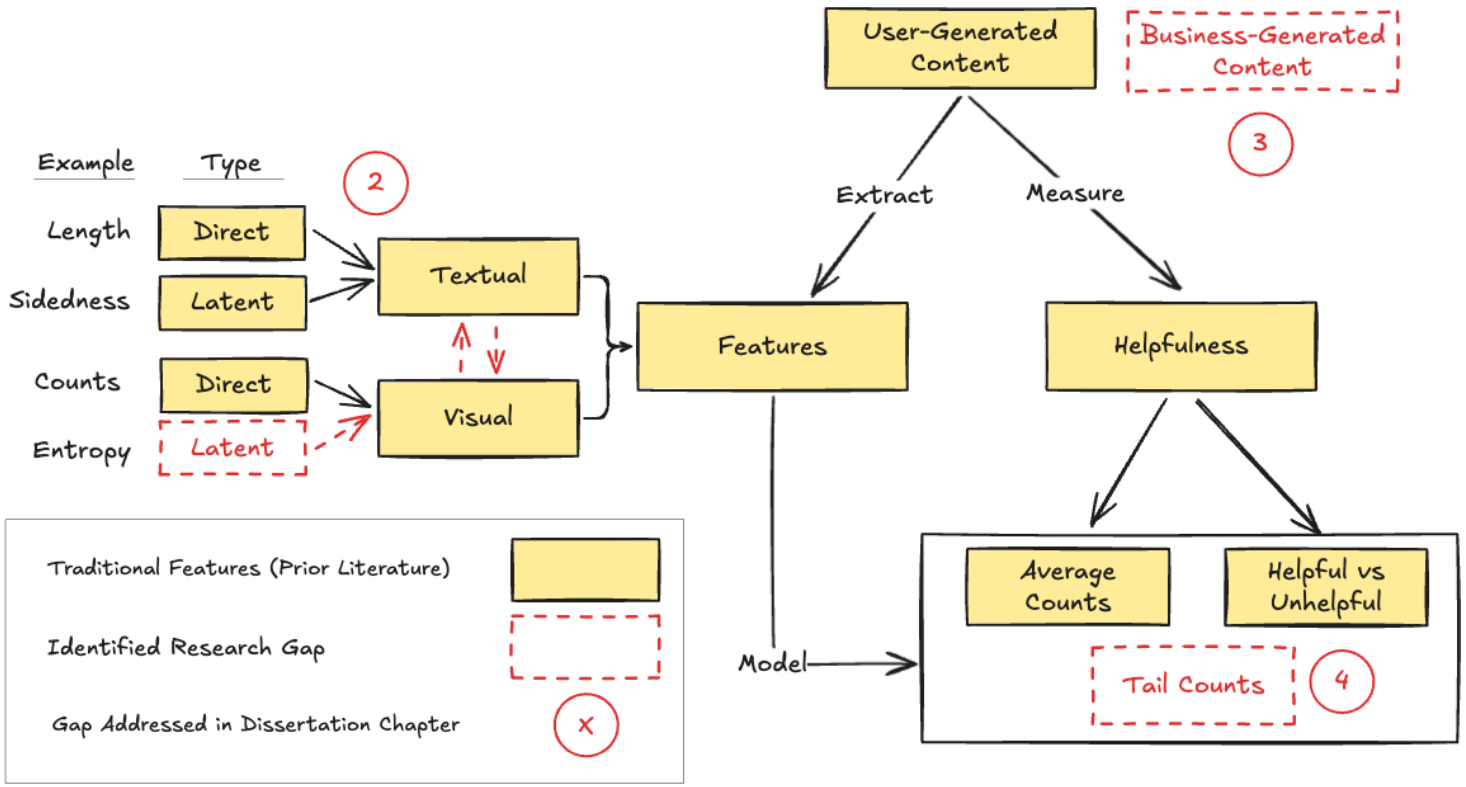
How Helpfulness Has Traditionally Been Studied



Adapted from Hong et al. (2017), Rietsche et al. (2019), and Saumya, Roy, and Singh (2023).

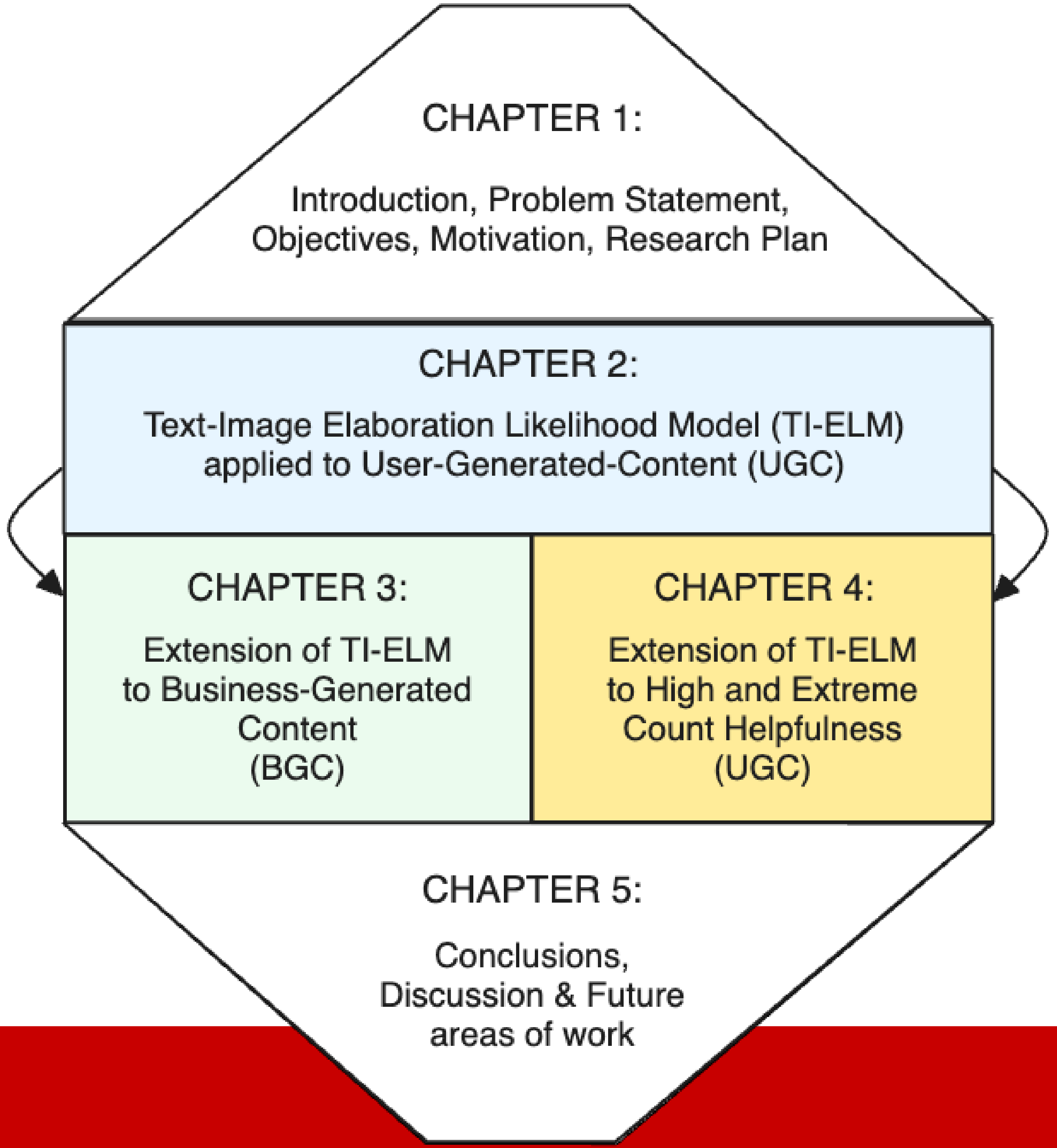
General Research Gaps in Review Helpfulness

- Lack of multimodal integration – Helpfulness models rarely include both text and image features comparably.
- BGC vs. UGC gap – Dynamics between business- and user-generated content remain underexplored.
- Extreme helpfulness – Limited understanding of disproportionately helpful reviews.



Ceylan et al., 2024 (Multimodal integration); Sukontip et al., 2024 (Business-Generated Content Helpfulness), Adil et al., 2021 (Extreme Helpfulness)

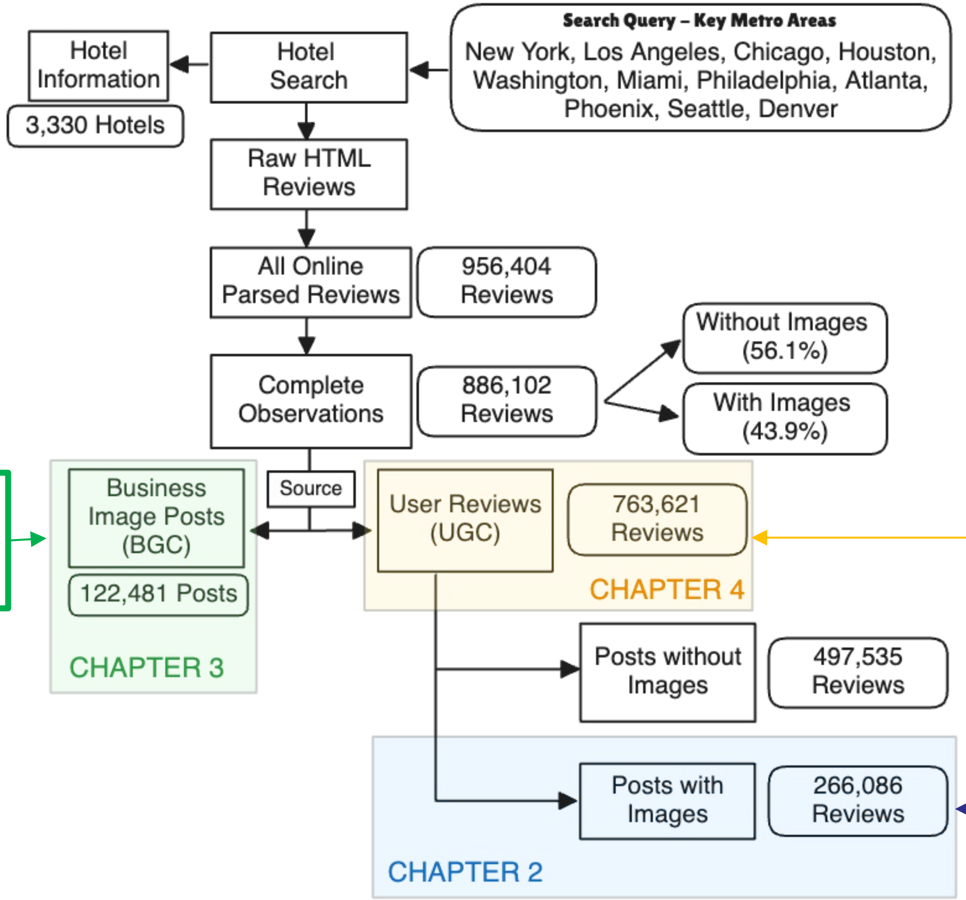
Dissertation Overview



Dataset Foundation used in the Dissertation

Custom-extracted multimodal online review data from a known hotel review platform

Business-Generated Content (BGC) Images



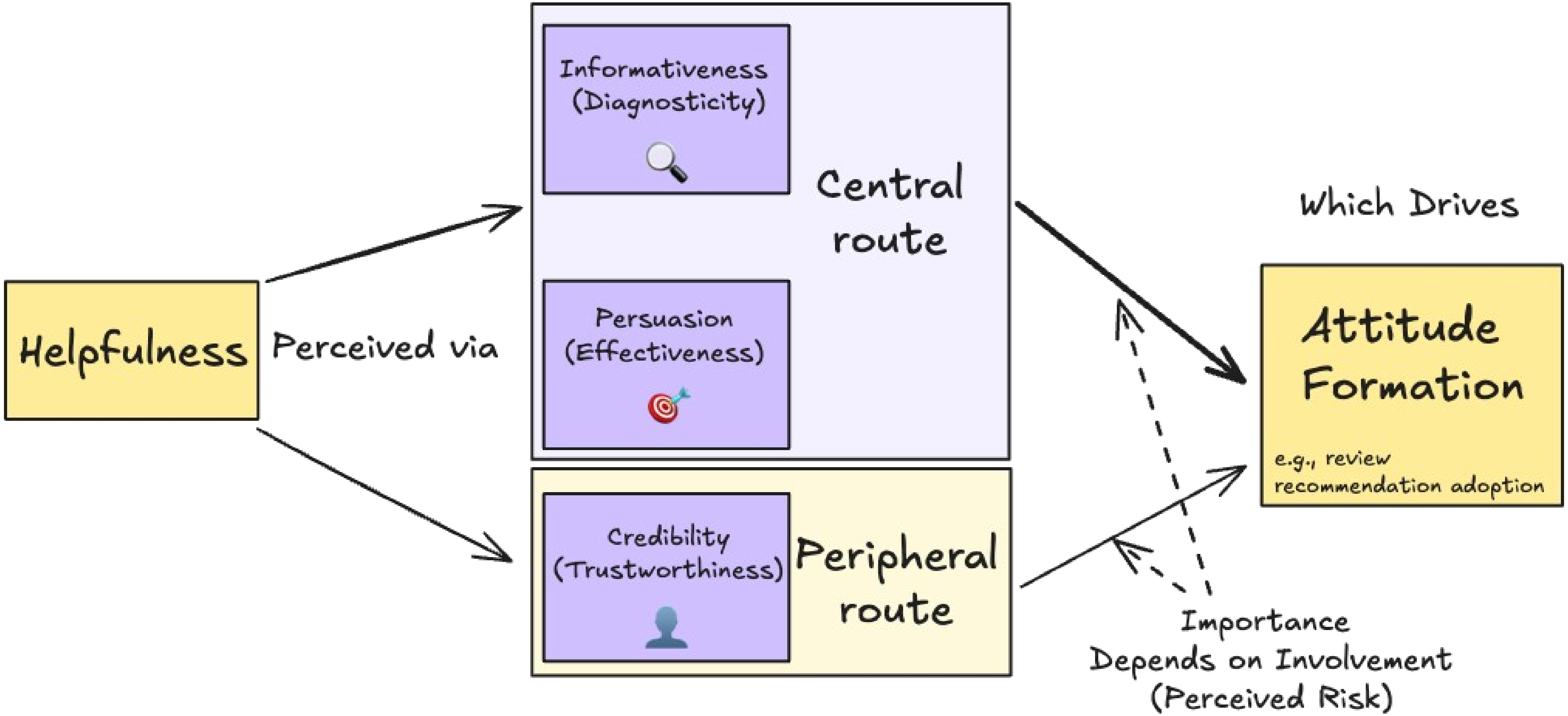
All User-Generated Content (UGC)

User-Generated Content (UGC) Images*

* Chapter 2 also tested UGC without images (Appendix C) to validate results and reduce bias.

Theoretical Framework: Elaboration Likelihood Model

This dissertation extends the Elaboration Likelihood Model to online reviews, framing helpfulness through multimodal elements



Adapted from Petty and Cacioppo (1983), Luo et al. (2013), and Srivastava and Kalro (2019).

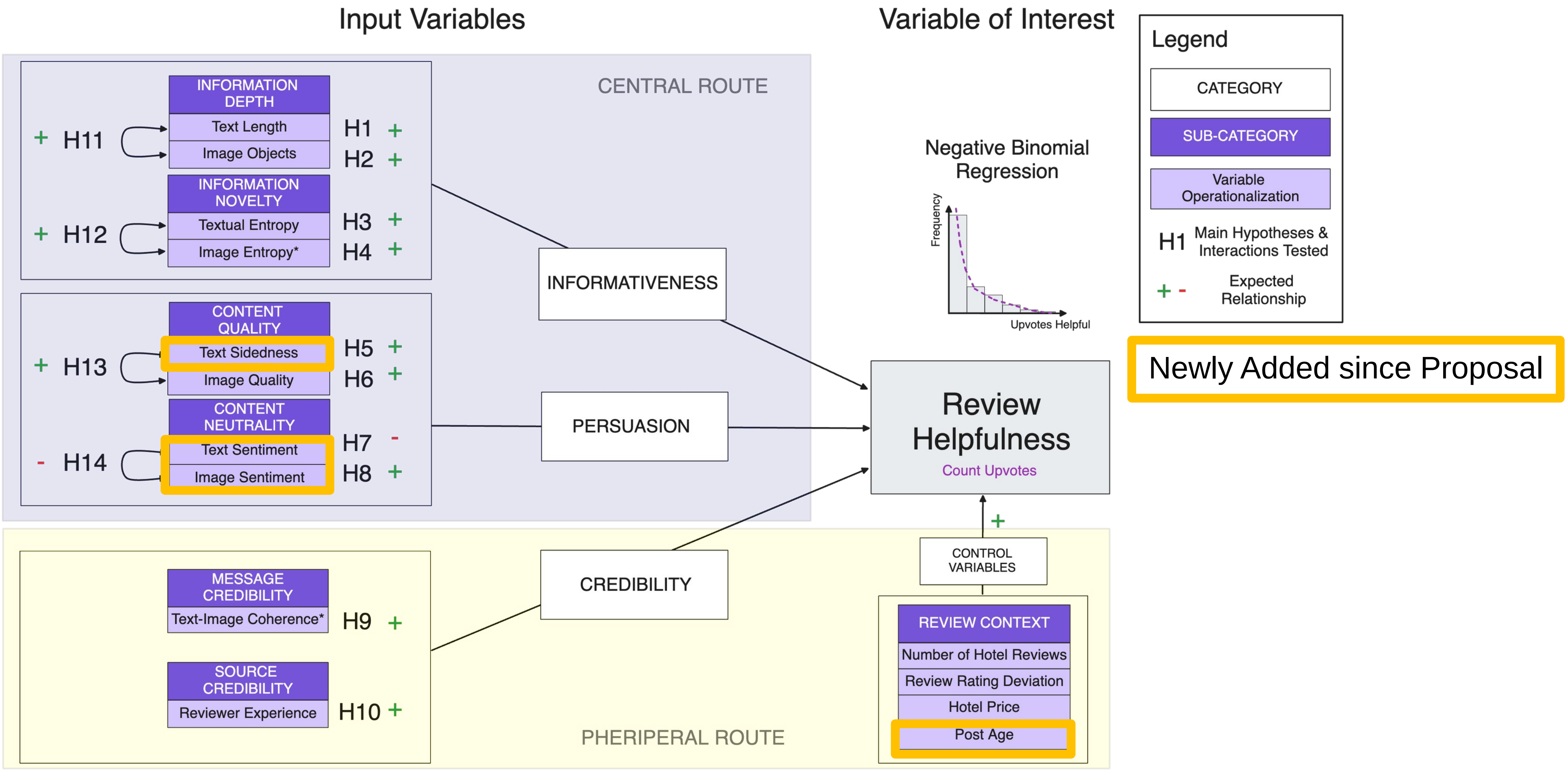
CHAPTER 2:

Synergizing Text And Images - A Multimodal Framework for Online Reviews

From Literature Gaps to Model Framework

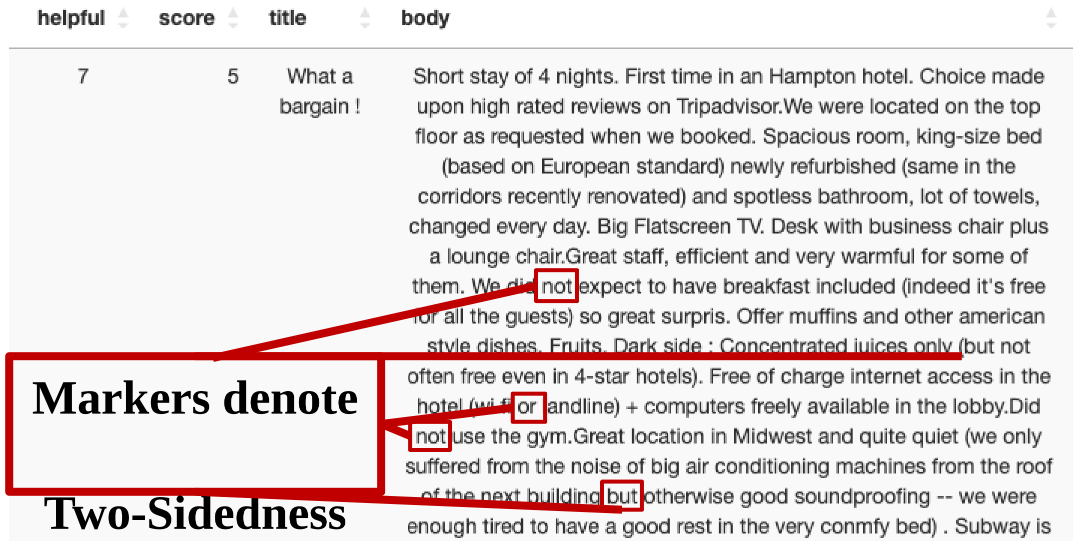
Research Area	Literature Insight (with citations)	Research Gaps
Multimodality	Most work analyzes text; reviews are now multimodal (Hong et al., 2017; Hu et al., 2022; Yang et al., 2023)	How users evaluate text vs. images vs. metadata in helpfulness
Multimodality	Text drives attention; images are more memorable and transfer to other elements (Bigné et al., 2020; Zinko et al., 2020)	Interplay between text and images and relative weight within ELM categories
Central vs. Peripheral	Central route more effective when product risk is high (Kübler et al., 2023; Mudambi & Schuff, 2010)	Importance of informativeness, persuasion, credibility, context
Information Depth	Consumers seek thorough info; images can be richer than text but not always more relevant (Srivastava & Kalro, 2019; Mudambi & Schuff, 2010)	Are images richer than text, and does it drive helpfulness?
Information Novelty	Novel/surprising info drives engagement and helpfulness (Fresneda and Gefen, 2019; Singh et al., 2017)	New visual information drive helpfulness as in text?
Content Quality	High-quality arguments increase persuasiveness (Cheung, 2008; Filieri, 2015; Srivastava & Kalro, 2019)	Are image quality cues stronger than textual argumentation?
Content Neutrality	Emotional intensity drives reactions; but users seek online reviews unbiasedness (Baek et al., 2015)	Is message neutrality processed differently in text vs. images?
Review(er) Credibility	Consumers trust reviews (and reviewers) that appear genuine, consistent, and unbiased (Racherla and Friske, 2012)	How do users detect credibility patterns in messages? Is message credibility > Author?
Contextual	Helpfulness depends on hotel type, price, rating, and review age (Fang et al., 2016; Filieri, 2015; Mudambi & Schuff, 2010)	-

Text-Image Elaboration Likelihood Model



Framework adapted from ELM Petty and Cacioppo (1983), Luo et al. (2013), and Srivastava and Kalro (2019).

Content Quality: Text Sidedness



helpful	score	title	body
7	5	What a bargain !	Short stay of 4 nights. First time in an Hampton hotel. Choice made upon high rated reviews on Tripadvisor.We were located on the top floor as requested when we booked. Spacious room, king-size bed (based on European standard) newly refurbished (same in the corridors recently renovated) and spotless bathroom, lot of towels, changed every day. Big Flatscreen TV. Desk with business chair plus a lounge chair.Great staff, efficient and very warmful for some of them. We did not expect to have breakfast included (indeed it's free for all the guests) so great surpris. Offer muffins and other american style dishes. Fruits. Dark side : Concentrated juices only (but not often free even in 4-star hotels). Free of charge internet access in the hotel (wi-fi or landline) + computers freely available in the lobby.Did not use the gym.Great location in Midwest and quite quiet (we only suffered from the noise of big air conditioning machines from the roof of the next building but otherwise good soundproofing -- we were enough tired to have a good rest in the very conmfy bed) . Subway is

Illustration of text two-sidedness markers in an online review.

- **High-quality arguments** promote deeper elaboration and attitude change.
- **Sidedness** = one-sided (only pros/cons) vs. two-sided (pros and cons).
- **Two-sided argument reviews** reduce uncertainty and build credibility — unlike social media, where content is more expressive.
- **Text sidedness** is measured using LIWC-22 “Cognitive Processes” markers (e.g., but, not, however, although).
- **Higher share of markers** = higher two-sidedness

Cheung et al., 2008; Watts & Zhang, 2008; Mudambi & Schuff, 2010; Chua & Banerjee, 2015; Lutz et al., 2018; Filieri, 2015; Park, 2018; Topaloglu & Dass, 2021

Content Neutrality: Image/Text Sentiment

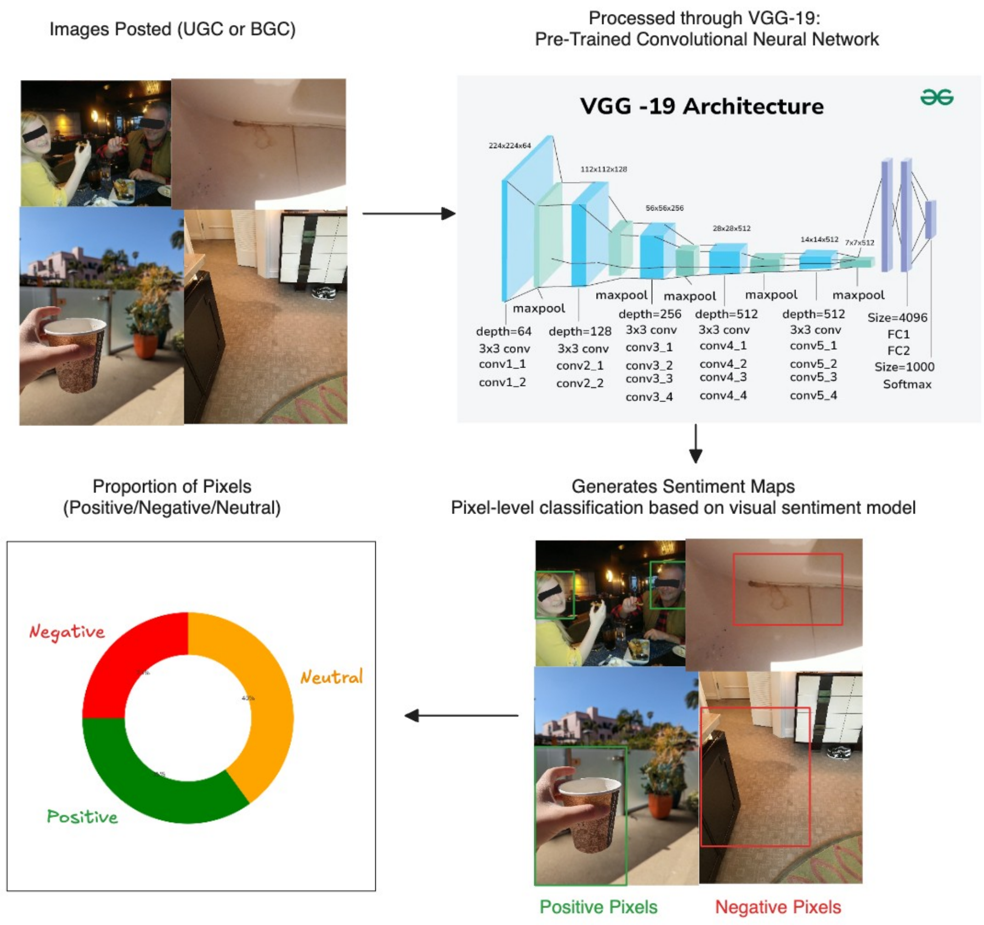


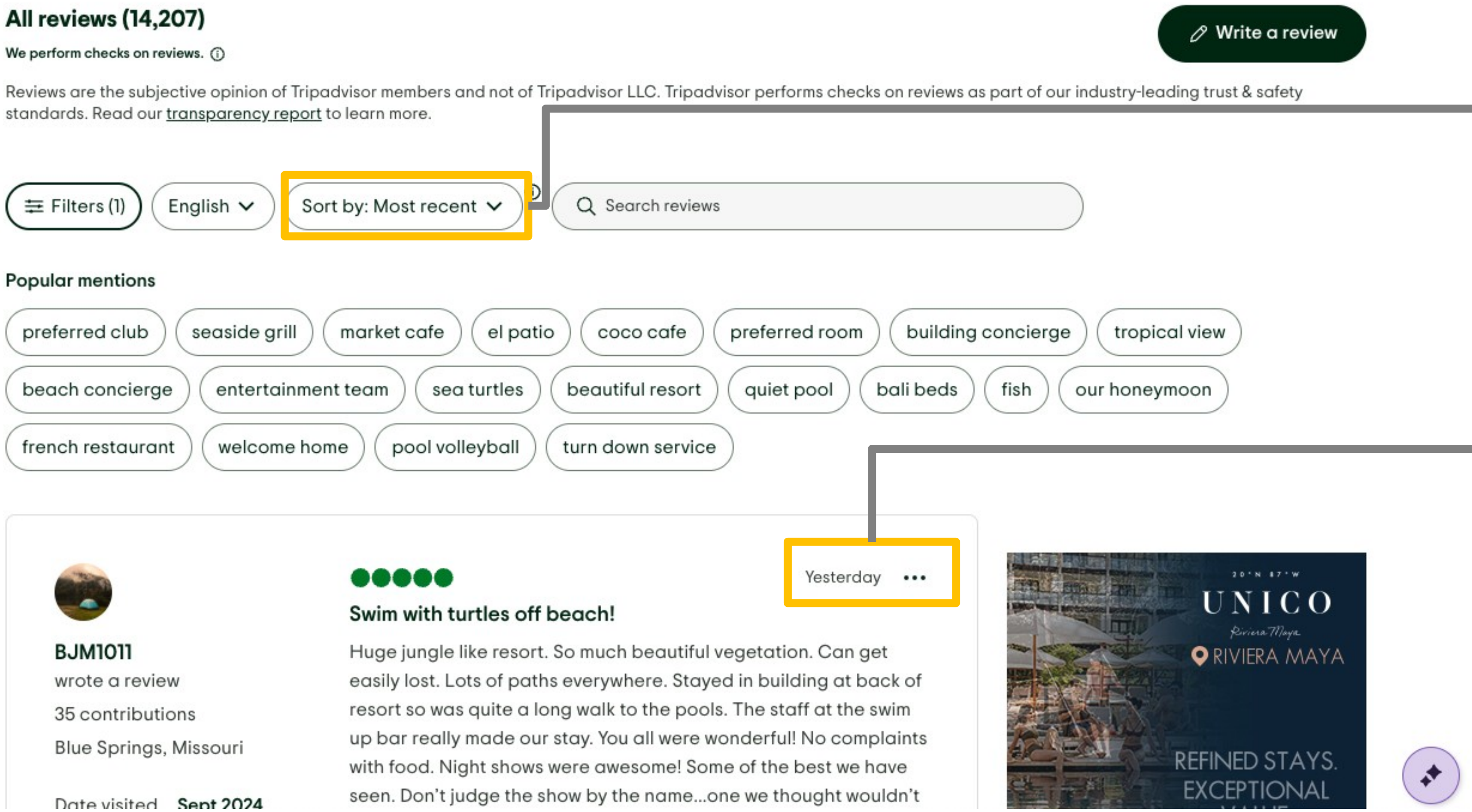
Image Sentiment Operationalization

- **Image/Text Sentiment** added across chapters (proposal: only in Ch. 3)
- **In text, neutral tone preferred** for balance (token-based).
- **Images seen as more trustworthy** experiential evidence, driving stronger affective responses in Online Reviews.
- Sentiment measured with VGG-19 CNN, using sentiment maps with pixel proportions of Positive, Negative, and Neutral:

$$\text{Sentiment} = \frac{|\text{Positive} - \text{Negative}|}{1 - \text{Neutral}}$$

Yin et al., 2014; Pan and Zhang, 2011; Salehan and Kim, 2016 (Text Sentiment)
M.-Y. Chen et al., 2019; H. Li et al., 2022; Chen, Teng and Chiou 2019; Guan et al. 2025 (Image Sentiment)

Review Context: Post Age



- Reviews are displayed by recency, also used in scrapping process.
- Recency increases visibility and perceived helpfulness.
- We control for age as time since post.
- This is consistent with literature and during proposal was in Chapter 3 and 4. Now in all Chapters.

Methodology Updates

Negative Binomial over Zero-Inflated

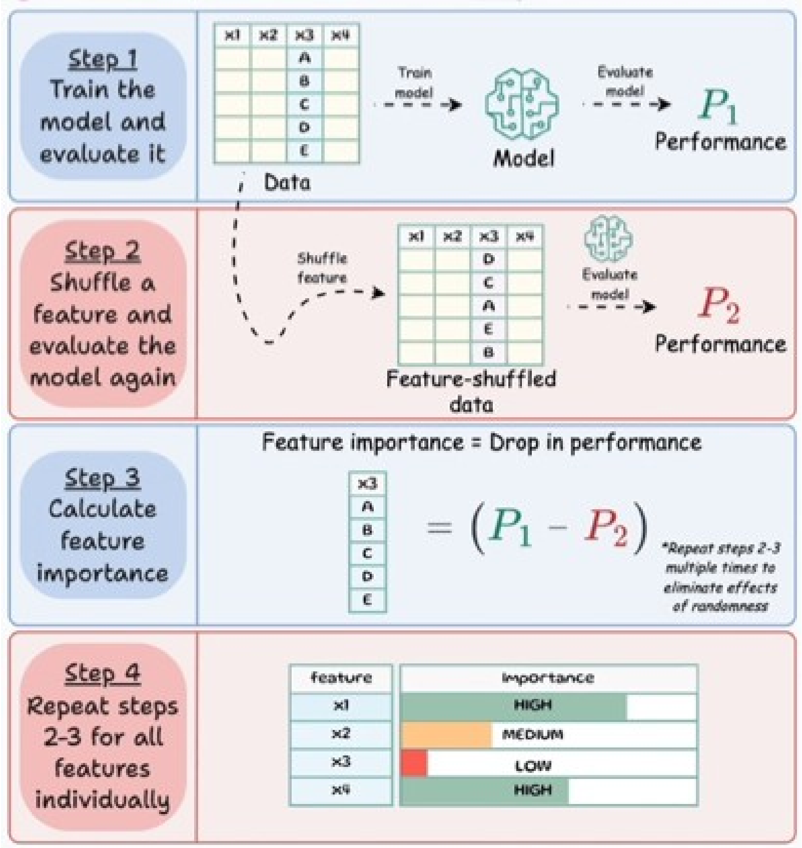
1. Alignment with prior Literature.
2. Theoretical Parsimony and Interpretability.
3. Modeling Simplicity and Analytical Flexibility.
4. Empirically our Negative Binomial fits well excess zeros:

Test	Check	Statistic	Results
Outlier Analysis (Cook, 1977)	Cook Distance	0.94	No Influential Outliers
Zero Inflation (Warton, 2005)	Theoretical vs Observed Zeros	1.01	No Excess Zeroes
Overdispersion (Cox and Lewis, 1966)	Dispersion Ratio	1.42	Some overdispersion (same as zero-inflated)
Probabilistic Integral Transform (PIT) (David and Johnson, 1948)	Over/under estimation	<0.05	**Within boundaries 95% uniform range**

5. **Recalculated heteroskedasticity-consistent variance (Huber–White)** following *twopartm* package implementation

$$\mathrm{Var}_{\mathrm{robust}}(\beta) = (X^T W X)^{-1} \cdot (X^T \hat{\Omega} X) \cdot (X^T W X)^{-1}$$

DALEX for Variable Importance



Source: Visual adapted from Daily Dose of Data Science (https://blog.dailydoseofds.com).

(1) e.g., Fang et al., 2016; Li et al., 2022; Park and Nicolau, 2015; Yang et al., 2023; (2) Cameron and Trivedi, 2013; (3) Hilbe, 2011; (5) Zeileis, 2004; Duan et al. 2022; DALEX = Model Agnostic Language for Exploration and eXplanation

Model Findings

R. Area	Finding	Test / Hypothesis supported	Result
Multimodality	Models including Visual features significantly improves fitting compared to Text-only models	AIC/Log-likelihood	Supported ***
Multimodality	Text-Image interaction are not significant except for Quality	P-value (H14 only)	Supported ***
Multimodality	Image > Text in terms of variable importance for the model	DALEX – Poisson Log Loss	↑ Relative Importance (%)
Informativeness	Information-Rich (both text and images) reviews are more helpful	P-value (H1, H2)	Supported ***
Informativeness	Information-Novelty (both text and images) reinforce helpfulness	P-value (H3, H4)	Supported ***
Informativeness	Image Entropy (Novelty) is more important than Text Entropy	DALEX – Poisson Log Loss	↑ Relative Importance (%)
Persuasion	Text Sentiment decreases helpfulness (Neutrality) while Image Sentiment increases it.	P-value (H5, H6)	Supported ***
Persuasion	Text Neutrality is more important than Image Sentiment in helpfulness	P-value (H7, H8)	Supported ***
Persuasion	Image Quality is the most important non-contextual feature	DALEX – Poisson Log Loss	↑ Relative Importance (%)
Credibility	Review and Reviewer Credibility increase helpfulness	P-value (H9, H10)	Supported ***
Credibility	Reviewer Credibility is more important than Message Credibility	DALEX – Poisson Log Loss	↑ Relative Importance (%)
Contextual	Control Variables: Hotel context and Time variables increase helpfulness	P-value (Control)	Supported ***
Contextual	Contextual Elements > Persuasion > Credibility > Informativeness in importance	DALEX – Poisson Log Loss	↑ Relative Importance (%)

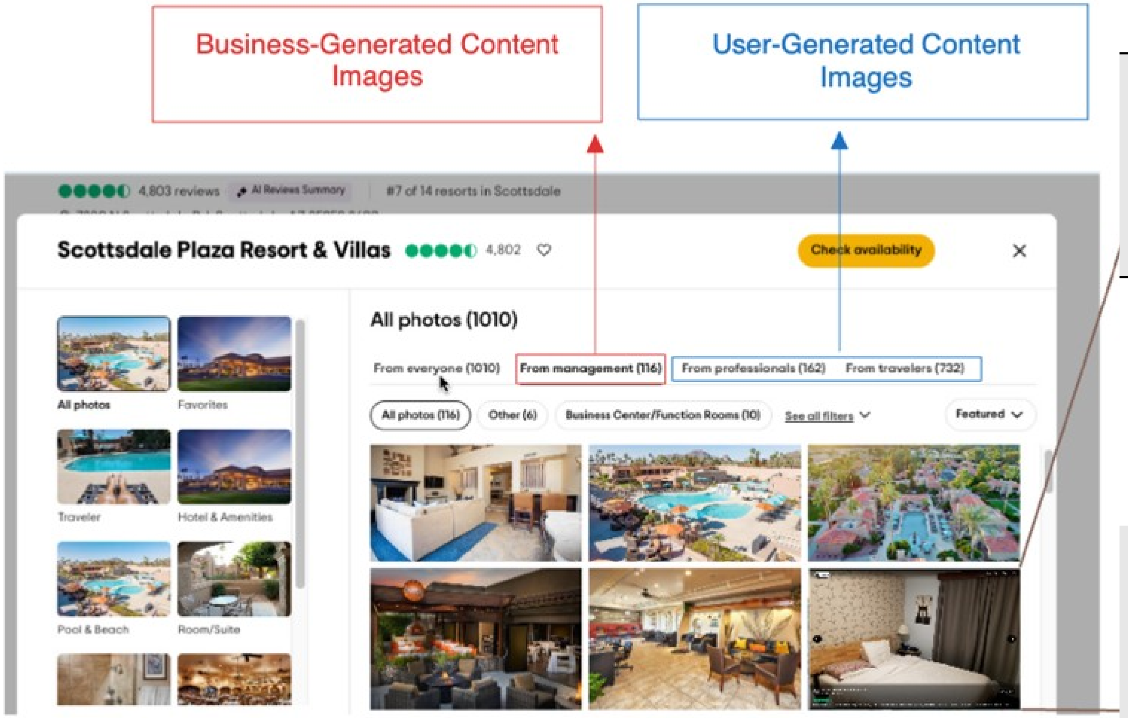
Note: *p<0.05; **p<0.01; ***p<0.001

DALEX = Model Agnostic Language for Exploration and eXplanation

CHAPTER 3:

Analysis of Business-Generated Content and Helpfulness Optimization

Business-Generated Content Key Research Questions



Research Area	Research Gap	Research Questions
BGC & Helpfulness	Determinants of BGC helpfulness remain underexplored, unlike UGC.	RQ1: What makes visual BGC helpful?
BGC vs UGC	Role of BGC in consumer decision-making is unclear.	RQ2: How do drivers of helpfulness differ between BGC and UGC?
Optimizing BGC	No research offers strategic guidance for managers on posting BGC.	RQ3: How can businesses optimize BGC to increase helpfulness?

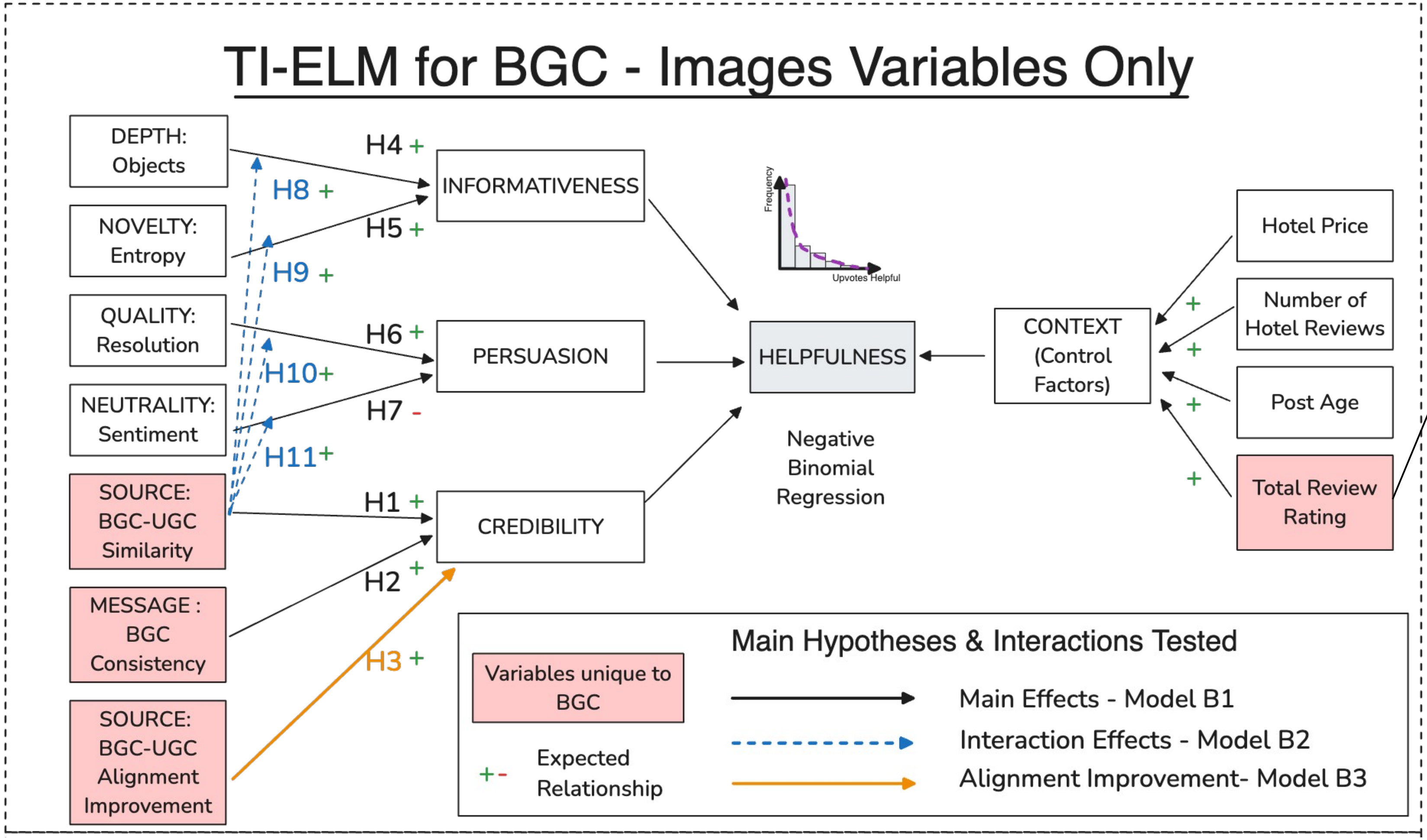
BGC = Business-Generated Content; UGC = User-Generated Content

Current Literature Insights and Gaps

Research Area	Literature Insight (with citations)	Research Gaps
BGC Helpfulness	BGC generates less trust/impact vs. UGC due to commercial intent (Friestad & Wright, 1994; Amazeen & Wojdynski, 2019).	Role of BGC in review helpfulness still unclear.
BGC through ELM	Drivers within central and peripheral route are well-documented in UGC (e.g., Chevalier and Mayzlin 2006; Luo et al. 2013; Wang and Karimi 2019	Not tested if mechanism apply in BGC review helpfulness.
Source Credibility	Aligning BGC with UGC reduces skepticism (Sukontip et al., 2024; Kwok and Xie, 2016).	Does UGC-BGC alignment increase credibility?
BGC–UGC Interplay	Personalized responses enhance helpfulness, esp. for negative/inconsistent reviews (Jin et al., 2023).	Do other BGC–UGC interactions enhance helpfulness engagement?
Credibility Moderation	UGC credibility can spill over to BGC (Chae et al., 2025; Hochstein et al., 2023).	Does spill over effect apply in online reviews?
Visuals in Helpfulness	Advertising studies show visuals influence helpfulness (Kwok & Xie, 2016).	BGC visuals in online reviews remain underexplored.
Brand Identity	Consistent imagery can build trust (Zarantonello, 2009) but in some contexts seem over-curated (Becker & Giessenberg, 2023).	**Tension between brand consistency vs. variety.**

BGC = Business-Generated Content; UGC = User-Generated Content

TI-ELM for Business-Generated Content



For UGC, rating deviation = review rating – hotel's average rating. For BGC, businesses do not self-rate; thus, the hotel's average UGC rating is used.

Variables operationalized for BGC as in UGC TI-ELM except for Credibility and Total Review Rating variables

TI-ELM = Text-Image Elaboration Likelihood Model; BGC = Business-Generated Content; UGC = User-Generated Content

Operationalization of BGC Credibility

*Businesses can deceive users through visually enhancing images or selectively posting favorable content**

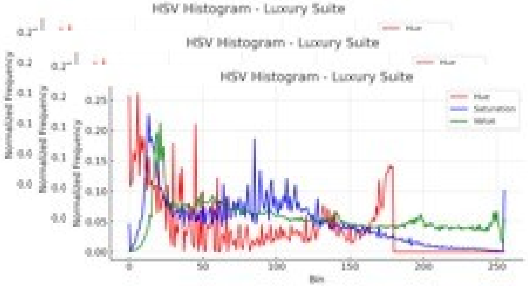
This approach captures these deception strategies compared to what regular users post:

Input Images | **Source**

UGC

Compute HSV histogram**

Average histogram across UGC



Compare**each BGC bin with the average UGC style

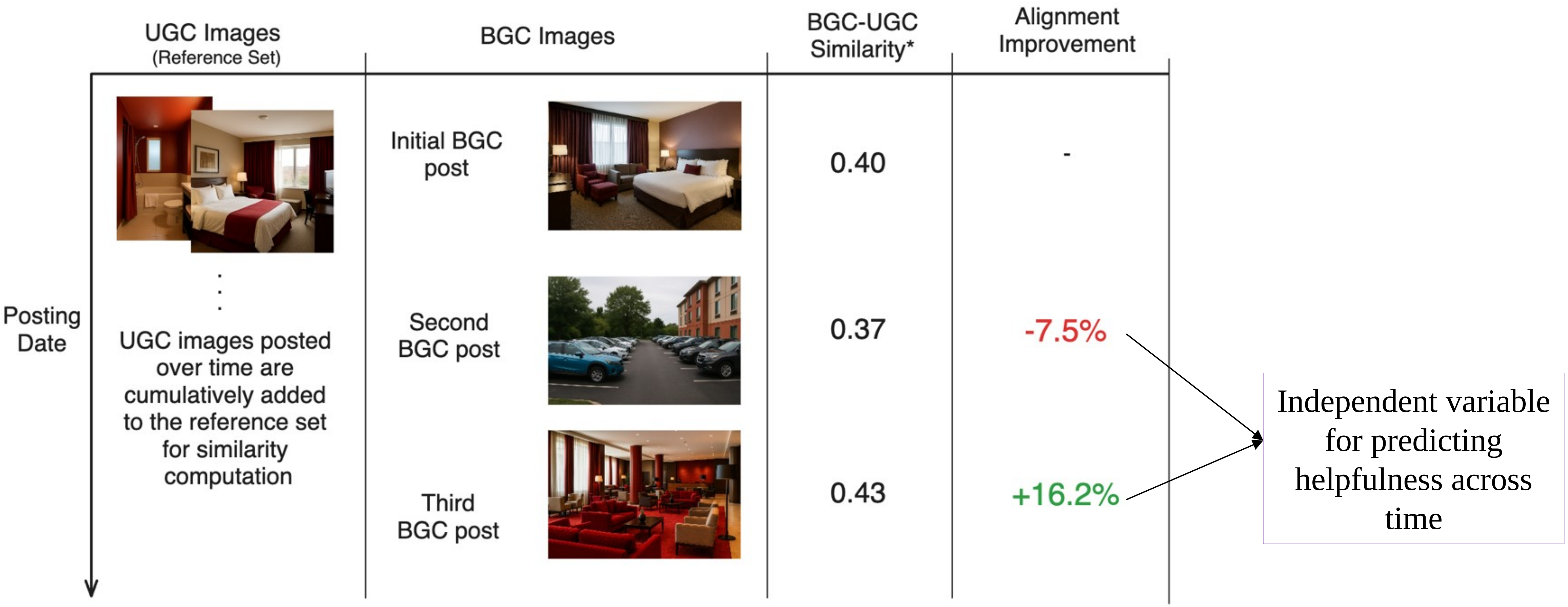
$$\cos(\theta) = \frac{\mathbf{A}\cdot\mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}}$$

BGC

***Reduced to 16 bins per histogram*

Is Aligning Content a Good Strategy for BGC?

Quasi-uplift* experiment to measure management posting strategy



* Adapted from Kontopantelis et al. (2015); Kim et al. (2024). Quasi-uplift experiment: longitudinal observational design that approximates uplift modeling by comparing BGC–UGC similarity changes over time with their effect on helpfulness.

Model Findings

RQ	Finding	Test (Hypothesis)	Result
RQ1: Drivers of BGC Helpfulness	TI-ELM drivers also apply to BGC (Significance across categories)	AIC/Log-likelihood	Supported ***
	BGC-UGC Similarity boosts credibility overall helpfulness fitting	P-value (H1)	Supported ***
	Diverse BGC more credible than consistent branding	P-value (H2)	Opposite Sign ***
	Informativeness Depth remains positive and significant despite BGC skepticism	P-value (H4)	Supported ***
RQ2: BGC vs. UGC Differences	Average Helpfulness for BGC is lower than UGC, and so is average coefficient impact	ANOVA P-Value	Supported ***
	While Significant, BGC Entropy, Quality, Sentiment have opposite signs to UGC	P-value (H5, H6, H7)	Opposite Sign ***
	Moderation effect of BGC-UGC Similarity has spill over effect flipping opposite sign variables (Entropy, Quality, Sentiment)	P-value (H9, H10, H11)	Supported **
	Moderation effect not significant for Informativeness Depth	P-value (H8)	Not Supported
RQ3: Optimizing BGC	UGC sets the narrative, BGC should follow. Findings show that when businesses adapt BGC to mirror UGC positioning, helpfulness increase >1.08x	P-value (H3)	Supported ***
	In BGC: Informativeness > Credibility > Persuasion opposite order as UGC	DALEX – Poisson Log Loss	↑ Relative Importance (%)
Contextual	Control Variables: Hotel context and Time variables increase helpfulness	P-value (Control)	**Supported *** **

Note: *p<0.05; **p<0.01; ***p<0.001

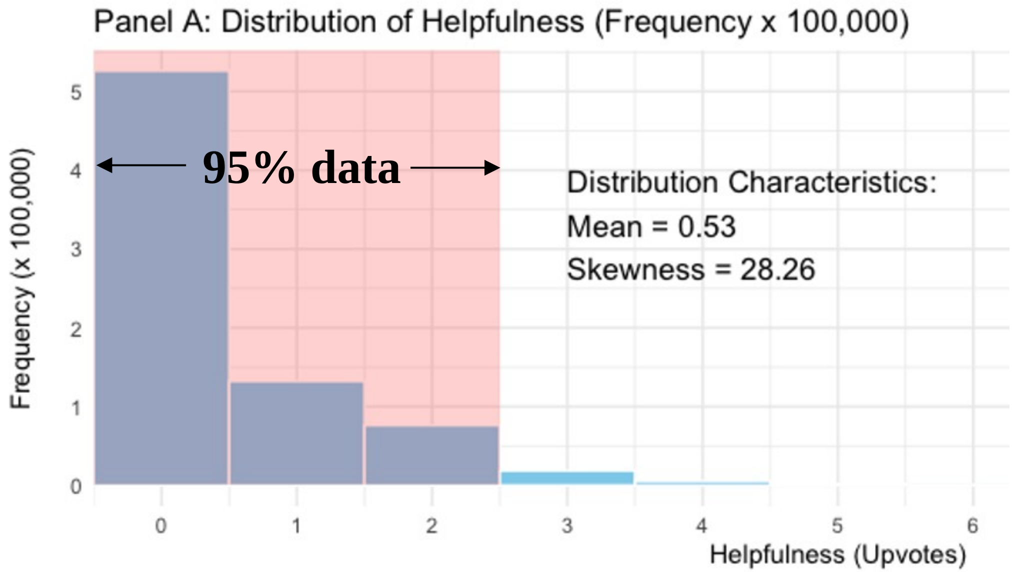
RQ = Research Question

CHAPTER 4:

Comparing Drivers of Extremely Helpful Vs. High- and Low-helpfulness Reviews in Hospitality

Zero-Inflation Bias in Online Review Helpfulness

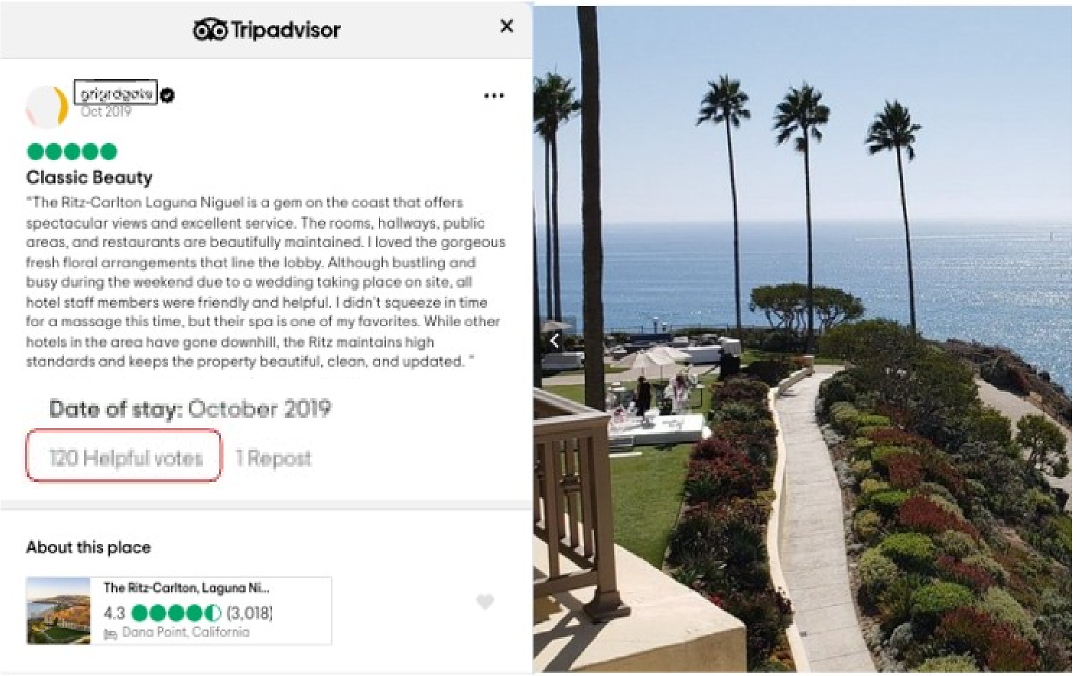
- Online review datasets are dominated by zero–helpfulness counts (70%+ in Hospitality) *
- Standard regression models focus on separating zero from low helpfulness (modal class).
- Then, high-helpfulness reviews are underestimated, and drivers remain unknown.
- While underexplored in online reviews, extreme helpfulness resembles "virality" in social media research.



Helpfulness distribution in our UGC dataset (95% of reviews clustered at zero/low counts)

* Estimated from studies listed in Hong et al. (2017); Rietsche et al. (2019); Saumya, Roy, and Singh (2023)

Virality Drivers of Extremely Helpful Reviews



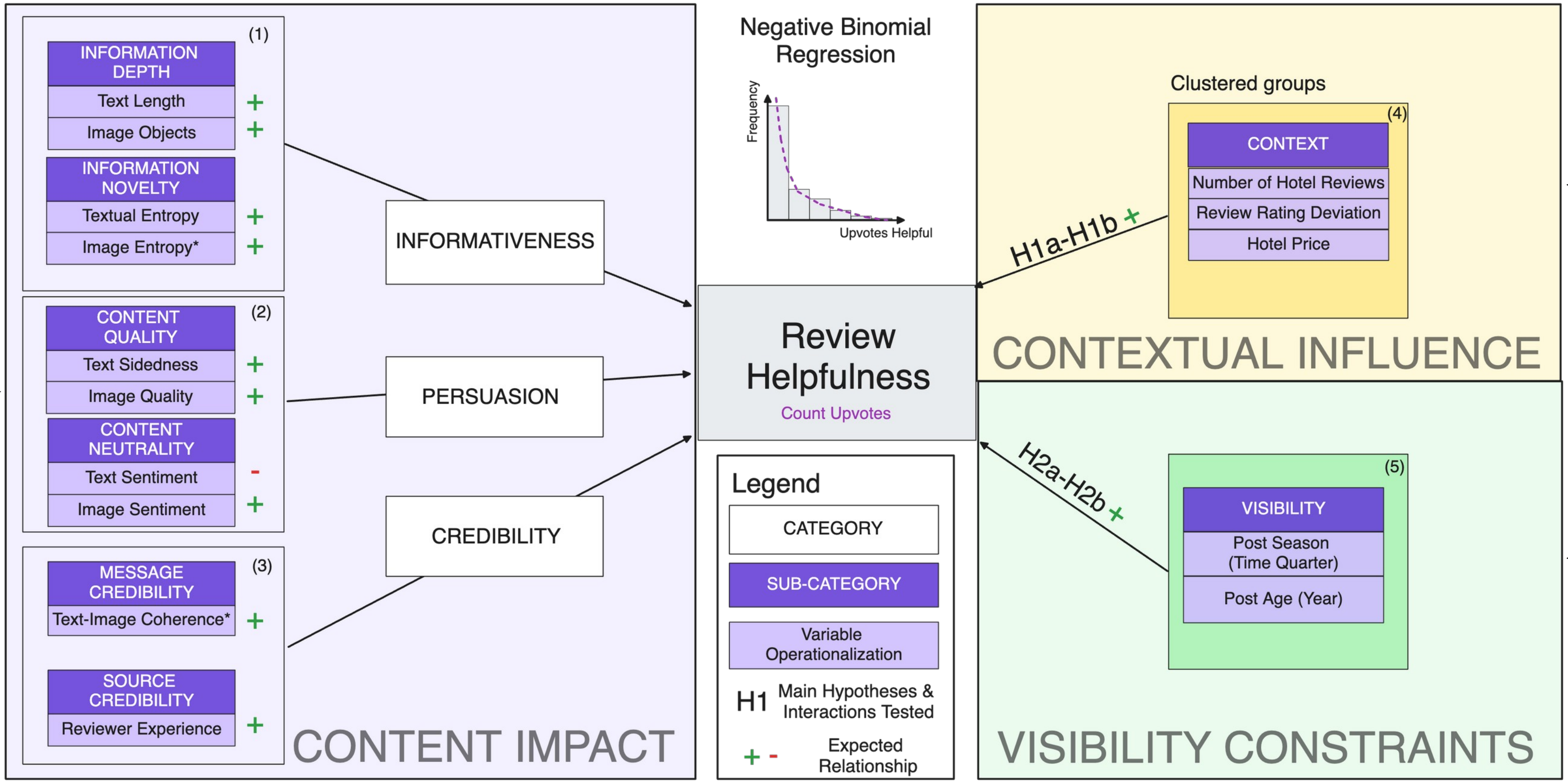
Literature Summary

Dimension	Platform	Insights
Content Impact	Social Media	Emotional arousal, novelty, and social cues drive engagement and sharing (Goel et al., 2016; Varol et al., 2017).
	Online Reviews	**Driven by content and source due to utility–Persuasion, Credibility, Informativeness.** (C. M. Cheung et al., 2008; Filieri, 2015)
Contextual Influence	Social Media	Trending topics, Platform reach and shareability guide attention (Berger and Milkman, 2012; Goel et al., 2016)
	Online Reviews	**Hotel type, price, and rating shape review reach and perceived risk and attention.** (Forman et al., 2008; Zhang and Tran, 2010)
Visibility Constraints	Social Media	Content is amplified via platform-wide algorithms, event relevance and follower networks. (Bakshy et al., 2012; Berger and Milkman, 2012; Goel et al., 2016)
	Online Reviews	**Helpfulness is amplified with time, relevance and search algorithms due to snowball effect** (Ghose and Ipeirotis, 2011; Luca, 2015)

Mapping Virality to TI-ELM Framework

And four key research hypotheses for each "Virality" dimension

(1) How content shifts across contextual and helpfulness levels (H1b)



(2) Differences for Low, High and Extreme count by Hotel Type (H1a)

(3) How Time & relevance drive visibility (H2a-H2b)

TI-ELM Framework adapted mapped to Virality components (Bakshy et al., 2012; Berger and Milkman, 2012; Goel et al., 2016; Varol et al., 2017)

Capturing Low/High/Extreme Count Effects

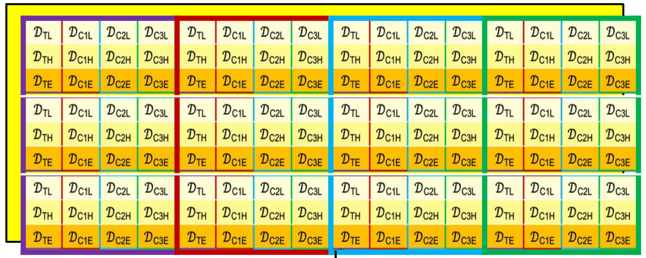
UGC

Total | C1 | C2 | C3

- Price
- Rating
- Num of Reviews

Determine Thresholds by Helpfulness percentile

75th	Low-Count (L)
95th	High-Count (H)
99th	Extreme-Count (E)

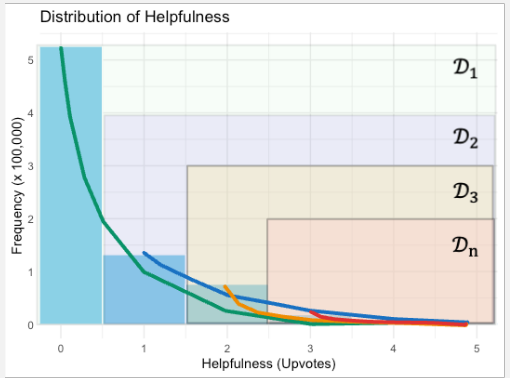


$\mathcal{D}_{ij}$ represent subsets by hotel cluster and helpfulness threshold

Group Differences (ANOVA) → Thresholds Diff. H1a

Quantile Neg. Binomial (INLA*) → Coefficient Trend H1b; Coefficient Rank H2a, H2b

1. **Cluster Data based on Contextual Characteristics**
2. **Separate datasets by cut-offs levels of Helpfulness**
3. **Regress by subset and compare differences (Average, Trend & Rank)**



Consistent with approach by (Filieri et al., 2018; Filieri & McLeay, 2014) and Cluster segmentation by (Forman et al., 2008; Zhang & Tran, 2010) Charrad et al., 2014 (NbClust); Hartigan and Wong, 1979 (K-means);
* INLA = Integrated Nested Laplace Approximation. Bayesian Negative Binomial Regression with Laplace approximations (Rue et al., 2009)

Model Findings

Research Area	Finding	Test / Hypothesis supported	Result
Extreme Helpfulness Thresholds	Thresholds vary significantly by Hotel tier. High-tier hotels need ≥5 upvotes; low-tier only ≥2 for “extreme” helpfulness.	ANOVA P-Value (H1a)	Supported ***
Contextual Influence	More reviews, higher price, and ratings → more helpfulness (esp. high-tier).	ANOVA P-Value (H1a)	Supported ***
Coefficient Trends	Coefficient importance shifts across thresholds, not always consistent.	P-Value (H1b)	Not Supported
Visibility Constraints	Seasonality & year rank top-3 drivers; relevance & snowball effects matter.	P-value (H2a, H2b); Rank	Supported ***
Visual Entropy	Entropy strongly affects high-end hotels; signals new visual info.	P-Value	Supported ***
Extreme Helpfulness Drivers	High influence from seasonality (visibility) and reviewer experience (content). In high-tier hotels content becomes more important	Rank	Supported
High Helpfulness Drivers	High influence from seasonality (visibility) and reviewer experience (content). In high-tier hotels content becomes more important	Rank	Supported
Low Helpfulness Drivers	Low helpfulness relies mainly on visibility (COVID, specific events, etc.)	Rank	Supported
Model Dispersion	Quantile Regression Overdispersion reduced within PIT boundaries	P-Value (PIT)	Supported ***
Context Variables	Contextual variables remain significant for low count helpfulness (consistent with research)	P-Value	Supported ***

Note: *$p<0.05$; **$p<0.01$; ***$p<0.001$

PIT = Probabilistic Integral Transform

Concluding Remarks

Key Contributions & Takeaways

1. **Theoretical Contributions**
   - Extend ELM with multimodal features (TI-ELM) for UGC & BGC (Ch. 2, 3 and 4)
   - Build a large-scale multimodal dataset (>880k reviews) (Ch. 2, 3 and 4)
   - Develop automated image-processing metrics (e.g., entropy, sentiment, credibility) (Ch. 2–3)
   - Empirically validate hypotheses with real-world hotel data (Ch. 2, 3 and 4)

2. **Empirical Findings**
   - Images boost helpfulness; text & images act independently except for quality (Ch. 2)
   - Persuasion > Credibility > Informativeness in UGC; reversed in BGC (Ch. 2–3)
   - BGC skepticism due to commercial intent; only depth escapes this effect (Ch. 3)
   - Virality driven by relevance & reviewer experience; only in high-tier hotels by content (Ch. 4)

3. **Managerial Implications**
   - UGC sets the narrative; BGC should align to increase engagement (Ch. 3)
   - Timely posts around events (e.g., COVID, concerts, seasonality) extreme helpfulness likelihood (Ch. 4)

NJIT

New Jersey Institute of Technology