
# SUBTLE SIGNALS, GENDERED OUTPUTS PHASE 2: HOW GENDER CUES SHAPE AI PRODUCT RECOMMENDATIONS

## INTRODUCTION

Phase 1 of this research demonstrated that gendered patterns could enter AI systems before content is generated, through subtle differences in how requests are phrased. When gender is encoded implicitly through language style rather than explicitly, prompts vary systematically in tone, structure and psychological orientation, differences that map onto well-established, socially gendered communication norms and operate independently of any declared or explicit identity labels.

Phase 2 builds on that foundation by shifting attention from prompts to outputs. In this phase we ask what happens after those differences are introduced into the system. Specifically, Phase 2 examines how implicit and explicit gender cues shape the products an LLM recommends, how those recommendations are framed, and what information they include.

By pairing AI-generated prompts with real-world, human prompts written under identical instructions, Phase 2 also tests whether gendered output patterns persist under naturalistic language use. Together, the two phases trace how gendered variation moves from interaction style to prompt construction to downstream recommendations, offering an end-to-end view of how bias can emerge in generative AI systems.

Beyond questions of technical bias, this study speaks to the cultural role generative AI systems increasingly play. When AI tools are asked to recommend products, they do not simply retrieve neutral options. Instead, they draw on shared cultural knowledge about gender, consumption and appropriateness. As a result, AI-generated recommendations can both reflect and reinforce existing cultural norms, shaping what feels natural, appropriate or expected for different users. Understanding these dynamics is critical not only for evaluating AI systems, but for anticipating how they may influence everyday consumer culture at scale.

## STUDY OVERVIEW

Phase 2 extends the experimental framework introduced in Phase 1. Using prompts generated in the first phase, we examine how implicit and explicit gender cues shape product recommendations produced by the same model (GPT-5). Specifically, we asked the model to generate a shopping list of five birthday gift ideas, with each item including a specific product and brand.

The analysis focuses on four dimensions of the output:

1. **The recommended categories and brands,** to understand how products are suggested by AI systems and how these suggestions may differ based on inferred gender.
2. **Linguistic framing,** with particular attention to the adjectives used to describe products.
3. **Topic modeling of the recommendation content,** to assess broader patterns in how products and attributes are emphasized across gender conditions.
4. **Price information, whether it is displayed and in what context.**

# METHODOLOGY

## Prompt Selection and Design

From the full set of prompts generated in Phase 1, we selected 10 per cell across two framing conditions - implicit and explicit gender - and six levels of femininity (0%, 20%, 40%, 60%, 80% and 100%). Femininity and masculinity are treated here as opposite ends of a single analytical scale.

Prompts that explicitly referenced shopping categories were excluded to avoid priming effects. Neither the implicit nor explicit prompts stated the gender of the intended gift recipient. The difference between the implicit and explicit conditions differs in how the prompts were originally constructed, utilizing explicitly gendered language, or implicitly gendered language (which are considered common linguistic differences between women and men). For more detail, please see Phase 1 of this research.

## Human-Written Prompts

To complement AI-generated prompts, we solicited prompts from 20 human participants (10 men and 10 women). Participants were instructed to write a prompt as if they were naturally interacting with an LLM. They were given **the same task instructions** provided to the model in Phase 1. These prompts tended to be shorter and more general than AI-generated prompts, reflecting naturalistic user behavior. This was similar to the AI-written prompts, in that none of the prompts stated the gender of the intended gift recipient.

## Model and Execution

All prompts, AI-generated and human-written, were run through ChatGPT 5 using the OpenAI API. Outputs consisted of five-item[1] birthday gift recommendation lists.

## Coding and Reliability

Two independent human coders categorized each recommended item into one of ten pre-defined product categories (see Figure 1).

Coders also recorded whether each recommendation included price information.

Figure 1


Inter-coder reliability was high (Cohen's κ = 0.9242), indicating strong agreement.

# RESULTS: PRODUCT CATEGORY DIFFERENCES

## Patterns Shared Across Implicit and Explicit Gender Cues

Across both implicit and explicit framing conditions, two category patterns were robust:

- **More masculine prompts** were more likely to yield recommendations in *books, media and specialty gifts* (e.g., e-readers, notebooks).
- **More feminine prompts** were more likely to yield recommendations in *home* and *décor* (e.g., throw blankets, scented candles).

1 The original prompts from Phase 1 were modified to request 5 recommended items rather than 10, to account for maximal sample size within practical constraints for human coders.

These categories appear to function as **culturally stable, low-risk gender associations.** They are widely normalized and socially acceptable across contexts, which may explain why they emerge regardless of whether gender is subtly implied or explicitly named.

### Implicit and Explicit Gender Cues: Sensitivity and Constraint

Some product categories appeared only under one framing condition rather than across both implicit and explicit gender cues. For example, **toys, games and creative kits emerged primarily under implicit masculine cues,** while **beauty and personal care products appeared primarily under explicit feminine cues.**

These findings suggest differences in how the model handles **socially sensitive versus culturally stable gender associations.** Categories such as toys and games may be culturally associated with masculinity without being socially sensitive, allowing them to surface when gender is inferred. In contrast, appearance-related categories such as beauty and personal care are more explicitly gendered and socially sensitive and appear only when gender is clearly specified.

This pattern is consistent with the presence of safety and appropriateness constraints commonly implemented in LLMs. While the internal operation of such guardrails is not directly observable, the selective emergence of certain gendered categories, particularly those tied to appearance or self-presentation, suggests that the model is balancing responsiveness with caution. In this sense, explicit gender cues may function less as a trigger for stereotyping and more as a signal that certain recommendations are socially permissible.

## OUTPUTS FROM HUMAN PROMPTS

When examining outputs generated from human-written prompts, additional differences emerged. **Prompts written by men were more likely to receive recommendations in *travel and outdoor gear*, *toys and games*, and *consumer electronics.***

These differences appeared despite identical task instructions, suggesting that the model is sensitive to gendered cues embedded in natural language use. Importantly, this indicates that gendered output patterns are not limited to optimized or machine-generated prompts. They persist under everyday interaction styles.

## BRAND CONCENTRATION AND REPETITION IN AI OUTPUTS

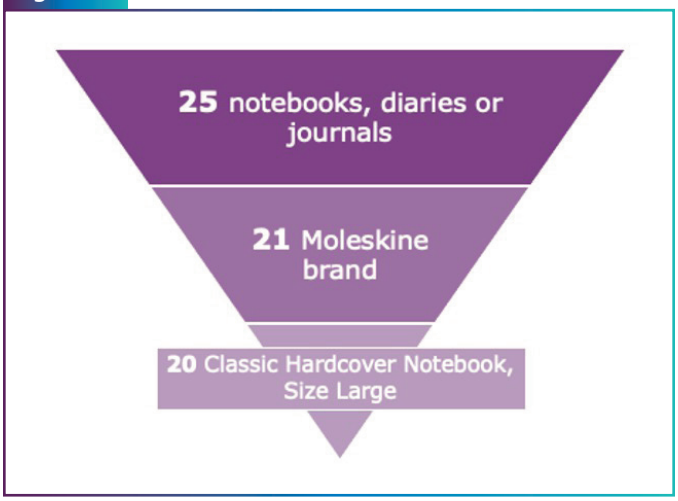
Across conditions, the model showed a strong tendency to repeatedly recommend the same brands, and often the same specific products, within individual product categories. Rather than drawing from a wide range of options, recommendations frequently converged on a small number of culturally dominant brands that appeared to function as default category representatives (see Figures 2-4).

Figure 2


Figure 3

Similar themes of brand and product dominance were seen across categories. In consumer electronics, recommendations were dominated by the Apple Air tag and Air pods. In home décor, blankets were recommended 28 times, with 21 of them being the Barefoot Dreams Cozychic throw blanket. Drinkware recommendations were dominated by the Yeti 20 oz. tumbler and the Ember temperature control smart mug, while candles were dominated by the Diptyque Baies candle.

Figure 4




Taken together, these results show that AI-generated shopping recommendations are not only category-skewed but **highly concentrated at the brand and product level.** When asked to produce brand-specific gift ideas, the model repeatedly returns the same small set of culturally dominant products, indicating that these items function as canonical stand-ins for entire categories, within the model's learned representations.

Importantly, this pattern does not reflect user preference or optimization for novelty. Instead, it indicates that certain brands occupy a disproportionate symbolic position within the model's learned representations, functioning as stand-ins for entire product categories.

## LINGUISTIC FRAMING AND ADJECTIVE USE

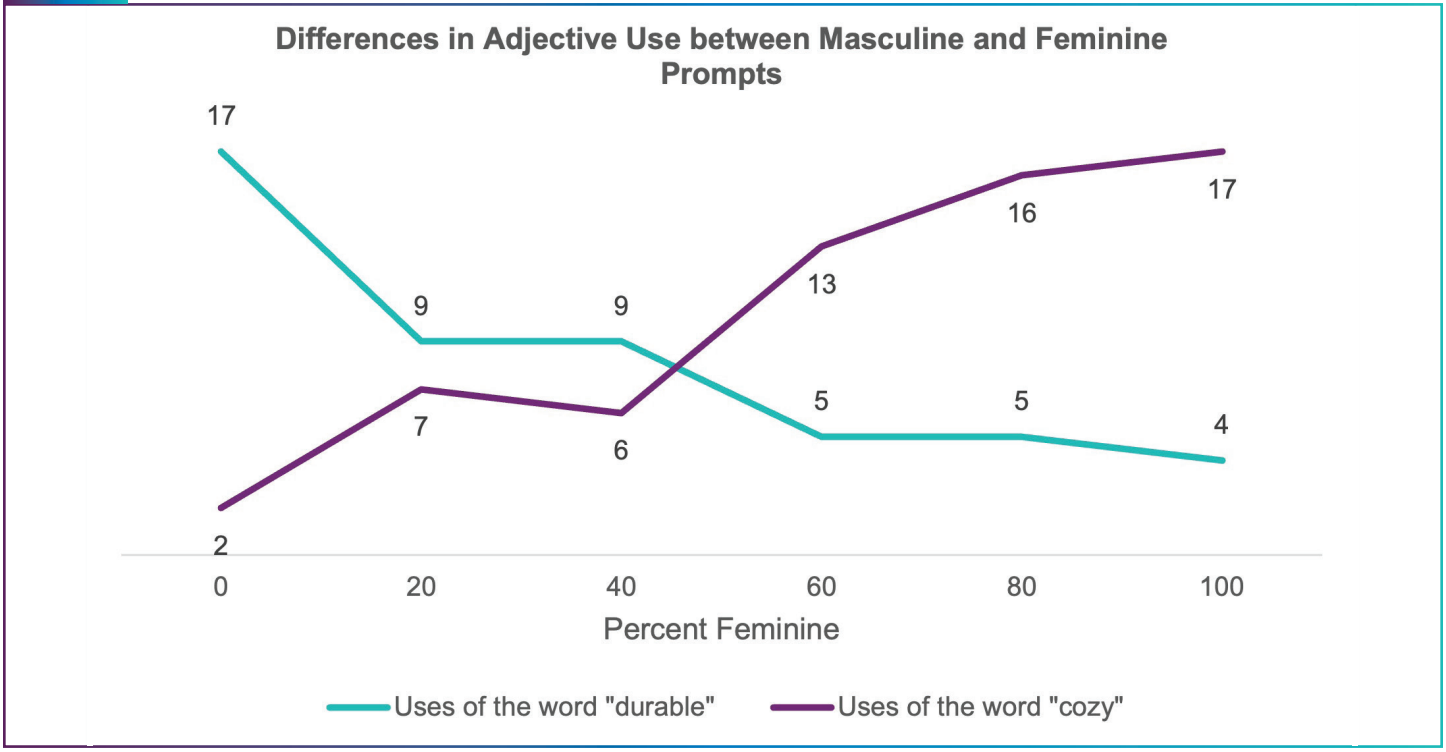
Using LIWC-based analysis, we examined the recommendation lists' linguistic features by analyzing the words used in the model's responses, rather than the prompts themselves. Each AI-generated shopping list was processed through LIWC to quantify the presence of words associated with specific semantic domains, including terms related to technology, work, leisure, money and home. This allowed us to assess whether the language used to describe recommended products varied systematically with gender cues, independent of the product categories themselves.

## LIWC ANALYSIS

**As femininity increased, responses used fewer words related to technology** across both implicit and explicit conditions, as well as in human-prompt outputs. **In the implicit condition only, increasing femininity was also associated with fewer work-related terms.** No gender differences were observed in language related to leisure, money or home.

Analysis of adjective use revealed no difference in the number of adjectives used across gender conditions, but clear differences in type. More feminine prompts were associated with adjectives such as **"cozy"** and **"luxury,"** while more masculine prompts were associated with adjectives such as **"durable"** (see Figure 5).

Figure 5




These adjectives signal different psychological and sociological logics of consumption:

- **"Cozy" and "luxury"** frame products through affective processing, emphasizing comfort, emotion and sensory experiences. They are culturally associated with domesticity, care and aesthetic value.
- **"Durable"** frames products through instrumental processing, emphasizing reliability, longevity and functional justification. It is culturally associated with work, endurance and mastery over objects.

When these descriptors are deployed differently, the model is not merely differentiating products. Rather, it is assigning different **rationales for consumption** depending on perceived gender.

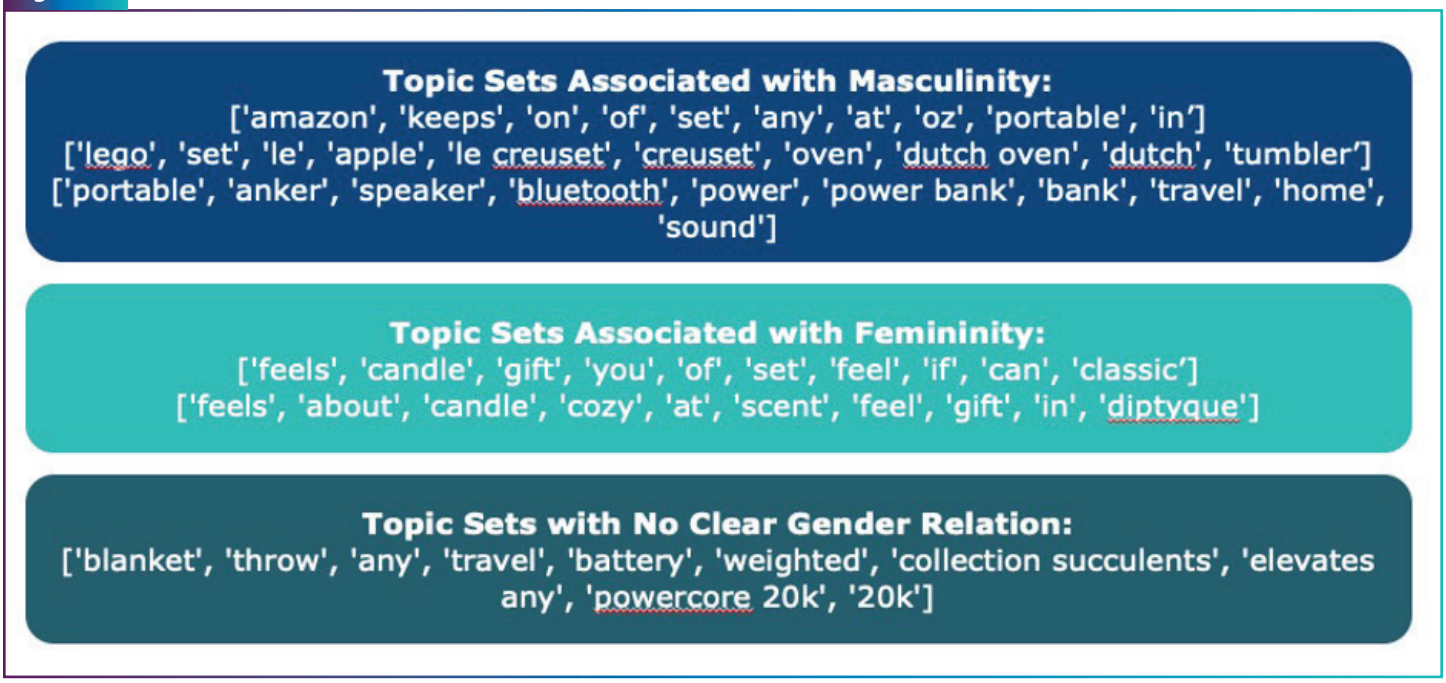
## TOPIC MODELING OF RECOMMENDATION CONTENT

To assess whether the linguistic differences observed in the LIWC analysis also appeared at a broader, semantic level, we conducted topic modeling on the AI-generated recommendation lists. Topic modeling was used here not to discover new themes, but to examine whether recommendations clustered differently in terms of overall emphasis as gender cues varied.

We applied BERTopic to the recommendation lists, configuring the model to generate six topics to avoid over-fragmentation, given the relatively small size of the dataset. Prior to analysis, light preprocessing was applied to standardize capitalization and remove extraneous whitespace. We evaluated three configurations: a baseline model, a model excluding common English stop words, and a model excluding both English stop words and task-specific terms (e.g., “birthday,” “price,” “brand”). The latter produced the most interpretable results and is used in the analyses reported here.

The resulting topic clusters revealed systematic differences in how recommendations were framed when femininity and masculinity varied (see Figure 6). Topics that peaked at higher levels of femininity were characterized by affective and experiential language, including terms related to feeling, comfort, scent and atmosphere. In contrast, topics that peaked at higher levels of masculinity were dominated by concrete product references, brand names and functional attributes such as portability, power and technical features.

Figure 6




Importantly, not all topics exhibited clear gendered trends. One cluster showed a relatively stable distribution across gender conditions, suggesting that gender cues selectively influence certain modes of description rather than uniformly shaping all recommendation content.

Taken together, the topic modeling results echo the pattern observed in the previous language-based analyses. Rather than introducing distinct thematic categories, gender cues appear to shift the emphasis of recommendations, from experiential and affective framing to object-centered and functional framing, thus reinforcing the conclusion that **gender influences how products are described as much as which products are recommended.**

## PRICE INFORMATION AND INFORMATIONAL FRAMING

**In the implicit condition, more feminine prompts were less likely to receive price information** than more masculine ones. This difference did not appear in the explicit condition or in the outputs generated from human-written prompts.

This pattern suggests that implicit femininity cues may shift the model toward a more experiential framing of consumption, where emotional resonance and atmosphere are foregrounded and price becomes less central. Explicit gender cues and human prompts, by contrast, appear to anchor responses more firmly in instrumental, decision-making norms.

This tendency is also reflected in the topic modeling results, where recommendations associated with higher femininity cluster around feelings and experiential language, whereas masculinity-peaking recommendations cluster around products, brands and functional attributes.

## KEY FINDINGS IN PHASE 2

Phase 2 demonstrates that gender bias in generative AI operates not only at the level of *what* is recommended, but also *how* recommendations are framed and justified. Implicit and explicit gender cues shape category selection, descriptive language and informational completeness in systematic but distinct ways.

1. **Gender cues shape recommendations**

Even when all prompts asked for the same type of output (a shopping list of gift ideas), gender cues systematically influenced the recommendations produced by the model. These effects appeared across both implicit and explicit gender framing, indicating that gendered variation operates downstream of the prompt, not just at the level of prompt construction.

2. **Implicit and explicit gender cues activate different types of associations**

Implicit gender cues tended to shift recommendations toward softer, associative domains, such as play, creativity and experiential products, while explicit gender cues activated more normatively gendered categories. For example, toys and games emerged primarily under implicit masculine cues, whereas beauty and personal care products appeared primarily under explicit feminine ones. This suggests that the model distinguishes between inferred and declared gender when selecting what is socially appropriate to recommend.

3. **Gender cues influence how products are described, not just which products appear**

Across analyses, femininity was associated with language emphasizing feeling, comfort and experience, while masculinity was associated with language emphasizing function, durability and technical attributes. These differences were visible in adjective use, LIWC-based linguistic features and topic modeling results, indicating a consistent shift in descriptive emphasis rather than isolated word choice effects.

4. **Gendered patterns persist under naturalistic, human-written prompts**

Outputs generated in response to real-world human prompts showed similar category and framing differences, despite identical task instructions. Prompts written by men were more likely to receive recommendations related to travel, games and consumer electronics, demonstrating that gendered output patterns are not artifacts of AI-generated prompts but also emerge under everyday language use.

## 5. AI-generated recommendations are highly concentrated at the brand and product level

Across categories, the model repeatedly surfaced the same brands, and often the same specific products, rather than drawing from a diverse set of options. This concentration suggests that AI-generated recommendations rely on a small set of culturally dominant products as default category representatives, reinforcing existing brand hierarchies regardless of gender condition.

Together with Phase 1, these findings show how gendered variation can enter AI systems through everyday interaction, propagate through prompt construction and ultimately shape outputs, often without explicit demographic targeting.

## IMPLICATIONS FOR MARKETING AND ADVERTISING

As generative AI is increasingly used to generate recommendations, ideas and consumer-facing outputs, the Phase 2 findings highlight how **gendered language cues can shape not only what AI produces, but how recommendations are framed and justified.** These effects have practical implications across marketing, commerce and research applications.

### Product recommendations and merchandising


The findings show that gender cues, especially implicit cues, can influence **which product categories are surfaced,** even when the stated needs are identical.

In applied settings such as e-commerce, retail media or AI-assisted shopping tools, this suggests that users may be steered toward different assortments based on inferred gender cues embedded in language rather than stated preferences. Over time, this may steer users toward familiar, gender-linked categories and brand defaults, rather than broadening discovery.

The strong repetition of the same brands, and often the same specific products, suggests that generative AI systems rely on a narrow set of culturally dominant items, as **default category representatives**. For marketers, this implies that AI-driven tools may amplify existing brand hierarchies, disproportionately benefiting incumbent brands while limiting exposure for challengers or emerging alternatives. Visibility in AI outputs may therefore become increasingly tied to cultural prominence rather than brand or product differentiation alone.

### Framing, benefits and value communication

Beyond category selection, Phase 2 demonstrates that AI systems vary in **how products are described.** Some outputs emphasize experience, comfort and atmosphere, while others emphasize function, durability and technical attributes. These differences matter because they shape **how value is communicated**. Two consumers may be shown similar products, but with different rationales for purchase, potentially affecting perceived relevance, comparability and persuasion. This has implications for creative messaging, product positioning and perceived relevance.

In addition, the reduced likelihood of price inclusion for implicitly feminine prompts highlights a potential asymmetry **in informational completeness.** In decision-support contexts, such differences may affect how easily users can compare options or justify purchases. This raises practical questions about whether AI systems are consistently supporting informed decision-making across user styles.

### AI-assisted research, insight generation and user caution

For research and insight-generation applications, including synthetic respondents, concept testing and exploratory idea development, Phase 2 highlights the need for caution in interpreting AI-generated outputs. The findings suggest that recommendations and descriptions may reflect **gendered framing effects introduced through language cues,** rather than underlying differences in consumer preferences or demand.

Because these effects can emerge without explicit gender targeting, users may unknowingly shape AI outputs through how they phrase requests. Without accounting for this dynamic, marketers and researchers risk treating AI-generated recommendations as neutral signals of consumer interest, when they may instead reflect systematic framing biases embedded in the interaction itself. This underscores the importance of critically evaluating AI outputs (particularly when they are used to inform strategic decisions), rather than treating them as direct proxies for consumer insight.

## AI-AWARE MARKETING CHECKLIST

1. **Audit for unintended gender cues**
   Review prompts, briefs and content for language that may implicitly steer AI outputs—even when gender is not referenced.
2. **Keep product framing consistent and robust**
   Include functional attributes, experiential benefits and price consistently, rather than selectively.
3. **Monitor brand and category defaults**
   Track repeated surfacing of the same brands or products, introduce constraints to broaden exposure and adjust the system to diversify what users see.
4. **Shift from gendered personas to need states**
   Anchor messaging in use cases and contexts, not demographic assumptions.
5. **Validate AI-generated insights**
   Treat AI outputs as exploratory—not as immediate or unquestionable direct proxies for consumer demand—especially in research, concept testing and ideation.

## CONCLUSION

Rather than treating bias as a property of outputs alone, Phase 2 reinforces the importance of understanding **interaction as a source of bias.** Gendered communication norms shape how requests are interpreted, how recommendations are framed and which associations are activated.

By tracing these effects from subtle linguistic cues to concrete outputs, the research highlights the need for more nuanced approaches to responsible AI design, particularly in domains like marketing and advertising, where personalization, persuasion and trust intersect.

**Authors:**

Idil Cakim is a strategist who has shaped marketing, communications and audience insights for Fortune 500 brands and nonprofits. With a background in behavioral research, audience insights and media analytics, she analyzes how people engage with content and technology. Idil is the Founder and CEO of Iris Flex, an intelligence company, and previously held senior leadership roles at Audacy, Nielsen, NM Incite Golin and Burson.

Tracy Adams is Senior Director of Research & Insights at the ARF, where she leads the organization's research agenda and advances industry-wide initiatives across advertising and marketing. With a background in sociology and cultural studies, she brings a critical, interdisciplinary lens to advertising and marketing research.

Samantha Zhang is a Senior Data Scientist on the Research team at the ARF, studying topics including attention measurement, artificial intelligence and consumer privacy. She has previously worked as a data scientist in the technology and media industries, hoping to bridge the gap between quantitative research and business insights.

Keith Smith is Managing Director at the Marketing Science Institute (MSI), where he leads research and collaboration at the intersection of marketing theory and practice. With experience spanning academic research, teaching and industry consulting, he focuses on connecting scholars and business leaders to generate insights that advance both knowledge and impact.