Novel Applications of Generative AI in Marketing

Arvind Rangasawamy
PennState Smeal College of Business
Bernd Skiera
Goethe-University Frankfurt, Germany
Michael Haenlein
ESCP Business School
Lukas Jürgensmeier
Goethe University Frankfurt
Sumon Chaudhuri
ESSEC Business School
Arnaud De Bruyn
ESSEC Business School
Wenlan Yu
The Smeal College of Business, Penn State University
Ning Zhong
The Smeal College of Business, Penn State University

Cite as:
Rangasawamy Arvind, Skiera Bernd, Haenlein Michael, Jürgensmeier Lukas, Chaudhuri Sumon, De Bruyn Arnaud, Yu Wenlan, Zhong Ning (2024), Novel Applications of Generative AI in Marketing. *Proceedings of the European Marketing Academy*, 53rd, (120021)

Paper from the 53rd Annual EMAC Conference, Bucharest, Romania, May 28-31, 2024

Novel Applications of Generative AI in Marketing

Session Chair(s):

Arvind Rangaswamy, The Smeal College of Business, Penn State University, University Park, PA 16802, USA

Discussant:

Michael Haenlein, ESCP Business School, Paris

Titles of each session paper, author(s), and affiliation(s), and the names of who will present the papers.

Paper 1:

Authors: Lukas Jürgensmeier, ***Presenter*** (Goethe University Frankfurt, Germany)
Bernd Skiera (Goethe University Frankfurt, Germany)

Title: Using Generative AI to Provide Scalable Feedback to Multimodal Exercises in Marketing Analytics

Paper 2:

Authors: Sumon Chaudhuri, **Presenter** (ESSEC Business School, Cergy, France)
Arnaud De Bruyn (ESSEC Business School, Cergy, France)

Title: Bots Bargaining with Humans: Building AI Super-Bargainers with Algorithmic Anthropomorphization

Paper 3

Authors: Wenlan Yu, ***Presenter*** (The Smeal College of Business, Penn State University, University Park, USA)
Ning Zhong (The Smeal College of Business, Penn State University, University Park, USA)
Arvind Rangaswamy (The Smeal College of Business, Penn State University, University Park, USA)

Title: Mining the "Mind of the Market" for New Product Ideas: A Prompted GenAI Model

Declaration:
Each presenter has agreed to register for the conference and to present the paper, if the proposal is accepted; and none of the papers has been submitted to other conference tracks, and none have previously been presented at EMAC.

Abstracts

Generative AI (GenAI) offers new opportunities for marketing academics to apply existing GenAI models, and/or adapt these foundational models for marketing applications and research. In this session, the presenters will discuss three applications of GenAI in marketing to highlight the opportunities and challenges associated with applying these technologies to address research issues in marketing. The discussant will summarize the common themes and insights across these applications.

Lukas Jürgensmeier and Bernd Skiera, "Using Generative AI to Provide Scalable Feedback to Multimodal Exercises in Marketing Analytics."

Detailed feedback on exercises helps learners become proficient in marketing analytics. However, such feedback is labor-intensive and expensive for educators. This study introduces a web application leveraging GenAI to autonomously provide feedback on multimodal exercises requiring coding, statistics, and economic reasoning. We compare the application's feedback with human expert feedback for 4,349 solutions. Using OpenAI's GPT-4, our application provides almost unbiased evaluations, correlates very highly with (r = 0.94), and deviates only 6 % from human evaluations. While GPT-4 outperforms GPT-3.5 by 31 – 39 %, it costs over ten times more. Compared to AI models, human evaluators are one to two orders of magnitude more expensive and require at least 20 times the evaluation duration.

Sumon Chaudhuri and Arnaud De Bruyn, "Bots Bargaining with Humans: Building AI Super-Bargainers with Algorithmic Anthropomorphization."

Bargaining is increasingly being automated by firms using AI. However, little is known about the psychological impact this would have on consumers. We train a bargaining AI within a Generative Adversarial Network (GAN) framework and task it with reaching superior economic outcomes while appearing "human" in doing so (a process we refer to as algorithmic anthropomorphization). We then experimentally compare it to two alternative bot specifications: a primitive bot that mimics human behavior and a purely economic-efficient bot. Our results suggest that (a) all bots perform poorly on subjective evaluations, (b) while superficial anthropomorphization helps portray a bot as a human, it does not improve subjective evaluations, and (c) algorithmic anthropomorphization offers the promise of a solution, albeit imperfect.

Wenlan Yu, Ning Zhong, and Arvind Rangaswamy, "Title: Mining the "Mind of the Market" for New Product Ideas: A Prompted GenAI Model."

We develop and test an automated model to generate novel and relevant new product ideas that helps overcome "fixation" in the creative process of idea generation. We leverage big data (user-generated and firm-generated), graph-theoretic modeling, and machine learning to simulate human analogical reasoning on a large scale. The model captures the links between products, product uses/benefits, and product features to create an aggregated "mind of the market." To generate new product ideas, we tap into this artificial mind to identify product-feature associations that have a high probability of occurrence but are not currently associated with a focal product. We test our model by generating new product ideas for four different mobile apps. We compare the ideas generated by our model with ideas generated directly by GPT-4. On average, the customers who participated in our study rank the ideas from our model as having much greater novelty than those generated by GPT-4.

In the short space available, we provide a few more details about the problem background, methods, and results from the three papers. We would be happy to provide more details, if needed.

Paper 1:

Title: Using Generative AI to Provide Scalable Feedback to Multimodal Exercises in Marketing Analytics

Authors: Lukas Jürgensmeier, ***Presenter*** (Goethe University Frankfurt, Germany)
Bernd Skiera (Goethe University Frankfurt, Germany)

Providing feedback is crucial to many aspects of our society. For example, learners receive *formative* and *summative feedback* in an educational setting (Dixson and Worrell 2016). Through formative feedback, learners solve exercises, receive feedback from educators, and improve. After such a learning process, educators commonly use *summative feedback* (or summative assessments) to evaluate a learner's progress through exams.

Providing learners with formative feedback during the learning process and summative feedback through exam grades are essential to higher education curricula. While essential, it is also time-consuming for educators. Nevertheless, automating feedback generation has seen little attention in the marketing context. One step in this direction is Czaplewski (2009), who discusses how grading rubrics can support educators in grading in a computer-assisted approach that promises efficiency gains for manual feedback generation by the educator.

Established automated grading systems work well with exercises with one correct answer or a specific answer pattern, as they can compare the learner's response with the correct answer or pattern. However, they may fall short with evaluation tasks that require higher-order cognitive skills, such as interpretation, reasoning, and decision-making. Indeed, we are not aware of automated grading systems in marketing that focus on similarly complex exercises as the ones we use in our empirical study.

Recent advances in Generative AI promise to solve these challenges because researchers feed massive data—representing a significant amount of all knowledge available—into training the models. Generative Pre-Trained Transformers (GPT) appear to have "general knowledge" consisting of significant subject knowledge across all domains. Hence, such models can rely on a foundational understanding of subject expertise without additional training. By providing additional context to the model through prompts, such models can perform specialized tasks across domains. Hence, GPT-based feedback systems promise to provide feedback on questions in all domains and could be useful without extensive training.

In this article, we explore the application of Generative AI in marketing analytics education. Overall, we aim to answer whether and how well Generative AI provides scalable feedback for marketing analytics exercises. We assess this question in different settings to evaluate under which conditions our application fulfills this aim to which degree—by testing two different Generative AI models and two degrees of supplied context. Specifically, we investigate how the application design choices impact the feedback's accuracy on an overall assignment level (including many different exercises) and by exercise type. This structure enables us to evaluate how accurately the application evaluates the learner's overall performance before drilling down on an exercise level to answer which exercise type the application can assess how well.

In our empirical study, we let the application generate feedback to 4,349 answers from 243 learners (here: undergraduate students) to 36 marketing analytics exam questions of varying difficulty and complexity and compare the Generative AI score (i.e., points achieved by exercise) to human scores. In the best-performing setting, the application's assessment of overall achieved points per learner correlates very highly with the human scores ($r = .94$), provides almost unbiased overall scores (mean error = - 2.4 points out of 90 achievable points), and produces a mean absolute error of 5.7 points (6 % of achievable points). Sixty-three percent, 2,766 of the 4,349 sub-exercises, received identical human and AI scores.

Evaluating one learner's exercise solutions to a 90-minute exam in our marketing analytics context costs 0.79 US Dollars for the best-performing setting. However, the relationship between cost and performance exhibits decreasing marginal returns with increased cost. While GPT-4 reduces the mean absolute error by 25 – 39 % compared to GPT-3.5 in our empirical study, the better-performing model costs more than ten times as much as the second-best. While this cost difference already encompasses one order of magnitude, human evaluators are one additional order of magnitude more expensive than GPT-4: the human evaluators cost approximately 15 times as much as

GPT-4, 170 times as much as GPT 3.5, and need more than 20 times as much time to provide way more limited feedback than our application.

Beyond evaluating the quantitative feedback, the presentation will describe the application in detail, feature additional expert assessments of the application's textual feedback, and present survey results on how useful learners find this application in class.

Full List of References

Barnett, G. Octo, Richard Winickoff, Joseph L. Dorsey, Mary M. Morgan, and Robert S. Lurie (1978), "Quality Assurance through Automated Monitoring and Concurrent Feedback Using a Computer-Based Medical Information System," *Medical Care*, 16 (11), 962.

Botelho, Anthony, Sami Baral, John A. Erickson, Priyanka Benachamardi, and Neil T. Heffernan (2023), "Leveraging Natural Language Processing to Support Automated Assessment and Feedback for Student Open Responses in Mathematics," *Journal of Computer Assisted Learning*, 39 (3), 823–40.

Czaplewski, Andrew J. (2009), "Computer-Assisted Grading Rubrics: Automating the Process of Providing Comments and Student Feedback," *Marketing Education Review*, 19 (1), 29–36.

Deeva, Galina, Daria Bogdanova, Estefanía Serral, Monique Snoeck, and Jochen De Weerdt (2021), "A Review of Automated Feedback Systems for Learners: Classification Framework, Challenges and Opportunities," *Computers & Education*, 162, 104094.

Dixson, Dante D. and Frank C. Worrell (2016), "Formative and Summative Assessment in the Classroom," *Theory Into Practice*, 55 (2), 153–59.

Koltovskaia, Svetlana (2020), "Student Engagement with Automated Written Corrective Feedback (AWCF) Provided by Grammarly: A Multiple Case Study," *Assessing Writing*, 44, 100450.

Zhang, Zhe (Victor) and Ken Hyland (2018), "Student Engagement with Teacher and Automated Feedback on L2 Writing," *Assessing Writing*, Special Issue: The Comparability of Paper-Based and Computer-Based Writing: Process and Performance, 36, 90–10.

Paper 2:

Title: Bots Bargaining with Humans: Building AI Super-Bargainers with Algorithmic Anthropomorphization

Authors: Sumon Chaudhuri, **Presenter** (ESSEC Business School, Cergy, France)
Arnaud De Bruyn (ESSEC Business School, Cergy, France)

The rise of artificial intelligence (AI) brings forth a disconcerting reality where human intelligence and skills may frequently be overshadowed by machines. Lee Sedol's retirement post-AlphaGo and the chess community's awe at AlphaZero illustrate the profound psychological impact of AI superiority. That said, AI is increasingly being used for games of strategic interaction like negotiation and bargaining. Companies like Pactum AI, Negobot and Aerchain have gained traction, yet research on the psychological aspects of negotiating with AI lags practical deployments. This is worrisome because a firm deploying AI negotiators could cause a negative psychological impact on the minds of their consumer, which in turn would lead to long term negative effects on the firm.

A documented psychological bias observed in various AI applications is algorithm aversion (Dietvorst et al., 2015). The mitigation of this bias prompted solutions like transparency of AI predictions, control over AI predictions, and anthropomorphization of AI (de Visser et al., 2016). However, these solutions are applicable to purely collaborative contexts. In contexts where AI can strategically interact with humans (like Go, chess, or bargaining games), these solutions may not apply. Moreover, even somewhat applicable solutions like anthropomorphization may not be a perfect solution as typically anthropomorphization is done through avatars, voice, or text (for example, Crolic et al., 2022), but in strategic interactions, the strategies used by the AI may divulge its identity.

In this paper, we develop the notion of algorithmic anthropomorphism, which we define as an agent that optimizes its objective outcomes in strategic interactions, given the constraint of appearing human-like in its behavior. We consider the context of bargaining games, and use Generative Adversarial Networks (GAN) to develop a bot that optimizes its economic outcomes in that game while retaining human traits in its strategies. We call it a SUPERHUMAN bot in the sense that it behaves like a human yet displays abilities (i.e., rationality, consistency) that exceed those usually found in humans. Prior research has either exclusively focused on optimization (Lewis et al., 2017) or incorporating human-like behaviour (Jacob et al., 2022; McIlroy-Young et al., 2022), but we make a methodological contribution by finding a balance between the two. We compare it to two other bot versions: a PRIMITIVE bot designed to mimic a sample of human negotiators (including their quirks and irrationalities) and an EFFICIENT bot focused on maximizing one's economic outcome exclusively.

We base our research on the game design proposed by Camerer et al. (2019). It is an unstructured bargaining game, played in continuous time, with a fixed deadline, and one-sided private information about the pie size. We asked 679 participants to play 13,580 unstructured, continuous-time bargaining games with asymmetric information, representative of many real-world bargaining scenarios. They were paired either with one of the bargaining bots or with another human being for the duration of the experiment. Following this, they responded to a questionnaire that contained the 16 item Subjective Value Inventory (Curhan et al., 2006), which have been shown to impact the long-term relationship between bargainers and the outcome of future negotiations (Curhan et al., 2009). Finally, participants responded to a Turing test question that is designed to measure to what extent the bots were identifiable as AI agents.

We find that the SUPERHUMAN and PRIMITIVE bots behave in ways that are indistinguishable from humans (i.e., they both pass the Turing test). In addition, while the SUPERHUMAN bot achieves objective outcomes in line with the EFFICIENT bot (i.e., number of agreements reached, monetary gains), it also receives superior subjective evaluations. Hence, our methodological approach appears to partly solve the conundrum faced by firms eager to deploy bargaining AI in their daily operations without having to choose between securing short-term gains and maintaining long-term relationships.

However, even though participants could not distinguish the SUPERHUMAN bot from actual human beings, human-vs-human negotiations still achieved the highest subjective evaluations, and no bot version compared favorably to humans in terms of perceived negotiation outcome, process, or relationship. Our research is a warning call that, even after reaching perfect anthropomorphism and excellent objective success metrics, asking an AI to negotiate on the firm's behalf may not be free from adverse long-term consequences on customer relationships.

Full List of References

Camerer, C. F., Nave, G., & Smith, A. (2019). Dynamic Unstructured Bargaining with Private Information: Theory, Experiment, and Outcome Prediction via Machine

Learning. *Management Science*, 65(4), 1867–1890. https://doi.org/10.1287/mnsc.2017.2965

Crolic, C., Thomaz, F., Hadi, R., & Stephen, A. T. (2022). Blame the Bot: Anthropomorphism and Anger in Customer–Chatbot Interactions. *Journal of Marketing*, 86(1), 132–148. https://doi.org/10.1177/00222429211045687/ASSET/IMAGES/LARGE/10.1177_00222429211045 687-FIG6.JPEG

Curhan, J. R., Elfenbein, H. A., & Kilduff, G. J. (2009). Getting Off on the Right Foot: Subjective Value Versus Economic Value in Predicting Longitudinal Job Outcomes From Job Offer Negotiations. *Journal of Applied Psychology*, 94(2), 524–534. https://doi.org/10.1037/a0013746

Curhan, J. R., Elfenbein, H. A., & Xu, H. (2006). What do people value when they negotiate? Mapping the domain of subjective value in negotiation. Journal of Personality and Social Psychology, 91(3), 493–512. https://doi.org/10.1037/0022-3514.91.3.493

de Visser, E. J., Monfort, S. S., McKendrick, R., Smith, M. A. B., McKnight, P. E., Krueger, F., & Parasuraman, R. (2016). Almost human: Anthropomorphism increases trust resilience in cognitive agents. *Journal of Experimental Psychology: Applied*, 22(3), 331–349. https://doi.org/10.1037/xap0000092

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, 144(1), 114–126. https://doi.org/10.1037/xge0000033

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative Adversarial Nets. *Advances in Neural Information Processing Systems*, 27. http://www.github.com/goodfeli/adversarial

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Jacob, A. P., Wu, D. J., Farina, G., Lerer, A., Hu, H., Bakhtin, A., Andreas, J., & Brown, N. (2022). Modeling Strong and Human-Like Gameplay with KL-Regularized Search. In *Proceedings of Machine Learning Research* (pp. 9695–9728). PMLR. https://proceedings.mlr.press/v162/jacob22a.html

Lewis, M., Yarats, D., Dauphin, Y. N., Parikh, D., & Batra, D. (2017). Deal or No Deal? End-to-End Learning for Negotiation Dialogues. EMNLP 2017 - *Conference on Empirical Methods in Natural Language Processing, Proceedings*, 2443–2453. https://doi.org/10.18653/v1/d17-1259

McIlroy-Young, R., Wang, R., Sen, S., Kleinberg, J., & Anderson, A. (2022). Learning Models of Individual Behavior in Chess. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1253–1263. https://doi.org/10.1145/3534678.3539367

Sutton, R., & Barto, A. (2018). Reinforcement Learning: An Introduction, Second Edition. The MIT Press. https://mitpress.mit.edu/9780262039246/reinforcement-learning/

Wang, W., Li, B., Luo, X., Wang, X., & Smith, R. H. (2022). Deep Reinforcement Learning for Sequential Targeting. Management Science, 69(9), 1–22. https://doi.org/10.1287/MNSC.2022.4621

Wierstra, D., Förster, A., Peters, J., & Schmidhuber, J. (2010). Recurrent policy gradients. *Logic Journal of the IGPL*, 18(5), 620–634. https://doi.org/10.1093/JIGPAL/JZP049

Paper 3

Title: Mining the "Mind of the Market" for New Product Ideas: A Prompted GenAI Model

Authors: Wenlan Yu, ***Presenter*** (The Smeal College of Business, Penn State University, University Park, USA)
Ning Zhong (The Smeal College of Business, Penn State University, University Park, USA)
Arvind Rangaswamy (The Smeal College of Business, Penn State University, University Park, USA)

It is often said that innovation is the lifeblood of business. And, innovation starts with ideas. In new product development, idea generation, often referred to as "the fuzzy front end," typically lacks well-defined processes, reliable information, and proven decision rules (Dahl and Moreau, 2002). One way to deal with this situation, as articulated by two-time Nobel Prize Winner Linus Pauling, is "to get lots of ideas and throw the bad ones away."

To generate lots of ideas, companies are expanding the range of individuals involved in the idea generation process. For example, "brainstorming" with a limited number of domain experts in the company is evolving to include broader audiences through approaches such as listening to "the voice of customers" (e.g., Lilien et al., 2002) and online crowdsourcing (e.g., Luo and Toubia, 2015). Concurrently, researchers are also developing methods to assist companies in efficiently filtering and selecting the most promising ideas (e.g., Toubia and Netzer, 2017). However, these approaches are subject to an inherent constraint, namely, the limitations of human cognitive effort. Consequently, idea generation by humans, regardless of the size of the team or crowd, is contingent upon their cognitive capacities, which may only capture a fraction of the myriad possibilities for novel and useful ideas.

Human creativity arises from the novel reassembling of elements from existing knowledge, but the brain is prone to the "fixation" problem that limits the range of possibilities considered. In comparison to human cognition, recent advances in

Generative AI highlight the potential for expanding the spectrum of creativity by leveraging the power of big data through computational methods. Employing GenAI has the potential to increase creative outputs across many domains, including new product idea generation.

Inspired by these developments, we develop a model-based automated approach for generating meaningful new product ideas in large numbers. Our model identifies non-obvious but relevant associations between an existing product and its potential new features by mining unstructured user-generated content (UGC) and firm-generated content (FGC). The essence of our model lies in its ability to identify the associations in the marketplace between products, product uses, and product features, based on UGC and FGC. We term this aggregated information set as the "mind of the market." Our model taps into this artificial mind to make predictions for potential new product features for any focal product within a data set.

Our modeling framework relies on natural language processing (NLP) methods to extract common contexts from user-generated product reviews, and product features from firm-generated product descriptions across different product categories. We then establish associations between usage contexts and products, as well as between product features and products. We model these associations using various network-theoretic constructs, including Graph Neural Networks (GNN). This modeling framework can be used to predict the probabilities of a focal product being associated with different product features currently available in the market. More importantly we can use the model to identify new product features based on product-feature associations that are predicted to have high probability of occurrence, but are not presently associated with the focal product.

To assess the value of the ideas generated by our modeling framework, we developed a prototype for generating new product ideas for mobile apps. Specifically, we scraped mobile app reviews from Google Play from the last week of 2022, as well as app descriptions generated by the app development teams. We also compiled app performance metrics, such as number of downloads, across the top 20 apps in 32 categories. Our model does a good job of predicting the associations between products and features, achieving an accuracy rate of 0.72 and a recall rate of 0.81. The hits@n, with n set at 10 and 20, are at 0.5 and 0.65, respectively. These performance metrics could potentially be improved by focusing on well-defined categories, but that may limit the novelty of the ideas generated.

As an initial test of the value of the model, we used the feature predictions from the model as prompts to GPT-4 to generate an ordered list of new product ideas. We surveyed 285 university students to assess their "intent to use" the new product features generated for four apps: Ticketmaster, Spotify, PayPal, and Amazon. On average, respondents express a greater likelihood of using the top-ranked features predicted by our model, as opposed to the bottom-ranked features. A more important test is to assess the model's value for new product managers. We are now in the process of working with industry executives to evaluate the perceived novelty (e.g., overcoming fixation) and utility (e.g., generating future revenue) of the ideas generated by the model.

Our research contributes to the field in three ways. First, we have developed a novel approach to identify non-obvious, but relevant, associations between products and features, which helps generate meaningful new product ideas in a cost-effective manner. The modeling framework is an instantiation of cognitive theories that have long hypothesized that customers organize information in an associative manner. However, the associations in our model are built at the market level, rather than at the individual level. Second, we empirically demonstrate the value of mining the "mind of the market" to generate new product ideas in the context of mobile apps. Lastly, our model offers a general framework for other creative tasks. By employing this framework to identify potential associations among entities and leveraging the capabilities of Generative AI, companies can deploy ideation processes across various domains such as new product development, advertising themes, and product recommendations.

References

Dahl, D. W., & Moreau, P. (2002). The influence and value of analogical thinking during new product ideation. *Journal of Marketing Research*, 39(1), 47-60.

Lilien, G. L., Morrison, P. D., Searls, K., Sonnack, M., & Hippel, E. V. (2002). Performance assessment of the lead user idea-generation process for new product development. *Management Science*, 48(8), 1042-1059.

Luo, L., & Toubia, O. (2015). Improving online idea generation platforms and customizing the task structure on the basis of consumers' domain-specific knowledge. *Journal of Marketing*, 79(5), 100-114.

Toubia, O., & Netzer, O. (2017). Idea generation, creativity, and prototypicality. *Marketing Science,* 36(1), 1-20