

# OpinionConv: Conversational Product Search with Grounded Opinions

**Vahid Sadiri Javadi**

Conversational AI and Social  
Analytics (CAISA) Lab  
University of Bonn

**Martin Potthast**

Text Mining and Retrieval  
(TEMIR) Group  
Leipzig University and ScaDS.AI

**Lucie Flek**

Conversational AI and Social  
Analytics (CAISA) Lab  
University of Bonn

## Abstract

When searching for products, the opinions of others play an important role in making informed decisions. Subjective experiences about a product can be a valuable source of information. This is also true in sales conversations, where a customer and a sales assistant exchange facts and opinions about products. However, training an AI for such conversations is complicated by the fact that language models do not possess authentic opinions for their lack of real-world experience. We address this problem by leveraging product reviews as a rich source of product opinions to ground conversational AI in true subjective narratives. With OpinionConv, we develop the first conversational AI for simulating sales conversations. To validate the generated conversations, we conduct several user studies showing that the generated opinions are perceived as realistic. Our assessors also confirm the importance of opinions as an informative basis for decision making.

## 1 Introduction

In order to elucidate the mechanics of conversational product search, [Kotler and Keller \(2015\)](#) delineated a five-stage process that encapsulates customer decision making (see [Figure 1](#), left). This process suggests that the customer: (1) recognizes a problem or need; (2) searches for information about potential products or services that could resolve the problem or fulfill the need, filtering them until a manageable set of alternatives remains; (3) evaluates and compares these alternatives against each other with regard to personal preferences and third party opinions to inform their decision making; (4) proceeds to make a purchase decision predicated upon this informed evaluation; and finally, (5) exhibits post-decision behaviors that reflect their satisfaction, which completes the process.

Typically, in-store shopping predominantly engages with the second and third stages of this customer decision process. Both the activities of reduc-

ing the number of alternatives and evaluating their merits and demerits are conducted in conversations between customers and sales assistants. The absence of such interactions in online environments is perceived as a deficiency in customer service especially with respect to the third stage ([Exalto et al., 2018](#)). Customers derive post-purchase satisfaction from personal exchanges, relating to others experience, and having the opportunity to ask questions ([Papenmeier et al., 2022](#)). The considerable number of online product reviews available are not a substitute for everyone, since many customers lack the patience to examine many of them, leading to post-purchase dissatisfaction and product returns. Conversational AI has been suggested as a solution ([Gnewuch et al., 2017](#)), with the goal of mimicking the conversational strategies of sales assistants ([Papenmeier et al., 2022](#)). But despite its importance, previous research on conversational product search almost entirely neglects the third stage, or rather its opinionated aspects ([Section 2](#)).

Recent advances in large-scale conversational language models, spearheaded by OpenAI’s ChatGPT, are driving a paradigm shift in the development of conversational technologies. Nonetheless, when it comes to expressing opinions pertaining to real-world events or entities, these language models lack the necessary grounding in tangible reality. For an individual to formulate an opinion on a particular subject matter, they require exposure to the subject to relate the new experience to past ones, and importantly, an emotional perception. A language model is only capable of generating what might be termed as a “statistical average” of third-party opinions, if they have been part of its training data. In the context of product search, such opinions would be deemed unauthentic as they are not based on real-world experiences or substantiated knowledge. This lack of authenticity poses challenges to the effective utilization of these models when (personal) opinions play an important role.

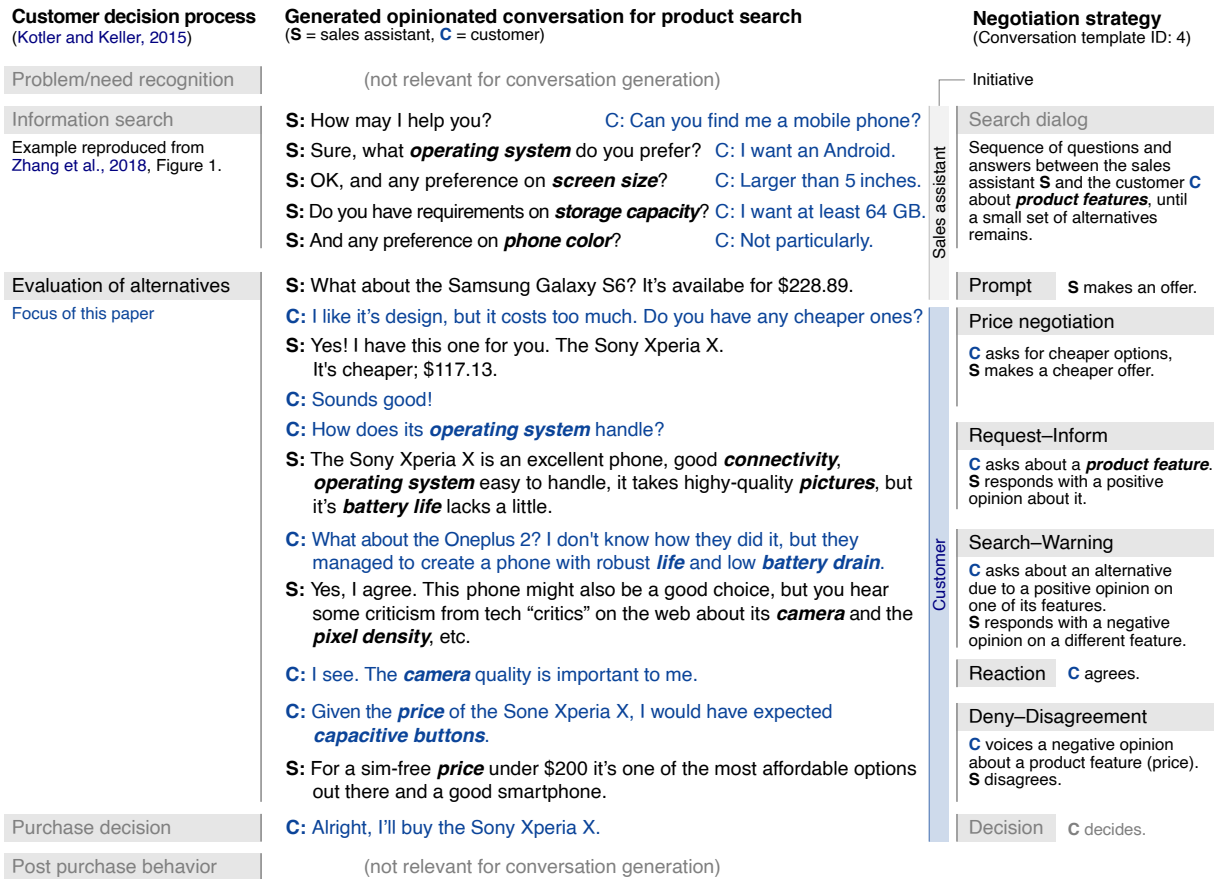


Figure 1: A grounded opinionated conversation generated by OpinionConv based on Conversation Template 4.

In this paper, we focus on the third stage of the customer decision process, for which we contribute the first approach to generate grounded opinionated statements (Section 3). We conceive and operationalize the generation of grounded opinions by positing that a grounded opinion about a product is an opinion which has been verifiably expressed by a minimum of one individual in a product review that specifically discusses the product under scrutiny. Our approach, OpinionConv, combines a product-specific index of reviews for a cohort of products of the same kind with a mechanism to generate realistic opinionated conversational exchanges. While carefully tuned, our approach must still be considered an early prototype. Consequently, before asking real customers to use it, its fundamental capabilities must first be established. We therefore simulate in-store dialogues between a customer and a sales assistant, where both parties incorporate grounded opinions. These conversations are then systematically evaluated in an experimental setup that ascertains the perception of human readers regarding the realism of these dialogs (Section 4).<sup>1</sup>

<sup>1</sup>Code and data: <https://github.com/caisa-lab/OpinionConv>

## 2 Related Work

Three lines of research are related to ours: opinionated question answering, conversational product search, and review-based conversation generation.

### 2.1 Opinionated Question Answering

While factoid Question Answering (QA) systems have a long tradition and some even outperform humans, non-factoid questions, such as opinions, explanations, or descriptions, are still an open problem (Cortes et al., 2021). Cardie et al. (2003) employed opinion summarization to help multi-perspective QA systems identify the opinionated answer to a given question. Yu and Hatzivassiloglou (2003) separated opinions from facts and summarized them as answers. The linguistic features of opinion questions have also been studied (Pustejovsky and Wiebe, 2005; Stoyanov et al., 2005). Kim and Hovy (2005) identified opinion leaders, which are a key component in retrieving the correct answers to opinion questions. Ashok et al. (2020) introduced a clustering approach to answer questions about products by accessing product reviews. Rozen et al. (2021) examined the task of

answering subjective and opinion questions when no (or few) reviews exist. Jiang et al. (2010) proposed an opinion-based QA framework that uses manual question–answer opinion patterns.

Closer to our work, Moghaddam and Ester (2011) address the task of answering opinion questions about products by retrieving authors’ sentiment-based opinions about a given target from online reviews. McAuley and Yang (2016) address subjective queries using relevance ranking, and Wan and McAuley (2016) extends this work by considering questions that have multiple divergent answers, incorporating aspects of personalization and ambiguity. AmazonQA (Gupta et al., 2019) is one of the largest review-based QA datasets. Its authors show that it can be used to learn relevance in the sense that relevant opinions are those for which an accurate predictor can be trained to select the correct answer to a question as a function of opinion. SubjQA (Bjerva et al., 2020) includes subjective comments on product reviews.

## 2.2 Conversational Product Search

Information is often gathered through conversations with a series of questions and answers. Conversational Question Answering (CQA) systems engage in such multi-turn conversations to satisfy a user’s information need (Zaib et al., 2021). Despite the attention this task has received in e-commerce (Ricci et al., 2011; Bi et al., 2019; Zhang et al., 2018), building a successful conversational product search system for online shopping still suffers from the lack of realistic dialog datasets for model training (Xiao et al., 2021).

## 2.3 Review-based conversation generation

Recently, multi-turn QA has grown more prominent (Cambazoglu et al., 2020). Product reviews are one of the sources of information that are being used for conversational product search. Penha et al. (2022) generate review-based explanations for voice-driven product search. Zhang et al. (2018) builds a dataset to answer conversational questions, as illustrated in Figure 1 (Stage 2 “Information Search”). They extract feature–value pairs from reviews and convert each review into a conversation based on the mentioned pairs, but omit opinionated statements. Xu et al. (2019) explores the possibility of turning reviews into a knowledge source to answer questions. Feature-related, non-opinionated statements in reviews are flagged and appropriate questions are formulated.

# 3 Grounded Product Opinion Generation

This section introduces the OpinionConv construction pipeline to generate grounded opinionated conversations for product search based on product reviews. Figure 2 gives an overview of the pipeline’s individual steps, grouped into preprocessing, information search dialog generation (Stage 2 of the customer decision process, which we reproduce from Zhang et al. (2018)), and evaluation dialog generation (Stage 3, our focus).

## 3.1 Data Source and Preprocessing

As a basis for grounded opinions, we utilize a crawl of Amazon product data including their reviews created by Ni et al. (2019).<sup>2</sup> The metadata enclosed includes product descriptions, multi-level product categories, and product information. For our proof-of-concept, we focus on one of its 24 product categories, *Cell Phones and Accessories*. As a first cleansing step, we reviewed the product data and added missing product details. We found the reviews to be of varying writing quality, especially with respect to basic syntax conventions, like the use of punctuation. We employed the model of Alam et al. (2020) to restore the punctuation, which enabled a more reliable sentence extraction and thus benefited the subsequently applied models, which were largely trained on “cleaner” data.

To extract the product features discussed in the reviews, we use the extraction model of Karimi et al. (2021). It is based on a hierarchical aggregation approach and was trained on the laptop review dataset of SemEval 2014 (Pontiki et al., 2014), performing best at that time. Given a review sentence, the model extracts feature terms on which an opinion has been expressed. On sentences containing such opinion statements, we then applied the sentiment analysis model of Zeng et al. (2019), which is based on self-attention to capture local context and global context features to determine the polarity score of the opinion.

## 3.2 Information Search Dialog Generation

To generate the information search dialog of the customer decision process (see Figure 1), we reproduce the approach of Zhang et al. (2018). Reproducing the original dialogs turned out to be straightforward, and we verified our success by direct comparison to the data supplied with the original paper. The dialogs are structured as follows: The sales

<sup>2</sup><https://jmcauley.ucsd.edu/data/amazon/>

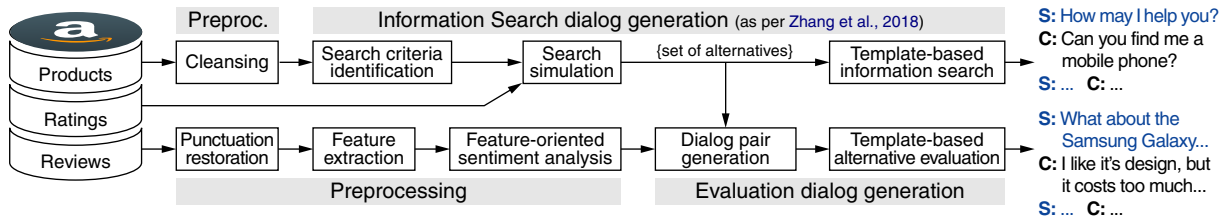


Figure 2: High-level overview of our approach in OpinionConv for generating opinionated multi-turn conversations.

assistant asks for preferences on product features, and the customer answers, narrowing down the set of alternatives. The resulting set of alternatives is fed to the next stage of the customer decision process, the evaluation of alternatives.

### 3.3 Evaluation Dialog Generation

The generation of an opinion-based evaluation of alternatives dialog is divided into two steps, the generation of pairs of talking points based on reviews of the alternative products, and their combination into a multi-turn conversation as exemplified in Figure 1. For lack of public corpora of in-store conversations, we resort to a template-based approach. The templates are derived from common conversational negotiation strategies from the literature.

**Dialog Turn Pair Generation** In each turn of an evaluation dialog the customer and the sales assistant discuss the relative value of product features, as well as their benefits and shortcomings compared to alternative products. Sales assistants, by training and/or experience, are usually well-equipped to provide customers with satisfying answers to their questions as well as to respond to opinions that customers express throughout the conversation. The most salient open question in this respect is: How should a sales assistant react to a customer’s opinion in the context of a negotiation?

Negotiations combine features of claiming and creating value. Each requires unique strategies and tactics for a negotiator to effectively achieve their objectives while creating the greatest value possible for all parties (Thompson and Hastie, 1990). We take inspiration from three negotiation tactics (Dwyer et al., 1987; Săvescu, 2019): (1) *Distributive negotiation*: This is a competitive win–lose situation. Any value claimed by one party is at the expense of the other. (2) *Integrative negotiation*: The parties create or generate value during the negotiation, and both parties may achieve mutual gains beyond what they would achieve independently, a win–win scenario. (3) *Compatible negotiation*:

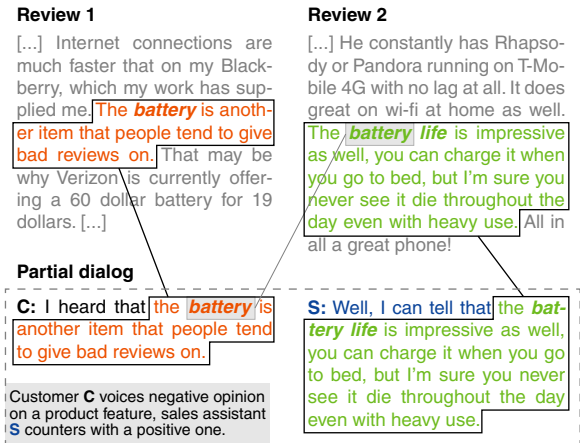


Figure 3: Example of a basic opinionated dialog pair generation step: Given a product feature like “battery”, opinionated statements are extracted from reviews of a given product to form part of a dialog between Customer C and Sales Assistant S.

The parties desire the exact same outcome, so that there is no need for any trade-off.

For instance, as illustrated in Figure 3, where Customer C’s remark about a product feature (left) is countered by an opinionated counterargument from Sales assistant S (right), we use the Deny–Disagreement tactic, where the customer expresses a negative opinion on a feature of a product in question, whereas the sales assistant disagrees and counters with a positive opinion on the same feature. This tactic may either correspond to a win–lose or a win–win situation, dependent on whose opinion applies more to the customer: If the customer is correct, they lose against the sales assistant, since the product is not switched. If the sales assistant is correct, they both win, since the customer still gets what they wanted, and the sales assistant may still get to sell the product in question.

A key constraint that we enforce by generating grounded opinions (i.e., opinions rooted in a real product reviews as exemplified in Figure 3) is that neither the customer nor the sales assistant can “lie” to each other, as their opinions are backed by a real person’s opinion about the product and its

Table 1: Negotiation tactics used in dialog pair templates (P=product, F=feature).

Dialog pair template	Description of negotiation tactic
<b>Request–Inform</b> Question: P-1, F-A, neutral Answer: P-1, F-A, positive	Customer asks about the sales assistant’s view on a feature of a product. Sales assistant expresses positive view on it.
<b>Deny–Disagreement</b> Opinion: P-1, F-A, negative Opinion: P-1, F-A, positive	Customer expresses negative opinion on a feature of a product. Sales assistant disagrees and expresses positive opinion on it.
<b>Deny–Switch Product</b> Opinion: P-1, F-A, negative Opinion: P-2, F-A, positive	Customer expresses negative opinion on a feature of a product. Sales assistant switches the product and expresses positive opinion on the same feature wrt. new product.
<b>Deny–Switch Feature</b> Opinion: P-1, F-A, negative Opinion: P-1, F-B, positive	Customer expresses negative opinion on a feature of a product. Sales assistant disagrees and expresses positive opinion on a different feature of the same product.
<b>Search–Agreement</b> Opinion: P-1, F-A, positive Opinion: P-1, F-A, positive	Customer expresses positive opinion on a feature of a product. Sales assistant agrees and expresses another positive opinion it.
<b>Search–Switch Feature</b> Opinion: P-1, F-A, positive Opinion: P-1, F-B, positive	Customer expresses positive opinion on a feature of a product. Sales assistant agrees and expresses positive opinion on different features of the same product.
<b>Search–Warning</b> Opinion: P-1, F-A, positive Opinion: P-1, F-B, negative	Customer expresses positive opinion on a feature of a product. Sales assistant warns the user and expresses negative opinion on different features of the same product.

feature. Thereby the dialog turns are more realistic, despite both parties being simulated. Moreover, our dialog turns enforce a conversational concept flow between (Li et al., 2023), as the product features and their attributes as key concepts are connected.

In Table 1, we list the templates for dialog pairs according to different negotiation tactics derived from the literature; seven patterns are devised, one question–answer pair, and six opinion–opinion pairs. The customer’s utterance consists of a feature-specific opinion with either positive or negative opinion for a certain product and one of its features, extracted from one of its reviews. The sales assistant’s utterance is a response that expresses either a positive or a negative opinion, not necessarily to the same product or feature.

As can be seen, depending on the type of dialog pair, different negotiation tactics may apply. The sales assistant is allowed to switch the polarity, feature under discussion, and product under negotiation. The mapping of a dialog pair to the negotiation tactics thus depends on the factuality of either opinion expressed by the customer or sales assistant, in case a product switch on the price changes (the price of the new product may be lower, similar, or higher), but also on whether the customer gets what they want. For instance, a switch to a pricier product is certainly worthwhile for the sales assistant, as long as the customer ends up with a

desired feature. However, given the necessity of interpreting each generated dialog with respect to its factuality, a mapping between dialog pair and negotiation tactic must be decided on a case-by-case basis, which is beyond the scope of this paper.

A constraint in cases where the product is being changed includes that the sales assistant is allowed to use only two types of products: (1) Retrieved products: The products from the set of alternatives retrieved in the information search dialog at the outset of a conversation based on customer preferences; (2) also viewed products: The products listed in the metadata that have been viewed by other customers. Customers thus can go beyond their original preferences and the set of alternatives.

**Template-based Alternative Evaluation** The last step of our pipeline generates conversations composed of multiple dialog pairs, based on a negotiation strategy. Considering the diversity of real dialogs and the fact that a coherent conversation should have a smooth transition between turns (Li et al., 2023), we define a diverse set of conversation templates inspired by past studies on negotiation in behavioral economics (Pruitt, 1981; Fisher and Ury, 1981; Thompson et al., 2010), including both high-level (e.g., insisting on your position: Disagreement) and low-level (e.g., focus on interests: Reaction) dialog acts.

Table 2: Example of the combination of dialog pairs in a conversation template.

Pair	Principle	Action	Example
Deny–Switch Product	Insist on position	Express negative sentiment	<b>B:</b> What I know about its battery is that the battery keeps draining because the phone is constantly looking for network signal.
	Invent options for mutual gain	Recommend a new product	<b>S:</b> If the battery is important for you, we can offer this product: Axon 7 is the same price as OnePlus 3, but it has slightly bigger battery.
Request–Inform	Focus on interests	Look for more information	<b>B:</b> What do you think about its speakers?
	Build trust	Express positive sentiment	<b>S:</b> It has dual front-facing speakers with good quality.
Search–Agreement	Focus on interests	Search for alternatives	<b>B:</b> I heard about this phone: Galaxy S4 that has a super-fast processor and a good battery life.
	Build trust	Confirm consumer’s preference	<b>S:</b> Yes, that’s true. This phone is also a good choice with the one premium hardware, great software and a reasonable price.

Table 3: Demographics of study participants.

Measure	Characteristics	Study 1 (N=100)	Study 2 (N=420)
Gender	Males	41.0%	31.0%
	Females	58.0%	69.0%
	Non-binary	1.0%	0.0%
Age	25 to 34 years	35.0%	38.0%
	35 to 44 years	28.0%	30.1%
	18 to 24 years	21.0%	15.7%
	55 to 64 years	6.0%	13.3%
	45 to 54 years	5.0%	1.8%
	65 years or older	5.0%	1.2%

We follow Zhou et al. (2019) and devise 14 conversation templates with different combinations of the generated question–answer and opinion–opinion pairs. Table 2 exemplifies one of them. We adapt the “CraigslistBargain” setting of He et al. (2018), where a buyer and a seller negotiate the price of a product. But unlike in their work, the sales assistant and the customer negotiate not only the price but primarily the relative merits of product features, whereas price may only be one of them.

## 4 Evaluation

A volunteer who is asked to pose as a customer in a laboratory user study, and who has no real intention of investing a fairly large amount of his or her own money in the purchase of a product, does not usually have the same information needs as a real customer. At the same time, we consider it unethical to confront real customers with an early prototype of a conversational sales assistant. Before a practical assistant can be developed, the basic means of generating informed opinions must first be established.

In our evaluation, we therefore decided to simulate full conversations between a hypothetical customer and a hypothetical sales assistant as described in the previous section. We then designed two user studies in which we specifically investigated whether human subjects consider these conversations realistic.

Study 1 investigates whether subjective narratives in conversational product search are considered important compared to purely factual exchanges. Study 2 investigates individuals’ perceptions of the quality and realism of the conversations generated by OpinionConv. We conducted the two studies by recruiting volunteers living in the US or UK on Prolific.<sup>3</sup> Table 3 shows the total sample size and key demographic information for both studies. As can be seen, we had fewer male participants than females and more than 60% of the participants are between 25 and 45 years old. Key to both our study design is that participants initially believed that the conversations are genuine transcripts of real sales negotiations recorded in a store, instead of generated ones. At the end of the questionnaire, it was revealed that they are not.

### 4.1 Study 1: Importance of Product Opinions

The first study started with the following instruction: “Below is an automatically generated transcript of a sales conversation. We show two variants: Variant 1 is focused on the customer’s preferences and requirements. Variant 2 starts similarly, but then continues with an opinionated discussion.” After reading both variants of the same dialog participants were asked “Which of the two variants would you as a customer hold with the sales as-

<sup>3</sup><https://www.prolific.co>

sistant while searching for a smartphone?” The survey concluded by asking participants an open-ended question to explain their judgment for the previous question, which also allowed to ascertain that they had actually read the conversations.

As a result, we find that 83% of the 100 participants of Study 1 prefer Variant 2 over the Variant 1, which confirms that they tend to prefer opinionated conversations when searching and evaluating a product rather than exclusively factual ones.

## 4.2 Study 2: Perceptions of Dialog Realism

For this study, the questionnaire consisted of two separate parts. In the first part, we again let participants believe they are reading a transcript of a real conversation by instructing them as follows: “Suppose you are in an electronics store. While browsing, you happen to overhear part of a conversation between a customer and a sales assistant. Both exchange opinions about the features of one or more products.” They are then asked to answer three questions using the 4-point Likert scale: (1) *Definitely yes*, (2) *Rather yes*, (3) *Rather not* and (4) *Definitely not*. As depicted in Figure 4, bottom, we ask questions for the following goals: *Customer understanding*, *Sales assistant answer sufficiency* and *Reasonableness of exchange*. To investigate whether participants’ perceptions change significantly after they learned that the conversation was generated, at the beginning of the second part, we reveal the truth and declare that the conversation they just read, was not a real but an automatically generated one. After the disclosure, they were asked answer Questions 4 to 6 using the same Likert-scaled responses as shown in Figure 4 in order to observe any changes of opinion. For each of our 14 conversation templates, we generated ten examples, and for each example, three participants were asked to answer the questions, a total of 140 questionnaires answered by 420 participants.

Figure 4 shows the distribution of participants’ responses to each question, outlining the alterations in perception after revealing the automated generation of conversations. Both user (Q1 & Q4) and agent (Q2 & Q5) utterances, as well as the overarching dialog (Q3 & Q6), were subjected to quality evaluation. The data reveals that prior to revealing the truth, over 66% of evaluators deemed the conversation reasonable (both “yes” answers combined), marginally reducing to 64% post-revelation (Q3 & Q6). Regarding participants’ assessment of

the customer’s understanding of the sales assistant (Q1 & Q4), over 78% affirmed it, which reduces to 77% after the disclosure, albeit almost half of participants switched from *definitely yes* to *rather yes*. Regarding the evaluation of the sales assistant’s response quality to customer inquiries (Q2 & Q5), over 60% of participants agreed that the responses were sufficient, while the disclosure incited a reduction in both *definitely yes* and *rather yes* responses to over 54%. Altogether, the responses indicate a generally positive reception of conversations generated by OpinionConv, with variation in assessment among different conversation templates. While a rough two thirds of participants agree with this outcome, more than half of participants improve their rating from *definitely not* to *rather not* considering overall reasonableness.

## 4.3 Participants’ Comments

In both studies, participants were asked to explain their judgment in about two sentences, in case of Study 2 once in each part of the questionnaire. With respect to assessing the conversation realism, we mostly observe positive comments. For instance, before disclosing the nature of the conversation one participants commented “The customer was recommended the phone by a friend, and the sales assistant was able to give further information on the phone. Likewise the sales assistant was able to inform the customer on a drawback associated with the phone.”, and after “The conversation appeared real as there was flow - i.e. the sales assistant was able to connect with what the customer said and elaborate upon it. Likewise the sales assistant was able to pick up on key details associated with the phone like the camera and OS.”

However, we also observe three key concerns raised: (1) Some features are of no interest to be discussed, e.g., “Why would the person asks the sales assistant about colors? That seems out of the ordinary.” (2) Some participants judge the conversations based on their personal experience with real sales assistants, e.g., “As always in marketing strategies, he [the sales assistant] was just trying to sell a phone not what he [the customer] wanted.” (3) A stronger argumentation is expected by some, e.g., “While the sales assistant did respond in a way that does answer the customer’s questions, their responses are not so direct and detailed as to be helpful towards the customer. For example, for the question about the screen, stating that it’s ‘bright

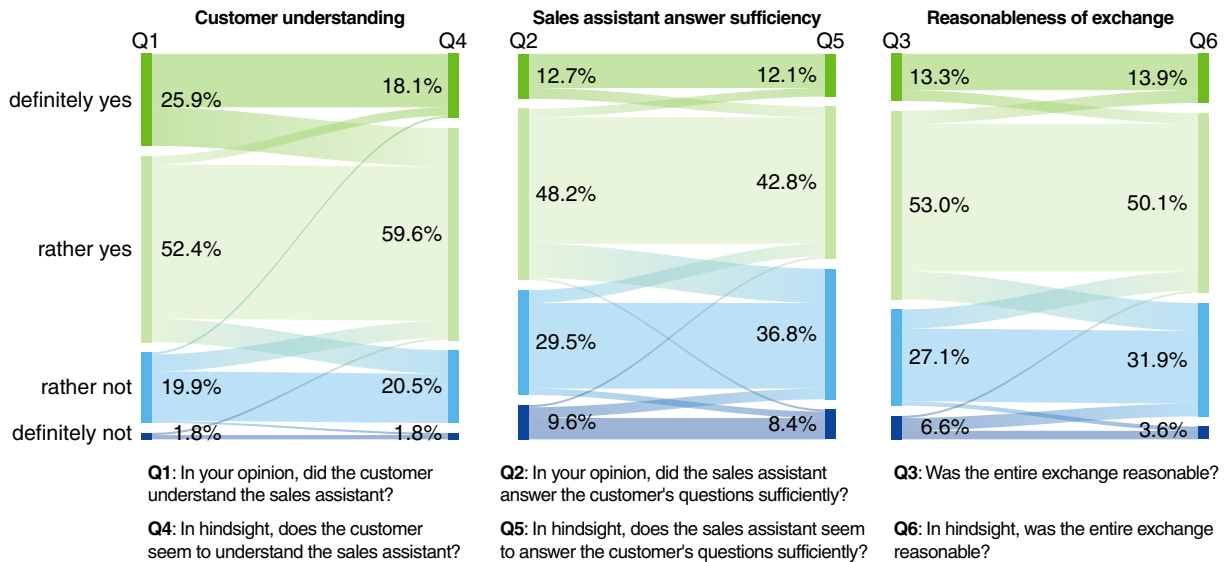


Figure 4: Evaluation results for Study 2; Questions Q1, Q2, and Q3 are asked in the first part of the questionnaire (before disclosing the conversations are generated), Q4, Q5, and Q6 are asked in the second part (after disclosure).

and good quality’ would not be convincing enough for me to want to buy the product.”

Reading the participants’ comments and observing the results of crowd-sourced qualitative evaluations have suggested several new research directions for future work relating to common sense product knowledge and argument generation.

## 5 Conclusion

We introduce a OpinionConv, a new conversation generation pipeline that generates opinionated multi-turn conversations for product search. OpinionConv was mainly designed to incorporate subjective narratives into conversational product search. The pipeline presented in this work can be easily extended to different domains. Recent progress in conversational systems, such as ChatGPT and YouChat, have shown tremendous improvements in natural language dialog between humans and conversational agents. However, when it comes to holding an opinionated conversation, specifically in product search, they are still limited for lack of grounding in real-world experience about products. This motivated the design of a pipeline to control both the dialog coherence and the information to be mentioned in the utterances. However, it should be mentioned that the trade-off between a coherent conversation and a more diverse conversation needs to be further studied. In order to validate the quality of the conversations generated by OpinionConv, we conduct two extensive human evaluations. Our results confirm the

conversational plausibility of the generated dialogs and reveal that people tend to exchange their personal opinions while searching for a product.

In future work, we envision customer-oriented assistant for buying products that assist customers in discussing the merits of products with a sales assistant, grounded in real-world reviews.

## 6 Limitations

As mentioned in Section 3.1, we focused on the *Cell Phones and Accessories* category of products. However, there is no inherent limitation of our design that prevents future work from including conversations related to other product categories.

Furthermore, an opinion is an observation or a belief that does not need to have evidence to support itself, whereas an argument requires premises. As we discussed in the Section 4.3, study participants expected to have stronger arguments in the generated conversation, rather than only expressing opinions. Therefore, future work should address this aspect utilizing argument mining techniques for generating argumentative dialogues.

## Acknowledgements

This work has been partially supported by the German Federal Ministry of Education and Research (BMBF) within the Junior AI Scientists program under the reference 01-S20060. We would like to thank the anonymous reviewers for their valuable feedback.



## Ethics Statement

Systems designed to influence humans via communication constitute a highly sensitive topic due to their intrinsically social nature (Stock et al., 2016). Any automated sales assistant comes along with the ethical risk of not only influencing customer opinion but doing so in ways undesired by customers, e.g., to their financial or otherwise personal disadvantage. Naturally, it is the company that deploys a manipulative sales assistance technology that is at fault, but the question of why research that may be misused in this direction has been undertaken in the first place is still pertinent.

Negotiation differs from persuasion in its goal. Negotiation strives to reach an agreement from both sides, while persuasion merely aims to change one specific person’s attitude and decision (Wang et al., 2019). Most human sales assistants have no interest in deceiving customers, since that very customer may come back to complain, or not come back to buy further products. Modern marketing strategies typically involve building a trustworthy customer relationship which includes the post-purchase stage of the aforementioned customer decision process, where customer satisfaction is to be maximized. We intend our research to serve as a step towards studying the capabilities of automated sales assistance with the goal of mutually beneficial negotiation. Nevertheless, if it turns out that it is easier for technology to manipulate its users with respect to a purchase decision than to consult them for mutual benefit, this must be found out, and publicly, or else no policies against such exploits can be enforced.

Moreover, an automatic sales assistant deployed by a marketplace must be considered separately from, e.g., an automatic sales assistant deployed by an independent third party (including open source variants). We imagine that not only the former will become available in the future, but also the latter, which will be more trustworthy overall.

## References

- Tanvirul Alam, Akib Khan, and Firoj Alam. 2020. Punctuation restoration using transformer models for high- and low-resource languages. In *Proceedings of the Sixth Workshop on Noisy User-generated Text (W-NUT 2020)*, pages 132–142.
- Aishwarya Ashok, Ganapathy Natarajan, Ramez Elmasri, and Laurel Smith-Stvan. 2020. [SimsterQ: A similarity based clustering approach to opinion question answering](#). In *Proceedings of the 3rd Workshop*
- on *e-Commerce and NLP*, pages 69–76, Seattle, WA, USA. Association for Computational Linguistics.
- Kejing Bi, Qingyao Ai, Yongfeng Zhang, and W Bruce Croft. 2019. Conversational product search based on negative feedback. In *Proceedings of the 28th acm international conference on information and knowledge management*, pages 359–368.
- Johannes Bjerva, Nikita Bhutani, Behzad Golshan, Wang-Chiew Tan, and Isabelle Augenstein. 2020. Subjqa: a dataset for subjectivity and review comprehension. *arXiv preprint arXiv:2004.14283*.
- B. Barla Cambazoglu, Mark Sanderson, Falk Scholer, and W. Bruce Croft. 2020. [A review of public datasets in question answering research](#). *SIGIR Forum*, 54(2):5:1–5:23.
- Claire Cardie, Janyce Wiebe, Theresa Wilson, and Diane J Litman. 2003. Combining low-level and summary representations of opinions for multi-perspective question answering. In *New directions in question answering*, pages 20–27.
- Eduardo Gabriel Cortes, Vinicius Woloszyn, Dante Barone, Sebastian Möller, and Renata Vieira. 2021. A systematic review of question answering systems for non-factoid questions. *Journal of Intelligent Information Systems*, pages 1–28.
- F Robert Dwyer, Paul H Schurr, and Sejo Oh. 1987. Developing buyer-seller relationships. *Journal of marketing*, 51(2):11–27.
- Marja Exalto, Maarten De Jong, Tim De Koning, Axel Groothuis, and Pascal Ravesteijn. 2018. Conversational commerce, the conversation of tomorrow. In *Proceedings of the 14th European Conference on Management, Leadership and Governance, ECMLG*, pages 76–83.
- R. Fisher and W. Ury. 1981. *Getting to Yes: Negotiating Agreement Without Giving in*. Houghton Mifflin.
- Ulrich Gnewuch, Stefan Morana, and Alexander Mädche. 2017. [Towards designing cooperative and social conversational agents for customer service](#). In *Proceedings of the International Conference on Information Systems - Transforming Society with Digital Innovation, ICIS 2017, Seoul, South Korea, December 10-13, 2017*. Association for Information Systems.
- Mansi Gupta, Nitish Kulkarni, Raghuvveer Chanda, Anirudha Rayasam, and Zachary C Lipton. 2019. Amazonqa: A review-based question answering task. *arXiv preprint arXiv:1908.04364*.
- He He, Derek Chen, Anusha Balakrishnan, and Percy Liang. 2018. [Decoupling strategy and generation in negotiation dialogues](#). In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, October 31 - November 4, 2018*, pages 2333–2343. Association for Computational Linguistics.

- Peng Jiang, Hongping Fu, Chunxia Zhang, and Zhen-dong Niu. 2010. A framework for opinion question answering. In *2010 6th International Conference on Advanced Information Management and Service (IMS)*, pages 424–427. IEEE.
- Akbar Karimi, Leonardo Rossi, and Andrea Prati. 2021. Improving BERT performance for aspect-based sentiment analysis. In *4th International Conference on Natural Language and Speech Processing, Trento, Italy, November 12-13, 2021*, pages 196–203. Association for Computational Linguistics.
- Soo-Min Kim and Eduard Hovy. 2005. Identifying opinion holders for question answering in opinion texts. In *Proceedings of AAAI-05 Workshop on Question Answering in Restricted Domains*, pages 1367–1373.
- Philip Kotler and Kevin Lane Keller. 2015. Marketing management. *New Jersey*, 143.
- Siheng Li, Wangjie Jiang, Pengda Si, Cheng Yang, Yao Qiu, Jinchao Zhang, Jie Zhou, and Yujiu Yang. 2023. Enhancing dialogue generation with conversational concept flows. In *Findings of the Association for Computational Linguistics: EAACL 2023, Dubrovnik, Croatia, May 2-6, 2023*, pages 1484–1495. Association for Computational Linguistics.
- Julian McAuley and Alex Yang. 2016. Addressing complex and subjective product-related queries with customer reviews. In *Proceedings of the 25th International Conference on World Wide Web*, pages 625–635.
- Samaneh Moghaddam and Martin Ester. 2011. Aqa: aspect-based opinion question answering. In *2011 IEEE 11th International Conference on Data Mining Workshops*, pages 89–96. IEEE.
- Jianmo Ni, Jiacheng Li, and Julian McAuley. 2019. Justifying recommendations using distantly-labeled reviews and fine-grained aspects. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 188–197.
- Andrea Papenmeier, Alexander Frummet, and Dagmar Kern. 2022. "mhm..." - conversational strategies for product search assistants. In *CHIIR '22: ACM SIGIR Conference on Human Information Interaction and Retrieval, Regensburg, Germany, March 14 - 18, 2022*, pages 36–46. ACM.
- Gustavo Penha, Eyal Krikon, and Vanessa Murdock. 2022. Pairwise review-based explanations for voice product search. In *CHIIR '22: ACM SIGIR Conference on Human Information Interaction and Retrieval, Regensburg, Germany, March 14 - 18, 2022*, pages 300–304. ACM.
- Maria Pontiki, Dimitris Galanis, John Pavlopoulos, Harris Papageorgiou, Ion Androutsopoulos, and Suresh Manandhar. 2014. Semeval-2014 task 4: Aspect based sentiment analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation, SemEval@COLING 2014, Dublin, Ireland, August 23-24, 2014*, pages 27–35. The Association for Computer Linguistics.
- D.G. Pruitt. 1981. *Negotiation Behavior*. Library and Information Science. Academic Press.
- James Pustejovsky and Janyce Wiebe. 2005. Introduction to special issue on advances in question answering. *Language Resources and Evaluation*, 39(2/3):119–122.
- Francesco Ricci, Lior Rokach, and Bracha Shapira. 2011. Introduction to recommender systems handbook. In *Recommender systems handbook*, pages 1–35. Springer.
- Ohad Rozen, David Carmel, Avihai Mejer, Vitaly Mirkis, and Yftah Ziser. 2021. Answering product-questions by utilizing questions from other contextually similar products. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2021, Online, June 6-11, 2021*, pages 242–253. Association for Computational Linguistics.
- D Săvescu. 2019. Some aspects regarding negotiation in business. In *IOP Conference Series: Materials Science and Engineering*, volume 514, page 012042. IOP Publishing.
- Oliviero Stock, Marco Guerini, and Fabio Pianesi. 2016. Ethical dilemmas for adaptive persuasion systems. In *Thirtieth AAAI Conference on Artificial Intelligence*.
- Veselin Stoyanov, Claire Cardie, and Janyce Wiebe. 2005. Multi-perspective question answering using the opqa corpus. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pages 923–930.
- Leigh Thompson and Reid Hastie. 1990. Social perception in negotiation. *Organizational Behavior and Human Decision Processes*, 47(1):98–123.
- Leigh L Thompson, Junwen Wang, and Brian C Gunia. 2010. Negotiation. *Annual review of psychology*, 61:491–515.
- Mengting Wan and Julian McAuley. 2016. Modeling ambiguity, subjectivity, and diverging viewpoints in opinion question answering systems. In *2016 IEEE 16th international conference on data mining (ICDM)*, pages 489–498. IEEE.
- Xuwei Wang, Weiyang Shi, Richard Kim, Yoojung Oh, Sijia Yang, Jingwen Zhang, and Zhou Yu. 2019. Persuasion for good: Towards a personalized persuasive dialogue system for social good. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5635–5649.

- Liqiang Xiao, Jun Ma, Xin Luna Dong, Pascual Martinez-Gomez, Nasser Zalmout, Wei Chen, Tong Zhao, Hao He, and Yaohui Jin. 2021. End-to-end conversational search for online shopping with utterance transfer. *arXiv preprint arXiv:2109.05460*.
- Hu Xu, Bing Liu, Lei Shu, and Philip S Yu. 2019. Review conversational reading comprehension. *arXiv preprint arXiv:1902.00821*.
- Hong Yu and Vasileios Hatzivassiloglou. 2003. Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In *Proceedings of the 2003 conference on Empirical methods in natural language processing*, pages 129–136.
- Munazza Zaib, Wei Emma Zhang, Quan Z Sheng, Adnan Mahmood, and Yang Zhang. 2021. Conversational question answering: A survey. *arXiv preprint arXiv:2106.00874*.
- Biqing Zeng, Heng Yang, Ruyang Xu, Wu Zhou, and Xuli Han. 2019. Lcf: A local context focus mechanism for aspect-based sentiment classification. *Applied Sciences*, 9(16):3389.
- Yongfeng Zhang, Xu Chen, Qingyao Ai, Liu Yang, and W Bruce Croft. 2018. Towards conversational search and recommendation: System ask, user respond. In *Proceedings of the 27th acm international conference on information and knowledge management*, pages 177–186.
- Yiheng Zhou, He He, Alan W. Black, and Yulia Tsvetkov. 2019. A dynamic strategy coach for effective negotiation. In *Proceedings of the 20th Annual SIGdial Meeting on Discourse and Dialogue, SIGdial 2019, Stockholm, Sweden, September 11-13, 2019*, pages 367–378. Association for Computational Linguistics.