This version of the contribution has been accepted for publication, after peer review (when applicable) but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections. The Version of Record is available online at: http://dx.doi.org/10.1007/978-3-031-60615-1_4. Use of this Accepted Version is subject to the publisher's Accepted Manuscript terms of use https://www.springernature.com/gp/open-research/policies/accepted-manuscript-terms

Optimizing conversational commerce involving multilingual consumers through large language models' natural language understanding abilities

Joseph Benjamin Ilagan[1[0000−0002−3748−6468]], Jose Ramon Ilagan[1[0000−0002−2111−3675]], Pia Ysabel Zulueta[1[0009−0005−6569−9902]], and Maria Mercedes Rodrigo[1[0000−0001−7881−7756]]

Ateneo de Manila University, Quezon City, Philippines
{jbilagan,jrilagan,pzulueta,mrodrigo}@ateneo.edu
https://www.ateneo.edu

Abstract. Due to the emergence of natural language processing (NLP) interfaces, there has been growing intent to use conversational channels for commerce. Beyond customer service, NLP-enabled AI agents are being integrated into various steps of the order-to-cash (OTC) process. Social media and messaging platforms such as Facebook Messenger have become pivotal for businesses, especially during and after the COVID-19 pandemic, but adoption has been limited. In addition, attitudes towards fully-automated conversational agents (CA) have been mixed, and there is room for human involvement in transactional conversations. A distinguishing contribution of this research is leveraging the inherent capabilities of Large Language Models (LLMs) in handling multilingual conversations and extracting transactional details through named entity recognition (NER). The study describes a hybrid human-AI setup augmenting agents with an auto-agent leveraging LLMs' natural language understanding (NLU) capabilities, designed using the OTC process pattern applied to conversational UX frameworks. A prototype of the setup aims to streamline operations and reduce errors by enhancing the user experience during key OTC steps through improved conversational design. Recognizing the irreplaceable essence of human interaction, the hybrid *human-in-the-loop* approach was chosen, mitigating the impersonal nature of full automation. A prototype handling customers and humans augmented by LLMs for NER handling of transaction, customer, and product information was built. Sample synthetic bilingual conversations between customers and sales agents were generated using ChatGPT and fed into the system for evaluation.

Keywords: Conversational user experience · conversational commerce · Large language models · Natural language understanding · Named entity recognition · Generative AI· Co-pilots

1 Introduction

Conversational commerce (CC) is the human-like dialogue between a business (or brand) and a consumer through messaging apps, live chat, chatbots (or digital

assistants), and voice assistants [11, 21]. The purported benefits of CC include increased engagement, reduced cart abandonment, obtaining helpful feedback, boosted (online) sales, and building brand loyalty [11]. Due to the emergence of natural language processing (NLP) interfaces, there has been growing intent to use a conversational agent (CA) for commerce. Beyond customer service, NLP-enabled AI agents are being integrated into various *order-to-cash* (OTC) processes: order entry (including product or service search and selection), order fulfillment, invoicing, and payment collection.

Social media and messaging platforms such as Facebook Messenger have become pivotal for businesses, especially during and after the COVID-19 pandemic. Messaging platforms have been used for general customer service [7], including answering inquiries and discovering product or service preferences [4, 25]. In countries like the Philippines, using social media and messaging platforms such as Facebook Messenger has allowed businesses to reach customers quickly and remotely, especially during and after the COVID-19 pandemic [4].

However, adoption has been limited due to challenges like lack of standards, insufficient natural language understanding (NLU), lack of personalization, absence of human touch, support for local languages, difficulty handling complex dialogue, and operationalizing conversations. As such, humans are still involved [11, 21].

Recent developments in artificial intelligence hold promise in addressing some of these challenges. Generative artificial intelligence (GAI) generates content using probability and statistics derived through training from existing digital content (ex., text, video, images, and audio) [2]. A large language model (LLM) is a GAI and statistical model of tokens in the large public corpus of human-generated text [20]. Through training, LLMs understand and produce human-like language [6]. The tokens involved are words, parts of words, or individual characters, including punctuation marks. They are generative because one can sample and ask them questions [20]. GPT (Generative Pre-trained Transformer), an LLM-based system, is designed to generate sequences of words, code, or other data from statistical distributions derived from training, starting from a source input called the prompt [9]. GPT is based on the transformer architecture [8, 19, 22], which trains large amounts of publicly-available data. Aside from generating content, LLMs have demonstrated improved performance in NLP and NLU, whether in English or other languages [24], but still make mistakes when following user intent [15, 16]. While LLMs continue to improve, human involvement may still be needed, whether to be the main business agent or representative or to provide feedback for further training and fine-tuning [16]. Still, selective intent mining leveraging LLMs' capabilities in named entity recognition (NER) will be useful for an AI-powered auto agent assisting the human agent.

A distinguishing contribution of this research is leveraging the inherent capabilities of Large Language Models (LLMs) in handling multilingual conversations. Unlike earlier tools and techniques, LLMs' advanced NLU and named entity recognition (NER) capabilities were harnessed to extract transactional details from unstructured conversation texts. This enabled the automated struc-

turing of transaction records for database entry, marking a significant departure from traditional methods. Online transactions already occur via messaging apps but predominantly with human agents involved. With their enhanced NLU and NER capabilities, LLMs promise to address these challenges more effectively. The scope of conversation types will be *service conversations* [14], where one party (the customer) seeks a product or service, and the other party (a representative of the organization) provides the product or service. Both parties involve real humans, with the automated AI-powered agent only as a co-pilot to the seller or provider.

1.1 Objectives and Research Questions

This paper explores designing, architecting, and evaluating CC systems leveraging LLMs with inherent multilingual pretraining. The study proposes a hybrid human-AI setup augmenting agents with an auto-agent leveraging LLM NLU, designed using the OTC process pattern applied to conversational UX frameworks.

RQ1) How may the LLMs be harnessed for their natural language understanding (NLU) and named entity recognition (NER) capabilities to extract transaction details automatically?

RQ2) How should the hybrid human-agent CC be evaluated for readiness to handle natural conversations?

The system aims to streamline operations and reduce errors by enhancing the experiences of customers and sales agents during key OTC steps. The hybrid approach was chosen to mitigate the impersonal nature of full automation by recognizing the irreplaceable essence of human interaction and not relying purely on automated agents [11].

2 Review of Related Literature

Conversational user experience (CUX) design is a new design discipline that utilizes natural language processing to provide meaningful user engagement experiences through emerging chatbots and virtual agent platforms [14]. Conversational designers must express the mechanics of human conversation, yet current patterns of visual UX do not help much with this articulation [14]. Aside from the accessibility of the messaging platform, factors like informativeness, assurance, and empathy play a significant role in the purchasing intent of the customers when using a messenger platform [4, 21]. Regarding informativeness, the architecture and design aspects involving conversational UX must acknowledge that agent knowledge is as important as knowing what the customer requests [14]. Therefore, part of the architecture must include knowledge database management and retrieval mechanisms (Including product, product description, price, etc.). The Natural Conversational Framework (NCF) consists of "(1) an underlying interaction model, (2) a library of reusable conversational UX patterns, (3) a general method for navigating conversational interfaces, and (4) a novel set

of performance metrics based on the interaction model" [14]. The intent of this study is not to develop a fully automated chatbot, so some of the work related to CUX cited earlier is not applicable. AI-based chatbots are also perceived to be impersonal, so humans are still involved [4, 21]. Instead, the conversation setup in this study is based on the concept of *human in the loop* and *mutual handover* [18]: small tasks handled by the machine, with the option to escalate to humans for more complex tasks. More importantly, regardless of the level of involvement of automation or humans, a task-oriented dialogue system specifically handles transaction tasks modifying state background (ex., recording a sale) [18]. Unfortunately, most task-oriented dialogue systems are information-seeking based and have quite limited transaction-based support [18], so this study hopes to contribute more to the transactional aspects of dialogue systems.

The evaluation of the efficacy of LLMs in extracting transaction data from natural conversations will require the presence of such conversations for testing, but preparing these conversations manually may be costly and time-consuming, so *data augmentation*, that is, the generation of new data by transforming existing data based on some prior knowledge [5]. The idea of generating and simulating natural conversations as a large language model task is nothing new, as featured in [13]. A similar technique of simulating conversations through synthetic data with the aim of *realism* was also featured in a framework for conversational recommender systems (CRS) by [10]. Related to this, while few-shot learning may be employed, some of the innate characteristics of humans acquired through how LLMs are trained are based on the idea of *homo silicus* [12].

Finally, with a prompting pattern or technique known as chain-of-thought (CoT) [23], LLMs can be induced to produce smaller, intermediate building blocks before answering [17]. The CoT approach to prompting has been found to yield significant performance improvements over other ways of prompting, and this study shall follow a similar approach.

3 Method and Implementation

The pilot case followed the ordering of merchandise (clothing of different sizes) with the buyers and sellers from metropolitan areas in the Philippines as the intended commercial location context.

3.1 Technical Architecture

The auto agent prototype for the pilot case was implemented as a Phoenix LiveView web application, receiving updates from a Viber chatbot webhook and executing actions based on the webhook messages. The choice of Phoenix was attributed to Elixir's capability to handle soft real-time updates, allowing the research to focus on the behavior of the LLM over implementation intricacies. OpenAI's gpt-3.5-turbo model served as the LLM via OpenAI's HTTP API. Supporting technology included Caddy, utilized as a reverse proxy web server, and Postgres, employed as the Phoenix application's database within a Docker

container. Initial evaluations used synthetic conversation scenarios generated by the LLMs, deferring the need for real conversation data and addressing ethics and data privacy concerns.

The evaluation of the system entails producing test multilingual conversation data, in this case, using the *taglish* (combination of English and *Tagalog*, a major local dialect in the Philippines). For this research, the approach used for data augmentation involving synthetic conversations is described in the following subsections.

3.2 Knowledge base.

A sample products table was generated manually for the pilot case to represent the standardized product catalog of a business. This database includes 20 rows. Each row represents one clothing product and contains the following attributes:

- Product ID (integer)
- Color (string)
- Name (string)
- UOM (string)
- Unit Price (integer)

3.3 Synthetic customer data preparation.

Five different Filipino customer personas were generated for the pilot case. GPT-4 was used to generate the personas. The basis of this approach is from [3] and [12].

The following properties were created for all five customer personas: *demographic attributes*, *preferred language (English, Tagalog, or Taglish)*, *conversational style*, and *monthly disposable income (in Philippine Pesos)*

This step involves having ChatGPT generate customer personas. Five personas were created for the initial round. ChatGPT will remember the personas in the session.

3.4 Synthetic seller persona creation.

One Filipino seller persona was created using a similar method to 3.3.

The following details were created for the seller persona:

- Demographic attributes
- Preferred language (English, Tagalog, or *Taglish*)
- Conversational style
- Monthly revenue in Philippine pesos

This step involves asking ChatGPT to generate a seller persona based on certain demographics and attributes, including conversational style, disposable income, etc.

3.5 Synthetic sales conversation creation.

Synthetic bilingual customer-initiated conversations were generated using ChatGPT and fed into the system for evaluation. Customers personas were initiated through prompts based on work done in [1, 3, 12] on the premise that LLMs have innate knowledge about behaviors based on the way they have been trained through large bodies of text.

GPT-4 was instructed to generate synthetic conversations with four characteristics intended to simulate the interaction between the customer as an agent that expresses relatively unstructured instructions and the seller as an agent that will need to perform structured bookkeeping of transaction details:

- First, that a conversation between a customer and a seller can end with either a confirmed transaction or a mere inquiry.
- Second, that the seller must ask for a customer's desired products, their delivery address, and an explicit confirmation of their order before the seller can consider the transaction to be confirmed.
- Third, that the seller knows about their own product catalog, but that the customer does not necessarily know the product catalog.
- Fourth, that the customer may sometimes change their desired products mid-conversation.

4 Results

Five synthetic conversations were generated using the approach detailed in 3. In this pilot case, the LLM component of the system is not given any power to affect the conversation directly; it is instead used to do background analysis of the conversation as it occurs. Such soft real-time analyses may be used as a basis for triggering parameterized jobs (e.g., inserting a new transaction record into the database upon determination that a transaction was confirmed solely from the conversation history contents).

4.1 General applied CoT approach

One pattern that emerged during the development of the pilot case was a pattern following CoT. The following sections will illustrate the efficacy of treating steps in conversation history analysis as black-box functions that accept a conversation history as a string and return semi-structured data (in this paper, JSON) to use in traditional programming control flow. The primary aim of this pilot case is to explore the ability of LLMs to serve as the implementation details of these black-box functions.

4.2 Presence of necessary conditions

The entry function of the conversation analysis tool determined whether the conversation history, up to that point, had satisfied three conditions: first, that

the customer had selected products; second, that the customer had given their delivery address; and third, that the customer had explicitly confirmed their order. The function returned one JSON object with three keys referring to these conditions as boolean values.

With this return data, decisions may be made regarding whether the assistant program should proceed to insert a record into a database, prompt the seller to ask for more information, and other relevant operations. In this pilot case, the second step involving 4.3 proceeds if all conditions are true.

4.3 Product resolution

In this pilot case, the assistant program is also tasked to parse the conversation data, which is relatively unstructured, into structured or semi-structured records of which products in which quantities are part of the transaction.

The first function for this task generates a JSON list where each record represents one conceptual line item in the transaction. At this stage, the function is only tasked to parse the product details as the customer and/or seller expressed them in the conversation.

The second function resolves each line item generated by the first function to an actual product in the products table. It was found that the LLM (GPT-3.5-Turbo) only performed acceptably once each column of the products table used the same terminology as the prompt (e.g., once the column **UOM** in the products table was renamed to "Size"). Only once each line item from the first function has been resolved to a known product can the transaction data be inserted into the database.

The execution of the pilot case has uncovered a potentially challenging and important step in this workflow. In this pilot case, the products table was small enough to include whole in the prompt of the second function; in larger cases, a proper search solution and further decomposition of tasks and functions may become necessary. Product resolution may become the subject of future work.

4.4 Database insertions

Generating SQL statements to insert a new transaction and its corresponding line items becomes trivial once the relevant product IDs have been parsed out of the unstructured conversation data. The pilot case program uses formatted strings to generate the necessary INSERT statements. A program to be used in production will more likely use parameterized queries to avoid SQL injection.

4.5 Sample case

This portion of the paper is an illustration of one of the five synthetic conversations and the analysis that the pilot case system performs on the conversation.

Example 1. The following conversation is a synthetic conversation generated by GPT-4. This specific conversation was chosen to illustrate that the assistant

program should be able to understand languages other than English in settings where other languages may be used. The original conversation includes emoji Unicode characters that were excluded in this document.

ELENA: Hi po! Tatanong lang po ako kung ano-anong klase ng damit ang meron kayo sa online shop ninyo?

MARIA: Hello po, ma'am Elena! Salamat po sa inyong pagtatanong. Mayroon po kaming iba't ibang uri ng damit, kabilang na ang silk shirts, polo shirts, at dresses. Meron din po tayong iba't ibang kulay at sizes. Ano po ba ang hinahanap ninyo na damit?

ELENA: Mga dresses po sana, yung mga bagay sa edad ko. Ano po ba ang available ninyo?

MARIA: Sigurado po akong mayroon tayong magugustuhan kayo, ma'am Elena. Mayroon po tayong "Sunday Dress" na puti and comes in sizes S, M, L, at XL, ito ay nasa 2450 PHP. At mayroon din po tayong "Cool Night Dress" na Midnight Blue, ito ay nasa 1990 PHP. Pareho po yung mga ito na babagay sa inyo.

ELENA: Yung Sunday Dress na puti po, gusto ko yun. Size M po ako.

MARIA: Magandang pumili po kayo, ma'am Elena! Ilan po ba ang gustong bilhin ninyo na Sunday Dress na puti, size M?

ELENA: Dalawa po sana. Pero teka, pwede rin ba malaman kung ano pa ang ibang kulay ng Sunday Dress?

MARIA: Pasensya na po, ma'am Elena, ngunit ang Sunday Dress po namin ay available lang sa kulay puti. Gusto pa rin po ba ninyo mag-order ng dalawang piraso?

ELENA: Ahh sige po, dalawang puting Sunday Dress na lang po, size M.

MARIA: Narinig ko po, dalawang Sunday Dress na puti, size M, para sa inyo. Para po matuloy natin ang order, pwede ko po bang malaman kung saan ito ipapadala o anong address?

ELENA: Sa Davao City po, pa drop off sa San Lorenzo Ruiz Gkk, Bacaca Road, El Rio Vista Village. Paki double-check yung order ko po ha? Dalawang puting Sunday Dress, size M.

MARIA: Tama po. Dalawang Sunday Dress na puti, size M na ipapadala sa address ninyo sa Davao City. Ang total po ay 4900 PHP. Kung tama na po ang lahat, pwede po bang i-confirm ninyo na gusto ninyong ituloy ang order na ito at ipapadala ko na po ito once may confirmation na?

ELENA: Sige na nga po, ituloy na natin. Iconfirm ko na po ang order. Salamat po.

MARIA: Maraming salamat po, ma'am Elena! Ipa-process ko na po yung order ninyo. Kung may iba po kayong tanong o kailangan, huwag po kayong mahiya na mag-message. Salamat po ulit sa inyong pagtitiwala!

Product resolution The first step in the product resolution workflow yields the following JSON from the conversation as shown in Listing 1.1.

Listing 1.1. Example JSON data for an order containing items.

```
{
    "items": [
        {
            "details": {
                "name": "Sunday Dress",
                "size": "M"
            },
            "quantity": 2
        }
    ]
}
```

The second step in the product resolution workflow resolves the data to a known SKU as shown in Listing 1.2.

Listing 1.2. JSON representation of an order with product ID and quantity.

```
{
    "items": [
        {
            "product_id": 10000009,
            "quantity": 2
        }
    ]
}
```

Generated SQL INSERT statements Passing the resolved product data and the conversation data (for additional details such as delivery address) through the approach described above results in the INSERT statements shown in Listing 1.3.

Listing 1.3. SQL commands for inserting a new transaction and its associated line items.

```
INSERT INTO app_transaction (created_timestamp,
    delivery_address) VALUES (NOW(), 'San Lorenzo Ruiz Gkk
    , Bacaca Road, El Rio Vista Village, Davao City')
    RETURNING id;

INSERT INTO line_item (created_timestamp, transaction_id,
    product_id, quantity) VALUES (NOW(), 1, 10000009, 2);
```

Assuming initially empty tables, these statements result in the tables as illustrated in Fig 1.

```
$ psql -U postgres

postgres=# SELECT * FROM app_transaction;
 id |       created_timestamp       |      delivery_address
----+-------------------------------+----------------------------
  1 | 2024-01-31 03:25:06.156323+00 | San Lorenzo Ruiz Gkk, ...
(1 row)

postgres=# SELECT * FROM line_item;
 id | ... | product_id | quantity | transaction_id
----+-----+------------+----------+----------------
  1 | ... |   10000009 |        2 |              1
(2 rows)
```

Fig. 1. Postgres and SQL commands for retrieving transaction details and associated line items from a database.

4.6 Drawbacks and limitations

If extended, the approach used in this pilot case may result in numerous calls to the LLM, where many of the calls include the whole conversation history. If used message-by-message, the number of calls to the LLM will be multiplied accordingly. The cost of calling available LLMs (i.e., in time, money, etc.) therefore becomes a consideration in whether this approach will be useful.

The specific approach used in the pilot case suffers from several limitations. First, the pilot case did not consider how to store session state (e.g., whether the current conversation session with the user already has an associated transaction row in the database). In a production-grade system, failure to consider session state may lead to critical errors such as duplicate transaction rows being inserted into the database. One possible solution is to include session data in the prompt alongside the conversation history and to edit the prompts to induce the LLM to consider the session data when making decisions. Second, the pilot case did not comprehensively consider safeguards for cases where the LLM incorrectly evaluates conversation data and thus mistakenly initiates side effects. One possible solution, consistent with the scope of the paper, is to require the seller and/or the seller's agent to confirm side effects (e.g., inserting a record into the database) before the side effects may be executed.

5 Discussions

RQ1) How may the LLMs be harnessed for their natural language understanding (NLU) and named entity recognition (NER) capabilities to extract transaction details automatically?

Development of the pilot case revealed that the CoT prompt engineering pattern, applied as chained black-box functions in an otherwise traditional server-side application, enables the structuring of data in unstructured conversation

logs. LLMs are used in the implementation details of these black-box functions. It was found that the best results were achieved when each function's scope was limited, but this results in a large amount of data transfer to the LLM, which may affect the economics of using this approach depending on which LLMs are available.

RQ2) How should the hybrid human-agent CC be evaluated for readiness to handle natural conversations?

The pilot case was executed on synthetic conversation data. It is a limitation of this paper that the prompts used to generate the synthetic conversations included broad instructions (e.g., the customer may not know the SKUs), which may have introduced bias that affects how well the synthetic conversations represent real conversations. Despite this limitation, the most prominent characteristic of conversational commerce, the weak structure (or lack of structure) in conversational data, was manifest in the synthetic conversations. The approach developed in the pilot case demonstrated an ability to handle such lack of structure, which is sufficient cause to perform future work to refine the approach on real conversation data.

6 Conclusion and Future Work

Transaction entities (products, quantities, prices) from early tests signify the effectiveness of the LLMs NER capabilities in capturing structured data needed to change database state automatically from unstructured chat transaction conversations. However, the scale of the tests performed in the pilot case was small. Future work may require more robust and more scalable methodologies than manual evaluation for evaluating the accuracy of LLM output, especially if real (and therefore potentially unlabeled) data is involved in future work.

After the initial evaluations using synthetic conversation scenarios generated by the LLMs, a significant direction for future work will be incorporating real human feedback. CRS and sentiment analysis will also be incorporated to further personalization and user satisfaction. While synthetic scenarios offer controlled, reproducible conditions and address ethics and data privacy concerns, they may not capture the full spectrum of human behavior and unpredictability. The anticipated outcomes encompass design guidelines, continuous model enhancement strategies, and best practices for integrating LLMs into hybrid human-AI frameworks, aiming to deliver natural CC experiences at scale across languages. While this paper touches upon technical implementation details, they are not its primary focus. These details will be discussed in subsequent papers as part of a larger project. Finally, any intermediate work will still involve humans-in-the-loop, and any attempts to transition to automated conversational agents shall be treated as an effort with a different research track.

Acknowledgments. We want to thank the Department of Quantitative Methods and Information Technology of the John Gokongwei School of Management, the Ateneo Laboratory fro Learning Sciences (ALLS), and the University Research Council of the

Ateneo de Manila University for making our participation and future presentation of this study possible.

References

1. Aher, G.V., Arriaga, R.I., Kalai, A.T.: Using large language models to simulate multiple humans and replicate human subject studies. In: International Conference on Machine Learning. pp. 337–371. PMLR (2023)
2. Baidoo-Anu, D., Owusu Ansah, L.: Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. Available at SSRN 4337484 (2023)
3. Brand, J., Israeli, A., Ngwe, D.: Using gpt for market research. Available at SSRN 4395751 (2023)
4. Casimiro, A.V., Chua, C., Pasquin, D.E., Grimaldo, J.R.: The Relationship of Facebook Messenger Marketing to the Purchasing Intention of the Consumers of Philippine MSMEs. Journal of Business and Management Studies **4**(2), 262–276 (2022), https://www.al-kindipublisher.com/index.php/jbms/article/view/3189
5. Chen, J., Tam, D., Raffel, C., Bansal, M., Yang, D.: An empirical survey of data augmentation for limited data learning in NLP. Transactions of the Association for Computational Linguistics **11**, 191–211 (2023)
6. Cooper, G.: Examining science education in chatgpt: An exploratory study of generative artificial intelligence. Journal of Science Education and Technology **32**(3), 444–452 (2023), publisher: Springer
7. Cui, L., Huang, S., Wei, F., Tan, C., Duan, C., Zhou, M.: Superagent: A customer service chatbot for e-commerce websites. In: Proceedings of ACL 2017, system demonstrations. pp. 97–102 (2017)
8. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 (2018)
9. Floridi, L., Chiriatti, M.: GPT-3: Its nature, scope, limits, and consequences. Minds and Machines **30**, 681–694 (2020)
10. Friedman, L., Ahuja, S., Allen, D., Tan, Z., Sidahmed, H., Long, C., Xie, J., Schubiner, G., Patel, A., Lara, H., Chu, B., Chen, Z., Tiwari, M.: Leveraging Large Language Models in Conversational Recommender Systems (May 2023), http://arxiv.org/abs/2305.07961, arXiv:2305.07961 [cs]
11. Hennigan, L.: What Is Conversational Commerce? (Jul 2023), https://www.forbes.com/advisor/business/conversational-commerce/
12. Horton, J.J.: Large language models as simulated economic agents: What can we learn from homo silicus? Tech. rep., National Bureau of Economic Research (2023)
13. Li, S., Yang, C., Yin, Y., Zhu, X., Cheng, Z., Shang, L., Jiang, X., Liu, Q., Yang, Y.: AutoConv: Automatically Generating Information-seeking Conversations with Large Language Models. In: Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). pp. 1751–1762 (2023). https://doi.org/10.18653/v1/2023.acl-short.149, http://arxiv.org/abs/2308.06507, arXiv:2308.06507 [cs]
14. Moore, R.J., Arar, R.: Conversational UX Design: An Introduction. In: Moore, R.J., Szymanski, M.H., Arar, R., Ren, G.J. (eds.) Studies in Conversational UX Design, pp. 1–16. Springer International Publishing, Cham (2018)

15. Naveed, H., Khan, A.U., Qiu, S., Saqib, M., Anwar, S., Usman, M., Akhtar, N., Barnes, N., Mian, A.: A Comprehensive Overview of Large Language Models (Oct 2023), http://arxiv.org/abs/2307.06435, arXiv:2307.06435 [cs]
16. Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A.: Training language models to follow instructions with human feedback. Advances in Neural Information Processing Systems **35**, 27730–27744 (2022)
17. Qin, C., Zhang, A., Zhang, Z., Chen, J., Yasunaga, M., Yang, D.: Is ChatGPT a General-Purpose Natural Language Processing Task Solver? (Feb 2023). https://doi.org/10.48550/arXiv.2302.06476, http://arxiv.org/abs/2302.06476
18. Quarteroni, S.: Natural Language Processing for Industry: ELCA's experience. Informatik-Spektrum **41**(2), 105–112 (Apr 2018). https://doi.org/10.1007/s00287-018-1094-1, http://link.springer.com/10.1007/s00287-018-1094-1, 22 citations (Crossref) [2023-12-18]
19. Rahaman, M.: Can chatgpt be your friend? emergence of entrepreneurial research. Emergence of Entrepreneurial Research (February 18, 2023) (2023)
20. Shanahan, M.: Talking About Large Language Models (Feb 2023). https://doi.org/10.48550/arXiv.2212.03551, http://arxiv.org/abs/2212.03551
21. Sidlauskiene, J., Joye, Y., Auruskeviciene, V.: AI-based chatbots in conversational commerce and their effects on product and price perceptions. Electronic Markets **33**(1), 24 (May 2023). https://doi.org/10.1007/s12525-023-00633-8, https://doi.org/10.1007/s12525-023-00633-8, 1 citations (Crossref) [2023-10-21]
22. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems **30** (2017)
23. Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q., Zhou, D.: Chain-of-Thought Prompting Elicits Reasoning in Large Language Models (Jan 2023). https://doi.org/10.48550/arXiv.2201.11903, http://arxiv.org/abs/2201.11903, arXiv:2201.11903 [cs]
24. Winata, G.I., Madotto, A., Lin, Z., Liu, R., Yosinski, J., Fung, P.: Language Models are Few-shot Multilingual Learners (Sep 2021), http://arxiv.org/abs/2109.07684, arXiv:2109.07684 [cs]
25. Yandug, J.S.G., De Francia, D.M.B., Paulo, J.O.: Assessment and improvement of Facebook business platforms for SMEs in the Philippines. In: Proceedings of the International Conference on Industrial Engineering and Operations Management, Dubai, UAE (2020), http://www.ieomsociety.org/ieom2020/papers/363.pdf