Original Article

Intelligent Neurological Based Email Categorization Using Svm To Increase Customer Support Efficiency

Harsh[1*], Inderpal Singh[2], Anshu Sharma[3], Chirag Sharma[4], Rachit Garg[5]

[1*]Department of Computer Science and Engineering, CTITR, Jalandhar
[2, 3, 4, 5]Department of Computer Science and Engineering, Lovely Professional University, Jalandhar

***Corresponding Author: -** Chirag Sharma
*Department of Computer Science and Engineering, CTITR, Jalandhar
DOI: 10.47750/pnr.2022.13.S08.525

Abstract

This paper is aimed to bring enhancement in customer support journey by automatically categorising emails into different categories using Artificial Intelligence. The major aspects of this research focus on neurological email based customer support processes. We examined that upto what neurological level the customer support process can be optimised. It explains an SVM based approach for email categorisation. Results of this research are providing approx 85% correct categorisation, precision and recall are also nearby 85%.

Keywords: artificial intelligence, svm, email categorization, customer support, neurological categorization

1 INTRODUCTION & BACKGROUND

This research is based on the topic of customer support. Customer support is the process of resolving any problems or queries raised by the customers on your product or service. Customer Support is a very important department for the success of any organisation. Every company differentiates itself from its competitors by virtue of its products & services features, but post-sales customer support helps in increasing the retention rate & customer loyalty, it also helps in decreasing customer acquisition cost (CAC) for the organisation. Technology is bringing revolution in all aspects of a business and hence customer support is also not untouched. There are various technical ways through which support is provided. However point to note here is lack of efficient use of technology may bring negative forces into branding whereas if used positively it can bring brand upliftment. Different ways have been evolved for providing customer support for e.g. - emails, call support, chat support, chatbots etc… Though there are tons of ways to provide support, many of them rely on manual work. To develop a customised support answer on a heavy flow of incoming queries in big organisations it becomes impossible to completely automatically build a support query's answer which provides solutions to a customer's problem by being inline with organisation policies.

Despite the fact that many communication channels exist for consumers to reach out to the organisation's customer support team and social media is rising as one big channel amongst them but emails have been a dominating mechanism for a good long and are not going to go away any sooner. Though there are various ways to write back solutions to those emails, eventually all of them lead to manual human work which raises a requirement of higher manpower & more costly implementations.

Customer support process and after sales service cost approximately 20-25% of the total expenses for any service based company, hence efficiency & optimization is much required after ensuring to provide right solutions to client's problem. "Data" is considered to be the oil for modern day organisation (specially tech driven organisation). A company in any stage, the way it handles data & leverages the power of it, decides the success of the company in thelong run. There are enough examples of data driven organisations including Facebook, Google, Amazon & many in India like - Zomato, MakeMyTrip etc… Customer support is one of those industries which can leverage data oriented technologies like machine learning, artificial intelligence and data science to improve by a significant margin. Some benefits of conducting further research over customer support is as follows:

- Clients will be benefited if the response to emails or calls would be faster & more up to the mark.
- Staff will be benefited out of it if these responses can be automated in some way, it'll allow organisations to operate on heavy volumes with lesser manpower.
- Branding of the company can evolve in a much different way.

This research discusses various components in a customer support journey and various challenges around them. Followed by that is implementation of SVM based design for automating the categorization of email queries received by organisations and results achieved by this approach.

2 LITERATURE REVIEW

According to research done by **Mujtaba et. al., 2017** major features to pick from an email which may help in detecting its category are as follows:

- Email Body Features
- JavaScript Features
- URL Features
- SpamAssassin Features
- Disconnected Features
- Online Features
- Social Features
- Network-based Features
- Stylometric Features
- Social Features
- Primary Features
- Lexical and Non-lexical Features
- Term-based Features
- State based Features
- Social Features
- Primary Features
- Lexical and non-lexical Features

Kannan et al. (2016) say that the Reply Generation System's main job is to figure out what the most likely answer is to a specific message. Deconstructing natural language dialogue in open spaces like Twitter has been done a lot, but most of the work has been on online media tasks like predicting whether or not a reply will be made, predicting only the next word, or putting strings in order. Before, full reaction forecasting was tried, which led to a challenge from the point of view of machine interpretation: given a Twitter post, use phrase-based quantifiable machine interpretation to "create an interpretation" of it as a reaction (SMT). But instead of SMT, an interpretation model for neural machines called "grouping-to-succession learning" can be used to improve accuracy. Arrangement-to-grouping recognition, which uses long transient memory organisations (LSTMs) to predict text successions, was first used in Machine Translation. Since then, it has been used successfully in other areas, such as picture subtitling and discourse recognition. Repeated neural networks (RNNs) or long short-term memory (LSTMs)have also been used in other studies to predict whole reactions. Some efforts use an SMT framework to find the n-best records, while others promote models that are completely generative. The main thing that makes our work stand out is that it was done in a production setting, which took into account issues like quality of response, usefulness, adaptability, and security. None of the related papers talked about these issues, so we made the changes we describe in the rest of this study. Also, an important focus on web-based media and film discourse is expected to help people understand messages well. In each of these situations, a simple connected or on-point answer is often enough. Email, on the other hand, often includes a request or goal that the reply should address.

In his study "Characterizing and Predicting Enterprise Email Reply Behavior," Liu et al. (2017) finds and mentions that by showing client answer practises like answer rate and answer time, we can add machine insight to email frameworks to help both email recipients and shippers. Answer predictions could help email recipients sort out messages that need a reply or need a quick response. This could help reduce the amount of email they have to deal with. During email structuring, answer habits for email senders could be predicted ahead of time.

In general, better response strategies could lead to better correspondence skills. For example, making it clear that an email is a solicitation (e.g., "Could we get together?") could help the recipient understand that he or she needs to respond quickly. Also, warning the sender that a response is likely to take longer if it is sent late at night or at the end of the week could improve the effectiveness of correspondence.

Previous studies looked at how people organise, reply to, and get rid of their email messages. Still, these evaluations are based on quick overviews or meetings. Some new studies suggest ways to pay attention to how clients are using messages with a lot of data. Both of these investigations are based on messages from Yahoo! Mail shoppers. Even though a few studies have shown that business email use isn't the same as consumer email use, Compared to the use of buyer email, it hasn't gotten as much attention. Some studies show, for example, that company clients send and receive twice as many messages as buyer clients. Other studies show that machine-made messages from corporate and informal communication destinations are now taking over shopper email. Kooti et. al. study .'s of how people respond to emails in Yahoo! Mail may be the most similar study to ours. Even so, they focus on private email and only look at a small subset of email transactions: those from coordinated (dyadic) email conversations between groups of clients who have traded more than 5 messages in buyer email. Focusing only on email conversations between two people is limiting, especially when it comes to large corporate communications. We looked at the email data from endeavour and found that 52.99% of messages were sent to more than one person other than the sender. So, it's important and more fair to focus on the bigger picture of how people respond to emails, which includes both one-on-one messages and messages sent from a group, with almost no restrictions on previous connections.

3 RESEARCH METHODOLOGY

Dataset required for this research is the received email, their corresponding replies, categorization done by executives for those emails and various steps taken further. The implemented web application suits best for such a data collection, as discussed earlier, all of this data was being stored in PostgreSQL database and can be processed and transformed for further processing by means of a scripting or programming language. Python once again is the chosen language for this and various tools and libraries used in Python for this step are as follows:

a) JSON extractor
b) UTF and XML decoders
c) Pandas and Numpy

XML decoders, JSON decoders, UTF-8 decoders etc… are then used to extract and decode features present in email messages and using various data transformation techniques are then converted into more readable formats. Some of the features in RFC format email messages include sender email ID, email subject, email body and attached files (if present), message format, delivery time and associated message ID etc…

This dataset is then passed through Numpy and Pandas to clean for missing values, remove outliers, convert datetime stamps from strings to UNIX values and categorical data into numerical values to get processed by the machine learning or deep learning algorithms. Cleaning and transformation converts all the data into uniform format which affects the prediction performance by a significant impact. Features extracted in this step are then prepared to be loaded into the artificial intelligence system.

Features extracted from the emails then acted as input to categorise emails into different categories e.g. need urgent reply or is a non-urgent message and various sub-categories. Various emails can be discarded to take any action based on the categories for e.g. promotional mails, spam mails, notification mails etc… However, few other types of mails such as job applications, freelancer bids etc… can be processed and saved into job applicant's data in case the company does not have a vacancy. Such cases will never be assigned to executives and will save a lot of overhead time. Other emails which require urgent / non-urgent response are assigned to executives in order of their detected urgency precedence and are expected to be solved by the executives.

In order to categorise the email here we have used a Support Vector Machine (SVM) classifier due to its performance and ability to categorise multidimensional data into different categories using the support vectors separation technique. Support Vector Classifiers are also complimented with the 'rbf' kernel to bring better performance. Outcome of using SVM classifier and performance on email categorization are discussed in the results section. Data was passed through various NLP (Natural Language Processing) processes before being passed through SVM. Various steps taken under NLP are intent recognition, similarity analysis etc…

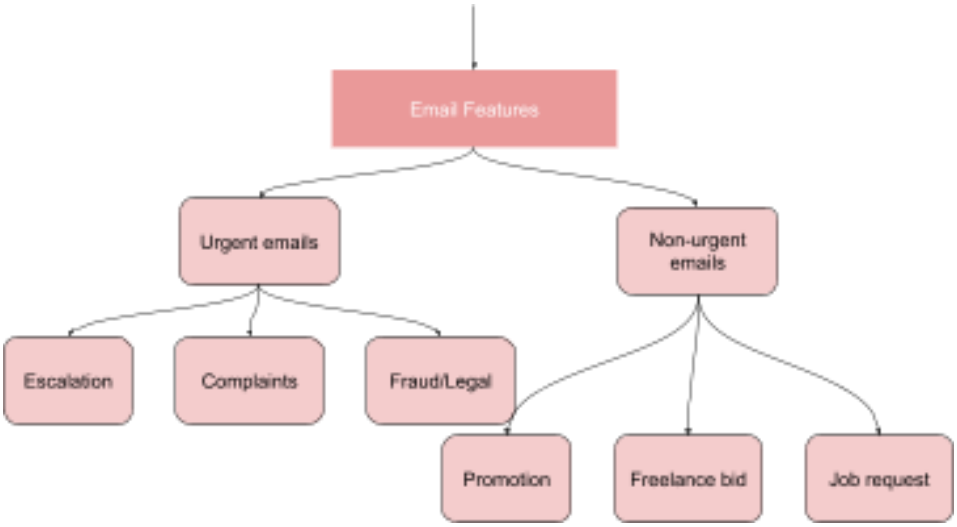

Figure 1: Categorization of emails into different categories (urgent and non-urgent) and sub-categories

4 RESULTS AND DISCUSSION

The experiments conducted using proposed designs to categorise emails into different buckets and mapping to right urgency score produced results which are discussed below. Note: Number of categories in the data used for this implementation are 10 categorised with a number in the range 0 to 9 (both inclusive).

Case 1: When research methodology is applied across industries

- Accuracy percentage is 54.7%

● Confusion Matrix

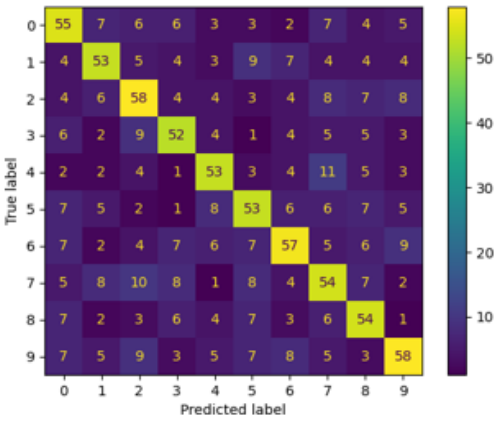

Figure 2: Confusion matrix for email categorisation with proposed methodology across industries

● Classification report

Table 1: Classification report including precision, recall, f1-score and support for various classes implemented across industries

	precision	recall	f1-score	support
0	0.528846153846154	0.561224489795918	0.544554455445545	98
1	0.576086956521739	0.546391752577732	0.560846560846561	97
2	0.527272727272727	0.547169811320755	0.537037037037037	106
3	0.565217391304348	0.571428571428571	0.568306010928962	91
4	0.582417582417582	0.602272727272727	0.592178770949721	88
5	0.524752475247525	0.53	0.527363184079602	100
6	0.575757575757576	0.518181818181818	0.545454545454546	110
7	0.486486486486487	0.504672897196262	0.495412844036697	107
8	0.529411764705882	0.580645161290323	0.553846153846154	93
9	0.591836734693878	0.527272727272727	0.557692307692308	110
accuracy	0.547	0.547	0.547	0.547
macro avg	0.54880858482539	0.548925995633642	0.548269187031713	1000
weighted avg	0.548485766657997	0.547	0.54712129667355	1000

Case 2: When research methodology is applied over only EdTech industry

● Accuracy percentage is 85%

● Confusion Matrix

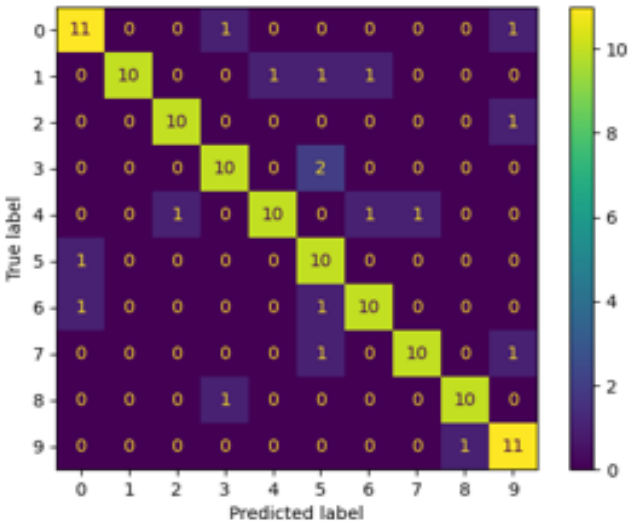

Figure 3: Confusion matrix for email categorisation with proposed methodology on EdTech industry

● Classification report

Table 2: Classification report including precision, recall, f1-score and support for various classes

	precision	recall	f1-score	support
0	0.846153846153846	0.846153846153846	0.846153846153846	13
1	1	0.769230769230769	0.869565217391304	13
2	0.909090909090909	0.909090909090909	0.909090909090909	11
3	0.833333333333333	0.833333333333333	0.833333333333333	12
4	0.909090909090909	0.769230769230769	0.833333333333333	13
5	0.666666666666667	0.909090909090909	0.769230769230769	11
6	0.833333333333333	0.833333333333333	0.833333333333333	12
7	0.909090909090909	0.833333333333333	0.869565217391304	12
8	0.909090909090909	0.909090909090909	0.909090909090909	11
9	0.785714285714286	0.916666666666667	0.846153846153846	12
accuracy	0.85	0.85	0.85	0.85
macro avg	0.86015651015651	0.852855477855478	0.851885071450289	120
weighted avg	0.862409812409812	0.85	0.851565403195838	120

It is evident that the results are significantly better in the case of EdTech based implementation in contrast with general implementation across industries because of varying industry dynamics. The results presented for both the cases shows that the model is not biassed towards any specific category and the confusion matrix is practically in a balanced score.

Accuracy of 85% is a good performance considering the risks involved. The domain is not critical and it is acceptable to have some form of deviation from correct answers against the much higher benefits of automation that is offered by the implemented design. Classification report mentioned in the result section presents precision and recall to be in line with accuracy. The weighted average of precision is 0.862409812409812 and that of recall is 0.85 which shows no-biases towards any particular category of emails.

5 CONCLUSION AND FUTURE WORKS

This research with presented literature review, methodology and results show that customer support processes can be and need to be automated for better branding and support to the customers on the basis of neurological parameters. The result section clearly shows that emails received by the business can be easily categorised into various categories and the process achieved 85% accuracy in doing so which can be enhanced further by more modifications. We also discussed various steps involved in building a reply generation mechanism for the received emails by any company which can drastically reduce the time between received email and sent response. Mentioning about the future works, there are various tangents this research can be extended upon. This design can be customised to make it generic and able to work across industry with acceptable performance metrics. Currently the accuracy is 54.7% for across industry implementations which can be improved by change in design, better sources for data collection, improved data processing etc…

6 REFERENCES

1. Adam, M., Wessel, M., Benlian, A., 2021. AI-based chatbots in customer service and their effects on user compliance. Electronic Markets 31, 427–445. doi:10.1007/s12525-020-00414-7
2. Ameen, N., Tarhini, A., ... Anand, A., 2021. Customer experiences in the age of artificial intelligence. Computers in Human Behavior 114. doi:10.1016/j.chb.2020.106548
3. Banday, M. Tariq & Qadri, Jameel & Shah, N.A.. (2010). A Practical Study of E-mail Communication through SMTP.
4. Barbosa, A., Godoy, A., 2021. Augmenting Customer Support with an NLP-based Receptionist. Sociedade Brasileira de Computacao - SB, pp. 133–142. doi:10.5753/stil.2021.17792
5. Carvalho, V.R., Cohen, W.W., 2005. On the collective classification of email "speech acts," in: SIGIR 2005 - Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. pp. 345–352. doi:10.1145/1076034.1076094
6. Cheng, Y., Jiang, H., 2022. Customer–brand relationship in the era of artificial intelligence: understanding the role of chatbot marketing efforts. Journal of Product and Brand Management 31, 252–264. doi:10.1108/JPBM-05-2020-2907
7. Chhabra, G.S., Bajwa, D.S., 2012. Review of E-mail System, Security Protocols and Email Forensics. International Journal of Computer Science & Communication Networks, 5, 201–211.
8. Cohen, W.W., Carvalho, V.R., Mitchell, T.M., 2004. Learning to classify email into "speech acts," in: Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing, EMNLP 2004 - A Meeting of SIGDAT, a Special Interest Group of the ACL Held in Conjunction with ACL 2004. Association for Computational Linguistics (ACL), pp. 309–316.
9. Daqar, M.A.M.A., Smoudy, A.K.A., 2019. THE ROLE OF ARTIFICIAL INTELLIGENCE ON ENHANCING CUSTOMER EXPERIENCE. International Review of Management and Marketing 9, 22–31. doi:10.32479/irmm.8166
10. Das, B., Majumder, M., ... Sekh, A.A., 2021. Automatic question generation and answer assessment: a survey. Research and Practice in Technology Enhanced Learning. doi:10.1186/s41039-021-00151-1
11. Di Castro, D., Karnin, Z., ... Maarek, Y., 2016. You've got mail, and here is what you could do with it! Analyzing and predicting actions on email messages, in: WSDM 2016 - Proceedings of the 9th ACM International Conference on Web Search and Data Mining. Association for Computing

Machinery, Inc, pp. 307–316. doi:10.1145/2835776.2835811
12. Huang, M.H., Rust, R.T., 2018. Artificial Intelligence in Service. Journal of Service Research 21, 155–172. doi:10.1177/1094670517752459
13. Kannan, A., Kurach, K., ... Ramavajjala, V., 2016. Smart reply: Automated response suggestion for email, in: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.Association for Computing Machinery, pp. 955–964. doi:10.1145/2939672.2939801
14. Khan, S., Iqbal, M., 2020. AI-Powered Customer Service: Does it Optimise Customer Experience?, in: ICRITO 2020 - IEEE 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions). Institute of Electrical and Electronics Engineers Inc., pp. 590–594. doi:10.1109/ICRITO48877.2020.9198004
15. Kreutzer, R.T., Sirrenberg, M., 2020. Fields of Application of Artificial Intelligence—Customer Service, Marketing and Sales. pp. 105–154. doi:10.1007/978-3-030-25271-7_4
16. Kühl, N., Mühlthaler, M., Goutier, M., 2020. Supporting customer-oriented marketing with artificial intelligence: automatically quantifying customer needs from social media. Electronic Markets 30, 351–367. doi:10.1007/s12525-019-00351-0
17. La Lau, R., 2021. Advanced Email, in: Practical Internet Server Configuration. Apress, pp. 339–366. doi:10.1007/978-1-4842-6960-2_14
18. Lee, S.B., 2020. Chatbots and Communication: The Growing Role of Artificial Intelligence in Addressing and Shaping Customer Needs. Business Communication Research and Practice 3, 103–111. doi:10.22682/bcrp.2020.3.2.103
19. Leone, D., Schiavone, F., ... Chiao, B., 2021. How does artificial intelligence enable and enhance value co-creation in industrial markets? An exploratory case study in the healthcare ecosystem. Journal of Business Research 129, 849–859. doi:10.1016/j.jbusres.2020.11.008
20. Li, C., Pan, R., ... Deng, Z., 2020. Research on Artificial Intelligence Customer Service on Consumer Attitude and Its Impact during Online Shopping, in: Journal of Physics: Conference Series. Institute of Physics Publishing. doi:10.1088/1742-6596/1575/1/012192
21. Madana Mohana, R., Pitty, N., Lalitha Surya Kumari, P., 2021. Customer support chatbot using machine learning, in: Advances in Intelligent Systems and Computing. Springer Science and Business Media Deutschland GmbH, pp. 445–451. doi:10.1007/978-981-15-5679-1_42
22. Mani, Abilash. (2020). Artificial Intelligence (AI) Chatbots revolutionizing the customer support service: An Overview of theories and Literature.
23. Masood, Gazala & Bansal, Prof & Jain, Vipin & Kapila, Dhiraj & Pandian, Asha & Jaiswal, Sushma. (2021). Application of Machine Learning to Improve Consumer Devices and Services. 12. 2307-2314.
24. McTear, M., 2010. The Role of Spoken Dialogue in User-Environment Interaction, in: Human-Centric Interfaces for Ambient Intelligence. Elsevier Inc., pp. 225–254. doi:10.1016/B978-0-12-374708-2.00009-7
25. Mujtaba, G., Shuib, L., ... Al-Garadi, M.A., 2017. Email Classification Research Trends: Review and Open Issues. IEEE Access 5, 9044–9064. doi:10.1109/ACCESS.2017.2702187
26. Nordheim, C.B., Følstad, A., Bjørkli, C.A., 2019. An Initial Model of Trust in Chatbots for Customer Service - Findings from a Questionnaire Study. Interacting with Computers 31, 317–335. doi:10.1093/iwc/iwz022
27. Pantano, E., Pizzi, G., 2020. Forecasting artificial intelligence on online customer assistance: Evidence from chatbot patents analysis. Journal of Retailing and Consumer Services 55. doi:10.1016/j.jretconser.2020.102096
28. Ping, N.L., Hussin, A.R.B.C., Ali, N.B.M., 2019. Constructs for artificial intelligence customer service in E-commerce, in: International Conference on Research and Innovation in Information Systems, ICRIIS. IEEE Computer Society. doi:10.1109/ICRIIS48246.2019.9073486
29. Prabu, M., Sai Tarun, T., ... Vijay, A., 2020. Enhancing Customer Service Using Chatbot Application through Artificial Intelligence. Journal of Computational and Theoretical Nanoscience 17, 1633–1637. doi:10.1166/jctn.2020.8414
30. Prabu, M., Sai Tarun, T., ... Vijay, A., 2020. Enhancing Customer Service Using Chatbot Application through Artificial Intelligence. Journal of Computational and Theoretical Nanoscience 17, 1633–1637. doi:10.1166/jctn.2020.8414
31. Rana, J., Gaur, L., ... Rasheed, M.I., 2021. Reinforcing customer journey through artificial intelligence: a review and research agenda. International Journal of Emerging Markets. doi:10.1108/IJOEM-08-2021-1214
32. Riabov, V.V., 2011. SMTP (Simple Mail Transfer Protocol), in: Handbook of Computer Networks. John Wiley and Sons, pp. 388–406. doi:10.1002/9781118256114.ch26
33. Satish Kumar Boguda, 2019. The Future of Customer Experience in the Information Age of Artificial Intelligence - Get Ready for Change. International Journal of Engineering Research and V8. doi:10.17577/ijertv8is060622
34. Sureswaran, R., Al Bazar, H., ... El-Taj, H., 2009. Active e-mail system SMTP protocol monitoring algorithm, in: Proceedings of 2009 2nd IEEE International Conference on Broadband Network and Multimedia Technology, IEEE IC-BNMT2009. pp. 257–260. doi:10.1109/ICBNMT.2009.5348490
35. Tyler, J.R., Tang, J.C., 2003. When Can I Expect an Email Response? A Study of Rhythms in Email Usage, in: ECSCW 2003. Springer Netherlands, pp. 239–258. doi:10.1007/978-94-010-0068-0_13
36. Wang, X., Lin, X., Shao, B., 2022. Artificial intelligence changes the way we work: A close look at innovating with chatbots. Journal of the Association for Information Science and Technology. doi:10.1002/asi.24621
37. Xu, Y., Shieh, C.H., ... Ling, I.L., 2020. AI customer service: Task complexity, problem-solving ability, and usage intention. Australasian Marketing Journal 28, 189–199. doi:10.1016/j.ausmj.2020.03.005
38. Yang, L., Dumais, S.T., ... Awadallah, A.H., 2017. Characterizing and predicting enterprise email reply behavior, in: SIGIR 2017 - Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval. Association for Computing Machinery, Inc, pp. 235–244. doi:10.1145/3077136.3080782