Leveraging Conversational Systems to assists New Hires during Onboarding

Praveen Chandar[1], Yasaman Khazaeni[1], Matthew Davis[1], Micheal Muller[1], Marco Crasso[2], Q. Vera Liao[3], N. Sadat Shami[4], and Werner Geyer[1]

[1] IBM Research, Cambridge, MA, USA
[2] IBM Research - SilverGate, Buenos Aires, Argentina
[3] IBM Research, Yorktown, NY, USA
[4] IBM, Armonk, NY, USA
{pcravich,yasaman.khazaeni,davismat,
michael_muller,sadat,werner.geyer}@us.ibm.com,
vera.liao@ibm.com, crasso@ar.ibm.com

Abstract. The task of onboarding a new hire consumes great amounts of resources from organizations. The faster a "newbie" becomes an "insider", the higher the chances of job satisfaction, retention, and advancement in their position. Conversational agents (AI agents) have the potential to effectively transform productivity in many enterprise workplace scenarios so applying them to the onboarding process can prove to be a very solid use case for such agents. In this work, we present a conversational system to aid new hires through their onboarding process. Users interact with the system via an instant messaging platform, to fulfill their work related information needs as if it were a human assistant. We describe the end-to-end process involved in building a domain specific conversational system and share our experiences in deploying it to 344 new hires in a month-long study. The feasibility of our approach is evaluated by analyzing message logs and questionnaires. Through three different measures, we observed an accuracy of about 60% at the message level and a higher than average retention rate for the agent. Our results suggest that this agent-based approach can very well compete with the existing tools for new hires.

Keywords: Artificial Intelligence, Conversational Agents, Personal Assistants, Evaluation, NLP

1 Introduction

Lave and Wenger [11] showed that people join a community or an organization slowly, through a process of "legitimate peripheral participation," in which they must first learn their new environment, then tentatively test their knowledge, and finally begin to act as fully competent members. Organizations and their new-hires would benefit from speeding up this process from "newbie" to "insider."

With all the challenges that come with being a new hire, the path for every employee to go from the "newbie" stage to an "insider" depends on how fast they acquire enough knowledge to feel comfortable at their new workplace. Obviously this depends on the size of the company and the direct group who will be working with the employee, the

job rank, his/her number of years of previous experience, whether there is a change in the type of job and his/her personality and ability to deal with uncertainty. In [14] seven different information seeking tactics (overt questions, indirect questions, third parties, testing limits, disguising conversations, observing, surveillance) are studied which are used by new hires in an organization to seek information.

This paper describes how a conversational agent can be used to augment the onboarding process and details our experience with building and deploying an end-to-end system. We explain the various steps involved from curating a domain-specific knowledge base to evaluating the system using message logs and questionnaires. The main advantage of our proposed system is that unlike humans, conversational agents can be available 24/7 to answer new hire questions and unlike existing web-based tools the agent provides a human-like interaction. Additionally, the agent can help with compliance by sending proactive messages to the user which could be helpful in reminding new employees of their tasks to be completed.

The primary goal of our proposed system is to cater to the work-related informational needs of the newcomers in an organization. The system is designed to provide an answer for questions on company policies and procedures, or search the company's Intranet. In addition, the system can help new hires connect with co-workers by allowing them to find experts on certain topics as well as looking up co-workers through an intelligent directory look-up service. While information retrieval is an essential objective for our agent, social behavior and interaction capabilities are equally important. A number of works have dealt with how conversational agents should interact and behave with humans [7,28,24]. As noted by previous works, we anticipated the anthropomorphic chit-chat behavior and implemented a chit chat module providing a more natural human/agent interaction.

The use of conversation in information systems goes back to the 1950s [25]. More recently, success in the field of natural language processing and machine learning has led to rapid adaptation of state of the art chat bot technologies for information access in various fields including education [22], health care [5], or accounting [15]. Conversational systems have been an active area of research with focus areas including user experience [23,18], personalization [10], and dialog modeling [29,16,26]. Most conversational systems proposed in the literature are evaluated on datasets created for artificial tasks or in a laboratory. Our work focuses on the process involved in building an end-to-end conversational agent evaluated in the field.

The rest of this paper is organized as follows. Section 2 describes the system design and the steps involved in building a conversation system for a domain specific application. Section 3 provides details of the study design followed by the results of our user study in Section 4. We conclude the paper with a discussion and directions for future work.

2   System Description

We designed a conversational agent named *Chip* to aid new hires with their information needs and ease their way into the organization. The agent was available on a company wide instant messaging service that the new hires could add to their buddy list and start interacting. We conducted a user study to gain insights into the user's experience with system and estimate system performance. In this section, we describe the capabilities of the system along with design in detail.

2.1 System Capabilities

Knowledge Base: The primary goal of the agent was to address the information needs of newcomers in an organization. We compiled a set of frequently asked questions on a variety of topics to address their needs (refer to Section 2.2 for more details on the methodology used to identify these questions). The knowledge base also contained question variations that enabled our system to provide the same answer to different versions of the same underlying question. The knowledge base consisted of answers for each question. An example question-answer pair is show below:

```
        User Request : How can I host a teleconference?
      Chip's Response: Before you can host a teleconference,
  you have to sign up for a host code. You can visit this link
        to register. For complete instructions, check out
                    this link on audio conferencing.
```

Chit Chat: Our goal was not only to build a question answering system but also to provide a human touch to the interactions. It has been shown that "anthropomorphism" can solicit social responses from users such as trust, empathy, etc., and also helps with conversation [2]. Casual conversations such as chit chat is one possible way to humanize an agent [12]. We developed a chit chat module to handle casual conversation pertaining to agent trait (e.g. *who are you?, what do you like?, etc*), agent status (e.g *how are you?, what are you doing?, etc*), compliments (e.g *you are great!, you are smart, etc*). The agent could even tell jokes.

Search Capabilities: The agent had the ability to search over a set of documents in the organization's intranet. We restricted the search index to contain documents related to company policies and HR related pages. An inverted index was used to rank documents (webpages) in decreasing order of relevance for a given information need.

Lookup Capabilities: In addition to the search, the users could use lookup services such as:

– *Employee Demographics Lookup* : Users could search the employee directory by asking natural language questions. For example, the users can ask the agent to "Look up John Smith's phone number and address" and the agent would understand the intent and extract names entity (e.g. "John Smith") to search the employee directory.
– *Experts Lookup*: Employees joining an organization often have a hard time finding domain experts within the company. To ease this problem the agent was equipped with the ability to search an internal social media portal to return a list of domain experts along with their contact information.
– *Wikipedia Search*: The agent allowed the users to search Wikipedia for a given natural language query.

Proactive Reminders: New hires joining an organization have to complete a set of tasks within a specified time frame. These tasks include signing up for health insurance, join relevant mailing lists, etc. Our agent was designed to send proactive messages to users reminding them of their tasks and answer question related to the tasks. Proactive reminders also served as a way to remind the users about the availability of the agent.

2.2 System Design

The process that we followed to develop the conversational system for new hires can roughly be divided into three different phases: (1) Curating the knowledge base; (2) Building a conversational agent; (3)Tuning the performance.

Knowledge Base Curation: In order to build a comprehensive knowledge base consisting of question-answer pairs relevant to our new hire use case, we used call records obtained from the company's employee service center (ESC). The call records were composed of email conversations between employees and customer service representatives organized in the form of threads. The ESC call records contained a broad range of topics, including topics that were irrelevant to our new hire use case. In order to refine the topics in the call records, we performed topic modeling to identify prominent topics in the call records using the Latent Dirichlet Allocation [6] technique. The topics guided us to manually define a diverse set of candidate questions to be included in the knowledge base. Once the questions were finalized, a subject matter expert went through them to provide a short and concise answer to each of these questions.

Conversational Agent: The goal of our conversational agent was to return a text-based response for a given user request in natural language. Prior work on short text conversational systems can be categorized into retrieval-based [8] and generation-based methods [17]. Retrieval-based methods select the best possible answer from a set of predefined responses whereas generation-based methods create a natural language response for a given question. We used a retrieval based method in this work and the steps involved in obtaining a response are explained below:

1. *Intent identification:* Identifying the intent for a given question is key to understanding the user's information need. We annotate every incoming user message with a predefined intent using a natural language classifier. The classifier used in this work was a multi-layer convolution neural network similar to the one used in [9]. The goal of the intent classifier was to categorize the user's information need into chit chat, knowledge base question, Wikipedia search, employee or expert lookup. In practice, we found that splitting the chit chat and knowledge base question category into more granular intents was more effective. For instance, we decompose the knowledge base question into a list of topics such as health insurance, benefits, IT Help, etc. The use of CNN models enabled us to generalize our approach beyond exact matches. For example, a user can ask for: "information on health-care benefits", or "where can I find my health benefits information?"; both will be matched to the answer containing the appropriate web page for health benefits.
2. *Entity Extraction:* Additionally, each user request was also annotated with named entities and keywords (noun phrases) present in the text. We use the AlchemyAPI to extract entities and keywords [1] to be used in the answer selection step.
3. *Answer Selection*
   - *Knowledge Base and Chit Chat*
     Once the intent classifier has identified the user request as a chit chat or knowledge base question, the next step was to retrieve the most likely response from a

pool of predefined responses. Similar to the previous step, we use a multi-layer convolution neural network based classifier to select the best response. The intent and answer selection classifier returned a confidence score along with the most likely class label. The agent returned the selected answer as a response only when both were above threshold t_1 for intent classifier and t_2 for answer selection classifier.

– *Inverted Index*
When the intent classifier's confidence was above a threshold t_1 and answer selection classifier's confidence was below a threshold t_2, we used the search index as a fall-back. A query was created by removing the stopwords from the user request to search an inverted index created using the Lemur toolkit [21]. A sequential dependency model [13] was used to find relevant documents, and the top three documents along with the snippets were returned as a response.

– *Employees and Expert Lookup*
An internal employee lookup service was used by passing in the extracted person entity as arguments. The keywords extracted are passed as an argument to the expert lookup service and the output was returned as a response.

– *Wikipedia Search*
The extracted keywords are passed to the Wikipedia Search API as a query. The top ranked document along with the snippet is returned as a response.

Finally, the agent had to handle questions that were out of scope. The thresholds t_1 and t_2 described above were used to determine out-of-scope questions and a generic "Sorry, I don't understand you question" response was returned.

Performance Tuning: The thresholds t_1 and t_2 have a direct influence on system performance because they determine when a question is out of scope or when the agent must search the inverted index.

Determining optimal values for t_1 and t_2 is a non-trivial task so as an ad-hoc solution, in this work, we divided our user study into three different cohorts and used the user data from first cohort to tune the values for thresholds t_1 and t_2.

3   User Study

The conversational agent was made available to 344 new hires of a large organization for 4-5 weeks. The participants were recruited in three cohorts and were mostly college graduates with different backgrounds, including software engineering, consulting, product design, administration, and sales. The participants were located at different work locations across a single country and they attended a mandatory new hire orientation program organized by the company. A member of our research team was present at the orientation to demonstrate the capabilities of the agent. Upon adding the agent to their instant messenger's "friend list", the participants were able to interact with agent. For example, a new hire could ask the agent about benefits – "how do I set up my health insurance", lookup directory information – "what is John's phone number" or even find experts in their field of interest – "find an expert in machine learning", etc. The system was introduced as an experimental tool and the participants were requested to provide feedback by sending a *#FAIL* message for incorrect agent responses.

3.1 Post Study Questionnaire

We followed up with a post study questionnaire sent to participants after about a month of usage. Similar to other studies [30,3,4], we developed a version of the questionnaire to capture user opinion. To understand the user's self assessed accuracy of the system, we included the following question: *What percent of questions was Chip able to answer for you ?*

The users were also asked to rate a series of statements on a Likert scale. The following statements were included in the questionnaire to estimate the user-perceived efficacy of the agent. We replaced the agent name below with Chip for blind review.

– The answers Chip gave me were of high-quality answers.
– I would continue to use Chip if he were available to me.

We also included the following question on comparing the agent to other sources.

– I tended to ask Chip before turning to a coworker for help.
– Chip was able to help me when I would otherwise have called the Employee Service Center.
– I thought to ask Chip for help before emailing my onboarding Specialist.

Additionally, we asked the survey takers to rank the following sources by their ability to answer questions about the company's knowledge and process that is required for a new hire: (1) Co-worker (2) Intranet Search (3) *Chip* (4) onboarding Specialist assigned to the employee by the human resources (5) Employee Service Center.

4 Results

4.1 System Usage

In this section, we present the descriptive statistics of all user messages to *Chip*. Overall, *Chip* received a total of 5984 messages throughout the study from 322 users (average of 18.58 messages per user). [5] The messages received could be a request with an information need, chit chat, a lookup for employee demographics, search for domain experts, Wikipedia search, or "#FAIL" reporting system failure. Among them, 75.8% users have 5 or more unique interactions with *Chip*. Table 1 shows the usages statistics across the three cohorts.

Table 1: Usage Statistics across all three cohorts

	tot. users	tot. messages	messages/user
Group 1	89 (82)	1383	16.9
Group 2	124 (115)	2513	21.9
Group 3	131 (125)	2088	16.7

---

[5] About 6.3% of the users never used *Chip*.

As an agent that aims to provide continuous interactions, a critical design goal of *Chip* is to encourage users' long-term engagement. We observed that 25% messages came in the second two weeks of the user's access time to *Chip* compared to 75% coming in the first two weeks. Considering the effect of introduction to a new tool in the first week or so, this shows users didn't only use to system as a toy tool but for actual retrieval of information. Additionally, we observed a considerable number of active users [6] during the second half of the study, 26.1% of were using the system in the latter half compared to 69.3% in the first half. This shows a much higher retention rate compared to what is reported for most chatbots at lower than 10% in the first month. [31]

4.2 System Effectiveness

To assess the accuracy of the system at the message level, we randomly sampled 149 of request-response pairs from the message logs. The pairs were annotated by a researcher with a four point scale: correct, needs improvement, wrong and out of scope. Table 2 shows the result we get from this annotation. Among 149 question, we found around 58% answered correctly while around 13% answers were wrong.

Table 2: Annotation Results

Total	Correct	Improvement needed	Wrong	Out of Scope
149	87	16	19	27
	58.4%	10.7%	12.8%	18.1%

As described in Section 3.1, we also measured user satisfaction through question-naires in which the user evaluated the quality of response, compared the agent with other sources, etc. We distributed the survey to all participants, however, only 37% completed the questionnaire. Table 3 shows the results for two types of questions in the survey to measure satisfaction. The users rated the statements related to *efficacy* and *comparison to other sources* on a Likert scale. We convert the nominals to continuous scale by assigning 7 for strongly agree and 1 for strongly disagree (neutral is 4).

Table 3 shows that users in general agreed that the answers were of high quality and showed a strong intent to continue using the agent. The results also show that, as the first source of answers, users preferred to start with *Chip* before contacting their co-workers, onboarding specialist or the ESC.

We also asked the users to rank the different sources by their general ability to answer questions on company policies. While it is evident from Figure 1 that co-workers are the preferred choice to obtain reliable and correct answer, *Chip* is in the same level as intranet, ESC, and onboarding specialists.

The users' ratings of system accuracy at the message level was 63%. which is very close to 58% we get from the annotation results. A second estimate of accuracy comes from the users' #FAIL messages. However, users do not always report system failures, nevertheless, it serves as a reasonable metric to measure system effectiveness, we observed that "#FAIL" was used for about 8.5% of the messages.

---

[6] Users with five or more unique interactions were considered active users.
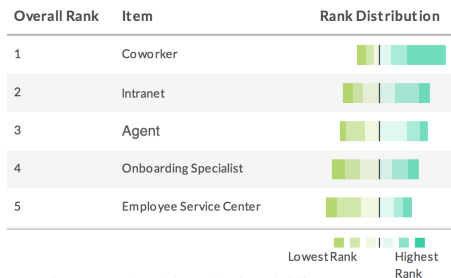
Fig. 1: Source of information ranked by their ability to answer questions for new hires from the questionnaire data.

Table 3: Questionnaire data analysis

Efficacy	Mean (SD)
The answers Chip gave me were high-quality answers.	4.75(1.6)
I would continue to use Chip if he were available to me.	5.44(1.63)
Comparison to other sources	
I tended to ask Chip before turning to a coworker for help.	4.59(1.98)
Chip was able to help me when I would otherwise have called the Employee Service Center.	4.8(1.69)
I thought to ask Chip for help before emailing my onboarding Specialist.	5.67(1.59)
What percent of questions was Chip able to answer for you?	62.2%(27.5%)

5 Discussion and Conclusion

In this paper, We presented a conversational agent (chat bot) that supports new hires with their informational needs during their onboarding phase. We described system capabilities and outlined design consideration in building such an agent. Evaluating domain specific retrieval-based conversational systems is a challenging open research problem, mixed results have been reported through objective vs subjective evaluations [19,20,27]. We approached evaluation through a live deployment and a field study with 344 new hires in a large organization and measured system effectiveness based on the data collected "in the wild," and post-deployment surveys. Our results are highly encouraging: The system we built can well compete with existing informational means such as call center or onboarding experts. We observed accuracies of around 60% (or 70% counting both correct answers and those that needed some improvement but still provided value) using both objective (message level annotations) and subjective (questionnaire) evaluations. Even with this average accuracy, subjective evaluation indicated a users' intent for continued use. In our experience, objective evaluations are effective in fine tuning the performance while subjective evaluations measure user satisfaction more effectively.

One of the most challenging and labor-intense aspects of building an informational agent is the curation of the knowledge base. It involves the manual generation of questions and their natural language variations with the guidance from subject matter experts from the employee service center. To automate this process, We have started applying text

analytic techniques to the content provided by the employee services center (emails, tickets, knowledge documents). We are currently working on improving this automated text analytics process so that knowledge bases for future chat bots in different domains can be built more effectively.

References

1. AlchemyAPI: http://www.alchemyapi.com (2010)
2. Bickmore, T., Cassell, J.: Relational agents: a model and implementation of building user trust. In: Proceedings of the SIGCHI conference on Human factors in computing systems. pp. 396–403. ACM (2001)
3. Bickmore, T., Pfeifer, L., Schulman, D.: Relational Agents Improve Engagement and Learning in Science Museum Visitors, pp. 55–67 (2011)
4. Bickmore, T.W., Pfeifer, L.M., Jack, B.W.: Taking the time to care: empowering low health literacy hospital patients with virtual nurse agents. In: Proceedings of the SIGCHI conference on human factors in computing systems. pp. 1265–1274. ACM (2009)
5. Bickmore, W.T., Utami, D., Matsuyama, R., Paasche-Orlow, K.M.: Improving access to online health information with conversational agents: A randomized controlled experiment. J Med Internet Res 18(1), e1 (Jan 2016)
6. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. Journal of machine Learning research 3(Jan), 993–1022 (2003)
7. Cassell, J.: Embodied conversational interface agents. Communications of the ACM 43(4), 70–78 (2000)
8. Ji, Z., Lu, Z., Li, H.: An information retrieval approach to short text conversation. arXiv preprint arXiv:1408.6988 (2014)
9. Kim, Y.: Convolutional neural networks for sentence classification. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014. pp. 1746–1751 (2014)
10. Kiseleva, J., Williams, K., Jiang, J., Hassan Awadallah, A., Crook, A.C., Zitouni, I., Anastasakos, T.: Understanding user satisfaction with intelligent assistants. In: Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval. pp. 121–130. ACM (2016)
11. Lave, J., Wenger, E.: Situated Learning. Legitimate Peripheral Participation. Cambridge University Press, New York, Port Chester, Melbourne, Sydney (1991)
12. Liao, Q.V., Davis, M., Geyer, W., Muller, M., Shami, N.S.: What can you do?: Studying social-agent orientation and agent proactive interactions with an agent for employees. In: Proceedings of the 2016 ACM Conference on Designing Interactive Systems. pp. 264–275. ACM (2016)
13. Metzler, D., Croft, W.B.: A markov random field model for term dependencies. In: Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval. pp. 472–479. ACM (2005)
14. Miller, V.D., Jablin, F.M.: Information seeking during organizational entry: Influences, tactics, and a model of the process. The Academy of Management Review 16(1), 92–120 (1991)
15. Pickard, M.D., Burns, M.B., Moffitt, K.C.: A theoretical justification for using embodied conversational agents (ecas) to augment accounting-related interviews. Journal of Information Systems 27(2), 159–176 (2013)
16. Ritter, A., Cherry, C., Dolan, W.B.: Data-driven response generation in social media. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing. pp. 583–593. EMNLP '11, Association for Computational Linguistics, Stroudsburg, PA, USA (2011), http://dl.acm.org/citation.cfm?id=2145432.2145500

17. Shang, L., Lu, Z., Li, H.: Neural responding machine for short-text conversation. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics,ACL 2015. pp. 1577–1586 (2015)
18. Shechtman, N., Horowitz, L.M.: Media inequality in conversation: how people behave differently when interacting with computers and people. In: Proceedings of the SIGCHI conference on Human factors in computing systems. pp. 281–288. ACM (2003)
19. Silvervarg, A., Jönsson, A.: Subjective and objective evaluation of conversational agents in learning environments for young teenagers. In: Proceedings of the 7th IJCAI Workshop on Knowledge and Reasoning in Practical Dialogue Systems (2011)
20. Silvervarg, A., JÃűnsson, A.: A.: Subjective and objective evaluation of conversational agents in learning environments for young teenagers. In: In: 7th IJCAI Workshop on Knowledge and Reasoning in Practical Dialogue Systems (2011)
21. Strohman, T., Metzler, D., Turtle, H., Croft, W.B.: Indri: A language model-based search engine for complex queries. In: Proceedings of the International Conference on Intelligent Analysis. vol. 2, pp. 2–6. Citeseer (2005)
22. Tegos, S., Demetriadis, S.: Conversational agents improve peer learning through building on prior knowledge. Journal of Educational Technology & Society 20(1), 99–111 (2017)
23. Trappl, R.: Your Virtual Butler. Springer (2013)
24. Traum, D., DeVault, D., Lee, J., Wang, Z., Marsella, S.: Incremental dialogue understanding and feedback for multiparty, multimodal conversation. In: International Conference on Intelligent Virtual Agents. pp. 275–288. Springer (2012)
25. Turing, A.M.: Computing machinery and intelligence. Mind 59(236), 433–460 (1950)
26. Vinyals, O., Le, Q.: A neural conversational model. arXiv preprint arXiv:1506.05869 (2015)
27. Walker, M.A., Litman, D.J., Kamm, C.A., Abella, A.: Paradise: A framework for evaluating spoken dialogue agents. In: Proceedings of the eighth conference on European chapter of the Association for Computational Linguistics. pp. 271–280. Association for Computational Linguistics (1997)
28. Wang, Z., Lee, J., Marsella, S.: Towards more comprehensive listening behavior: beyond the bobble head. In: International Workshop on Intelligent Virtual Agents. pp. 216–227. Springer (2011)
29. Williams, J., Raux, A., Ramachandran, D., Black, A.: The dialog state tracking challenge. In: Proceedings of the SIGDIAL 2013 Conference. pp. 404–413 (2013)
30. Xiao, J., Stasko, J., Catrambone, R.: An empirical study of the effect of agent competence on user performance and perception. In: Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems-Volume 1. pp. 178–185. IEEE Computer Society (2004)
31. Yao, M.: 4 critical steps to maximize chatbot retention & engagement (Dec 2016), http://www.topbots.com/4-critical-steps-to-maximize-chatbot-retention-engagement/