Towards A Six-Level Framework of Emotional Intelligence for Customer Service Chatbots

Qicheng Ding[1], V. Ivan Sanchez Carmona[2], Mingming Liu[1], Fangzhen Peng[1], Yu Zhang[1], Changjian Hu[2]

[1]Technical Strategy and Innovation Platform, Lenovo Research, China

[2]AI Lab, Lenovo Research, China

dingqc1, vcarmona, liumm9, pengfz1, zhangyu29, hucj1@lenovo.com

Abstract

A chatbot with emotional intelligence can improve the interaction experience and efficiency with human users. However, there is no standard scheme or methods to evaluate the degree of emotional intelligence of chatbots. Based on our industrial practice, this paper aims to propose a preliminary framework to evaluate to which extent a chatbot is emotionally intelligent. We divide it into two parts: emotional understanding and emotional strategy. Each part contains several factors that are important to emotional intelligence. Furthermore, we propose a six-level scheme as a guideline to evaluate the emotional intelligence. This paper provides valuable suggestion in the entire lifecycle of a chatbot including design, implementation, evaluation, and maintenance.

1 Introduction

Emotional intelligence is a fundamental part of the overall intelligence of a human being. Emotions mediate the cognitive processes related to individual intelligence. In other words, a person's capacities such as memory, decision-making, and reasoning are affected by his or her emotional states. Furthermore, emotions also affect social interactions. It is easier to achieve a consensus when group members are in pleasant moods rather than bad ones. To achieve these pleasant moods, it may be necessary for group members to identify and manage emotions, abilities that are the building blocks of emotional intelligence.

Chatbots, which refer to dialog systems that aim to solve task-oriented problems or provide social interactions, are always inevitably embedded in human-computer interactions. Chatbots with emotional intelligence can correctly identify users' emotions and provide concise responses, which represent a stronger interaction ability and better user satisfaction. For example, when a user is unsatisfied with the chatbot, the user may not speak out, but his face will clearly show frustration. Ideally, the chatbot would make a parenthesis to let the user know that his frustration has been noticed, and that such an emotional expression has triggered a request to verify that he is understanding the information clearly. This ideal solution arises from a chatbot with emotional intelligence.

Chatbots that have no model of emotional intelligence may be at risk of jeopardizing the whole interaction since different users may have different preferences for the chatbot's behavior. For example, some users may prefer a repetitive chatbot that frequently rephrases its replies in different ways for the sake of clarity over a chatbot that never repeats information; some users may feel comfortable interacting with a succinct chatbot that formulates statements in as short and concise sentences as possible; some users may prefer a chatbot that is constantly checking whether the user is still engaged in the interaction and is understanding every reply from the chatbot. Emotional indicators such as facial gestures, voice tone, body gesture, or specific sentences are natural indicators of confusion, boredom, frustration, and other emotions that show the dissatisfaction from the user. Chatbots failing to perceive, attend, and act upon these emotional signals may lead the user to a stressful situation, and to waste a lot of time to solve a problem, or even worse, to abruptly quit the session and leave with problem unsolved.

In this paper, we propose a 6-level framework of emotional intelligence for chatbots. This framework aims to characterize the abilities required by chatbots on different levels of emotional intelligence. To do so, we first provide a model of emotional intelligence for chatbots, which we build based on the human-centric model of emotional intelligence of (Salovey & Mayer, 1990; Salovey P. G., 2005; Mayer J. D., 1999; Mayer J. D., 2016). Similar to the work of Salovey, we propose a two-branch model of emotional

intelligence where each of the branches correspond to broad abilities that we believe chatbots should possess, namely Emotional Understanding and Emotional Strategy. The first branch corresponds to the ability of chatbots of perceiving and recognizing emotional indicators from the users; the second branch corresponds to the ability of formulating and carrying out a strategy to deal with the emotional states of users in order to successfully solve the target task the user requires the chatbot to solve.

Chatbots can then be classified based on our 6-level framework. For example, while a *null-intelligent* chatbot that falls at the Level 0 in our framework works with no capabilities to handle emotions, a *highly-intelligent* (Level 3) chatbot can perceive, recognize and respond to users' emotions by integrating information across multiple channels (such as text, voice, images from video camera) during the entire conversation while carrying out strategies to handle the users' emotions to solve the target task. At the top of our framework we categorize a chatbot as *fully-intelligent* if it can fully understand, remember, predict and respond to users' emotions by integrating information from all visual, auditory and tactile channels while carrying out the appropriate strategy to both keep the user satisfied and successfully solve the target task.

Our 6-level framework can benefit in advancing research in chatbots in different perspectives. For example, designers and engineers can use our framework to check the abilities of their chatbots; researchers can build roadmaps to guide future research directions for emotional intelligence in chatbots based on our definition of intelligence levels; industry developers can use our emotional intelligence model to visualize the technology required for chatbots to perceive and recognize the emotional states of users.

2 Related Work

The lack of objective and automatic measures of emotional intelligence is one of the major limitations across diverse fields (Miners, Côté, & Lievens, 2017). In the field of Psychology, a number of researchers have attempted to develop self-reported scales of emotional intelligence; however, their emotional intelligence definitions are often inconsistent with each other (Bar-On, 1997; Roger & Najarian, 1989; Salovey & Mayer, 1990; Schutte, et al., 1998).

Mayer et al., (Mayer, Caruso, & Salovey, 1999) constructed a Multi-factor Emotional Intelligence Scale (MEIS) that is objective (has correct answers), reliable and less associated with personality. After that, they established a more reliable scale named the Mayer-Salovey-Caruso Emotional Intelligence Test, Version 2.0 (MSCEIT, V2.0) (Mayer, Salovey, & Caruso, 2002). MSCEIT is one of the most widespread measure of emotional intelligence (Fiori, et al., 2014; Papadopoulos, Gkintoni, Halkiopoulos, & Antonopoulou, 2018). For example, Brackett et al. used MSCEIT, V2.0 and found out that emotional intelligence in males was associated with everyday behavior (Brackett, Mayer, & Warner, 2004); Rosete & Ciarrochi argue that higher emotional intelligence prompted workplace performance (Rosete & Ciarrochi, 2005).

In the field of Artificial Intelligence, efforts have been made to both standardize (Sedoc, et al., 2019; Miller, et al., 2018) and improve (Sedoc & Ungar, 2020) the evaluation of chatbots' dialogue capabilities using both human judgments and automatic metrics.

Also, research in the field of Human-Computer Interaction has pinpointed the importance of AI applications to match social norms while avoiding social biases (Amershi, et al., 2019). Moreover, it has been discussed the importance of emotional intelligence in chatbots to better engage users into a more fruitful conversation (Chaves & Gerosa, 2019).

However, to our best knowledge, there is a lack of standard methods to evaluate the AI-based implementation (e.g., call center chatbot) from an emotional intelligence perspective. Therefore, this paper aims to provide a preliminary guideline to support stakeholders to better understand and evaluate how emotionally intelligent their AI products are.

3 Emotional Intelligence

We stick with the definition of emotional intelligence from the work of Mayer and Salovey (Salovey & Mayer, 1990; Mayer J. D., 2016), due to its wide acceptance on both disciplines Psychology and Affective Computing (Picard, 2000). According to Mayer and Salovey, emotional intelligence is characterized across four branches: 1) perception and expression of emotion, 2) using emotion to facilitate thought, 3) understanding emotions, and 4) managing emotions.

The first branch, *perception and expression of emotion*, refers to the ability of recognizing one's own emotions and to accurately display them through different means such as physical actions. This ability extends to recognizing other's emotions also through different means such as language, facial expressions, and behavior.

The second branch, *using emotion to facilitate thought*, refers to the ability of using emotions to leverage cognitive processes, such as planning, reasoning, making decisions, and so on. Moreover, this branch also includes the ability to adequately choose problems according to one's own emotional state in order to take advantage of such a state; for example, happiness may stimulate a person's creative thinking, so choosing to solve problems that require creativity when one is happy may be deemed as an intelligent decision.

The third branch, *understanding emotions*, refers to the knowledge that a person has about different types of emotions: how they may develop throughout time, how they may

interact with each other and what the possible causes and effects of emotions are.

The last branch, *managing emotions*, is the ability to self-manage one's own emotions, as well as to manage the emotions of other people, in order to achieve a desired goal. For example, when a dad is upset with his kid due to an improper behavior, he knows that to have an effective conversation with his kid he should first manage his own emotions and elicit a more calm and peaceful emotional state; moreover, if the kid is crying, the dad will try to comfort him in order to have a fruitful conversation.

Important to notice is that each of the branches in this framework are subject to social and cultural backgrounds; variations in emotional indicators across different cultures must be acknowledged and respected. For example, an overt smile may be interpreted with different levels of happiness (such as happy vs. very happy) in different cultures. Thus, for an individual to be emotionally intelligent, she must be aware of and understand the social and cultural context.

Branch name	**Dimension**	**Description**
Emotional understanding	Breadth	The range of emotions the bot can identify.
	Variety	The number of channels of information the bot can utilize to understand emotions.
	Contextual-wise	The contextual information that the bot can utilize to understand emotions, e.g., from one conversation turn to the entire conversation.
Emotional strategy	Breadth	The range of emotions that the bot can deal with, e.g., from dealing with one emotion to seven emotions.
	Variety	The different ways in which the bot can deal with one type of emotion, e.g., from pure a text-based strategy to a multi-media strategy.
	Contextual-wise	The contextual information that the bot can utilize to deal with emotions, e.g., from one conversation turn to the entire conversation.
	Social norm	How well the bot's strategy fit into the customer's social and cultural background and avoid biased response, e.g., from being unaware of any customer privacy to deal with emotions with proper privacy consideration.

Table 1 Two-branch model of emotional intelligence for chatbots. Each branch decomposes along different dimensions which characterize in a finer-grained detail the abilities required for chatbots.

4 Chatbot-Centric Model of Emotional Intelligence

Based on the four-branch model from Mayer and Salovey, we propose a two-branch model of emotional intelligence for chatbots where we call these two branches as Emotional Understanding and Emotional Strategy. While the branch of emotional understanding refers to the ability of a chatbot of understanding users' emotions, the branch of emotional strategy focuses on the capability of a chatbot in dealing with the users' emotions. These two branches are characterized across several dimensions, as is shown in Table 1.

4.1 Emotional Understanding (EU)

Emotional understanding refers to chatbots' ability in understanding customers' emotions while solving their issues. Chatbots need to know customers' emotional status before acting properly. It includes the number/range of emotions chatbots can understand (breadth), the number of channels of information (variety) and the contextual information chatbots can utilize (contextual-wise).

4.1.1 Breadth

The breadth of emotional understanding refers to the range of emotions that a chatbot can identify. Generally, researchers treat emotions as either discrete (Ekman, 1994) or dimensional (Mehrabian, 1980), and they both have pros and cons (Gunes, 2016). We regard both classification methods as a suitable choice to define the breadth of emotions to be recognized by chatbots since we believe both methods can accurately capture users' emotions. To characterize the breadth of emotions that a chatbot should identify, we propose a simple logic: the more types of emotions a chatbot can identify the smarter it is and thus the higher in our scale it is positioned. Table 2 shows this logic. This scale can be effectively taken by stakeholders as a criterion to qualitatively evaluate the level of emotional intelligence of a chatbot in terms of emotion identification.

Level	Capacity	Description
0	No	Cannot understand user's emotion.
1	Basic	Can understand limited (one or two) emotions.
2	Partial	Can understand several emotions.
3	High	Can understand major emotions, e.g., Ekman's seven emotions

		(Ekman, 1994) or PAD model (Walter, 2011).
4	Very High	Can understand most emotions.
5	Full	Can fully understand emotions.

Table 2. Six-Level Scheme for Breadth of EU.

4.1.2 Variety

Variety refers to the number of channels of information a chatbot can utilize to perceive and recognize users' emotions. Similar to our logic in Section 4.1.1, the more channels a chatbot uses to identify customers' emotions, the more intelligent can be deemed. Thus, the main challenge remains in integrating the information across the different channels rather than merely augmenting the number of channels. Nevertheless, the more channels are used, the more the privacy concerns to be considered: a multi-media chatbot does not mean that it can open and use all capable channels every time; when a customer expresses his preferred channel(s), the chatbot should adjust its variety to fit the customer's choice. Table 3 shows the degrees of complexity in handling different channel types as the number of channels increment.

Level	Capacity	Description
0	No	No usable channel.
1	Basic	Can use one channel to understand emotion e.g., text, audio etc.
2	Partial	Can use multiple channels separately but cannot integrate the information together.
3	High	Can integrate multiple channels, i.e., can combine multi-channel emotional information together.
4	Very High	Can acquire and integrate multi-channel emotional information; also, can distribute different weights for each channel.
5	Full	Can use all channels (visual, auditory, tactile etc.) flexibly and properly to understand emotion.

Table 3. Six-Level Scheme for Variety of EU.

4.1.3 Contextual-wise

Contextual-wise refers to the contextual information that the chatbot can utilize to understand emotions, e.g., from using a single turn from the conversation to using the entire conversation. The contextual information the chatbot can use include, but is not limited to: conversation history, user profile, the interaction with other customers, etc. Table 4 shows the incremental capability of a chatbot in using contextual information for understanding users' emotions. While a chatbot with a basic ability can extract emotional indicators from a single turn at a time, a highly capable chatbot is able to use the whole interaction with the user to extract and identify emotional states from the user. In this way, a Level 1 chatbot may recognize that a user is angry at the current turn, but it will not be able to use this information to build up a global model of the user's emotional state; however, a Level 4 chatbot is not only able to build such a model, but it can also recall information from previous interactions to build the emotional model in a finer-grained way.

Level	Capacity	Description
0	No	No continuity.
1	Basic	Can identify emotion from the ongoing conversation turn.
2	Partial	Can continually identify emotion from multiple turns.
3	High	Can continually identify emotion from the entire ongoing conversation.
4	Very High	Can continually identify and remember emotion from the entire ongoing and historical conversation.
5	Full	Can fully identify, remember emotion as well as anticipate emotion from the entire ongoing and historical conversation.

Table 4. Six-Level Scheme for Contextual-wise of EU.

4.2 Emotional Strategy (ES)

Emotional strategy describes the chatbots' ability to deal with customer emotions. It includes the range of emotions (breadth), the expression channels (variety) and the contextual information it can use, and the ability to fit into the social and cultural background (social norm).

4.2.1 Breadth

Breadth in emotional strategy refers to the range of emotions that a chatbot can respond to, e.g., from one emotion to seven emotions. Table 5 shows our characterization of this dimension across varying levels of complexity. For example, a chatbot capable of dealing with customers' major emotions can be deemed as very-highly intelligent.

Level	Capacity	Description
0	No	Cannot deal with emotion.
1	Basic	Can deal with up to two emotions.
2	Partial	Can deal with several emotions.
3	High	Can deal with major emotions.
4	Very High	Can deal with most emotions.
5	Full	Can fully deal with emotions.

Table 5. Six-Level Scheme for Breadth of ES.

4.2.2 Variety

Variety refers to the different ways in which a chatbot can deal with one type of emotion, e.g., from a pure text-based strategy to a multi-media strategy. Here we define two types of strategy: emotion-focused and task-focused. An emotion-focused strategy refers to a chatbot dealing only with the user's emotion without progressing on the target task (e.g., fixing a broken smartphone). A task-focused strategy, on the other hand, refers to a chatbot dealing with the user's emotion and problem at the same time. We notice that while a chit-chat chatbot only requires an emotional-focused strategy to deal with the user's emotions, a customer service chatbot should be able to solve the user's problem while handling his or her emotions; in other words, for a task-oriented chatbot to be considered emotionally intelligent it should be capable of carrying out a task-focused strategy. Table 6 shows our characterization of this dimension across levels of intelligence. Thus, according to this characterization, a customer service chatbot should be at least at Level 3 in order to effectively embody an emotional intelligence model of the user by leveraging the user's emotion in order to solve the task at hand.

Level	Capacity	Description
0	No	No strategy.
1	Basic	Can use one channel to deal with emotion by of one or two emotion-focused strategies.
2	Partial	Can use multiple channels separately to deal with emotion by several emotion-focused strategies.
3	High	Can use multiple channels together to deal with emotion by varied emotion-focused strategies and be able to generate task-focused strategies occasionally.
4	Very High	Can use multiple channels flexibly to deal with emotion with varied emotion-focused and task-focused strategies leveraging current condition and history.
5	Full	Can use all channels properly to deal with emotion leveraging current condition and history.

Table 6. Six-Level Scheme for Variety of ES.

4.2.3 Contextual-wise

It refers to the contextual information that the chatbot can utilize to deal with emotions, e.g., from using one conversation turn to using the entire conversation (see Table 7).

Level	Capacity	Description
0	No	No continuity.
1	Basic	Can deal with emotion from the ongoing conversation turn.
2	Partial	Can continually deal with emotion from multiple turns.
3	High	Can continually deal with emotion from the entire ongoing conversation.
4	Very High	Can continually deal with emotion from the entire ongoing and historical conversation.
5	Full	Can fully deal with emotion from the entire ongoing, historical and predicted conversation.

Table 7. Six-Level Scheme for Contextual-wise of ES.

4.2.4 Social Norm

Social norm refers to how well a chatbot's strategy fit into the customer's social and cultural background, and how well it avoids biased responses: from not considering any socio-demographic, ethnicity and racial factor to fully and world-wide modeling all sensitive factors by which a user may be identified. Based on the premise that a chatbot should both respect the uniqueness of each type of background and avoid any offensive behavior, we propose our characterization of the social norm dimension across increasing levels of social knowledge as shown in Table 8.

Level	Capacity	Description
0	No	No consideration of social and cultural background.
1	Basic	Can consider limited social and cultural background, e.g. avoiding mocking/insulting customer.
2	Partial	Can consider one or two major social and cultural backgrounds (e.g., Western/Eastern cultures).
3	High	Can consider common social and cultural background world widely.
4	Very High	Can consider most social and cultural backgrounds world widely.
5	Full	Can fully consider all social and cultural backgrounds.

Table 8. Six-Level Scheme for Social Norm of ES.

5 Six-Level Framework of Emotional Intelligence for Chatbots

Table 9 shows a summary of our 6-level framework of emotional intelligence. This scale ranges from Level 0, where a chatbot possesses no ability related to emotional intelligence, to Level 5, where chatbots are said to be fully emotionally intelligent. The overall logic underlying our 6-level characterization is that the smarter a chatbot is, the more kinds of emotions it can understand, the more information it can utilize (multiple turns, multiple channels etc.), and the

more proper strategies it can generate (from just emotion-focused strategies to task-focused strategies).

Level	Emotional Capacity	Description
0	No	Cannot understand nor deal with customer emotion.
1	Basic	Can understand and respond to a limited number of emotion types expressed through one channel type in a one-turn basis, and deal with it by a limited emotion-focused strategy.
2	Partial	Can understand and respond to several emotion types expressed through multiple channel types separately across multiple turns, and deal with it by several emotion-focused strategies.
3	High	Can understand and respond to major emotion types by integrating information from multiple channel types during the entire conversation, and deal with it by varied emotion-focused strategies while being able to generate task-focused strategies under some conditions.
4	Very High	Can understand, remember and respond to most emotion types by integrating information from multiple channel types during the entire ongoing and historical conversation, and deal with it flexibly leveraging current condition and history.
5	Full	Can fully understand, remember, predict and respond to all emotion types by integrating information from all channels, and deal with it properly leveraging current condition and history.

Table 9. Summary of the 6-Level Scheme. The descriptions with keywords "understand"," predict" and "remember" are related to the Emotion Understanding branch; those with keywords "deal with" and "strategy" correspond to the Emotion Strategy branch.

We refrain from proposing specific channel or emotion types for each level of intelligence, as well as specific tasks and metrics to evaluate the accuracy of chatbots' abilities, since these aspects may be culturally specific to the target regions where chatbots are to be deployed and may be better defined by the chatbots' developers and designers.

Level 0 includes chatbots which are neither equipped with any ability to perceive, recognize or identify any emotional signal from a user such as text keywords showing frustration- or anger-related emotions, nor capable of executing any strategy to attend the emotional state of the user by, for example, identifying what triggered the user's emotion and what actions may alleviate this emotional state (if required).

Chatbots at Level 1, on the other hand, are considered as *basic-intelligent* chatbots since they possess basic abilities to produce the corresponding behavior that takes into account one or two user's emotions such as anger or happiness. This behavior is carried out through a single channel type in both directions to perceive the emotional signals from the user and to respond to them. Moreover, chatbots Level 1 can only model users' emotions in a single-turn basis; this implies that they are not able to aggregate emotional signals throughout the whole conversation leading to temporal or local models of emotion which are independent of each other across turns. Nevertheless, these chatbots are able to constrain their responses based on a simple model of social norms according to the region they are deployed at; these responses are constrained to avoid the most common or simple social- or cultural-related biases such as gender biases. Even though this type of chatbot can build a very basic model of emotions, it cannot leverage this model to solve the target task the user expects the chatbot to solve.

Moving one level above in our scale, we find the category of *partially-intelligent* chatbots—Level 2—which can be seen as a more advanced version of *basic-intelligent* chatbots. The number of emotion types which Level 2 chatbots can identify and respond to increases by one or two orders of magnitude, i.e. they can deal with up to 3 or 4 (even 5) types of emotion. Dealing with users' emotions can be done through multiple channel types independently of each other; for example, while the visual channel is dedicated exclusively to deal with happiness-related emotions, the text channel is dedicated to distressful-related emotions; the pairing of channel type with emotion type is selected based on the feasibility to capture emotional patterns (it may be easier to capture certain types of emotion by text patterns than by visual cues). The information captured through the channels cannot be integrated or aggregated into a single emotional model, however, by the chatbots which are thus required to possess different emotional strategies to deal with each type of emotion separately. These strategies are intended to modulate the users' emotions to keep them in a stable or neutral point that can lead to a maximum customer satisfaction. However, the emotional strategies are carried out independently of the target task to be solved which may lead chatbots to be seen as partially intelligent by apologizing whenever a facial expression denoting anger from the user is recognized or by displaying grateful messages such as *"thank you for patiently waiting"* whenever keywords signaling distress from the user are read; these emotional signals may not be related to usability or performance factors such as a slow reply from the chatbot or a wrong interpretation of the user's query, they may just be due to the user's mood, but chatbots at this level have no capacity to

integrate the user's emotional state with the state of the task. In addition, a key feature of Level 2 chatbots is the ability to cope with social norms; their behavior faithfully aligns to the norms of one or two cultures avoiding thus any social, cultural or gender bias from the geographical region they are deployed at.

At the next level in our hierarchy we observe an inflection point in the emotional intelligence of chatbots. Level 3 chatbots not only increase in the range of emotion types, channel types and the scope of the context information from which to extract emotional signals (the entire conversation with the user) with respect to chatbots at Level 1 or 2, but more importantly, they are able to integrate information and inferences into a global emotional model of the user that can be further integrated with the task's model. Emotional signals from diverse channels can be aggregated to resolve the emotional state of the user; for example, visual cues can be used to infer an emotion type which may differ from the emotion type inferred from text or vocal signals, but the chatbot is able to resolve any conflict and to decide what is the most likely emotional state. Moreover, chatbots can align this inference to the state of the task in order to resolve further possible conflicts such as the causes of the user's emotion; for example, resolving that the main user's emotion is distress requires the chatbot to resolve afterwards if usability factors, such as taking too long or proposing too complex strategies to solve the user's problem, are the determinants of this emotional state.

Resolving the conflicts described above is done while considering the entire conversation with the user. Thus, another degree of freedom to be resolved by the chatbot is the continuity of determinants of emotion across the conversation; for example, the chatbot may infer distress as the user's emotional state from the initial states of the interaction up to the current turns due to the perceived emotional signals; thus the chatbot may resolve that distress is both the user's emotional state and the baseline mood of the user since it was identified from the initial turns and it was maintained throughout the conversation; in this way, the chatbot discards distress being an emotional state caused by factors from the interaction; however, while the user's baseline mood may be distress, he may indeed be induced to a strengthen distress state by factors related to the chatbot. Chatbots at this level can accurately resolve this type of problems and thus recognizing that the user's baseline mood is distress while some interaction factors are further distressing him.

Furthermore, this type of chatbot is able to generate and carry out task-focused strategies that leverage the emotional state of the user while considering the determinants of this state to solve the target problem; for example, if the chatbot is not able to understand the user's query after several attempts, and the user's emotion turns to anger, then the chatbot should first modulate the user's emotion by, for instance, displaying an apology message, and then stop asking the user to reformulate her query and change to a strategy that may further alleviate the user's emotional state while properly solving the target task, such as displaying a list of queries posed by previous users to let the current user manually select the closest one.

Finally, another useful feature of *highly-intelligent* chatbots is the scope of the model of social norms; these chatbots can conform to most of the common social and cultural rules across the world.

We consider this level in our framework to be the baseline level of emotional intelligence; chatbots with the characteristics described above better align to human-centric frameworks of emotional intelligence in psychology than chatbots at lower levels by not only perceiving and identifying users' emotions but by applying the corresponding strategies based on such emotions to solve the target problem.

The next category of chatbots in our framework are v*ery-high intelligent* chatbots which elaborate on top of the skills of *highly-intelligent* chatbots. The number of emotion types to be handled increases as well as the number of social and cultural rules to take into account. Furthermore, Level 4 chatbots cannot only model the user's emotions based on the entire ongoing conversation, they can use previous conversations to resolve the user's emotional state; for example, taking the case posed above, if across the entire ongoing conversation the chatbot is able to capture emotional dependencies that when aggregated lead the chatbot to resolve for distress as the emotional state of the user, the chatbot still needs to resolve for the emotional determinants such as user's baseline mood or usability factors; thus, the chatbot can consult emotional models from previous conversations to compare to what extent usability factors, for instance, lead to the observed emotional dependencies.

At the top of our framework, we propose the category of *fully-intelligent* chatbots which represent a second point of inflexion. Chatbots at Level 5 can handle all types of emotion across all world regions while respecting all social and cultural norms of each region. Furthermore, they can handle emotions by means of any channel type. This type of chatbot can resolve a user's emotional state by extracting emotional dependencies from the current conversation and from previous conversations archived. Also, it can resolve for the factors that influenced or induced the user's emotional state while trying to both alleviate such emotional state (if required, such as in the case of anger) and optimally solving the user's problem. Moreover, Level 5 chatbots are able to predict the possible emotional state resulting from applying specific emotional- and task-focused strategies; for example, if the chatbot resolves for anger as the current emotional state of the user where time was the factor that influenced the user to achieve this state, the chatbot then will try to choose for both emotional- and task-focused strategies that will alleviate the anger while reducing the time to solve the problem; therefore, choosing or generating strategies that optimize these two variables (reducing both anger and time) is resolved, indeed, as an optimization problem where the chatbot can predict the structures and features of the

emotional- and task-focused strategies that maximize the probability of success. In this way, we conceptualize Level 5 chatbots as ideal dialogue systems. Thus, it may be tempting to compare the emotional capacities of this type of chatbots to that from trained humans in customer service; but such a comparison implies a great complexity and we leave this task as future work.

6 Discussion

This paper aims to provide valuable suggestions regarding emotional intelligence in the entire lifecycle of a chatbot including design, implementation, evaluation, and maintenance. At the design stage, this paper could support stakeholders to draw a feasible blueprint the industry based on factors derived from our framework. During the implementation and evaluation stage, our framework could be used as a look-up table to assess the current intelligence level of a target chatbot to guide the overall development schedule. At maintenance stage, after the launch of a target chatbot, this paper could continually support to offer insights about next-generation design.

Apart from the industrial areas, this paper could also inspire academic researchers to develop a quantitative toolkit to evaluate the overall intelligence performance of chatbots, as well as to tackle the difficulties that stop us from achieving smarter chatbots.

Even though this scale is not in its final version, as it will nurture from future research from the AI and Psychology fields, it is self-contained, and it comes from our vision after working at the trenches deploying chatbots.

In the future, we will include more factors that contribute to the emotional intelligence of chatbots; for example, we will present a quantitative method to evaluate each factor's performance. In addition, a case study will be performed to illustrate how to use our framework to get insights about intelligent chatbot design and development.

References

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., . . . Horvitz, E. (2019). Guidelines for Human-AI Interaction. *CHI '19: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems.* Association for Computing Machinery.

Bar-On, R. (1997). *The emotional intelligence inventory (EQ-I): Technical manual.* Toronto, Canada: Multi-Health Systems.

Brackett, M. A., Mayer, J. D., & Warner, R. M. (2004). Emotional intelligence and its relation to everyday behaviour. *Personality and Individual Differences,, 36*(6), pp. 1387-1402.

Chaves, A. P., & Gerosa, M. A. (2019). How should my chatbot interact? A survey on human-chatbot interaction design. *arXiv:1904.02743.* arXiv.

Ekman, P. E. (1994). *The nature of emotion: Fundamental questions.* Oxford University Press.

Fiori, M., Antonietti, J.-P., Mikolajczak, M., Luminet, O., Hansenne, M., & Rossier, J. (2014). What is the ability emotional intelligence test (MSCEIT) good for? An evaluation using item response theory. *PLoS One, 9*(6), p. e98827.

Gunes, H. a. (2016). Is automatic facial expression recognition of emotions coming to a dead end? The rise of the new kids on the block. *Image Vis. Comput., 55*(1), 6-8.

Mayer, J. D. (1999). Emotional intel-ligence meets traditional standards for an intelligence. *Intelligence*, pp. 267-298.

Mayer, J. D. (2002). *Mayer-Salovey-Caruso emotional intelligence test (MSCEIT) item booklet.* MHS Assessments.

Mayer, J. D. (2016). The ability model of emotional intelligence: Principles and updates. *Emotion review, 8*(4), 290--300.

Mayer, J. D. (2016). The ability model of emotional intelligence: Principles and updates. *Emotion Review*, 290-300.

Mehrabian, A. (1980). *Basic dimensions for a general psychological theory: Implications for personality, social, environmental, and developmental studies.* Oelgeschlager, Gunn \& Hain Cambridge, MA.

Miller, A., Feng, W., Fisch, A., Lu, J., Batra, D., Bordes, A., . . . Weston, J. (2018). ParlAI: A Dialog Research Software Platform. *arXiv:1705.06476.* arXiv.

Miners, C. T., Côté, S., & Lievens, F. (2017). Assessing the Validity of Emotional Intelligence Measures. *Emotion Review, 1*(9).

Papadopoulos, D., Gkintoni, E., Halkiopoulos, C., & Antonopoulou, H. (2018). A Computational Approach of Consumer Decision Making Process Emotional Intelligence and Their Effects on E-Marketing. *6th International Conference on Contemporary Marketing Issues (ICCMI)(2018: Athens, Greece) 6th International Conference on Contemporary Marketing Issues (ICCMI): June 27-29, 2018, Athens, Greece/Co-Organized by Alexander Technological Educational Institu*, (p. 166).

Picard, W. R. (2000). *Affective Computing.* MIT Press.

Roger, D., & Najarian, B. (1989). The construction and validation of a new scale for measuring emotion control. *Personality and Individual Differences, 10*(8), pp. 845-853.

Rosete, D., & Ciarrochi, J. (2005). Emotional intelligence and its relationship to workplace performance outcomes of leadership effectiveness. . *Leadership & Organization Development Journal.*

Salovey, P. G. (2005). The Science of Emotional Intelligence. *Current Directions in Psychological Science*, pp. 281-285.

Salovey, P., & Mayer, J. D. (1990). Emotional intelligence. *Imagination, Cognition and Personality, 9*(3), pp. 185--211.

Schutte, N. S., Malouff, J. M., Hall, L. E., Haggerty, D. J., Cooper, J. T., Golden, C. J., & Dornheim, L. (1998). Development and validation of a measure of emotional intelligence. *Personality and Individual Differences, 25*(2), pp. 167-177.

Sedoc, J., & Ungar, L. (2020). Item Response Theory for Efficient Human Evaluation of Chatbots. *Proceedings of the First Workshop on Evaluation and Comparison of NLP Systems* (pp. 21-33). Association for Computational Linguistics.

Sedoc, J., Ippolito, D., Kirubarajan, A., Thirani, J., Un-gar, L., & Callison-Burch, C. (2019). ChatEval: A Tool for Chatbot Evaluation. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (Demonstrations)* (pp. 60-65). Association for Computational Linguistics.

Walter, S. a. (2011). Multimodal emotion classification in naturalistic user behavior. *International Conference on Human-Computer Interaction* (pp. 603--611). Springer.