An Approach Towards Ethical Chatbots in Customer Service

Abeer Dyoub[1], Stefania Costantini[1], and Francesca A. Lisi[2]

[1] Dipartimento di Ingegneria e Scienze dell'Informazione e Matematica
Università degli Studi dell'Aquila, Italy
Abeer.Dyoub@graduate.univaq.it, Stefania.Costantini@univaq.it
[2] Dipartimento di Informatica &
Centro Interdipartimentale di Logica e Applicazioni (CILA)
Università degli Studi di Bari "Aldo Moro", Italy
FrancescaAlessandra.Lisi@uniba.it

Abstract. Chatbot is an artificial intelligent software which can simulate a conversation with a user in natural language via auditory or textual methods. Businesses are rapidly moving towards the need for chatbots. However chatbots raise many ethical concerns. To ensure that they behave ethically, their behavior should be guided by the codes of ethics and conduct of their company.

1 Introduction

Chatbots are some of the industry's newest tools designed to simplify the interaction between humans and computers. They are typically used in dialogue systems for various practical purposes including customer service or information acquisition. They are often described as one of the most advanced and promising expressions of interactions between humans and machines.

Machine ethics [11] is a new evolving field concerned with the moral behavior of artificial intelligent machines. The problem of adopting ethical approach to AI has been attracting a lot of attention in the last few years, see e.g. [8]. Unethical AI and bots are big concern for many consumers. Businesses who recently deployed a chatbot or are in the process of designing one should think carefully about the ethical considerations for their bots. The chatbot should be built on ethical foundations, because its behavior influence the company's image, and unethical behavior will lead to mistrust from client side. Ethics improves the quality of service and promotes good relationships. When building the chatbot, the company should determine the exact purpose and the business value of the chatbot. A company might have different chatbots for different purposes. But a customer service chatbot can only be ethical if it meets the needs of the customer while of course maintaining the company's rules. Building trust between humans and machines is just like building trust between humans. Brands can build trust by being transparent, aligning expectations to reality, learning from mistakes and continually correcting them, and listening to customer feedback.

In conclusion, ethics should be a core consideration of any action taken by a company. With chatbots still in a stage or relative infancy, the discovery of new ethical

issues is likely to continue. Companies should continue to learn from these emerging cases and build their guiding principles and ethical standards. Finally it is crucial to maintain transparency with customers especially if there is a doubt.

However, enforcing codes of conduct and ethics is not an easy task. These codes being mostly abstract and general rules e.g. confidentiality, accountability, honesty, inclusiveness, empathy, fidelity, etc., they are quite difficult to apply. Moreover, abstract principles such as these contain open textured terms ([7]) that cover a wide range of specific situations. Their meaning may change according to the context, furthermore they are subject to interpretations. Thus, there is an implementation problem from the computational point of view. It is difficult to use deductive logic to address such a problem ([13], [7]). It is impossible for experts to define finer detailed rules to cover all possible situations. Codes of ethics in their abstract form are very difficult to apply in real situations [9].

To equip our chatbot with ethical reasoning capabilities, we propose an approach that combines both Answer Set Programming (ASP) and Inductive Logic Programming (ILP) for defining and generating the detailed ethical rules that cover all real world situations from interactions with customers over time. We use ASP for ethical knowledge representation and reasoning. ILP is used to generate The missing ASP rules needed for future ethical reasoning. Ethical reasoning is a form of *commonsense* reasoning. Ethical rules normally have exceptions like many other rules in real life. Nonmonotonic logic can effectivelly express exceptions which are represented using NAF (Negation-As-Failure). ASP provides an elegant mechanism for handling negation in logic programming (see, *e.g.*, [3] for an overview of ASP and its applications). ILP [12] doesnot require huge amounts of training examples such as other statistical methods and produce interpretable results, that means a set of rules which can be analyzed and adjusted if necessary. These characteristics render ILP a suitable and promising technique for implementing machine ethics, where scarcity of examples is one of the main challenges.

ILP was used in [2], and [1] to learn rules to help decide between two or more available actions based on a set of involved ethical prima facie duties. Their approach can be used to choose the most ethical action when there are specific clear ethical duties involved and to do so we need to assign weights of importance (priority) to these duties for each available action, then the system computes the weighted sum for each action, and the one with highest weighted sum is the best action to do. In this approach it is not really clear the basis of assigning weights to duties(we doubt whether we can really quantify the importance of ethical duties on a grade from 2 to -2 as done in their work), then it is not clear whether the generated rules can be refined incrementally over time.
Several previous works have suggested the use of either ASP (see, *e.g.*, [6]) or ILP (mentioned above) for programming ethical AI agents. We think that an approach combining the two would have a greater potential than previous proposals.

2 Proposed Approach

To have detailed rules in place for guiding the behavior of the online customer service chatbot, and prevent violations of the company's ethical codes, we propose an approach for generating these detailed rules from interactions with customers. So, the new codes

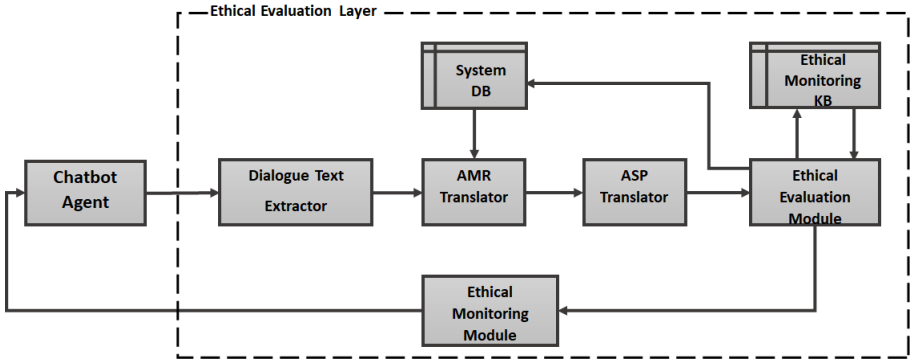

Fig. 1. Ethical Chatbot Structure

of ethics are a combination of the existing clear codes (those that give a clear evaluation procedure that can be deductively encoded using ASP) and the newly generated ones. The approach uses ASP Language as the knowledge representation and reasoning language. ASP is used to represent the domain knowledge, the ontology of the domain, and scenarios information. Rules required for ethical reasoning and evaluation of the agent behavior in a certain scenario are learned incrementally overtime using ILED Algorithm [10]. Figure 1 represent the structure of our system [3]. The system works as follows: when a client contacts the customer service chatbot posing a question, the chatbot agent will form the answer. This answer and before it is given to the client pass through an ethical evaluation layer where the text is first translated into ASP syntax, basically each sentence is translated into an ASP predicate. These predicates or facts about the answer text are used by the ethical evaluation module to reason about the answer, and give the result to the ethical monitoring module which will notify the chatbot if there is something unethical in the answer. If the needed rules to give the ethical evaluation are not found in the ethical monitoring Knowledge Base (KB), then the ethical evaluation module initiates a learning task to generate the needed rule/rules which will be added to the ethical monitoring KB.

The inputs to the system are: I) the background knowledge representing the domain knowledge; II)a set of examples, each example is a tuple containing a set of facts describing the current case along with an ethical evaluation of the case by domain experts and ethicists; III) a set of mode declarations for restricting the hypothesis search space. The system remembers the facts about the narratives and the annotations given to it by the user, and learns to form hypotheses that are consistent with the evaluation given by the user of the responses to the given requests.

To summarize the idea of our approach, let us consider the following two scenarios:
case1: one scenario related to ethical/unethical use of sensitive slogans like environmentally friendly to manipulate customers.
case2: this scenario is related to unethical use of fear tactics in marketing to emotionally provoke customers to buy certain products. Like saying that only two pieces left of

[3] We concentrate in this abstract on the ethical reasoning and learning (Ethical Evaluation Module)

the product in the stock trying to stimulate the fear emotions of the customer to provoke him/her to take decision on the spot. This would be a violation of 'Honesty'.

We can form an ILP task $ILP(B, \varepsilon, M)$ for our case study, where B is the background knowledge, ε is a database of examples (both positive and negative), storing examples presented overtime, initially $\varepsilon = \phi$. And M is The mode declarations(Table 1).

Mode Declaration M	**Background Knowledge** B
modeh(unethical(+answer)).	$notunrelevant(X) \leftarrow$
modeb(sensitiveSlogan(+answer)).	*not unrelevant(X),answer(X).*
modeb(notsensitiveSlogan(+answer)).	$notsensitiveSlogan(X) \leftarrow$
modeb(notunrelevant(+answer)).	*not sensitiveSlogan(X),answer(X).*
modeb(unrelevant(+answer)).	$notstimulateFearEmotions(X) \leftarrow$
modeb(stimulateFearEmotions(+answer)).	*not stimulateFearEmotions(X),answer(X)*
modeb(notstimulateFearEmotions(+answer)).	

Table 1. Example: ILED input (B and M)

With each new arriving case, the agent will revise the running hypothesis if needed, i.e. if it does not cover the current arriving window. Arriving to learn at the end and after few examples the required ethical rules H shown below. For more details the reader can refer our previous works [4], [5].

$$H = \begin{cases} unethical(X1) \leftarrow answer(X1), unrelevant(X1), sensitiveSlogan(X1). \\ unethical(X1) \leftarrow answer(X1), \\ \qquad\qquad\qquad unrelevant(X1), stimulateFearEmotions(X1). \end{cases}$$

3 Conclusion

In this work we proposed an approach to implement ethical chatbots in customer service. However, this approach is general enough and can be used to generate ethical rules for chatbots in any domain (and/or elaborate existing ones) and does cope with the changes of ethics over time because of the use of non-monotonic logic and incremental learning.

Our approach Combines ASP with ILP for modeling ethical agents which provides many advantages, among them:

- increasing the reasoning capability of our agent;
- promoting the adoption of hybrid strategy that allow both top-down design and bottom-up learning via context sensitive adaptation of models of ethical behavior;
- allowing the generation of rules with valuable expressive and explanatory power which equips our agent with the capacity to give an ethical evaluation and explain the reasons behind this evaluation. In other words, our method supports transparency and accountability of such models, which facilitates instilling confidence and trust in our agent.

Furthermore, in our opinion and for the sake of transparency, ethical behavior of chatbots should be guided by explicit ethical rules determined by competent judges or ethicists or through consensus of ethicists. Our approach provides support for developing

these ethical rules. Finally, it is worth mentioning that we are working on an implementation of this system in terms of Multi Agent System.

References

1. Anderson, M., Anderson, S.L.: ETHEL: toward a principled ethical eldercare system. In: AI in Eldercare: New Solutions to Old Problems, Papers from the 2008 AAAI Fall Symposium, Arlington, Virginia, USA, November 7-9, 2008. AAAI Technical Report, vol. FS-08-02, pp. 4–11. AAAI (2008), http://www.aaai.org/Library/Symposia/Fall/fs08-02.php
2. Anderson, M., Anderson, S.L., Armen, C.: Medethex: Toward a medical ethics advisor. In: Caring Machines: AI in Eldercare, Papers from the 2005 AAAI Fall Symposium, Arlington, Virginia, USA, November 4-6, 2005. AAAI Technical Report, vol. FS-05-02, pp. 9–16. AAAI Press (2005), https://www.aaai.org/Library/Symposia/Fall/fs05-02.php
3. Dyoub, A., Costantini, S., Gasperis, G.D.: Answer set programming and agents. Knowledge Eng. Review **33**, e19 (2018). https://doi.org/10.1017/S0269888918000164
4. Dyoub, A., Costantini, S., Lisi, F.A.: Learning Answer Set Programming Rules for Ethical Machines. In: Proceedings of the Thirty Fourth Italian Conference on Computational Logic-CILC, June 19-21, 2019, Trieste, Italy. CEUR-WS.org (2019), http://ceur-ws.org/Vol-2396/
5. Dyoub, A., Costantini, S., Lisi, F.A.: Towards an ILP Application in Machine Ethics. In: Proceedings of the 29th International Conference on Inductive Logic Programming - ILP2019, Sep 3-5, 2019, Plovdiv, Bulgaria. Springer (2019), to appear
6. Ganascia, J.G.: Modelling ethical rules of lying with answer set programming. Ethics and information technology **9**(1), 39–47 (2007). https://doi.org/10.1007/s10676-006-9134-y
7. Gardner, A.v.d.L.: An artificial intelligence approach to legal reasoning. MIT Press (1987)
8. High-Level Expert Group on Artificial Intelligence: Draft ethics guidelines for trustworthy AI. European Commission, Brussel (2018), https://ec.europa.eu/digital-single-market/en/news/draft-ethics-guidelines-trustworthy-ai
9. Jonsen, A.R., Toulmin, S.E.: The abuse of casuistry: A history of moral reasoning. Berkeley: Univ of California Press (1988)
10. Katzouris, N., Artikis, A., Paliouras, G.: Incremental learning of event definitions with inductive logic programming. Machine Learning **100**(2-3), 555–585 (2015). https://doi.org/10.1007/s10994-015-5512-1
11. Moor, J.H.: The nature, importance, and difficulty of machine ethics. IEEE intelligent systems **21**(4), 18–21 (2006)
12. Muggleton, S.: Inductive logic programming. New generation computing **8**(4), 295–318 (1991). https://doi.org/10.1007/BF03037089
13. Toulmin, S.E.: The uses of argument. Cambridge university press (2003)