Cezanne-ai: A Conversational AI Framework for Emerging Languages and Limited Data

Florin Coman
NLP Researcher, Screenwriter &
Philosopher

We hereby present Cezanne-ai, a hybrid NLP/AGI framework (not limited to Natural Language Processing, but including non-specific language elements and original concepts of Artificial General Intelligence) for developing conversational AI bots. The objective is to implement a solution that can cover all the requirements of a virtual conversation, focusing on interactions between non-specialists (the user) and experts (the bot). In order to achieve this task, we analyzed more than one hundred complex conversational topics, tested different architectures and existing state-of-the-art pipelines available in NLP, but not limiting to them.

After analyzing advanced functionalities, such as providing advisory and the "art of conversation" in a limited data environment, we are concluding that only a complex model, built from scratch (using the newest NLP models), and based on linguistics fundamentals, can create premises for reaching the human baseline in the conversational AI tasks.

Therefore, we are proposing an open-source framework to further corroborate our research and we recommend it to be used for emerging languages and small companies, due to their data limitations.

1. Introduction

All three main layers of a virtual conversation were analyzed (Natural Language Understanding – NLU, Dialogue Policy Learning – DPL and Natural Language Generation - NLG), and we are going to propose **6 models** (with **50+1 pipelines**) by combining Deep Learning (DL) algorithms with fundamentals observed in authentic creativity processes and with cybernetics systems theory (especially of Gregory Bateson's). This way, bots will not only have a neural model background but also an intuitive & holistic side (more resembling a labyrinth) and will have the instruments to cover complex conversations adapting to the level of knowledge and interest of the user.

One of the most important findings of the research is the effectiveness of using a structured understanding,[1] which is having a beneficial impact on the entire NLP/AGI framework. This new approach to NLU is constructed on four factors:

• Understanding the user core-input as a structured sentence (sentence-intent) that includes all the functions of a basic syntax: 1 Subject+1 Predicate+1 Complement+1 Attribute, and not as an arbitrary retrieval, a labeled intent, or a parsing method.

1 We are dismissing the General Conversational Intelligence methods (Kolonin, 2020) from the start, as we do not intend to represent linguistics and semantic models by replacing real languages with surrogates that eventually will impact our models. Furthermore, an incremental knowledge model is not helpful, as our model is confronting with complex inputs from the beginning.

• Understanding the entire input, in context, by splitting the user utterances in small pieces and then reconstructing them based also on fundamentals. Nothing that the user is transmitting is for granted.

• Considering more efficient premises for long-term conversational memory, as an alternative to models that have huge datasets requirements like 1. retrieval-augmented generative models (Lewis, et al., 2021; Shuster, et al., 2021); and 2. the read-write memory-based model that summarizes and stores conversation on the fly (Xu, et al., 2021).

• Changing the embedding principles. We are going to refer to Gregory Bateson and use his perspective to implement word vectors: "we don't have 5 fingers at one hand, but 4 relations between them and this is what counts for the brain".

Even though the model is proposed for implementation in emerging markets/languages, for small and medium-sized companies due to limited corpora/datasets requirements, we are also targeting the state-of-the-art models used in developed countries and large companies (in terms of the data availability) that deal with multiple challenges:

• Advisory limitations of the current sales-oriented/client-service models and frameworks that work either as question-answering NLP tasks with limited conversational turns, chatbots built on encoder-decoder models, or as open-domain conversational agents with datasets restrictions;

• Limited frameworks to adapt bots on different industries at the same time (multi-task domains bots);

• Not using books instead of datasets for deep conversational[2] purposes (please be advised that we are not referring to pre-training processes, retrieval, answering extraction or transfer learning), thus limiting conversational scenarios;

• Huge number of resources spent with preparing, labeling, processing, training and testing databases, if pre-trained models don't meet the expectations;

• Grounding and lack of intuition, which is rising cultural and social walls between users and machines;

• Limitations of the conversational AI solutions that are using pre-trained models with self-supervised data (like BERT,[3] ERNIE[4]), which are more suited for downstream tasks assimilated to conversational AI/chatbots, such as: question-answering and Text2Text Generation;

• Highly data-dialogue driven models that are limited to chit-chatting tasks (PARL.AI[5]);

• No self-corrected systems in the current models, as the feedback concept from cybernetics is used by few frameworks and mainly for learning[6] not for self-corrected purposes;

• Still no complete solution to the long-term memory problem, as the recent solution on retrieval-augmentation (Lewis, et al., 2021; Shuster, et al., 2021) and summarization-augmentation (Xu, et al., 2021) don't take into consideration past ambiguities between the user and the bot;[7]

• Limitations in using intents and multi-class classification algorithms as an open-source conversational AI.[8] By using predetermined scenarios, you risk affecting the fluidity of a conversation even if you have a lot of real conversations included in the model.

• Regulatory & ethics requirements, and

2 Discussions related to day-to-day human's existential issues or even the well-known small talk, essential in conversations.

3 J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. (arXiv eprint, 2018).

4 Z. Zhang, et al., ERNIE: Enhanced Language Representation with Informative Entities (arXiv eprint, 2019).

5 A.H. Miller, et al., ParlAI: A Dialog Research Software Platform (arXiv eprint, 2018).

6 B. Hancock, et al., Learning from Dialogue after Deployment: Feed Yourself, Chatbot! (arXiv eprint, 2019)

7 To be further elaborated in Fundamentals and in the pipelines Memory update and Reset.

8 We will further approach this issue in Fundamentals on https://github.com/Cezanne-ai.

• Limited progress in duplicating a real human-to-human conversation (by not focusing more on soft & conversational out-of-the-box AGI solutions).

To answer these limitations, we built a framework with a fundamental architecture that takes into consideration most of the possible conversational AI tasks/needs:

• Natural Input Understanding (NIU – extension of NLU). We will start from the hypothesis that we don't have an efficient pre-trained model that we can use for transfer learning, and we will investigate two possible solutions:[9] 1. How to train models with limited corpora, available for some languages?[10] and 2. How to understand natural language without over-using transformers that are expensive? We will introduce the augmentation-through-conversation proposition by learning the LM (language model) not only from supervised or self-supervised datasets, but also from the conversation itself with the user.

1. Pirkin[11] 1 model. Core inputs are separated from numbers, punctuation, emoji, intentions enablers, adverbs, specific adjectives, chit-chats and replies (all these are treated separately, in context).

2. Pirkin 2 model. The core-input (including core-inputs from the past interactions) is coded in a "sentence-intent[12]" (different in structure and content from the current intents or from a summarization task as in Xu et al. [2021]), composed of 1 Subject + 1 Predicate + 1 Complement + 1 Attribute. By exception, utterances that are identified as replies and chit-chats will be handled by using existing intent-based solutions (or other DL algorithms).

3. Back-up model.[13] In order to evaluate our model and for the bot to have alternatives when it doesn't understand the user's input, we will implement 3 advanced models and try to optimize them for our limited corpora/datasets using existing XLM models[14] and augmentation models. These models will have dense layers and they will be fit in the overall architecture: 1. Encoders-decoders/ T5 model for chatbot task (Raffel, et al., 2020); 2. Open domain dialogue models with summarization memory-augmentation (Roller, et al., 2020, Xu, et al., 2021); 3. Open-Source Language Understanding and Dialogue Management (Bocklisch, et al., 2017).

• Conversational Policies Learning (CPL – extension of DLP).

Pirkin 3 model. Achieve human-level conversational AI by using intuition and a labyrinth model (feedback and creativity policies), not limiting to dialogue, but covering all known states. We consider that a neural network approach in dialogue management is not a solution for our model (which is addressing low resources and small to medium-sized businesses, plus we do not want to direct the policies towards business objectives). On the other hand, fundamental CPL rules-based (by opposition to understanding methods) are in line with human behavior (see the Etiquette code).

• Natural Output Generation[15] (NOG – extension of NLG). In order to formulate an output, we will implement two additional AGI models: NLG enhancements and self-generative bot model. NOG models will work with three (plus one) trained sources:

9 We will not consider implementing data augmentation through translation or pre-training on English to achieve LM and then transfer the learning on the specific task in the desired languages, because these solutions are fundamentally wrong, as we will explain later.

10 Like Oscar and Common Crawl datasets.

11 *Pirkin* is the name of our proposed six models. *Pirkin* is the phonetic NLU of the word virgin that we found out to be used in Bali, Indonesia.

12 This new approach can also solve situations with similar intents.

13 This model will limit the need for fallback policies.

14 Cross-language models.

15 Even if in many cases the answers/outputs are extracted and not generated, we have chosen to name this layer NOG as it describes the overall solution better.

1. Pirkin 4 model. NOG from databases provided by experts,[16] not from crowdsourcing or 3rd parties.[17]

2. Pirkin 5 model. NOG from books, based on a labyrinth model instead of a neural network model.

3. Pirkin 6 model. NOG from intuition by using self-generative bot policies and behaviors (part of AGI).

4.Back-up/chit chat model. Subsequent to the NIU/NLU back-up model, this model will provide generated or extracted answers.

By successfully implementing these models we will be able to have a framework that can fulfill all the 12 expectations resulting from the brainstorming presented in the next chapter. A primary objective is to develop a conversational bot that is adapting to the client's expectation, and not the other way round. We are also concentrating on creating a framework that allows easy customization for many types of businesses that need conversational bots in their digitization strategies, as we want to show the adaptability component. For this reason, it is mandatory to complete our value proposition with AGI enhancements.

We disclosed detailed information on how our research can be implemented in practice together with detailed fundamentals, context, data prerequisites and inspirations.[18] We are naming this subsequent project Cezanne-ai[19] . Feel free to implement it yourselves in the language you desire and with your own corpora and datasets.

2. Components and objectives

Conversational AI/NLP component:

Conversational AI bots are split in 3 categories:[20] 1. Q&A (with one or multiple turns); 2. Task-oriented (or virtual personal assistants); 3. Chatbots with recommendation capabilities or social/empathetic models.

Going back to fundamentals we need to establish, however, what the common user's expectations are. 1. Info, 2. Answers, 3. A personal assistant to solve the user's needs, 4. Chit-chat in an empathetic/ social environment, 5. Sales-oriented recommendation, 6. the bot's own views on some issues, 7. Confirmation, 8. Understanding, 9. a close friend, 10. Advice, 11. Knowledge through books, 12. the art of conversation?

Depending on the culture of each country or the user's education you will get different responses for sure. Everybody is using Google or Bing in order to search for information and some are starting to use their multi-turn conversational capabilities also to get more specific answers. In US and more developed countries personal assistants like Alexa, Siri & Google Home are more in trend with people's expectations (including, for example, AWS Lex or wit.ai for NLU and digitization). In China, XiaoIce[21] (a Microsoft social chatbot with task-oriented capabilities) is extremely used, but it is a big question if a multi-environmental conversation

16 We will refer not only to answers given by experts in the core-intent field, but also to the playwriters' assessments of the conversational states that are translated into possible: questions, disclaimers, change in topics...

17 Even if text generation methods like unlikelihood training and adding minimal length constraints can be solutions, we will not use these methods here in order not to affect the accuracy of the expert output. Humans are experts when it comes to circumventing filters and safe-guards (Dinan, et al., 2019). But we will keep an eye on every new method that can improve our NOG model.

18 https://github.com/Cezanne-ai

19 The name is inspired by a painting of Cezanne: "The Conversation".

20 J. Gao, M. Galley, L. Li, Neural Approaches to Conversational AI (arXiv eprint, 2019). There are other classifications, but as we didn't find them in the research paper, we will not mention them. Most of them are somehow intuitive about the future.

21 L. Zhou, J. Gao, D. Li, H.-Y. Shum, H.-Y., The Design and Implementation of XiaoIce, an Empathetic Social Chatbot (arXiv eprint, 2019).

will attract users in cultures where a conversation is seen as more personal. Facebook Blender Bot 2.0[22] may be the perfect argument, as it hasn't had so much success until now, even if it is targeting all types of users.

But let's go to Webster's definition of a conversation: "a. Oral exchange of sentiments, opinions or ideas" or "b. An informal discussion of an issue by representatives of governments, institutions or groups". This means that the first 5 definitions which are now in the scope of Conversational AI bots are not even mapping the definition.

Firstly, the conversation needs to be informal, and the personality and education of participants are in discussion. The fact that machines are trained with huge datasets that belong to different persons (not to mention the pre-trained models used mostly for the language model) makes the dialogue non-personal and q.e.d. not a conversation.

Secondly, users expect the machine's own input through sentiments, opinions, or ideas, not facts or a compilation of sentiments, opinions, ideas from many other persons. As we are not there yet in terms of self-generated artificial brains, the only solution (if we want to use the term 'conversational') is to give machines the personality and knowledge of a human (using AGI), with its limitations in terms of answers and background, but also with his accountability/expertise/recognition as a differentiator between conversational bots. We intend to use this over-strict definition/interpretation to cover the missing pieces from the puzzle.

Conversational AGI component:

We will now introduce our view on soft AGI that is different from the current trends. It's interesting that one of the main researches on AGI (Baum, 2017; Fitzgerald, et al., 2020) is sponsored by Global Catastrophic Risk Institute that insisted on having their sponsorship on the Baum paper with bold - uppercase letters. On the first page the word "catastrophic" is more visible than the name of the working paper, but if you carefully read the information on the 72 AGI past and current projects that were researched, there is nothing catastrophic about the AGI impact especially when fundamentals are taken into consideration. It's worth noticing that between 2017 and 2020 there is a decrease in new projects and in AGI interest, as the same research concludes. One possible theory is that Artificial General Intelligence needs **not** to be regarded as a risky/terrifying concept especially when it is not used for core-inputs and is fundamentally based and used for research purposes (please be advised that although we will include implementation of commercial bots in our scope, we will not use AGI in these situations).

Furthermore, the idea that a time will come when robots have their own brains and they will not be controlled by humans, can create unlimited passionate discussions and this is not in the scope of this article, as we try to stay out from philosophical debates. But, in order for a bot to have a real conversation it is necessary to make him more actively involved, not only reactive to the user's needs/questions/topics. We need a soft model, limited to conversational topics that will not allow the bot to have an impact on the answer, nor to act, making the safety issues requirement redundant.[23]

Quick classification of AGI original concepts integrated in our framework:

• intuition provided in some NLU instances in order to make sure that the bot understands what the user is saying,

• machine education[24] - give conversational bot a similar education with humans by using AI algorithms (but not using Turing or Baby Turing tests),

22 S. Roller, et al., Recipes for building an open-domain chatbot (arxiv eprint, 2020).

23 We will, however, have a chapter on AGI's safety issues.

24 Not to confuse machine education with the processes of language models and transfer learning, even if the pe-trained models are fine-tuned to cover the individualization objective. The main difference is similar with the one between knowledge and memorization, or between the concepts of short-memory and long memory (from neuroscience).

• self-generating policies for non-core-inputs/states,
• self-generation model for behavioral/mood instances that do not affect the answers/output,
• labyrinth model based on fundamentals, used in different situations to structure a book, for query purposes,
• enhance NLG on the non-verbal/non-written components to be more adapted to human conversations.[25]

This proposed approach has no intention to win any race for AGI, but to enhance the hybrid model.

Additional objectives are in scope of this article. For instance, everybody should have access to experts in domains like programming, food, medicine, legal and other, not only the rich, the ones with connections, the ones that are qualified or the ones that are living in more developed countries. Conversational bots can fulfill this desire and can bring this precious resource of experts into people's houses. Of course, there are books available, and we have YouTube, but not everybody can understand a book written by an expert or a presentation on YouTube. Many are more comfortable with asking specific questions, using their own vocabulary, and finding answers in an exhaustive conversation instead of searching through huge amounts of available information that are contradictory many times.

3. Fundamentals and assumptions[26]

1. The important results in NLP and ML are also based on a good understanding of neuroscience. However, this success can have an inversely proportional impact on Conversational AI. Let's explain why below.

Simple Neural Networks are not much different from the short-term memory from neuroscience theory[27], which is saying that some impulses can have a chemical reaction to our brain. For example, we are starting to play ping-pong and after a week of practice we can believe that we are good at ping-pong. The truth is that if we go to the Olympics, we will not win any medal for sure, or after one year of not playing anymore we would have forgotten everything we've known. The same is with one epoch of forward and back propagation of the input in NLP/ML. It helps us achieve some basic tasks (for example, that week of practice will be beneficial for our health) but will not make us achieve complex results.

The scientific community understood this limitation quickly and proposed different solutions to go from the short-term to long-term memory, hoping to achieve physical changes to the brain. This is the meaning of introducing solutions like: more layers, epochs and batching, sequential models, attention models, transformers, fine-tuning, pre-trained models with big datasets, continual learning, knowledge graphs etc. All these are very similar to educational techniques used to grow the intellectual capacity of a human. And this is why many NLP tasks have now efficient solutions, especially in the most used languages.

Now, we want to use these incredible developments in NLP to implement conversational AI bots. Further solutions were found, like implementations of intents, entities, forms, better quality of the data, better frameworks with optimized pipelines, DL models in Dialogue Management, testing GRU[28] more in NLP, Information Retrieval hybrid solutions, better extraction and gener-

25 We will consider visual NLU integration as well in the future.
26 A detailed version of this chapter can be found on the project page on github. Kindly note that this article doesn't have a related work chapter.
27 For more on this topic, we are proposing Lara Boyd's research related to Neuroplasticity – see Bibliography.
28 Gated Recurrent Units have been shown to exhibit better performance on certain smaller and less frequent datasets - Wikipedia

ation of the output, fine-tuning by using additional datasets or optimizing the pre-trained model etc.

And still there is not a lot of progress in conversational AI, especially in deep-conversational. WHY? There are some observations from real life that can help us understand the problem. Did you ask yourself why top managers are not very talkative in interviews and in conversations that are not linked with company PR or marketing? Why are applied scientists not very open to do conversation on topics that they cannot handle, or even on their main specialty topics with people that do not understand the subject very well? The answer is very simple. Better experience, an analytical approach, increased intellect or substantial knowledge do not mean better intuition, increased creativity and greater openness to new things that are also essential to a conversation.

So, to achieve conversational AI efficient solutions, we need to add an AGI perspective to the architecture. Having better results in question-answering tasks (also in open-domain/close-domain Q&A) is an important part of the conversation, however, if resolving analytical issues prevails over the systemic ones (for example, we cannot link the existing questions with past interactions), it can have a detrimental result on the conversation. Everybody hates that smart guy who is barging in a conversation and gives the right answers, but is not interested in the context, implications, and is not sympathetic to others.

2. Another important aspect that needs to be taken into consideration is the wrong assumption over the understanding principles inside a conversation. When you are in an exam, for example, it is mandatory to read the question very carefully, because that question is supposed to cover also the context and everything needed in order to express a response. This is also the assumption of NLP tasks, especially question-answering and text-to-text generation tasks. This is completely right and the fact that we have this huge results in benchmarks for the core-languages is mind-blowing, even if the success story is different for emerging languages.[29]

However, in a conversation (not in an exam) there are few people that can synthesize efficiently and the scope of the conversation is to build the questions, answers and argumentation on different turns, so the context is not present in a single interaction and is formulated better. Furthermore, questions in exams are built in a way that allows the student to understand them with the knowledge that he already has (not understanding the question and having additional ones, is not a good sign for the student). In Chatbots and conversational AI bots, the environment is completely different. Most probably the user will not utilize the same language model that the bot is trained on, so additional and confirmation questions are mandatory (the more experienced users will probably not use any chatbots in the field in which they are experts). Consequently, the relationship between users and bots is similar to the one between non-experts and experts and the language model is supposed to be different. Having pre-trained models on huge datasets can solve this issue but will have limitations as well.

Let's take some examples. Somebody is going to a doctor, or a lawyer, or to an IT developer because they need help. One part is the expert and the other doesn't have the right vocabulary in order to discuss and present his needs clearly. How do they understand each other? One solution is to ask the non-expert to go home and read a book in order to present his case better (this is pretty much the pre-trained model in NLP tasks, which is helping the expert understand the language model of the non-expert). However, in reality, the understanding process is done differently, through conversation and detailed presentations of the query/input/intent. So, why not learn the language model through conversation and only basic semantics (which requires limited corpora) through neural network training? It is possible that the best conversations that you had in your life are the ones in which you didn't have many preconceptions, you learned from

29 We define the emerging languages as those languages that have limited resources in terms of corpuses, datasets, pre-trained models

the conversation and you adapted to it. This is the augmentation-through-conversation method that we will apply in our NLU model.

For a complete image of the assumptions and fundamentals that we considered for backing up our proposal we will direct you to the project page.[30]

3. Another path that is currently followed in Conversational AI is: open-domain conversational agents. These agents fulfill the users' needs to share information about their own findings, chit-chatting on different topics, gossiping etc. and in order to keep users engaged they are using advanced methods of continual learning, available datasets from narrow domains and Reinforcement Learnings (RL) models. This way, they are providing a source of information and also a source of sharing information as an alternative to social media and different chit-chatting subjects that you usually have with colleagues, family or friends. We will make some comments regarding this approach, asking you again to go as well through the detailed fundamentals:

a. The term "conversational" isn't used properly in this situation, according with Webster's dictionary.[31]

b. For this task you need large updatable corpora not available for most of the languages. As some unsupervised data are available for different languages, dialogue datasets in specific languages are impossible to find.

c. This approach can be extremely bias in terms of the source of information.

d. Using rewards and RL can be detrimental in conversation.[32]

e. Seeing the conversation as an interaction between two random persons is divergent with our scope that is targeting non-specialist vs expert.

f. Open-domain is not the same as multi-domain for the same argument provided above.

g. To achieve the "art of conversation" we need to focus on the personality of the bot, not on providing him with more info in order to be an "agent". There are sufficient solutions for this type of needs outside the Conversational AI world.

4. In order to fix the long-term memory issue in a conversation, recent studies have proposed two solutions with important results over the standard encoder-decoder transformer. We've already talked about them: retrieval-augmentation and summarization-augmentation. They have a correct assumption of the fundamentals in conversation, but they have a drawback in the implementation: the unstructured summarization or retrieval makes it impossible to know what the user's past reaction to the already discussed topics was and how we can take into consideration these reactions. Let's take an example: We are at a second meeting with a client and in order to prepare properly we take a look at the minutes of the last meeting (this is exactly what the models in question are actually doing by summarizing and/or retrieving). At the first look, you can consider that it is a good approach and a fundamental one, but when you look at the minutes you are also observing some discussion topics that have already been dismissed completely or partially and this has important implications to the present. These models can be adjusted to the "dismissed completely" part by implementing an additional NLP task to the model, but they can hardly be a solution to this issue as the summarization and retrieval are done in an unstructured way. By opposite, our Pirkin 2 model is proposing a structured summarization (sentence-intent) and applying two functions: memory update and reset.[33] Our principle is that in a conversation, and especially in "the art of conversation", humans do not record and memorize everything as we are doing in business meetings, but rather they start the conversation with a simplified idea

30 https://github.com/orgs/Cezanne-ai/repositories
31 See the above chapter.
32 See detailed Fundamentals.
33 Both are pipelines in the NIU/NLU – Machine Learning architecture – see further.

(sentence-intent as a core-input), they elaborate on it and they leave the conversation also with a simplified idea (which can be exactly the same or different if the conversation was effective).

5. Two solutions are used frequently to find workarounds for the lack of corpora, datasets, and pre-trained models in most of the languages (besides GRUs that haven't been researched too much in NLP so far): data augmentation through back-and-forth translations and using pre-trained multilingual models (XLMs) for the language model and then fine-tuning with additional labeled datasets in the desired languages. These methods will provide for sure much poorer results than the core-languages for the tasks in question, but are solving some problems: you don't need to spend huge resources on data and on training. However, these models are wrong in terms of fundamentals and, thus, they are not supposed to be called solutions. Why? If you don't have corpuses and resources for your NLP task, the translation and pre-training are definitely done with the same lack of resources for the language in question. A similar discussion is with small and medium-sized companies (with limited customers/users) vs corporations.

4. Conversational AI fundamental-framework[34]

We are proposing six solutions based on fundamentals[35] and the analysis made on existing conversational AI architectures and pipelines. AGI concepts will be used on deep conversational,[36] not on core-input.[37]

Functionalities:

• AGI proposition- allowing the framework to have self-generated policies & behaviors based on its own intuition, grounding, external factors and the current conversational states;

• Incorporating EQ (empathy) and a sensible side of bots into conversation;

• Virtual personal assistant capabilities;

• Automatic training (Continuous AI pipelines) & conversational analysis with live impact;

• Understanding the user's opinions/ideas in a long line dialogue and keeping also the context of the conversation;

• The bot is positioned as an expert in a multi-domain environment;

• Deep conversational topics through books queries;

plus

• No interfaces with other languages platforms;

• No 3rd party databases are used for training, processing, testing or validation (only for pre-training, language models and transfer learning purposes);

• No external embeddings/weights are used in the Pirkin model;

• Three adapted Deep Learning models will be used as back-ups. The processing algorithms (Machine Education) and CPL will be correlated, however, to assure consistency.

In which domains is a Conversational AI/AGI bot more suited to be implemented?

The main objective of the conversational AI framework is to be able to replace an expert in different domains. For this reason, we will target the following domains:

• Programming, by implementing natural language instructions + conversational methods instead of coding;

• Restaurants recommendation/reservations;

• Travel & accommodation advice;

34 All practical and detailed information in order to implement the models in the desired language and with your own corpuses and datasets are at https://github.com/Cezanne-ai.

35 Meaning that everything we are proposing has a fundamental background.

36 We will define deep conversational either as an existential topic, or as the need of the user to find additional information in available treaties or books (not limiting to databases).

37 We will define core-input as the input of the user that is correlated with the main specialization of the bot.

• Medical consultancy;
• Legal consultancy;
• HR, both for recruiting and employees' satisfaction;
• Art & Culture events proposals;
• Fashion advice;
• Investment advisory & banking services;
• Books reading through queries/conversation*. Ex: Novels, specialized books or screenplays;
• Documents queries;
• Social/empathetic/sentiment driven bots.
• Operating systems;

*This is not intended to replace the actual reading of the books (as we don't recommend reading books summarization, especially the belletristic ones), but it is helpful in 4 situations:

• You have already read the book, but you need a refresh, or you missed a detail;
• You need more info/details before deciding to actually read the book;
• In order to have a more interactive way of finding what you need in scientific books and a search (even using advanced retriever-reader models) is not being helpful;
• In order to interact on some ideas/topics by making book queries.

Table 1
Simulation of a conversation inside a Conversational AI/AGI front end:

User input	**Bot input**	**Comments**
Response to salute	Hello!	The bot is initializing the conversation. A possible input from the user can occur.
Possible detailed answer Possible question/task/response	Please give me a detailed description of your need/request	The bot is trying to avoid Q&As or chit-chat This triggers advisory CPL (Dialogue policy). Trigger the contextual dialogue.
Possible additional input Possible question/empathy/response	Can you provide additional information?	The bot wants to receive as much info as possible. The bot is doing a memory update. Trigger the contextual dialogue.
Possible reply to the answer Response to confirmation question	Bot answer/empathy/chit chat Possible confirmation question	After assessing the real need/empathy/ If Context validation matrix is against Possible back-up method.
Possible task (ex: reservation)	Reevaluation of the context Task implementation	Triggering Conversational policy. Bot queries modules initiations.
Possible Additional questions	Context update Answers/ Confirmations Changing the topic	Constant update of LSTM (memory) and Conversational policies considering not to duplicate answers. Triggering Hierarchical discussion topics.
Possible follow up questions Possible chit chat Possible ending of discussion	Reevaluation of the context Automatic Conversational analysis	The bot is prepared for unlimited conversation either on core topics, either conversational or deep conversational inputs.

This model is recommended for bots that do not require user authentication; consequently, we don't have implications related to personal data regulations.

Even if Pirkin models bring a fresh approach in terms of architecture and language understanding, current frameworks and "state-of-the-art" models are used in different pipelines. Thus, Pirkin models cannot be regarded as traditional/rule-based NLP models, but as fundamental models that are using contextual and up-to-date DL algorithms.

Layers (or modules, or pipelines) of a complete conversational AI framework:

• Natural Input Understanding (NIU – extension of NLU concept/layer);

• Conversational Policy Learning (CPL – extension of DPL concept/layer);

• Natural Output Generation (NOG – extension of NLG concept/layer).

4.1 Natural Input Understanding (NIU)

4.1.1 Pirkin 1 model. Machine Education (M/DE) in 11 pipelines.. M/DE methods that we are proposing are very similar with school's curricula for teaching language classes and consequently the architecture can be regarded as an AGI concept and can be part of a developmental robotics approach.

We will break NIU/NLU in two.[38] First part of NIU is Machine/Deep **Education** (M/DE) that is also addressed in the detailed fundamentals. The next part of NIU is the Machine/Deep **Learning** (M/DL) - similar to NLU in terms of format, content - in which the robot must learn specific things useful in his, let's say, job.

Based on fundamentals and the current NLP algorithms specific to conversational AI, we are proposing effective pipelines of a machine education process and then presenting a structured architecture of Pirkin 1 model, which is further detailed in the Cezanne-ai project. Keep in mind that Machine Education cannot be substituted by using pre-trained models or existing tokenizers (even if we have good results for specific languages), as the objectives are different and it's important to build the framework from scratch. Additionally, we are developing a conversational bot in which a non-expert user is talking to an expert bot and this approach has particularities.

Please be advised that a vectorized model in M/DE is not desired, as the order of the pipelines and the retrieval of some information needs to be done in a certain way/order. Don't worry, we are not returning 50 years in time. To implement some NLP tasks, we will use vectorization, but at the same time it is important to keep the raw version of the input for different pipelines to be effective and not miss even a single letter from the user's input.

Necessary information regarding corpora, datasets, language influences over the model, are included on github together with all the steps, in order for the coding to go smoothly. We are just building the model architecture at this point.

Auto- Correct[39]. Why is this algorithm important and, moreover, from the beginning?

• Garbage in – garbage out principle;

• Even if auto-correct might be available in the front-end, many do not use it, or it is customized for other type of interactions/domains;

• Users tend to write very fast and they are making many mistakes;

• In reality, humans use this approach and they ask questions only if the targeted, unknown word, is similar with different words, morphologically or/and syntactically;

• Transferring this issue to CPL would make the conversation too complex.

Pre-processing is needed for implementing this algorithm but should not impact the main model. Existing algorithms are suited for this pipeline (spell-check pipelines, masking tasks. . .)

38 Check Fundamentals for further details.

39 No entity mapper will be needed, if it's done from the beginning, for example.

Input Processing I.. Input/text processing needs to be implemented at two different levels:

• Main level. The objective is not to eliminate/stem/lemma too much data/info, in order not to lose precious information. Also, the chronological steps are important to be implemented gradually, in order for different algorithms to use the current state of the processed text/input. The sub-words tokenization can be a solution but depending on the language can affect the model. The same with stop-words or keywords that can impact our objective to have a multi-domain conversational bot.

• Local/Algorithm level. More elaborate processing depending on the objectives.

The first processing needs to be focusing on better understanding of the language and conversational specificities. At the same time, we need to separate the information transmitted by the user in five categories:

1. Words, numerical data and characters not needed in the main NIU model in order to understand them (ex: visual inputs or links that are in scope of future developments).

2. Numerical data needed especially for task-oriented user requests (ex: reservation for 3 persons on Monday, 5th January). As an exception, specific words need to be transmitted to DER (Data Entity Recognition- to be further detailed).

3. Emoji that will be analyzed separately.

4. Words in the core-input that will be processed.

5. Punctuation together with other information that must be analyzed in context.

Composed words. One of the biggest risks in NLP is the fact that everything is trained without considering the exceptions. From our experience in using language interfaces (platforms), many of the issues were in these situations. The next general categories need to be taken out from the core NLP/DL algorithms and treated differently:

• Expressions, ironies, metaphors, quotes. Including them in a training process is not necessary as they must be reproduced exactly (word by word) and they have impact on reactions (ironies) and not on the core-input.

• Specific adjectives with superlatives & adverbs (even the ones in one word). For example, “one of the best” must be treated as an exception from training, especially if you don’t have a huge database to cover all of them in many different situations.

• Description/definition of words in the dictionaries (optional). Every lexicon has short descriptions of words in order to understand them better. Using these definitions and replacing different sentences found in the user’s input with the corresponding lexicon word, is in the scope of this research. There are situations when you cannot find the right word in a conversation, but you manage to give an approximation of its definition.

• Specific entities (for example in the restaurant industry: name of chefs, name of restaurants, types of food) that are written in more than one word, but the entity must be treated as one.

Named Entity Recognition (NER) . We need to identify the commercial/core domain in every input, in order to see if we are dealing with core-input. For this reason, we will search for NERs and classify them in NER1, NER2, NER3.... For example, in a case study of restaurants recommendation digitalization, NER1=name of the restaurants, NER2=locations, NER3=types of cuisine, NER4=types of food, NER5=names of chefs, NER6= word in the gastronomic dictionary, NER7= other generalities... For multi-domains/industries, the customization process will include providing the databases split by NER and importance of NERs.

This pipeline is very important to determine the Subject and Complements of the core-input. Once the domain is determined, NER will remain unchanged for the next inputs of the user, if no other NER domains are inputted.

NER can include some generalities (general inputs of the user - GER) that can be common with other domains and can create confusion as it might have different meanings in different

domains. The model implies a prioritization and if the user is unsatisfied, either he will give additional information, or the bot will search for back-up answers. Most probably the user will understand that it is a multi-domain bot, and it is important to give a complete input which will be captured by the sentence-intent, which has also the benefits of the multi-domain conversational bot implementation.

Emoji/ Abbreviation (EER). As we are not developing a chit-chat but a virtual conversation between a user and an expert, we will make some assumptions:

• The user will not use emoji or abbreviations in the same sentence with an important request.

• The user might use emoji/abbreviations as a reaction/reply or in a follow up discussion.

• The user might use emoji/abbreviations in a more complex input and a bot response might be regarded as emotional and, thus, might help the conversation to go forward.

Considering the above, we will implement a secondary flow that will have specific answers/policies and can exist together with the main flow, or standalone if no other input is provided by the user.

Important: Abbreviations that are not considered specific to chit-chat needs will be treated in the main flow, together with the sentence-intent. For example: abbreviation for institutes that are NER.

Grammar/Semantics (Tagging). This algorithm usually has a huge impact on understanding and on the conversational policies/states. As we are dealing with semantics also in other pipelines it is important to find the right steps and resources to complete it correctly and in here, we are preparing the field.

To find the specific-domains semantics, a database is needed (can be also books related to the domains that are used for query purposes, or site's corpuses, or any unstructured data related to the domain) together with the available corpuses that we have for the language model. It's not important for corpuses to be large. In our research, 100-200 sentences for each type of conversation (chit-chat/ reactions/ restaurants advice/ legal consultancy etc.) are enough at this point.

We don't want to use pre-trained models that are coming with their own tokenizers and semantics from different domains.

Data Entity Recognition (DER) + adverbs. Adverbs, some specific numerical or data information are not important for understanding the core-input in a more complex utterance. They are essential for task-oriented purposes, for example. That's why people usually note them down to use them in the final phase of the conversation, after they understand the core part of the input.

We will not include these slots in the main model as well. We will see in the NOG/answers layers if we have all the information needed to accommodate task-oriented needs of the user. If not, additional questions will be asked.

For architectural reasons we will not refer to DER as slots, as NER is separated in another pipeline with different NIU/CPL/NOG implications, and we are including also a relatively *(please accept this word instead of relative)* vector in DER (based on the type of used adverbs[40]), which is not usually treated as a slot and we are using it for deep conversational purposes.

40 The categorization of the adverbs and implication over NLU is detailed at https://github.com/Cezanne-ai

Splitting sentences. "Divide et impera" is one of the strategies of our Pirkin 1 model. We will split the input (that can be a very elaborate description or can contain both sentence-intents and reactions) in short sentences.

At the same time, in case of the same consecutive POS separated by space, comma, "or", or "and" we will keep only one POS and all the others will be sent to Input Processing 2. From our experience, the existing language platforms have issues with this kind of situations as well. They are poorly administered by language models, especially the bi-directional ones (like BERT). In the example: 'I do not want to go home or to the cinema', the masking model could easily miss that "not" is referring to the cinema as well.

There are NLP solutions like sentence-based tokenizers or libraries with sentence recognizers that can be useful.

Context Validation Matrix (CVM). Using NLP question-answering solution to conversational AI can be tricky if we are seeing the conversation for what it really means.[41] The same is with intent-based solutions. Let's take some examples:

"I am going home" vs "Am I going home?"

"I do not intend to go home and watch tv" vs "I do intend to go home and not watch TV"

In question-answering or chatbots the hypothesis is that the user is addressing a question which is generally built in a positive sentence. This way the similarity scores will be high with all the available functions and probably will not have a huge impact if you forget to put the question mark. In a conversation, the two pairs of examples have completely different meanings and, even if you use dense layers for embeddings and a very accurate similarity score, you can fail to understand what the user is saying (especially if you don't have a large corpus). For this reason, we will build a matrix (CVM) to constantly double check the context and the user's intention. Another way to do it is to check your pipelines and optimize the tokenizer and the sub-word function that you are using. This method can have some effect for huge corpora and languages like English, however this is not in the scope of our article.

This matrix is crucial for layer 2: Conversational Policies Learning (CPL) that must use this matrix (together with other matrices) to decide the policies and the states of the dialogue/conversation. The following information needs to be retrieved and kept in the matrix:

- If the input is an affirmation, a negation or a question (or a request).
- If the user is talking about the present, the future or the past.
- If the user is speaking in first, second or third person.

Principles:

- Each sentence will have a CVM.
- The priorities are as follows: negation, question, affirmation.
- The principal CVM will be chosen in the SPCA[42] algorithm.
- CVM will be updated in the memory update algorithm.
- For deep conversational inputs without subject, CVM will complete SPCA with the subject.

The dependencies with other validation matrices are in scope of layer 2.

Input Validation Matrix (IVM). The second matrix that will be used for policies in the Conversational Policies Learning is IVM. This matrix tries to identify the kind of input the user is entering:

- A description of his needs (as asked by the bot in the first interaction);

41 See Context and Fundamentals for the definition and fundamentals of a conversation.

42 The structured sentence-intent with 1 Subject + 1 Predicate + 1 Complement + 1 Attribute.

• A short utterance (a question related to the main topic, for example);
• A reaction to the bot questions or answers (or lack of answers);
• A chit-chat with an emotional reaction/input;
• An input that cannot be trained (EER/Composed word).
The following information needs to be imported in IVM:
• No. of short sentences, no. of complex sentences, no. of inputs from EER & Composed word;
• No. of turns;
• What the bot is receiving: a description/sentence-intent, a chit-chat input or a reply (reaction of the user to past interaction).

Answer Validation Matrix (AVM). The final matrix that will be initialized in NOG and used in CPL is Answer Validation Matrix (AVM). Even if the info is obtained in NOG, we chose to include this pipeline here. Three main information are part of AVM, depending on the actions of the bot:
• Answers and types of answers;
• Questions and types of questions;
• Other actions (bot doesn't understand, changing topic, disclaimers, 3rd flow interactions used for empathy).
CPL will evaluate each time the states of these three matrices (CVM, IVM and AVM) and apply the necessary policies in the conversation.

Sentences Classification/ Assumptions of the Machine Education. We are going to make the following assumptions based on fundamentals and our background (plan B will be available if the user reacts differently from our expectations):
• The descriptions are expected to be in the first interaction, or if the bot is asking additional questions, or when the user is changing the topic being satisfied with the past answers (or not). For sure they are common in the consultancy conversation and the user will be using longer sentences or phrases that are very hard to be labeled & trained also with the present state-of-the-art DL/RL models that are using big databases. This is also in the scope of the research.
• A short utterance can appear at any time from the user's part. The existing models can easily accommodate a conversation based on this Q&As (including task-oriented capabilities) if you have a decent database and budget available for training computation. The issues that we have with choosing this method are the following: we are expecting the user to variate the type of utterances (not only Q&As), we are expecting the existing capabilities not to keep track of past interaction, also because of the first statement, and we are in the emerging AI environment when also decent databases are hard to be obtained on different domains.
• A reply is usually built in a shorter sentence and a decent sentiment analysis algorithm is pretty much doing the job. So, if the bot is expecting a reply, then we will implement an existing NLP algorithm that will have the best results on our database. The result of the sentiment analysis/or intent-based algorithm will be kept in the IVM and together with other validation matrices will decide the conversational policies. If the user is giving longer replies, we will also try to determine it by using the Pirkin 2 model and CVM. In this situation, if the user includes an elaborated reply in a single input and a Q&A or a description, there is the risk of missing one or another. The user can understand the situation if he doesn't consider the bot much smarter than a human.
• A chit-chat with or without an emotional reaction is expected at any time but will not be encouraged. The bot will try to respond to two consecutive chit-chats inputs (that do not have a specific sentence-intent) but after that will try to direct the discussion on the main topics with various questions. If the input of the customer contains sentence-intents and chitchats as well

then, the bot will reply to both, first to chit-chat and then to the core- input. The short sentences will be labeled and trained and responses will be given to the user, if found. If not, the input will be considered in the main model (Pirkin 2 model).

• Emoji/Abbreviation and Composed word are marking inputs that are intended for direct responses without training. A database will be provided for mapping and the responses will be provided in the second flow, as it is possible for the input to contain also other types of conversation. If the bot gives an answer in the chit-chat conversation, then NO untrained answers will be given. If more than one Emoji/Abbreviation or inputs from Composed word is received, then the bot will respond to the last known input.

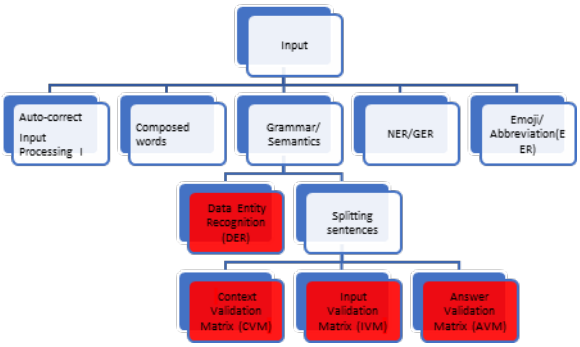

Figure 1
NIU/NLU Machine Education Layer

These are the principles behind a Machine Education layer and we are proposing the above flow, (red indicates outputs from NIU, the order and direct lines are important for an improved retrieval of the data needed from the user's input).

4.1.2 Pirkin 2 model. Machine/Deep Learning in 12+1 pipelines.. Now that the bot has an education (a foundation) we need to provide "on the job" training. This is the main model of the framework, which is responsible for input understanding. We will start with a secondary flow, named like this in order to differentiate from the flow where the core-input is.

Reply. Replies/Reply is referring to the action of the user resulting from other actions of the bot (answers, questions, other). The bot expects these replies to be in short utterances and will evaluate all the input sentences for sentiment analysis (after they went through splitting sentences), but will take into consideration also the cases when the replies are pretty complex (in scope of SPCA pipelines). In this pipeline, we analyze only the short utterance replies. Consequently, we can make exceptions from the fundamentals and use as well data augmentation through translation if there aren't any available datasets or pre-trained models for sentiment analysis tasks. We will choose to have as well our own created datasets for fine tuning, as we have experience in screenplays. An intent based solution (Rasa) is suited for this pipeline, since we need to know the kind of reply the user is giving, and we can track different intents.

We will go deeper in evaluating the negative/positive sentiment and split these sentiments in three:

• Confirmation and negation;
• Satisfied and unsatisfied;
• Understanding and misunderstanding.

Please note that the bot is not responding to the utterances identified as the user's replies. It is sending only the sentiment analysis evaluation for the CPL layer. They need to be evaluated

in context in order to understand why the user is having a positive sentiment or a negative one. Using an existing model for the sentiment analysis task, pre-trained on social media datasets and fine-tuned on BST (without the generative model part) can be a back-up solution.

Chit-chat/Emotional Reaction. Our intention is not to build a chit chat bot because we don't believe chit-chat is fundamentally a conversation, on one hand, and on the other the bot should not replace the need of a human presence in our life.

At the same time, in every conversation (even with an expert) chit-chats or emotional reactions are for sure present. As we are working in a limited datasets environment and in an expertise domain, we will restrict the conversation to two consecutive turns. We've already done a cleanup and we are working with short sentences which gives us room for using an intent based model similar to the one from the previous pipeline. For more complex chit chats, we will implement sentence-intent that doesn't need a huge database and is not computational expensive. We will limit chit-chats in this case as well for the same reasons stated above.

This pipeline is very similar with the previous one also in terms of solutions. The differences are the following:

• We will generate a response in a different flow in order to address also the possible core-input in the utterance.

• Some chit-chats are linked with the deep-conversational (elaborate and feedback) and for that we need to take into consideration the Labyrinth model from the CPL layer.

• It's important to add your own scenarios to be consistent with the bot's built-in personality.

Untrained NIU. As we stated in Composed word and Emoji/Abbreviations, part of the user's input must not be included in any training process, but everything that the user is saying needs to be addressed. This is the purpose of this pipeline, to find a middle ground in order not to disturb/ limit the conversation.

We will not use complex algorithms, but only retrieve Composed words and Emoji/Abbreviations, which will be addressed in the NOG/NLG layer.Pirkin 2 model – main flow (core input + deep conversational + back-up)

Pirkin 2 model. Main flow:

For transparency reasons, we will provide detailed description of the next pipelines as well, but SPCA main algorithm is not part of the current research in order to allow other developers to be creative. However, it's important to say that it is an original model, which can be included more at AGI instead of DL or NLP based rules algorithms. We will further comment for an easy development and better understanding.

First, a quick recap of what we have at this moment:

• corrected words,

• CVM/IVM/AVM and DER/NER that are capturing important information,

• eliminated inputs that are addressed in a secondary flow,

• sentences split into complex and simple,

• tuples with words and their POS.

The issues that we have:

• No vector embedding and no stemming, nor uppercase removal;

• Possible UNK words (unknown) in the user input;

• All kinds of POSs, some of them more important than others;

• Cleanup needed and only words are included (without numerical data, punctuation etc.);

• Multiple CVMs that can give divergent information;

• Missing words, as people tend not to use words that are considered implied.

What is our objective? Summarize core-input in a four syntactic-functions sentence-intent with Subject, Predicate, Complement & Attribute.

To achieve this, we will use the following resources/ know how:

• A database for each domain (ex: restaurants expertise, books . . .);

• Existing auto-complete and processing algorithms/models;

• Fundamentals in terms of how the human brain is actually summarizing a possible complex/long input. Some of these fundamentals are specific for each language/culture.

Even if we are not fully disclosing the next pipelines, the essential part of our framework is the architecture.

All this hypothesis pushes us to a middle ground model between DL and traditional (based rules), adding more importance on architecture and on human way of understanding, which will probably not be surpassed by the machine. Not soon, anyway.

Input processing II. Everything that has already been retrieved by another pipeline will not be used in the main algorithm. At the same time, we need to work with stem/base words.

Database processing . In reality, a person understands you based on past conversations that he had with other people (this is the fundamental principle of labeling and it is right). Therefore, familiar words are the first that stand out when trying to understand the other person's utterance. For this reason, we will process the limited database (with interactions in the same environment and going through the same pipelines, as we are doing for inputs) in order to give the bot some prioritization and additional data for word embedding. Using additional resources like the internet or other similar conversation is helpful as you want to cover more topics/wording, but in the end can affect the understanding, as a different environment or grounding gives different meaning to words.

Book's processing. In order to be used for NIU and NOG, all the books that are uploaded in the bot knowledge need to be processed first (going through the same process as the input[43]) because some data is irrelevant and others must be structured. Keep in mind that the model allows for future adding of books and the bot needs to use automatically the added knowledge in the understanding and in the generation processes. This layer is transforming books in databases pretty much automatically (unstructured data into structured), but keeping the interdependencies and the chronological order chosen by the author.

Other significant factors:

• The type of book is important. The model will be able to work with three types: novels, scripts & screenplays (even if scripts are not actually books) and scientific books (all types – psychology and philosophy included).

• The names of the book characters in the novels and scripts are not important in the conversation because we don't need the actual action from the book, but ideas. Deep conversation is not intended to replace the actual reading, but to have intelligent discussions based on some topics and promote the book reading.

• As we don't have sufficient databases to do training it's important to find a way to translate the book into a database as a strategic task.

• For the answer to be correlated with the user utterance, rows 2 and 3 of the answers CVMs need to correspond to the CVM of the input. For example, if the user is talking about the past, the searched answer should be in past tense. If the user is addressing in the first person, then the bot should reply in the second person, and vice versa.

A possibility is to do the same with site corpuses, meaning to process, train and then use them as databases without the need of API integration, if information from these sites is needed.

43 See Cezanne-ai correspondent pipeline for details.

But, as Bateson suggested, "the arbitrary division can be misleading if we don't apply scientific methods". Books have a prerequisite defined structure, they have either literary or scientific value, and we can use these fundamentals to walk through books' labyrinths. On the opposite, sites are heterogeneous in terms of the structures and have a value depending on the users and products/services.

Auto-Complete & Augmentation-through-conversation.. An important part in a conversation is the intuition (other types of intuition will be used in CPL and NOG). Even if some words are not explicit (many times the subject is implicit in some languages, for example), the other person can easily understand the utterance in context and finding the implicit or missing words can be more effective than a past utterance's summarization when we try to fix the long-term memory. An algorithm that tries to discover the implicit words is also useful for the model, as statistically we will also deal with UNK words at this point.

Labeling can solve this problem because machines can be trained to have some deviations (for example 80% understanding of the intent can be considered safe). We believe, though, that an NLP- DL algorithm for auto-complete & augmentation-through-conversation will do a better job because we will direct the training upon a specific job/task, not to mention that we will have a solution for goldfish memory.

We will be using the bi-directional masking task from BERT (as an exception to our fundamentals) with the existing labeled and unsupervised data (having books processed and trained in the model, suited for the domain in question, can be very effective for language models instead of pre-trained models with huge uncontrolled datasets). In order to accommodate our needs and subtract specific context- long-term memory-words especially for the SPCA pipelines, which will come next in our architecture, we will apply masking in 4 directions (in the order presented) on each user conversational turn that implies core-inputs:

• We will mask the first token of the utterances that do not have a subject/NER with the scope of finding the implicit Subject. The extracted word will be searched in the past utterances and if found, will become the new subject. In the next pipelines, we will have additional back-up methods for this task.

• We will mask the UNK words in terms of both vocabulary and POS and apply the same training flow.

• We will mask the last token and apply the same flow. GPT3 can be more effective for this task and then we will apply the same training flow.

• We will mask the last verb of the core-input. If after masking task we are going to have the same verb (or similar) we will consider this verb as the Predicate, if not we will also search in the past utterances, applying the same training flow from the above tasks.

This is the part of the solution in which we use neural networks, but we will also apply other methods for the same tasks in the following pipelines. Having limited resources and the datasets, books and the input already processed, will have as result shorter training times, however this point needs to be tested and depending on the results to make exceptions and use pre-trained models.

Input processing II will be repeated, if necessary.

Embeddings. Now, we need to transmit data to the bot for computation and understanding. We have already transmitted important info through matrices and entities and we must transmit the core input (which is at this moment the remaining corpus of the input). As we don't have a huge database and we created this detailed/unique architecture, it's only natural to make our own word embedded vector & rules.

Let's say we want to describe the pointer finger functions. A solution is to mask the finger (as in the left picture of Figure 2) and then reconstruct it based on similarities with what we

have in the datasets (this is pretty much the current state of the art solution). Another solution is to show the finger, independently, or in a relation with the previous finger, like RNNs (see the middle picture). As we previously mentioned, we are going to refer to Gregory Bateson's work to better understand the concept. A word doesn't have only intrinsic or extrinsic representation, and it's meaning can be extracted due to the relations with other words (the same is with the pointer finger). This solution limits the size of corpora needed to build an accurate embedding if used in a sentence-intent model that has already extracted the important meaning in terms of words and their syntactic functions.

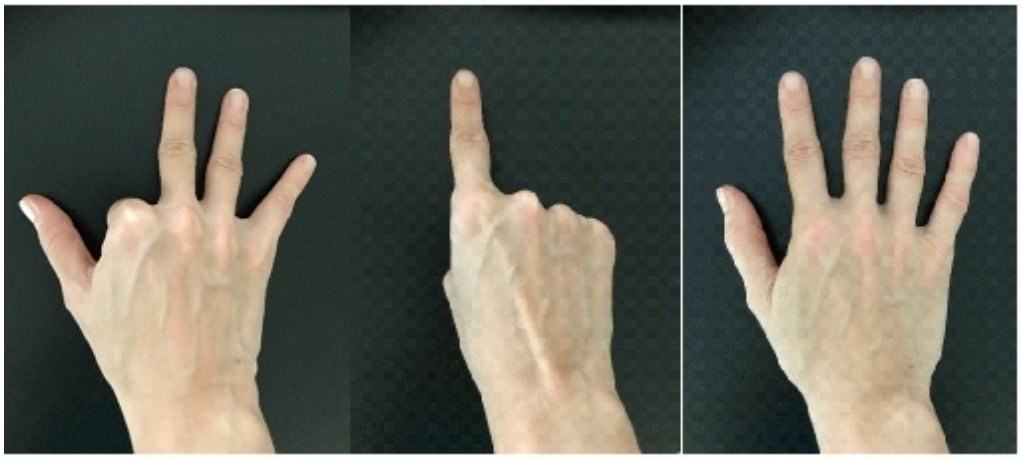

Figure 2
Pointer Finger — Three Perspectives

Word has 4 dimensions in terms of understanding:
- Its intrinsic meaning;
- Its meaning for other words;
- Its meaning inside the sentence/phrase/description;
- Its meaning for the persons/bot to whom you conversate.

Consequently, in our Pirkin 2 model, a word will be vectorized through a special vector[44], which covers the relations between the words in the entire input, as fundamental for the real meaning of the word. The same vectorization process will be applied on database and books. Doing this process will allow us to apply simple similarity functions between embedded word vectors from the input and the ones from the databases in order to extract/generate the answers.

SPCA – will be shared at request.

Triggering/Common interest. In reality, the user's input can be from the most complex to the simplest one. Even if our architecture is doing all kinds of cleanups and is separating core-inputs from replies or chit-chats (if identified), it can be the case that we have a complex and long input in this state, which contains Q&As, opinions, long descriptions, replies, chit-chats... or any combination of this type of inputs. The fact that we summarize everything in a simple SPCA sentence-intent with one subject +one predicate +one complement + one adjective can be seen as limitative in terms of understanding, even if this summarization is effective. Another issue that

44 Please check the step-by-step implementation in Cezanne-ai project.

might occur is to have an input which is incomprehensible or too simple to find a complete SPCA (and the past interactions didn't fill in the missing info).

On the other hand, our objective is not to make a super-machine, but a bot that has the capacity to achieve the human level in conversations. Our bot will make the next decision in these situations, very similar with the human behavior:

• If the input is too short it will initiate further interactions.

• If the input is too long it will trigger the common interest. It's natural for a person not to memorize/understand everything that he is hearing/reading.

• If the input is incomprehensible, it will ask additional questions.

• If the input contains several types (replies, chit-chats, Q&As...) it will do a prioritization. If the user considers that the bot didn't understand everything, he will probably do a more specific follow up and the model is prepared for that.

Domain Validation. Now that we have the SPCA, we will also have the unique CVM that can validate our understanding. The fact that somebody is talking about a domain (let's take as an example: restaurants) doesn't mean that he has an intention (meaning a question, a need, a request). We must take into consideration the fact that the words used by the user aim to describe a situation, or reply to the bot's actions, or chit-chat, and the intention has not been identified in the past pipelines... That's why we need to include the words into a context, by using CVM and results from the secondary flow analysis.

We will split the SPCA domains in 4 categories (plus SPCA0 – if no domain is identified).

The bot will also be able to have deep conversational discussions (SPCA4) based on its own personality defined by the books available in the database. The latter can be increased by adding new books on the open-framework – without having a data scientist to prepare the database for training. Specific training will be defined in NOG for each type.

This pipeline differs from existing query classifiers due to the CVM evaluation in the processes.

SPCA- memory update. One of the main reasons for choosing this model is the possibility to keep past interactions in a structural way and update the sentence-intent/SPCA/input with every turn, covering the multi-session-conversations. Usually, a user will not repeat himself and many times the subject/predicate or even complement/attribute remains unchanged throughout multiple turns or until a final answer is provided by the bot in the consultancy session. A reset pipeline in CPL will decide the information that needs to be kept, depending on AVM/IVM, and will trigger an algorithm in the present pipeline, SPCA-memory update, which is choosing how the final SPCA will look like, by considering the present SPCA and the past one. Implicating the reset pipeline in the memory update will be beneficial, in order not to create redundancies or use already dismissed past utterances that a summarization or a retrieval will not be able to take into consideration.

In order to implement a long-term-memory update on a backup model that is using an E2E model we will add to this pipeline a summarization-augmentation (Xu, et al., 2021).

Chronologically, this pipeline must be after CPL and NOG, but because it is an important pipeline for the NIU we will present it here. Please note that at this point, the pipeline augmentation-through-conversation has already made some steps towards the long-term memory update.

Main flow/ Back-up model. The three models that we will implement as back-up[45] by using the same datasets as for SPCA model are:

• Encoders-decoders model for chatbot task (meaning T5);

• Open domain dialogue model with summarization memory-augmentation (meaning PARL.AI);

• Open-Source Language Understanding with Dialogue Management (meaning RASA).

As we dismissed the solutions offered by data augmentation methods and we are working in a very limited environment in terms of pre-trained models, emerging languages and datasets/ corpora, using E2E models, even the most recent and advanced, would not give us satisfaction. But there are some arguments to use them anyway, besides the fact that they will give us a clear picture of our framework efficacy:

• Implementing these models after a clean-up process (the pipelines from Machine Education and 4 pipelines in Machine Learning) will have a beneficial impact and diminish the need for huge datasets.

• Using three models instead of one increases the chances for success.

• We will use the validation given by the three matrices (CVM, AVM and IVM) also for the back-up models.

Steps:

• Use the core-input as it is, before the Input-processing II, stemming/lemma processes.

• Verify if the back-up model was initiated by pipelines in CPL or NOG.

• Check if one of the back-up models (the order is the one presented above) is providing answers to the current core-inputs.

• Apply in the reset pipeline (from CPL) different methods to accommodate long-term memory: like summarization-augmentation, retrieval-augmentation.

• Validate the possible answers provided by the back-up model inside the CPL layer, by using the CVM/IVM/AVM.

• Adapt these E2E models in order to be suited to the big architecture of our framework.

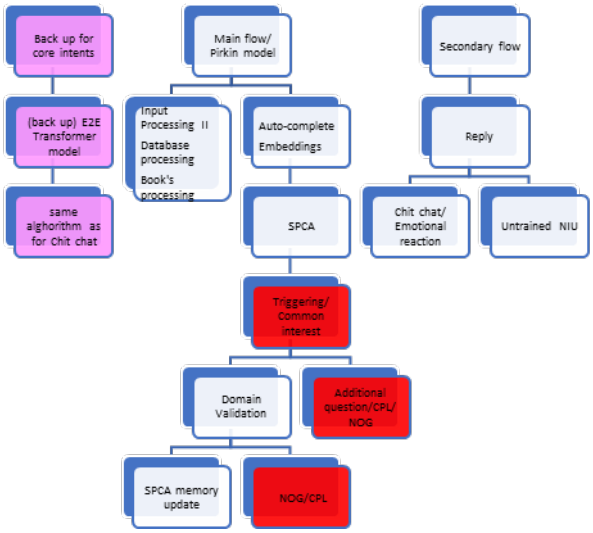

Figure 3
NIU/NLU Machine Learning Layer

45 More details about these models are available in Cezanne-ai project.

These are the principles behind a Machine Learning layer and we are proposing the above architecture (red indicates outputs from NIU).

4.2 Conversational Policy Learning (CPL)

4.2.1 Pirkin 3 model in 11 pipelines.. A correct terminology of Dialogue Policy Learning, that is frequently used in the conversational AI papers, is Conversational Policy Learning, as dialogue is only part of a conversation. We will use a fundamental approach as well in order to implement CPL. In every human conversation there are rules and people do not conversate telepathically. Because the conversation is based on rules, we will not consider current approaches in dialogue management (Bocklisch, et al., 2017), as using trained/pre-trained models to establish right policies might corrupt the conversation. The fact that the majority of users agree with some conversational policies, doesn't mean that they are right.

There are 4 conversational known types/states:

• Discourse (we referred to it as description; a consultancy session);

• Dialogue (we will refer to dialogue when we have specific sentence-intents validated by CVM);

• Diatribe (in here we will include chit-chats/emotional reactions);

• Debate (this is the opposite of the dialogue, because the bot will have the initiative: asking questions or trying to deal with the user's reactions).

We will add three additional ones to the four above mentioned:

• User Queries (similar to dialogue, but the user has more specific questions, which can be addressed by a specialized chatbot). Examples:

1. The user wants to know more about a specific NER1 (for example, the restaurant's name) and the conversational bot will be replaced by a commercial/ restaurant chatbot developed inside the same framework.

2. The user has specific discussions topics (deep conversational) and book queries can be a solution.

• Bot Queries (similar to debate, when the bot has multiple questions, for example: he wants to complete the reservation to a restaurant. Also, in here the bot will try to use its intuition to open different discussions based on the self-generative policies; for this reason, we will not refer to them as prompts or forms);

• Special policies. This is a state where the bot doesn't need to understand the input (but to request it and send it forward) and/or doesn't need to provide an output (for example when the user's input is rhetorical or is a review/opinion). Additionally, policies like no action from the bot's perspective, or no reaction from the user's perspective are included here.

In order to be empathetic with the user, we will design a 3rd interaction flow (inside the debate state).

States will be assessed using the following data:

• matrices AVM/IVM/CVM + DER;

• type of the SPCA: SPCA1/ SPCA2/ SPCA3/ SPCA4/SPCA0 or type of the back-up model;

• implications of the DL training model from pipelines: Reply/Chit-chat + untrained NIU.

As CPL is depending on NOG too (the past actions of the bot can influence the policies also in multiple turns conversations), it is mandatory to go through with the NOG layer in parallel to understand better the implications/correlation.

Reaction analysis. Reaction in this case refers to the feedback/responses of the user due to bot's actions. The bot must carefully analyze these reactions that can be standalone inputs or part of a more complex input/user turn.

Firstly, it needs to classify the reactions as positive or negative and, in some instances, to see more specifically what type of negative or positive reaction is transmitting in order to move forward with the conversation.

For this reason, we need to assess 5 NIU pipelines:

• Ironies (composed words);
• Emoji/Abbreviations (classified as smiley face and sad face);
• Untrained NIU – metaphors, quotes/expressions;
• Replies;
• SPCA2 (replies identified by Pirkin 2 model).

Assessing the last two means a training & labeling database that is presented in the NOG layer. CPL needs the information now because the bot's reactions are not generated but are hypotheses for NOGs.

Reset . After every turn, a decision regarding the information that needs to be kept from the previous turn has to be taken. This process is complex because we have 3 types of NIU models in place: Pirkin 2 main, back-up and secondary. Depending on the type of model used in the previous turn we have three different structures to work with. After the reset pipeline, the Memory update pipeline will decide how to combine past and present inputs of the user.

States. Now that the bot chooses what to use from the past interactions/turns, we need to assess the current input and the previous state (if the case) in order to know the state of the conversation. For each of these states we will have further policies that will help us in the NOG layer.

Possible states:

• Discourse – with 3 levels of advisory depending on the importance;
• Dialogue – with 3 levels of Q&As depending on the importance;
• Diatribe – chit-chat;
• Debate – art of conversation with controlled policies;
• User Queries – with 2 categories (commercial chatbot (1) & deep conversational through book queries (2));
• Bot Queries – with 2 categories (virtual assistant/forms (1) & AGI with intuition (2));
• Special policies – with four categories (the user is giving a review (1) & the bot is asking for a review (2), no action (3), no reaction (4)).

Discourse policies. As the discourse is a straightforward state, the 3rd flow (empathetic flow) will not be used in this phase.

We have 3 types of discourses initiated in the previous pipeline:

• Discourse 0 – we are not 100% sure that we are dealing with a discourse.
• Discourse 1 – we know that the user just inputted a discourse.
• Discourse 2- the user is making consecutive discourse inputs.

After assessing the CVM, we will apply the following policies:

• Give an answer – NOG1.
• Request to keep the review/opinion from the user.
• Ask for a confirmation.
• Ask for additional info.
• No action.
• Respond to the user that the bot does not know the answer to his input.

Dialogue policies . They are very similar with discourse policies, with 3 important differences:

• We will not ask for additional information or confirmations – the policies are straightforward.

• The bot will ask for detailed reviews, not for existing.

• This state has many interferences with user queries. User queries are defined as successive dialogues.

Diatribe policies. The line between a conversation and a deep conversational dialogue is relative. That's why we are going to use CVM to make some actions inside an input that was assigned to SPCA3 or chit-chat (diatribe state) in NIU/CPL:

• Give answers NOG3.

• Cancel the answer.

• Redirect to Debate.

We will propose 10 chit-chat categories to perform a better assessment.

Debate policies. How to handle the cases when the user doesn't agree with you (or is not satisfied with the bot's answers/actions) or agrees with no other continuation of the conversation which is crucial in a real conversation. For this reason, it is important to show empathy and try to change the state of the conversation or return to a previous one. The debate policies are applying also to the cases where the user is responding to the bot questions.

The bot needs to be proactive in debate and take charge of the conversation. If not, it will lose the user's interest. To do this the bot will apply the following strategies:

• Initiate a 3rd flow conversation – NOG4;

• Change topic;

• Continue the process + answer;

• Back-up answer.

We will have 2 different approaches to debate states depending on the originated state/condition:

• Initiated from other states (dialogue, discourse, diatribe, user queries) due to CVM assessment;

• Initiated due to different reactions of the user to the bot actions (answers/questions. . .).

User queries policies. This state addresses the situations when the user discloses his exact intentions, and he wants to elaborate on them. It is very similar to Q&As but more specific on the topic chosen by the user. Two main queries:

• Detailed information on a specific domain (a restaurant bot for example or a bot that is specialized on a front-end development of a chatbot) – user queries 1;

• The user wants to talk more about the topic of books. . . - user queries 2.

Chit-chats/Diatribes are limited to 2 consecutive ones. Dialogues are unlimited but can easily be transformed into queries. However, queries are limited to the user choice or database capabilities. Another particularity of user query is related to the state exit rules, which becomes more limited in comparison with other states. Once you enter a user query, the bot will understand that the subject and also the predicate (in many cases) remain unchanged.

Labyrinth network model – also an AGI concept, applied for user queries 2. In a conversation, especially in deep conversational, many sensitive topics can occur, and the current models have important limits to deal with this kind of situation. Being politically right, for example, has an important impact in the way we are conversating, as people started to see even a simple conversation with increased attention to sensitive aspects. Even the existential or philosophical conversations became challenging due to many changing and different optics and the fact that numerous people with basic knowledge have strong opinions about most of the subjects. A solution is to implement some rewards for E2E/DL conversational AI (specific to RL) in order for the bot to adapt. Another solution is to implement supervised learning and carefully use labeling

to be as professional as you can and not take into account the dynamic environment in which the conversation is taking place. Finally, a solution is to build the policies as strict as they can possibly be, even if this will translate into an important limitation to the conversation.

As we are looking to build a human-level AI for deep conversational (replies and chit-chats are not included here) and thus we introduced two additional states, we designed the labyrinth model to cover the following situations (please note that the possible answers will come from different types of books that will be trained – the training methods will be presented in the layer NOG from books):

• User's feedback, which is different from reactions, similar to bot intuition presented in the next pipeline. This feedback will have impact on the bot's behavior – see Self-Generative model:

1. Have doubts regarding some ideas/topics;
2. Show interest in the bot/book writer opinion;
3. Show interest in the writer as a person;
4. Wants to share his thoughts.

• The user wants to elaborate on the subject and wants to see the bot's answers which are retrieved from the book (see pipeline NOG from the book on how books are classified and how the training is done). We will use specific fundamentals from literary art:

1. Expositions - The user wants to go back to the introduction/title of the chapter/subchapter. The initial input that was classified in other states was already answered;
2. Plot – The user wants to know the intrigue related to the subject;
3. Core action – the main corpus of the chapter;
4. Highlight – the most frequent topic in the chapter;
5. Outcome – The conclusions for the subject in the book.

• The user is interested in the main subject/theme of the book.

• The user wants different approaches to the subject from other books or chapters.

• The bot considers that it is a good time for promoting the book.

All these situations are very much similar to a labyrinth and will represent the deep conversational foundation. The proposed solution comes from fundamentals.

"Art is an extension of the masterpiece that the nature is." – Gregory Bateson.

Bot queries policies (prompts/forms) – part of the AGI model. As we previously argued, conversation doesn't mean only Q&As, role plays and chit-chats. At the same time, conversational AGI doesn't necessarily refer to an answer generation, but can also mean generating a policy in a specific conversation that can ultimately make the user more engaged. Our focus will be on the latter. Conscious thinking accounts for limited percentages, the rest is unconscious thinking and doubt is the main driver of existence. This translates into a high probability that any conversation sooner or later will become unintelligible and not only due to the bot's limited capabilities, but also due to the specificity of the user's behavior. To tackle this issue, at some point, we need to have more than debates policies, where the bot is having initiatives. We need to have a structural thinking, in more than one step, in order to drive the discussion on a controlled path.

Characteristics of the pipeline:

• Assessing DER. Not only in terms of information/data but also in terms of relatively vector in order to assess the "certainty" of the user related to the main subject.

• Identifying criteria for the cases the user wants to exit the bot (cascade) queries.

• For core inputs/forms, for example, the policies will be controlled and the objective is to have all the information needed from the user (DER assessment). The model will be similar with the appropriate slots thinking.

• In case of Bot queries 2 the bot will decide the policies based on DER- relatively vector (all these percentages need to be further revised in alpha/beta testing for each language prototype).

• If the relatively vector is high (>90%) or low (less than 10%), no queries are initiated considering that the user is sure/certain regarding his intentions/request.

• If relatively vector is within 10 and 90%, the bot will intuit that it's important to have a deeper understanding of the user's needs.

Intuition.

A more substantial input regarding our view on AGI will be disclosed in the next layer (Natural Output Generation), but in this pipeline we will lay the foundation by initiating policies that could enact the bot's independence without randomizing or programming. The bot intuition will be based on the level of the relativity that specific adjectives/adverbs can imply over the sentence-intent. Depending on the relativity factor, the bot will engage into a more complex conversation by having 4 possible policies:

• Has doubts regarding some ideas/topics;

• Has interest in the user's opinions;

• Has interest in the user as a person;

• Wants to share his thoughts.

All these policies are symbolic of human behavior in a conversation. For example, doubts will make the bot as close to a human as it can possibly be.

Special policies (Reviews/Opinion/No action/No reactions). This state doesn't have a NOG correspondent, because the bot doesn't need to give an output. At the same time, prior to initiating this state, the user must agree to offer a review/opinion, responding to NOG- Review questions.

This state is only initiated by other states and doesn't have additional activation conditions.

In terms of review/opinion, we have two different situations depending on the user active-ness/reactiveness approach to the matter. In both cases, this state can only be initiated from the debate because we need the acceptance of the user in order to capture the review anonymously (or it is a decision of the user to sign the review/opinion or use a pseudonym after being informed of the implications) due to personal data regulations.

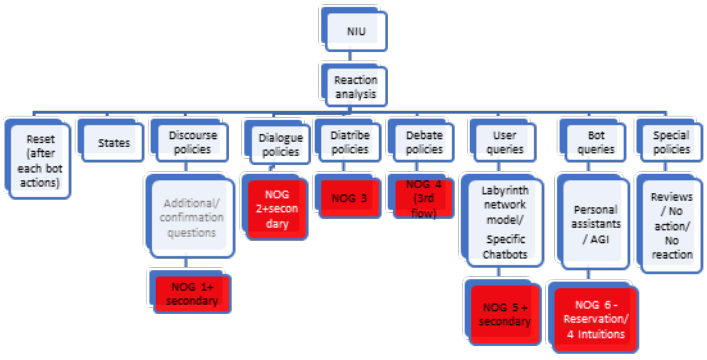

Figure 4
CPL layer

These are the principles behind a CPL layer and we are proposing the above architecture (red indicates outputs from CPL).

4.3 Natural Output Generation (NOG) in 16 pipelines

As one principle is clearly stating,[46] the verbal/written part account for a small percentage in communication. Limiting the conversation to understanding the user's language and then generating a natural language response (for example through NLG/seq2seq algorithms, most of them also used in translation) will not have the expected effect even in a business/commercial environment. The social/emotional behavior of humans will in the end determine a gap in the user-bot communication.

For this reason, we will not use Natural Language Generation (NLG) to define this last layer. We consider Natural Output Generation (that incorporates NLG) to be more exhaustive. The components/objectives of NOG (some of them already introduced in the previous layers, others will be defined further) are the following:

• Respond in accordance with the user's formal/informal communication. If the user utilizes formal terminology/language, the bot's generated answer/reaction needs to be formal and so on. Seq2seq can accommodate this requirement, but the conversation can become non-consistent. It is a risk to alternate the formal/informal approach depending on each sequence of the conversation. Furthermore, it will seem odd to the user if the bot is sometimes informal and other times informal, depending on the understanding of each turn in conversation. We will propose a more consistent solution to this requirement in the next pipeline that applies for core inputs.

• Have a pessimistic/optimistic reaction based on the self-generation model. To duplicate human behavior, a neural solution might not be enough, as humans react not only based on the brain's pragmatic inputs. External factors and mood can play an important part. Thus, we will propose a solution that applies for deep conversational inputs.

• Make visual contact. Many times, an inflexion in the voice, or a facial reaction can transmit more than 1000 words. Current models that are based on text and speech recognition are limiting the conversation by not having specific solutions for this requirement. Also, in the NLG enhancements pipeline we are proposing a solution, using avatars.

• Behavioral changes over time in the bot's answers. The same questions addressed in different time frames could have adjusted answers depending on developments in the word and on realities. Those will not impact core inputs of expertise but will impact deep conversational inputs that are defined by us as questions regarding human day-to-day issues linked with their existence. See also NLG enhancements.

• Confidence understanding. To generate the answer for the SPCA model a confidence answer indicator is computed. Answers will be given only if minimum 2 syntactic functions (from the 4 possible: Subject, Predicate, Attribute or Complement) have correspondents in the database training, taking into account the final SPCA (after memory update and possible additional questions from the bot). If one syntactic function identified in NIU doesn't have a correspondent in the databases, then the confidence indicator is 75%, in case of two, the confidence indicator is 50%. In these two situations the NOG answer is preceded by a disclaimer explaining the user that the bot is not 100% sure that the answer is the right one. The model has back-up functionalities for these situations in order to recover, but we consider that this kind of approach will benefit the conversation and the user will have sufficient understanding as in numerous cases the answer is wrong due to many causes which do not depend on bot capabilities. More detailed information regarding this enhancement will be provided in next pipelines.

• Reacting to emoji, abbreviations, expressions, ironies, quotes, expressions, etc. We've already discussed this, and our solution is the secondary flow.

46 See detailed 'Fundamentals' on github

• Intuition. The bot needs to have in specific moments the possibility to have its own impression/impact on the conversation topics/states. Thus, we have previously introduced Bot queries 2, which will be detailed also in this layer in terms of NOG from intuition. At the same time, if the user has some intuitive inputs, the bot will comply with his behavior in NOG from book queries or from other bots.

• Understanding implied words (not part of the inputs). We proposed a solution as well in Auto-Complete pipeline.

• Simplification of labeling for NOG that corresponds to Pirkin 2 model. Due to research purposes, we are going to use simple-to-complex utterances for labeling, in order not to favor our models. Using SPCA for summarizing the input of the user, implies not having a specialized labeling (including for instance screen-players in the process). Labeled utterances can be basic sentences, not necessary with all syntactic functions specified in SPCA. This is an important advantage for our model and is due to the fact that we are not limited to language understanding and language generation and we are analyzing the entire input (with a principle of divide et impera) to generate NOG.

• Generation of natural language responses (NLG). As we've previously argued, we are not interested in using the open-internet and huge databases to accommodate this requirement, even with a qualitative clean-up of the databases or even if we had copyrights. At the same time, we consider that using other users' conversation or making interfaces with other external bots can affect the rights and privacy of users. We will use past conversations for testing/automatic training but not for NLG purposes. Instead, we will use experts, published books with copyrights and in-house databases built by screenwriters to accommodate NLG requirements.

4.3.1 Training and additional models..

Database training[47]. This is a complex part in the model because we are dealing with 4 different types of database training (some types of training are not, however, similar with the AI concept of training):

• Pirkin 2 model database training by using experts' answers (core inputs). Experts in different fields (programmers, AI/NLP engineers, doctors, lawyers, HR specialists, investment advisors, restaurants evaluators...) will provide answers to different possible questions or to advisory/consultancy sessions (which can be addressed by the user in a million possible ways, in short inputs or complex phrases). The bot must identify if the user's input, summarized in a SPCA sentence (received by the bot in 2 vectors: k-word + embedded vector for each word in the SPCA) has a correspondent in the expert's answers. For this scope, we will need a database with questions/answers or situations/information that can be obtained from a specialized site, from an interview with the expert or even specialized books. The bot cannot interfere in the answer and generate a different/adapted answer (in SPCA answer, slots included in NER and DER will be treated as slots types). The labeling can be done with very simplistic utterances as an advantage for Pirkin 2 model. This will provide a very smooth training process without an important focus on testing, which anyway will be in scope of Cezanne-ai project when we put side by side the results from the 2 models' implementation (SPCA vs back-up).

• Back-up answer for core inputs is generated by one of the three E2E models. We made the setup in NIU, so there is no need for further clarifications.

47 The training steps will be presented in Cezanne-ai project and in *Experiments* we will make a comparison between databases used for SPCA and the ones used for back-up models.

• Pirkin 2 model database training of chit-chats (in a similar way, we will use databases for reply to feed up the Reaction analysis from CPL). This process of training will be close to the first one, but NER and DER are not an issue.

• Secondary flow training (chit-chats and replies) is a proposed intent-based model that comes first in the NOG flow.

All utterances in the database will be trained and we are going to allocate a notation to each correspondent answer to check the duplications in the conversations.

It's important to state that the training process will be ongoing and will be automatically updated - see Auto-training pipeline.

Book training. Using books for query purposes can be regarded as a continual learning process of the bot. Since the bot is positioned as an expert, what better way to further improve his knowledge base?

We will use the Book processing results in terms of paragraphs, which are keeping the original chronology and dependence on the book/chapters/subchapters. This process is similar with training databases for Pirkin 2 Model- SPCA, except the labeled answers which will be in scope of Labyrinth model and described more in the NOG from books.

NLG enhancements – can be assimilated with AGI concept. To define the personality of a bot (meaning general personas) we can use current custom weights (Li, et al., 2016) and configure a bot that has conversational capabilities, duplicating in this way the personality of a human by using advanced AI models and methods. It is, however, a debate if the users find a good use of time in speaking with unknown duplicate-humans (which do not impress in terms of character, but mostly in terms of data/information/topics), especially when our objective is "the art of conversation".

• Formal/informal speech for core inputs;

SPCA3 or secondary flow inputs can trigger informal responses/addressing and this can depend on a scoring model. Once an informal type is triggered, it will not be changed in the interaction with the same user, one of the reasons being the fact that conversation is defined as informal in Webster dictionary and this is one of the objectives of the interaction user-bot. All NOGs (answers/questions, other) related to core inputs will have both formal and informal alternatives in order to adapt to the user vocabulary/addressing. For this task a writer/screenwriter input is needed. Other forms of extracting or generating answers will not be helpful to tackle this issue.

• Pessimistic/optimistic behavior for deep conversational inputs (SPCA4);

This enhancement will be addressed in the next pipeline. Depending on the research results, these functionalities can be extended also to core inputs, but at this moment, we cannot evaluate the assimilated risk if we let bots impact responses that are coming from expertise field/persons.

The bot will offer answers matched with the pessimistic/optimistic indicator computed in the previous pipeline.

• Avatars – visual state;

Besides having a scoring related to informal/formal speech, SPCA3 and secondary flow can change the face reaction of the bot through the avatar, which will replace the bot's icon every time the scoring changes, letting the user know the bot's expressions and state of mind. The visual impact can be higher than the one of words in many situations. Avatars will not have an impact on NOG/DPL but can determine the user to take actions if he feels that there is a gap between his state of mind and the bot's.

• Behavioral changes over time in the bot answers;

This enhancement will be also addressed in the next pipeline.

• Confidence understanding;

More details regarding this enhancement will be presented in the SPCA Answers.

Self-generative bot model — an AGI concept. We will go into more details regarding our objectives presented at the beginning of this article. What is our scope/purpose?

• Building bots with the intention of making money? Bots with commercial/business value?

• Deploying robots that can write books/songs or can make jokes and create art or ideas based on provided databases? Furthermore, the idea that a robot can go deeper than the code and the databases and become alive can be very attractive.

• Making expert robots with knowledge on different fields that can provide advice/education or with who you can have an interesting conversation?

We have chosen the last option, as we believe that everyone has the right to the top legal expertise, best doctor advice, most influential CEOs/ politicians/ models/ influencers and so on. Another argument is our belief that this objective is realistic and we can add human characteristics to the bot. In this layer we will extend our enhancements to build the self-generative part of the bot. But not every user will have access to this part, as he needs to check some cumulative boxes:

• The conversation must be informal (see more in the next pipeline);

• The user must open at least one discussion classified as SPCA4;

• The feedback was used previously in the conversation (see Labyrinth model).

The principle behind the self-generative bot model is that the external factors/interests/profile and time (as an impact factor) can have influence on the behavior, defined by us as pessimistic/optimistic or trustworthy/doubtful towards the bot's answers. The bot's mood/behavior will count in NOG answers generation. It can start to have doubts (for example) in its own database answers, if the self-generation model provides the necessary incentives.

4.3.2 Pirkin 4 model. NOG from databases .

Answers (NOG1/NOG2 or NOG3 without deep conversational from NOG books). If CPL initiates NOG1/2 or 3, it means that the following type of answers will be given by the bot:

• Answer to SPCA (core inputs or chit-chats);

• Back-up answer for core inputs;

• Secondary answer for chit-chats;

• Answer to SPCA + secondary;

• Back-up answer + secondary;

• "I don't know" reply – at least 20 types of answers that will be provided randomly and without duplicates.

Before answering, the bot will have to take the following actions:

• Check if an answer has already been given (check AVM). If yes, check if a back-up answer is available, if not make a disclaimer for duplicate answers.

• Assess all NLG enhancements and the Self-generative model in order to give an exhaustive answer. Use disclaimers if the case. Assess DER (specially the budget vector that may have implications on the answer). Apply important criteria:

1. If a secondary answer is given, the answer for SPCA3 is not provided (if the case);

2. Conversational analysis will have impact on the training to be done first (for the back-up or SPCA);

3. If the SPCA database training doesn't identify an answer, then training for back-up flow will be done. The same applies when back-up flow is initiated first.

Chit-chat & Untrained Answer (NOG3)[48]. This answer comes first in the conversational flow, if the bot identified either in the Secondary flow, or in SPCA3 a correspondence in the database training answers. We've already addressed many implications in the NIU and CPL related to this topic and now we are describing how the output generation will work.

SPCA answer (NOG 1 and NOG 2 from database). The main difference between NOG1 and NOG2 is the need for fluency of the flow in terms of dialogues (NOG2 correspondence) and the need for better assessment of the user discourses by using additional questions and confirmation questions (meaning NOG1).

If we have already had a chit-chat/untrained answer, the SPCA answer will be the second in the conversational flow, because it's important to have the core answer last in the conversation, as the user tends to focus on the last thing they are reading.

The standard for our model is to have Pirkin 2 main model/SPCA checked first for a possible answer and the back-up to be the E2E/DL models. But, in time, the Conversational analysis pipeline can automatically change the order and make it similar to the Chit-chat answers where one of the E2E/DL models is the first choice and SPCA is the back-up.

Back-up answers (only NOG 1 and NOG 2 from database due to no solution for books training). More important in this pipeline is to integrate a possible back-up answer provided by the machine in the general model. The answers will be given as part of the E2E model initiated in the NIU, similar to secondary flow chit-chats. Two important aspects impact the model:

- The type of the back-up answer (linked with task-oriented requests, related to main NER 1.0 – another core input);
- Not to duplicate responses and have a common classifying rule for states with a consistent AVM for all types of reactions/actions of the bot.

Conversational analysis (AGI). Conversational analysis is an AGI concept, which we are proposing, and has two main objectives that are common to human behavior in a conversation: live self-correcting and automatic adaptation. Both have the same characteristic: change, because humans are always trying to adapt not only to the environment and to social constraints, but also to different conversational situations that cannot be captured by a programmed CPL state. Let's take as an example a call center employee that is going through training and role plays and is prepared by the employer for different situations. What will the employee do in case of an unexpected situation? A common solution is to ask for the supervisor's help or to come back later with the response. But all these solutions are on the client's time and many of us had bad experiences with call centers, not to mention that rethinking and restructuring the activity to the rapidly changing environment means a lot of time and money for the businesses. Our solution is to educate the bot and prepare it better in order to be more efficient and solve an increased percentage of the user's needs.

Without change, humans cannot evolve and sometimes post-administration or post-training of a conversational bot is either coming too late or is addressing other commercial needs and not the need of the machine to better perform in a conversational environment.

The fact that we are choosing to implement two simultaneous models, Pirkin 2 & AGI concepts, on one hand, and an E2E DL model on the other, allows the bot to immediately adapt and make automatic changes. This approach is also helpful for the multi-domain open-framework, because different industries and databases can work better with a model or the other, and this doesn't necessarily mean that Pirkin 2 model is better than E2E DL model (or the

48 Detailed steps in Cezanne-ai project.

other way around) but it means that one model is more suited than the other, depending on the configuration of the bot.

The conversational analysis will have two possible outcomes and both impact only core-inputs:

• Offering a live back-up answer if the user has a negative reaction to the bot's initial answer (the solution has already been presented in the previous layers);

• Choosing the best suited custom model for a configured bot depending on the user's reactions in time.

Auto-training (AGI). "Learning never stops." – Gregory Bateson

One important issue is post-roll-out training of the conversational bot. We are saying that the customer is always right, but the current NLP platforms allows the administrator/supervisor to change the labeling of the unanswered utterances or to change the labeling of some intents, if he considers that the given answer was not correct, without understanding the client's reaction, the context or the impact of the changes, constantly adding to the database. The hypothesis is that in this way the bot will become more and more intelligible and efficient, but please find details in github- *Fundamentals* regarding our position related to the sophisms and the misleading conclusion that more is better in every situation.

Having a supervisor or several, that sometimes have different evaluations of the same situation, is creating more confusion. Furthermore, the evaluations are made annually, or quarterly, or monthly or due to the client's reaction. You risk being a bad manager if you perform evaluations constantly and you do not have all the context. The same is with bots and with post-training. Our research is focusing on the current MLOops and continuous AI pipelines to understand better what an efficient post-roll-out process should be in terms of training and labeling. As we are discussing digitization, our proposal is to automate completely the post-administration of the conversational bots and assure a standard in terms of results. If the results are not in line with the company's expectation, a better solution is to change the model, than to risk damaging the entire process.

We have three principles of our Auto-training proposal model:

• Everything the user is inputting will be automatically considered for NIU for machine education purposes, processing and embeddings. This way, the bot will learn semantics, better relations between words and will adapt to the right frequencies of words in the domain they are used.

• CPL and NOG are not affected at all by auto-training. The companies will have reviews/opinions captured by the bot, which can be analyzed.

• Every month the bot will provide a list with SPCAs that were retrieved from the user's utterances with an unsatisfactory outcome for the user. The administrator/supervisor will choose to train them by including them in the post-roll-out databases.

Questions/Change Topic/Disclaimers. We will apply four principles:

• Not to duplicate the questions;

• Take into consideration the context;

• Excellent build of CPL, especially debate state and reset, in order to take into consideration the user's reaction and the history;

• Choose the action that needs to be taken.

4.3.3 Pirkin 5 model. NOG from books queries or from specific/commercial bots. A solution is to query the books using sparse or dense retrievers, which are searching through a document and finding answers by using advanced readers. Furthermore, there are solutions that provide additional questions linked with the topic, if the user has additional questions.

Why do we choose not to use existing models directly on the user's utterance or on the summarized SPCA?

• They cover only limited languages and, from our experience, this is affecting drastically the results, even if we adapt the code.

• It is very important to keep the same conversational UI as much as possible and not to redirect the discussion on external UI, which can affect the conversation.

• It is not a fundamental choice, as we explained in the Books processing pipeline. Slicing or dividing without a scientific method will give subjective results, which are not in line with what deep-conversational is supposed to be. On the other hand, we are proposing the same principles applied by the author when he wrote the book.

Deep conversational answers (NOG1/NOG2 or NOG5 only from books NOG). There are 3 types of books that our model will accommodate. The books were processed and by this time we have structured databases, and we need to accommodate answers based on a labyrinth model, a self-generative model and CVM.

NLG enhancement and DER will not have an impact on these answers, but the self-generative model will have and the bot will search for optimistic/pessimistic answers, depending on the model output.

Answers from specific/commercial bots. Our framework allows us to integrate a commercial bot, which is mainly a Q&A bot (the task-oriented capabilities are consolidated in the Bot queries 1) and has the benefits of becoming a sales-oriented environment integrated with the advisory environment.

Let's take some examples in order to understand the benefits:

• A restaurant wants to integrate a chatbot. This allows the restaurant to offer more details to customers such as: parking capabilities, menus, Q&As, paying methods, smoking and children's policies.

• A specific medical cabinet wants to integrate a chatbot that offers medical assistance. In this way it can give the user additional information about the program, prices, type of services etc.

• Same for law firms, companies that are hiring, banks, broker houses and so on.

Specific chatbots become active when user queries 1 state is initiated, due to NER1 identification in the user input. This state has also exit conditions provided by CPL.

4.3.4 Pirkin 6 model. NOG from intuition. A conversation in which the machine is defensive and asks questions only for a better understanding of what the user is saying, will be very short and probably not so productive. On the other hand, it's not appropriate for the bot to become aggressive and detour the discussion from the initial objective. We are choosing a middle ground by allowing the bot to use intuition in three situations:

• Task-oriented or bot queries 1. The user is letting the bot know that he wants his urgent help, even if he hasn't provided full data. The bot should understand the user's needs and complete this task/form.

• Clarifications – bot queries 2. Many times, humans are using adverbs, superlatives, or adjectives that have a relative meaning and can change the understanding of the core-input. We have already talked about this issue and the solution is DER relatively vector.

• Debates. The bot needs to understand when the user is entering a debate and be intelligible in this conversational state by being able to make pros/cons and keeping in mind at the same time the initial objective, in order to have a conclusion to this debate. This is the biggest challenge of a conversational AI bot and we are not aware of solutions provided by AI/AGI until now.

NOG 6 – bot queries 1. The present forms pipelines are sufficient for this one. We will make, however, some comments.

Keep in mind that we had designed the bot without the need for authentication, eliminating all the data that has implications in the personal data regulations. For this reason, workarounds need to be implemented depending on the businesses. A codification system can be a good solution as in the future we will have increasing concerns over the user's personal information.

For example, in order to make a reservation at a restaurant it is not necessary to provide the name and phone number, you can confirm the reservation by using a random code or pseudonym that both parties are aware of. If something changes in the reservation status, the communication can be done through conversational bots, and not through any other channel as the phone.

As the bot is built for businesses that are using advisory and not aggressive selling approaches, the forms can be limited to:

- Make a reservation (at a restaurant, hotel, art & culture event. . .);
- Schedule a meeting/appointment (with a doctor, lawyer, banker, broker).

NOG 6- Bot queries 2. One solution for implementing bot queries is to provide the bot with engaging content. This solution comes with two problems. 1. We are still missing a strong ability to transfer to a new task, one of the most fundamental open problems in machine learning today (Roller, et al., 2020). 2. Letting the bot initiate new subjects for discussion can lead to negative reactions from the user and the bot could face a risky environment.

As an alternative, we will use Socrates' maieutic approach and guide the user, not influence him.

Bot queries 2 has two objectives:

- Eliminate the need to have huge databases to cover all possible implications of a relatively DER.
- Make the conversation more interactive and clarify the user's expectations.

NOG 6 will provide the steps to have these clarifications and, at the same time, not lose the initial core-input that will be put on hold until things are clarified or the user is exiting the queries.

NOG 4 – 3rd flow. The fundamental question that we need to ask in order for the machine to formulate a NOG4 strategy to handle the debate is "why has it come to a debate?". If we have a debate due to positive reactions from the user, then it is not so concerning and an empathetic reaction can be enough. The problem is when the user didn't get what he wanted. Usually there are two reasons for that, correlated with the 6 types of actions that were presented in the debate state- 3rd flow:

- The user doesn't agree with the bot's answers;
- The bot provided an answer without having all the pieces of the puzzle.

Some random questions based on the policies are not enough in order to solve these issues. We need to address the matter very directly and link the NOG4 with the actual conversation by using the user's vocabulary and intention against him (metaphorically speaking).

We are proposing the above NOG layer architecture.

5. Experiments

One of the most important objectives of the research is to find solutions for limited resources available for emerging languages and small/medium-sized companies in terms of data and computation. As our framework is covering a multi-language and a multi-domain implementation, Cezanne-ai project (which is a work-in-progress project throughout 2021) is covering only two domains and one language and will not be enough for exhaustive experiments. Therefore, we

Figure 5
NOG layer

are also initiating the Cezanne-ai challenge[49] in order to obtain results in a wide range of domains and languages and with different datasets requirements. After achieving results from our implemented model and also from Cezanne-ai challenge, we will return with further conclusions.

Some comments are necessary at this point:

• We will look at the bot as a partner in conversation and it is not in our scope to pass the Turing test of discovering if a bot is human or not.

• As we implemented live-feedback pipelines in our framework, an Acute (Li, et al., 2019) or Likert evaluation (Joshi, et al., 2015) are redundant.

• We will not compete with open-domain frameworks (like dodecadialoque of Shuster et al. [2019]) that do not have the same objective: conversation between a non-specialist user and an expert-bot.

• Even if there are specific pipelines in the current NLP frameworks that are covering an important number of languages, our framework is among the first that targets a complete multi-language framework for Conversational AI.

What we need to further investigate:

• The average numbers of turns in a real conversation, as a result of our proposed models;

• Our models vs back-up models or versus other state-of-the-art models;

• The dynamics of the sentence-intent. If the sentence-intent that is summarizing the core-input is not changing over conversational turns, it can be a proof that the model is not accurate, as the user's perspective is not changing due to the conversation.

However, we can conclude the following and we are going to argue further below:

• At this moment we don't have a complete solution for Conversational AI and even if state-of-the -art solutions from other AI branches are having important results, they are leaving important gaps in the puzzle. Our proposed model has the benefits of fundamental inspirations

49 Due to the complexity of the proposed models and the interest that they might have, we are asking scientific community (not limiting to it) to get involved together with us in order to demonstrate the efficiency of a fundamental framework for at least 10 languages (at least 2 – core-languages) by using different types and sizes of corpuses/datasets. Our research will continue at https://github.com/Cezanne-ai. We are also posting links of the related work, if similarities in the model's architectures are met.

and an exhaustive architecture created specifically for conversational AI bots (not an adaptation of NLP tasks or from other AI branches models).

• By using as a back-up model three of the most utilized models in conversational AI, we will achieve two things: 1. The overall result of our framework should be at least equal in terms of results with current solutions on the market; 2. We will have a clear image on the scale-up implications in terms of datasets.

• It is not sufficient to have only one implemented project in order to have a clear image of the results of Pirkin model.

• Using as alternatives to our framework, solutions built on natural dialogue datasets or domain driven datasets can have collateral costs: proprietary, privacy and biases.

Possible conversational AI expectations taking into account fundamentals presented in this article and on the github page:

• A multi-domain conversational AI. Having a specialized bot for each need, with different specificities, can be too much for a user to handle.

• A bot that can understand the meaning of what the user is saying, not to make a simple keywords search.

• To return trustable answers/outputs, not to give thousands of choices, like Google search, that could be time consuming. At least when the user has a medical question, the first answer will be from a doctor.

• The previous request can be accommodated by a specialized forum, but the user might need fast advice 24/7.

• To be available in the user's native language or in a language in which he is comfortable with.

• To understand also long utterances if the user needs are more complex.

• To be able to have a fluid conversation that might include utterances that are not requests, but chit-chats, replies, deep-conversational topics. . .

• Memory. To be able to keep track of past interactions, not to repeat everything all over again.

• To be able to resolve task-oriented requests from the user (like reservation, schedule a meeting. . .).

• And some additional expectations from the developer's point of view.

• For the bot to work with available labeled/structured data or unstructured data and practical optimizations.

• Limited resources for training.

• Comply with laws, data- protection requirements, copyrights, anti-discriminatory requirements.

• Limitation in the number of tokens for queries/summarization/answers that the current models have.

We will compare the three back-up models (based on their paper conclusions/ experiments: Raffel et al. [2020- v3], Roller et al. [2020], Xu et al. [2021], Bocklisch et al. [2017]):

• Encoder-decoder for chatbots. We took into consideration building multi-turn transformers from scratch and also fine-tuning the pre-trained model with additional datasets created on the core-input scenarios.

• Open-Domain dialogue models with summarization-augmentation and dialogue datasets.

• Open-Source Language Understanding and Dialogue Management. We are also taking into consideration an E2E model, with specific pipelines inside, in order to enhance the intent-entities-forms solution.

*If the NLU/DPL is accurate

A possible and preliminary interpretation (many conclusions are still in the TBD state) of this table is that we are dealing with a vicious circle if we are searching for practical solutions

Table 2
Comparison between our model and the three back-up models

Expectations	Encoder-decoder chatbots	Open-Domain dialogue models	Open-Source Language Understanding and Dialogue Management	Cezanne-ai open-framework
Multi-domain	Deeply challenging scenarios	NO	NO	TBD
Understand the meaning	Yes	Yes	TBD	TBD
Trustable answers	Yes*	TBD	Yes*	TBD
24/7	Yes	Yes	Yes	TBD
Emerging languages	TBD	NO	TBD	TBD
Long Utterances	TBD	TBD	NO	TBD
Complex multi-turns	Deeply challenging scenarios	TBD	TBD	TBD
Memory	TBD	TBD	TBD	TBD
Task oriented	Deeply challenging scenarios	TBD	Yes	TBD
Limited datasets	NO	NO	TBD	TBD
Limited training	NO	NO	TBD	TBD
Compliance requirements	NO	NO	Yes	TBD
Solution for tokens limitations	NO	Yes/ limited	NO	TBD

in the current conversational AI environment, from the user's and the developer's point of view, at least. Our proposed open-source framework has the intrinsic value of relying on fundamentals and on the 360 degrees approach.

6. AGI component safety issues

We argued throughout the article on five directions linked with the impact of the AGI component in our framework:

- We used soft AGI solutions with limited impact.
- A conversation with 2 parties that not are proactive is not a conversation, and thus the AGI component is mandatory in Conversational AI.

• The existing models are riskier than our framework as they cannot control the biases from the datasets used by pre-trained models or the extraction/generation of the output from the web/3^{rd} party data environment.

• AGI is not used on core-inputs.

• We are using AGI for intuition over the states and policies related to the conversational flow and not as an instrument for answers generation.

Even so, we find it is our responsibility to put in place safety issues procedures related to the implication that an AGI model could have on the users. For this reason, we will allocate a stream in the beta-testing of Cezanne-ai project that will determine possible reactions of the users that could have negative repercussions on different levels (psychological, social...).

7. Conclusions

After a thorough linguistic and conversational analysis on five different domains (restaurant recommendation, legal consulting and deep conversational: scripts, novels and philosophy) and more than one hundred complex conversational topics we came to the conclusion that current models are not suited for advisory and the art of conversation (as main objectives for a virtual conversational task) due to architectural gaps. Our proposal has essential arguments to close these gaps in a limited data environment and reach the human baseline for the conversational AI tasks.

Furthermore, the Cezanne-ai open-source framework creates perspectives (without business objectives assimilated to the project) to deploy conversational AI bots in all languages (not only covering main languages and providing work-around solutions), on multi-domains and by using existing data, limited or not. If the framework overall is not giving satisfaction, each of the proposed 50 pipelines provides creative solutions for different tasks.

8. References

Bateson, G., 2000. Steps to an Ecology of Mind: Collected Essays in Anthropology, Psychiatry, Evolution, and Epistemology. New York: University of Chicago Press.

Bateson, G., 2002. Mind and Nature: A Necessary Unity. New York: Hampton Press.

Bateson, G., 2005. A Sacred Unity: Further Steps to an Ecology of Mind. New York: Hampton Press.

Bateson, G. & Bateson, M., 2005. Angels Fear: Towards an Epistemology of the Sacred. New York: Hampton Press.

Baum, S. D., 2017. A Survey of Artificial General Intelligence Projects for Ethics, Risk, and Policy. SSRN Electronic Journal. https://www.researchgate.net/publication/321381830_A_Survey_of_Artificial_General_Intelligence_Projects_for_Ethics_Risk_and_Policy

Bocklisch, T., Faulkner, J., Pawlowski, N. & Nichol, A., 2017. Rasa: Open Source Language Understanding and Dialogue Management. arXiv eprint. https://arxiv.org/abs/1712.05181

Carey, L., Nillson, M. & Boyd, L., 2019. Learning following Brain Injury: Neural Plasticity Markers. Neural Plasticity. https://www.researchgate.net/publication/335583735_Learning_following_Brain_Injury_Neural_Plasticity_Markers

Chaitanya, K. J., Fei, M., Boi, F., 2015. Personalization in Goal-Oriented Dialog. arxiv eprint. https://arxiv.org/abs/1706.07503

Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K., 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv eprint. https://arxiv.org/abs/1810.04805v2

Dhingra, B. L. L. L. X. G. J. C. Y.-N. A. F. a. D. L., 2017. Towards end-to-end reinforcement learning of dialogue agents for information access. arXiv eprint. https://arxiv.org/abs/1609.00777

Dinan, E. et al., 2019. Wizard of Wikipedia: Knowledge-Powered Conversational agents. arXiv eprint. https://arxiv.org/abs/1811.01241

Fitzgerald, M., Boddy, A. & Baum, S. D., 2020. 2020 Survey of Artificial General Intelligence Projects. Global Catastrophic Risk Institute. http://gcrinstitute.org/papers/055_agi-2020.pdf

Gao, J., Galley, M. & Li, L., 2019. Neural Approaches to Conversational AI. arXiv eprint. https://arxiv.org/abs/1809.08267

Hancock, B., Bordes, A., Mazaré, P.-E. & Weston, J., 2019. Learning from Dialogue after Deployment: Feed Yourself, Chatbot!. arXiv eprint. https://arxiv.org/abs/1901.05415

Joshi, C. K., Mi, F. & Faltings, B., 2015. Personalization in Goal-Oriented Dialog. arXiv eprint. https://arxiv.org/abs/1706.07503

Kolonin, A., 2020. Controlled Language and Baby Turing Test for General Conversational Intelligence. arXiv eprint. https://arxiv.org/abs/2005.09280

Lewis, P. et al., 2021. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. arXiv eprint. https://arxiv.org/abs/2005.11401

Li, J. et al., 2016. A Persona-Based Neural Conversation Model. arXiv eprint. https://arxiv.org/abs/1603.06155

Li, M., Weston, J. & Roller, S., 2019. ACUTE-EVAL: Improved Dialogue Evaluation with Optimized Questions and Multi-turn Comparisons. arXiv eprint. https://arxiv.org/abs/1909.03087

Miller, A. H. et al., 2018. ParlAI: A Dialog Research Software Platform. arXiv eprint. https://arxiv.org/abs/1705.06476

Pasikowski, S., 2017. Gregory Bateson's cybernetic methodology: The ecosystem approach in empirical research. In: N.Bateson & M.Jaworska-Witkowska, eds. Towards an ecology of mind: Batesonian legacy continued. Dabrowa Gornicza: Scientific Publishing University of Dabrowa Gornicza, pp. 63-84. https://www.researchgate.net/publication/324039850_Gregory_Bateson's_cybernetic_methodology_The_ecosystem_approach_in_empirical_research

Raffel, C. et al., 2020. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. arXiv eprint. https://arxiv.org/abs/1910.10683

Roller, S. et al., 2020. Open-Domain Conversational Agents: Current Progress, Open Problems, and Future Directions. arXiv eprint. Directions. https://arxiv.org/abs/2006.12442

Roller, S. et al., 2020. Recipes for building an open-domain chatbot. arXiv eprint. https://arxiv.org/abs/2004.13637

Serban, I. V. et al., 2016. A Hierarchical Latent Variable Encoder-Decoder Model for Generating Dialogues. arXiv eprint. https://arxiv.org/abs/1605.06069

Shuster, K. et al., 2019. The Dialogue Dodecathlon: Open-Domain Knowledge and Image Grounded Conversational Agents. arXiv eprint. https://arxiv.org/abs/1911.03768

Shuster, K. et al., 2021. Retrieval Augmentation Reduces Hallucination in Conversation. arXiv eprint. https://arxiv.org/abs/2104.07567

Vaswani, A. et al., 2017. Attention Is All You Need. arXiv eprint. https://arxiv.org/abs/1706.03762

Wu, J. L. M. a. L. C.-H., 2015. A probabilistic framework for representing dialog systems. arXiv eprint. https://arxiv.org/abs/1504.07182

Xu, J., Szlam, A. & Weston, J., 2021. Beyond Goldfish Memory: Long-Term Open-Domain Conversation. arXiv eprint. https://arxiv.org/abs/2107.07567

Zhang, Z. et al., 2019. ERNIE: Enhanced Language Representation with Informative Entities. arXiv eprint. https://arxiv.org/abs/1905.07129v3

Zhou, L., Gao, J., Li, D. & Shum, H.-Y., 2019. The Design and Implementation of XiaoIce, an Empathetic Social Chatbot. arXiv eprint. Social Chatbot. https://arxiv.org/abs/1812.08989