Crime at the palace
Verbal Interaction

Daniel Rodrigues Mendes
Instituto Superior Técnico, Universidade de Lisboa, Portugal

Abstract—A layer was developed above the existent dialog system EDGAR (Fialho et al., 2013) in such a way that the virtual agent acts as a suspect in a crime and is therefore in a police interview. EDGAR, by itself, is a chatbot with several plugins which picks user input and matches it with the closest question/answer pair in its corpora. Once the question is matched, the new layer calculates which is the most appropriate answer based on emotional factors from the virtual suspect. Which emotional factors and how strongly they impact the answer decision was set according to several rules from the model of the Cognitive Interview used by the Portuguese Escola da Polícia Judiciária as well as observations of enacted interviews and other researched psychology theories. The emotional factors of the virtual suspect begin with a static default state which is modified throughout the interactions.

Index Terms—Natural dialog systems, tutoring systems, serious computer games, police interview training.

1 INTRODUCTION

The emergence of chatbots is large and clear. Nowadays, these are made popular at the consumer level and are aimed at several different domains: entertainment, real state, customer service, health, etc. Better and more human-like chatbots means an opportunity for serious games. Particularly, several serious games have been developed in order to train police social skills [2]. Police training interviews are mainly done with the use of actors [3]. This training delivers the reality components which machine interaction lacks but are usually arduous, non scalable and expensive to be arranged [4]. This leads to a mostly theoretical training with only a few selected police officers being able to pratice their skills. Serious games allow police officers to be immersed in a controlled universe while getting regular feedback on their behaviour. Moreover, having non-realistic scenarios provides more freedom to discover new ways to impact the suspect which would otherwise not been used. This because in real scenarios the interviewer cannot afford to make mistakes and try new approaches. The absence of duty to do everything correctly stirs the focus of these games from performing towards flexibility and experimentation, [11]. With more training, feedback and experimentation, the police academy should be able to better detect certain behaviours in real interactions and improve their ability to handle the situations. Realizing how the officer’s behaviour impacts the interaction will increase his social-awareness [2]. This is the essential motivation behind this project: to create a dialog system to train police officers on the best ways to achieve their interviewing goals. On each interaction the application should understand if the input is being compliant with the interviewing theories and reward or penalize the user. This includes a user interface where the interviewer can see and interact with a 3D virtual suspect. In order to create this application, several features are changed and added to the EDGAR dialogue system [5].

2 BACKGROUND

2.1 Edgar, the virtual butler

The already existent EDGAR [5] is a question-answer architecture. This embodied virtual agent answers questions about the palace of Monserrate. EDGAR’s game environment is developed in Unity. The speech is recognized by AUDIMUS [6] which translates it to a sentence with a confidence value. This system selects one answer if it finds a match for the input, or uses string matching and string distance strategies to find the best answer.

The question-answer pairs which make the knowledge base are stored in XML files which make it easier for corpus enrichment.

Fig. 1. EDGAR’s XML corpus example.

Figure 1 shows an example of EDGAR’s corpora. The first block is for different ways of asking the same question and the second block has possible answers. Each set of question-answers contains several questions with the same meaning, together with a set of answers from which the system will pick one to display. The XML is enriched with an ID tag for each question-answer set. The answers may, or may not have the tag emotion and intensity which lets the front-end know which expression and with what intensity Edgar has to produce (i.e the expression tag can take a value such as “happy” which tells the front-end that Edgar should be smiling).

1. Palace of the 19th century located in Sintra, Portugal


2.2 Rapport and Entrainment

Rapport in an interview is decisive for a successful cooperation from the suspect [14]. It is a psychological term which refers to the ability one has to create levels of trust and understanding towards another. This impacts the process of responsiveness at the unconscious level. This is why most interrogation techniques involve rapport. It creates a level of compliance valuable to obtain confessions. It must be built and maintained. A common way to tell if the level of rapport is high in a conversation is through mirroring. Mimicking or mirroring is when a person tends to suit his/her body language (posture, gestures, mannerisms, etc.), tone of voice and choice of words to the person he or she is talking to. When pitch, loudness, words and speaking rate are similar in both speakers, there is what is known as acoustic-prosodic entrainment. That is, a sort of mimicking at acoustic and prosodic levels. Rapport plays a major role in the interview and therefore must have impact on my virtual suspect decision making. In fact, the whole interview is based on creating as much rapport as possible.

2.3 PEACE Model and Cognitive Interview

The virtual agent has a story and information which is accessed by the officer if the rightful behaviour and course of the interview are met. This is done by theories from social psychology. There are various models and guidelines for police interviewing. One of these models was developed by the police forces of England and Wales [15] and has the acronym of PEACE [16] which stands for:

- Preparation and Planning
- Engage and Explain
- Account, Clarify and Challenge
- Closure
- Evaluation

The model is based upon the fact that if the interviewer manages to create rapport with the subject it will eventually lead to a more cooperative conversation in contrast to an atmosphere charged with aggression and intimidation.

Besides PEACE there is the Cognitive Interview developed by Geiselman e col. (1984). Its objective is to get the most of information without distorting it [17]. Cognitive Interview is recognized as one of the best models of interview [18] [19]. The Cognitive Interview was developed into what is now called Enhanced Cognitive Interview [18]. This work is based on the Enhanced Cognitive Interview adaptation by [16] which is also influenced by PEACE and the outlines of the document “Application Scenario Outlines: EPJ-MJ use case “Interview Skills for Police Officers (ISPO)” from Escola de Polícia Judiciária.

Enhanced Cognitive Interview (ECI) has nine phases:

- Establishing Rapport
- Explaining the objectives of the interview
- Free speech
- Questioning
- New information recovery strategies
- Important questions to the investigation
- Summary
- Closure
- Evaluation

Throughout all the interview the use of silences and open questions are encouraged. Even during phase one the interviewee should be asked open questions so he/she gets more accustomed to a more elaborated and active speaking style of communication while not being intimidated [16]. The PEACE model and Cognitive Interview are the guidelines used in this project to understand what must be reward or prejudice to the interviewer. The use silences, use of rapport, following of the stages of interview and the avoidance of leading questions impact the suspect’s mood and its decisions on what answer to pick.

2.4 Police Interview Analysis

Three staged interviews were kindly provided by Escola da Policia Judiciária (EPJ). Analysis of these recorded enacted interviews yielded in several important points. In the first two interviews, it is noticeable the presence of the stages of the PEACE Model. Efforts to create rapport are noticeable throughout the use of of back-channeling (words such as “claro”, “está certo” or “pronto”) and the use of words from the suspect (even slang words). These serve as an example for what a normal interview should be. The third interview is a bad example on what not to do as interviewers fail to follow the PEACE model and establish rapport altogether. What is also taken out of these enacted interviews is that, in reality, these are compiled with a lot of “back and forth”, mumbling, audible fillers and overlaps which are difficult to produce in a virtual suspect.

3 BUILDING THE VIRTUAL SUSPECT

3.1 General Overview

In Figure 2 one can have a general overview of what is the application’s architecture. Firstly, we have the audio or text input. The sentiment underlining the text is analyzed and this will impact the suspect’s mood. This text input (the question) is then linked to a knowledge base. If there is a question which resembles what the user inputs, the possible answers from that question are sent to be picked from. If the application is not able to find resembling questions, it answers saying it did not understand what was asked (this answer is not always the same). From the list of possible answers the application picks the one most suitable according to its tags and the suspect’s current mood. If there is no input for a number of seconds it is considered as a silence and the suspect tries to add information to the last given answer.
The application on which this is based - Edgar - simplified the implementation as it provides the linking between an input question and an existent question in the knowledge base.

3.2 Speech Recognition

In order to capture audio from the user a third party software called AUDIMUS\(^3\) is used. This software is specifically made for the European Portuguese. AUDIMUS combines the temporal modeling capabilities of Hidden Markov Models (HMMs) with the pattern discriminative classification capabilities of multilayer perceptrons [6]. During runtime, the user presses a push button to record voice in the form of binary data. After release, the recording is sent to an external server which sends back the recognized text.

3.3 Sentiment Retrieval

For the decision of which answer is the most appropriate given a situation, the application needs some input on what is the sentiment behind each interaction by the police officer. To extract this metadata, a sentiment analysis tool was created. This tool uses WEKA libraries [7] in order to train a model, more specifically using a Naive Bayes Classifier.

To create a model it was necessary a suitable data set. Three data sets were tested. Each model was validated using a 10 fold cross validation. This means that the data is split into 10 parts in which 9 are used to train the algorithm and the 1 is used to evaluate the algorithm. This process is repeated, allowing each of the 10 parts of the split data set a chance to be the held-out test set. The validation results are combined for an estimate of the model’s predictive performance.

The first model was created making use of IMDB’s\(^4\) review data set [8] which contains 50,000 highly polar movie reviews (25,000 positive and 25,000 negative). This data set was translated into Portuguese using Google’s translation tools.

A second model was created resorting to a much different data set called sentiCorpus-PT [9]. This is a compilation of 2,625 manually annotated comments posted by the readers on the website of the Portuguese newspaper Público. The comments were made regarding a series of 10 news articles covering the TV debates on the 2009 election of the Portuguese Parliament. The data set is closer to what one would hear in a conversation and therefore better appropriated for police interviews.

To further approximate the model classification to the interviewing theories a third model, which is used by the application, was trained using an edited version of the data set previously described. There was a need to add, change and erase many instances. Some instances in the data set were either repeated, added nothing or made no sense altogether. Therefore, the whole data set was filtered manually for sentences such as “À Sócrates!” “Está decidido, votarei PS”, Spanish instances, personal names, shortened words (“pf” meaning “por favor”). Another problem regarding this data set is the amount of negative instances relative to the positive ones. An attempt was made to have the latter closer to a 50%-50% ratio. Overall, the data set was trimmed to roughly a third of its original 2,625 comments and around 100 new instances were added. The adding, erasing and changing of instances are all made so that the model correctly tags sentences according to the psychology behind the police interviews. For instance, sentences such as “would you care to explain” or “ask me anything you would like” must be positively tagged whereas sentences such as “you don’t need to say anything else” (going against the idea of letting the suspect talk as much as he can) should be negatively tagged. Verbs such as “diga” (say), “falar” (talk), “explicar” (explain) are highly positive whereas leading questions are highly negative. Several examples of these were added. Moreover, in attempting to enrich the data set several instances were taken out of a Portuguese dictionary of slang [10]. The model using the filtered data set yielded the results shown in the table below.

<table>
<thead>
<tr>
<th>Model Evaluation - General Overview</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Correctly Classified Instances</td>
<td>74.0%</td>
</tr>
<tr>
<td>Incorrectly Classified Instances</td>
<td>25.9%</td>
</tr>
<tr>
<td>Kappa statistic</td>
<td>0.3</td>
</tr>
<tr>
<td>Mean absolute error</td>
<td>0.3</td>
</tr>
<tr>
<td>Root mean squared error</td>
<td>0.4</td>
</tr>
<tr>
<td>Relative absolute error</td>
<td>90.1%</td>
</tr>
<tr>
<td>Root relative squared error</td>
<td>105.8%</td>
</tr>
<tr>
<td>Total Number of Instances</td>
<td>2625</td>
</tr>
</tbody>
</table>

TABLE 1

Naive Bayes Classifier’s results for sentiCorpus-PT edited data set.

Like all the models, there is a need to test whether it would or not rightfully rate common police interview questions. Overall, the results to normal interview questions were good but a lot more instances should be added to the data to have a more robust decision maker. This is a strenuous task but such an important one that something was needed to tackle it. A new feature was added to make the system better with the more use it gets. One can type a sentence and include one of two special tags: \_pos\_ for positive instance and \_neg\_ for negative instance. By doing so, one is adding knowledge to the system which might have had wrongly rated a certain instance. After a predetermined number of tagged instances added the system retrain the model once it restarts. The default value was set to retrain every 20 new instances. Choosing the best classification threshold is a problem of maximizing the F1-score (harmonic average of the model’s precision and recall). However, in the sake of this project, false positives and false negatives have a far greater impact than getting the classification correctly. Thus, the classification threshold is set to an inflated value of 0.9. This means that only instances classified between [-1;0.9] and [0.9;1] are considered. Another reason to set the threshold high is that the model is remade every 20 new instances as mentioned above. This means this value needed to be changed every time a new model comes into place.

3.4 Rapport detection

Rapport is a powerful tool in an interview and its establishment ought to be tried. Rapport attempts are detectable though speech entrainment. The latter is hard to detect.

\(^3\) https://www.l2f.inesc-id.pt/w/AUDIMUS
\(^4\) http://www.imdb.com/
However, one easy way to detect it partially is by analyzing the sharing of similar words in a conversation by both the subjects. A tool was created to test the similarity between what the user types and the suspect’s history. Every time there is an input, words are compared to the history of answers by the suspect. Words such as pronouns and determiners were discarded for these are used for language purposes rather than rapport itself. On each input by the user, the interviewee’s empathy is increased based on the number of similar words.

3.5 Knowledge Base

3.5.1 Knowledge base structure

The knowledge base (compilation of sets of questions and answers) is made resorting to XML documents. These documents represent different topics of conversation.

Figure 3 shows how the data is structured for each set of question and answer.

Fig. 3. Question-Answer example in XML.

The new XML has an ID to prevent duplicates and to further allow direct access from the application to that specific set. The XML tags <q> represent the set of questions. The multiple questions in Figure 3 are basically many different ways of asking the same thing. The tags <a> represent the set of answers. The application picks one of these as the answer from the interviewee. Each element may have the following added attributes: An ID which uniquely identifies the set of questions/answers; A correctTiming which states on which period of the interview should this question be asked; An importantInfo which states how relevant is the question; An empathyRequired which represents the empathy required for the use of that answer; A truth tag which represents whether the answer is a lie, truth or neither; A silence tag indicating that the new answer should be a continuation of a previously answered question on the same set; A repeat tag in case the question has been repeated more than once, the application looks to pick these repeats. Its value is either 0 or 1.

3.5.2 Tags

An important feature about the structure is the knowledge-base tagging. Research was put into this with the endeavour of which theories and concepts better endorse our reality.

Four different tags were taken into account according to above:
- **Empathy Required**: Throughout the interactions, an interpersonal state/mood is developed which entails a value for empathy. The answer picked by the interviewee depends on the answer’s empathyRequired. For example, the police officer asks “How long have you worked for Miss María?”. If the suspect carries a great deal of rapport with the police officer, the application picks the answer with the greatest empathyRequired. Usually this answer is a more cooperative and extensive one. Conversely, if the empathy on the conversation is low means the application will pick a low empathyRequired answer which usually translates into a short, bad humour response. By doing so, the application is rewarding or penalizing the interviewer with important or inessential information respectively.
- **Important information** tag is a way to understand how much evidence and important information is being asked at a certain point. The virtual agent needs to understand the importance of the question and how much is he being exposed by answering the question (if guilty). This is also highly correlated with how aggressive a question is in terms of information. If the tag has a high value of Important Information it means the question is trying to retrieve information is really crucial and possibly incriminating.
- **Correct timing** tag serves to identify the correct place of the question in the interview. For instance, open questions are usually best for information retrieval in the middle of the interview, after rapport has been established, whereas closed questions are used at the very end to confront the interviewee. What is taken into account is the order of these tags. In other words, if the order of the questions’ correct timing tag is maintained, the application increments the interviewee’s empathy whereas if the order is violated it takes out empathy.
- **Truth/lie/omission** tag allows for the application to choose a lie or omit information. For instance if the interviewee is under a great deal of pressure and the empathy levels are not sufficient for the information to be given. In this case the application will pick an answer with the omission or lie tags.
- **Silence** tag is used to reward the use of silences from the interviewer. This allows the application to use a specific answer after the detection of a silence. This way, it looks as if the interviewee is just continuing developing his/her answer leaking more information.

More tags can be added at the cost of revisiting every question of the corpus and adding its impact on the application’s answer decision.

3.6 Choosing an Appropriate Answer

Upon matching with a question in the knowledge base, the application chooses the answer based on two factors:
- **Silence**: In the front end (Unity), when the interviewer passes a number of seconds without asking a new question, Unity automatically asks the application with “_SILENCE_” indicating there was a silence/pause. The application keeps the first answer with silence tag set to 1 from the previous interaction. Upon receiving “_SILENCE_”, the application throws the answer tagged with a silence=1 from the last interaction (it is cleared every interaction and updated according to the current interaction as it only makes sense that “_SILENCE_” makes the suspect talk solely about the last interaction).
- **Empathy Required**: If not a silence, the application gets all the answers and picks the one with the maximum empathy required which is still less than the current Interpersonal State’s empathy. In case no answer’s empathyRequired satisfies the previous condition, the system picks the answer with the least empathyRequired even if above the current empathy.
3.7 Interpersonal State

The interpersonal state is a module which contains general information about the virtual suspect's current mood. This is being calculated during the interview. In order to make the character closer to reality it must have a personality. Some people are edgier, some are less prone to stress. The closer one wishes his virtual suspect to be closer to reality, the more variables one must define. Ideally, a PAD state model or Five Factor model should be used, but it incurs in unnecessary complexity which eventually drives the project out of its scope. A much simpler approach was made. This is the current emotional state of the suspect which varies dynamically during the interview:

- **Empathy**: current value of the suspect's empathy;
- **Pressure**: current value of the suspect's pressure;
- **Expression**: what is the suspect's body language suppose to be (e.g. happy);

Both empathy and pressure are on a scale from 0 to 100. The player is given the opportunity to define the scope before the conversation takes place or it is set to default. This is made by the tagging a question with _set_ followed by a value for empathy and a value for pressure.

The Expression is calculated by taking into account empathy and pressure which enables the front end to enact the characters emotions as angry, sad, relaxed, nervous or happy. These are the possible values of the interpersonal state Expression and their mapping to the empathy and pressure values are shown below:

- **Relaxed**: If \( (70 \geq \text{empathy} \geq 30) \) && \( (70 \geq \text{pressure} \geq 30) \)
- **Happy**: If \( (\text{empathy} \geq 50) \) && \( (\text{pressure} \leq 50) \)
- **Sad**: If \( (\text{empathy} < 50) \) && \( (\text{pressure} \leq 50) \)
- **Nervous**: If \( (\text{empathy} \geq 50) \) && \( (\text{pressure} > 50) \)
- **Angry**: If \( (\text{empathy} < 50) \) && \( (\text{pressure} > 50) \)

Let's now move on to how the tags from answers and questions interact with the interpersonal state.

If the suspect picks an answer with the tag _truth_ set to -1 or 0 it means the interviewee has chosen to lie or omit respectively. This incurs in greater pressure on the interpersonal state. Specifically, lying (value of -1) increases pressure in 10 points whereas omitting increases in 5 points. Telling the truth is debatable as to whether simply doing nothing (value of 0) or rewarding by subtracting pressure. On one hand, telling the truth psychologically diminishes the suspects thinking effort of making things up because it acts as a solid pillar of facts to localize oneself and construct a new, more solid, web of lies. On the other hand, telling the truth at a certain point of the interview does not eradicate the thought process necessary to keep up all the omissions and lies been said so far. The introduction of truths into a set of lies might even provoke more effort of thinking as lies and truths are difficult to be tied together. For this very last reason, telling the truth is set not to increase or decrease pressure.

A question with importantInfo valued with 0 means this is a question more intended to make the subject talk and create rapport. No crucial information is being asked and the suspect is at ease, relaxing. These reward the suspect’s interpersonal empathy by 5 points. Conversely, if importantInfo is valued with -1, meaning the questions is completely out of scope, the suspect’s interpersonal empathy drops by 5 points. Lastly, if importantInfo is valued with 1 it means the questions has important, delicate information that might incriminate the suspect. These increases the suspects interpersonal pressure by 5 points.

If a question is classified as positive (by the sentiment analysis discussed above), the suspect’s interpersonal empathy increases by 5 points. On the other hand, if the question is negatively classified, the suspect’s empathy drops by 5 points. In case the classification is below the classification threshold, there is no impact at all.

If the order of questions, regarding the tag correctTiming, is violated, the suspect’s empathy drops by 5 points. For instance, if the suspect is presented with a question with correctTiming valued at 2 followed by questions with correctTiming valued at 4, the suspect’s empathy drops. Conversely if the correct order is respected, the suspect’s empathy is rewarded with 5 points. If the value is zero it means the question should never be made in the first place. If a correctTiming valued 0 is read, the suspect’s empathy drops by 10 points.

Lastly, as explained above, a method was created which slightly detects the attempt of rapport creation by using the same vocabulary as the suspect. The method returns the amount of repeated words. To reward the attempt of rapport, 1 point of empathy is given for each word repeated detected. Note that it searches the whole conversation and not just the last interaction.

Bear in mind that the values are found empirically based on sense. These might need fine tuning. More important than their values is the differences between each other. For instance, inverting the order of a question (regarding correctTiming) should not be as impactful on the suspect’s interpersonal state as, let’s say, lying.

3.8 User Interface

The interface was developed within the cross-platform Unity3D Editor. This tool allowed to develop and manage objects in hierarchy which are controlled via C# scripts. The whole application starts with a main menu following an options menu and finally the game mode the user chooses.

Three different game modes were built:

![fig4](image)

**Speech + Text Mode.**

Speech + Text mode is the simplest mode (Figure 4). This mode is a basic text chat with the suspect with display of its current state of its personality and time elapsed. It enables the user to tell the system which questions are positive or negative if the Teaching Mode is on.

**Video + Speech + Text** mode which is, as the name implies, the addition of video and speech to the previous

5. www.unity3d.com
mode (Figure 5). The camera is locked which means the user cannot look elsewhere. The user can choose to speak or type the questions. The feedback panel is brought up on the PC’s display.

Fig. 6. Virtual Reality Mode.

Virtual Reality mode (Figure 6), unlike the previous mode, has no lock on the camera, only on the movement of the player (the camera is sitting on the chair). There are several differences on the set to make it more reality like as this mode is more intended for the use of goggles (VR). In this mode the user cannot input text, only speech.

4 Character Animations

The character body is named Winston and was imported from Unity’s Store\(^6\) which was developed by Reallusion using their Character Creator 2 software\(^7\). Winston was an ideal character to work with due to its strong appearance, highly detailed textures but ultimately because a fully rigged body eases the animation process. Despite being free for Unity, it was used with Reallusion’s permission. Overall, around 40 different facial expressions were created for Winston. The expressions are divided into 5 different groups: relaxed, happy, sad, nervous and angry. Each group is an overlay of different animations. Another aspect to the animations is the lip-sync which is made by creating a viseme to each phoneme that comes out of the text to speech engine on the server side. The head and eye movement are all randomized with certain constraints. The processes of creating, importing and controlling the animations as well as lip-sync are detailed in appendix B of the Thesis.

5 Evaluation

The evaluation of the whole system is made in two experiments.

In experiment 1 it was analyzed if values are changing according to the expected, whereas in experiment 2 the results are discussed at a more fundamental level (i.e. if the values and the suspect’s mood swings make sense according to the real world). Note that there was no evaluation with real users mainly due to linguistic variation and the sheer number of possible questions. In order to have a more realistic conversation there had to be many types of questions in the knowledge base multiplied by the many different ways one can say each one of these questions. A more real conversation would be tangible with a knowledge base in the thousands of sets of questions and answers. This would require massive amounts of tagged data driving the thesis out of its scope. Consequently, to evaluate the application, especially in a more fundamental level, it was necessary to come up with a scenario with a simpler knowledge base. A simple story was created and a knowledge base drawn out of it.

Several conclusions were drawn from the performance of the application. Experiment 1 shown that the values are in fact changing correctly to what was expected. On the other hand, with experiment 2, it seems that the suspect’s empathy is correctly being updated and it only increases when good things are happening (i.e the script is made to comply with all that adds empathy). However, the suspect’s pressure, with no reason not to decrease, always increases. There should be a way to decrease the pressure, for instance, with time elapsed. One other conclusion is that, despite being a positive script, the interpersonal state should vary up and down other than constantly increasing its values. This implies that there should be more balance between the number of points added and the number of points subtracted. This means it needs tuning as far as tags and their values are concerned, at a more fundamental level.

6 Conclusions and Future Work

Overall, the demo proposed was built successfully. With a reasonable knowledge base, the application will answer the interviewer’s question in a rewarding or punishing way if
the correct stance towards the theories behind the Cognitive Interview and PEACE Model are taken. However, the way the application processes each interaction is made in a very simple way and the rewarding and punishing of some attributes is not accurate. Furthermore, the application is still very reliable on the knowledge base. Which tags to use, their values and in which ways they impact the suspect’s choices are three different aspects that amount to a great deal complexity in the non exact and very often biased world of psychology. Furthermore, the adding of tags makes the extension of knowledge extremely hard. This is due to the fact that each tag must be carefully thought of. Overall, the application correctly identifies which answer to pick, regardless on which assumptions the decision is made.

In the future, several improvements could be made into this virtual suspect so as to make it closer to reality and the cognitive interview. These possible improvements include expanding the knowledge base, giving the interviewee a more detailed personality, taking into account the interviewer’s mood via sentiment analysis speech and video, additional tags on the knowledge base and a more complex approach to picking the answer other than being based on the current empathy.

References