

# Proposal to the Challenge

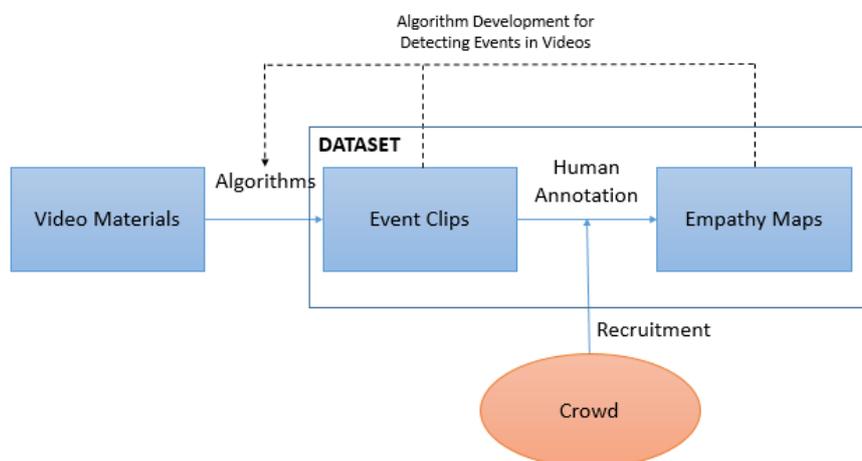
## **Built-in Artificial Empathy in AI**

### **1. Empathy and Artificial Empathy**

To response to the challenge entitled “How would you teach AI to be kind?” issued by EthicsNet, the proposed solution by the challenge Solver (the Solver) is to create a dataset for building artificial empathy in AI. Empathy can be defined as the capacity to understand or feel what another person is experiencing from his or her position, and respond to it with a proper emotion. The former and later is so-called cognitive empathy and affective empathy respectively. Artificial empathy is to develop ability of AI systems, such as companion robots, to detect and respond to human emotions. Building artificial empathy in AI systems to encourage them to take pro-social behavior and avoid anti-social behavior based on human emotions could be a more feasible way compared to the effort of teaching AI the general principles of good and bad.

### **2. Overview of the Solution**

The process for creating a dataset for building artificial empathy in AI was summarized in Figure 1. The dataset will comprise two parts, clips of events and corresponding human annotated empathy maps. The clips containing events which could be described by an empathy map will be extracted from video materials. For automating this process to produce mass event clips, algorithms for detecting these events in videos will be developed based on samples of event clips and empathy maps. The extracted event clips will be further annotated by the crowd to create corresponding empathy maps for each. The event clips and empathy maps, therefore, form a dataset which can be used for training AI.



**Figure 1.** The Process for Creating a Dataset for Building Artificial Empathy in AI.

### **3. Materials and Methods**

- 3.1 Video Materials: Videos from public or private domains with proper licenses can provide raw materials for creating this dataset. Video hosting services, such as Youtube and Vimeo, might be good resources for providing these raw materials.
- 3.2 Event Clips: For the purpose of training, segments containing specific events which fit the pattern of the empathy map will be extracted and make clips.
- 3.3 Empathy Maps: An empathy map is a tool for helping teams develop deep, shared understanding and empathy for other people. It was proposed as part of a human-centered design toolkit called Gamestorming. It can be applied to various purposes and help people to improve customer experience, to navigate organizational politics, to design better work environments, and a host of other things. A case of empathy map was shown in Figure 2. It contains several sections covering what is the goal of analysis, what the objects see, say, do, and hear, and what the objects think or feel (see Figure 2 and also <http://shorturl.at/dwV39>). For better fitting to the goal of the proposal, a revised empathy map might be considered if necessary.
- 3.4 Event Detecting Algorithms: To provide sufficient training materials, the process of making event clips must be automated. Algorithms for detecting events in videos could be developed based on a sample set of event clips with or without the corresponding annotated empathy maps. In this step, annotated empathy maps might be unnecessary as input, if the algorithms can just detect the event-like patterns directly from the clips.
- 3.5 Human Annotation: Each event clips will be annotated manually in the format of empathy map. Multiple annotations for a single clip is necessary to avoid bias and generate nuances of people's responses to a given event.
- 3.6 Crowd Recruitment: Recruiting enough participants to enable mass collaboration on annotating event clips to create the corresponding empathy maps is important. To achieve this goal, an empathy testing toolkit which can be delivered independently or as part of psychological tests/games and for serious tests or entertainment will be developed. For example, an independent website implementing this toolkit can be developed to provide empathy testing and also provide API so that other website or APP can incorporate the toolkit.
- 3.7 Other Methods: On the other hand, a possible alternative way is to use Generative Adversarial Networks (GAN) to amplify the dataset and reduce the number of needed participants, for example, to automatically generate empathy maps with nuances for sample events. Furthermore, an empathy maps-based automated ethical Turing test simulator might be developed for other methods for the effort to teach AI to be kind.

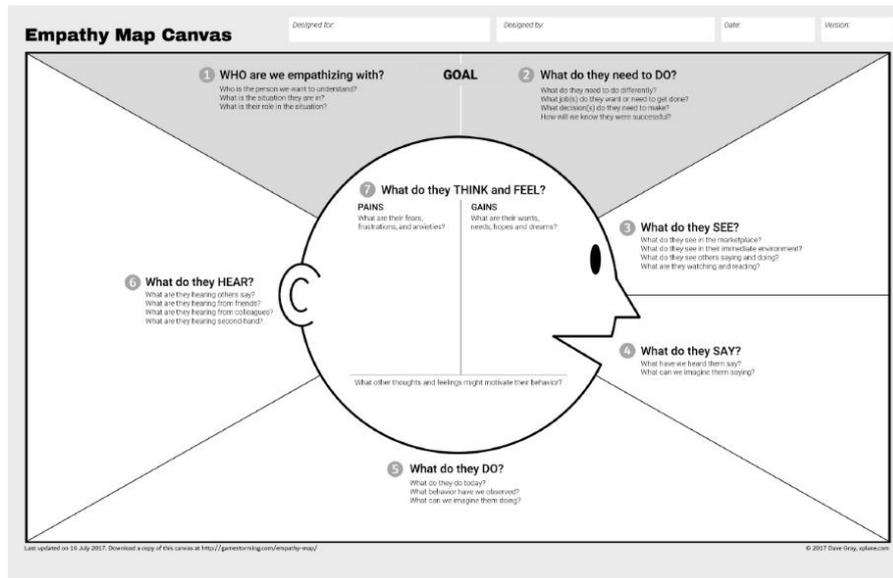


Figure 2. Empathy Map. Taken from <http://shorturl.at/dwV39>

#### 4. Discussion

Empathy is considered one of the key differentiators of being human. The fact that we were born with empathy makes us different from other species and artificial products. While human societies have evolved, rules and taboos for managing social interactions may change from time to time, empathy sort of set a bottom line of social behaviors for human societies. Hence, despite trying to teach AI all the rules and taboos might be a way to teach AI to be kind, building artificial empathy in AI is more fundamental. In this proposal the Solver pointed out this strategy and further described a process to create a dataset in need for building artificial empathy in AI. This dataset is composed of video clips of events and human annotated empathy maps. The Solver believes that it provides a great beginning for the purpose of building artificial empathy in AI and therefore teaching AI to be kind, though how to recruit enough participants to annotate the clips in the format of empathy map might remain challenging.