An Initiation Rite for Intelligent Machinery

Orkan Telhan

MIT Media Lab
orkan@media.mit.edu

Abstract

This paper discusses an experimental software project that interactively trains an artificial intelligence program with user input. The project is a commentary on the human-machine relationship inspired by the infamous Turing 'imitation' game. The software is an anthropomorphic play modeled on the different stages of an initiation rite (separation, margin and aggregation) observed by Van Gennep in early African societies.

We ask users to interact with the software in order to teach it what is needed to become intelligent. The users reply to a series of questions such as: what kind of things should an intelligent machine have, what kind of things should intelligent machines know about, and so on. By comparing and contrasting the answers given by different participants, we identify the underlying assumptions, desires, and expectations projected onto the attempts to build artificial agency. By pointing out a series of similarities between the birth of agency in human infancy and the processes for building human-like intelligent machinery, we comment on how the perception of intelligent machines and newborns are configured by the established norms and values in society.

Introduction

The history of artificial intelligence research showcases a number of studies of human infants to model and simulate for AI systems. Researchers studied how newborns develop sensory motor skills, acquire language and develop cognitive models (Steels 1997) to understand the foundations of human intelligence and to guide research on artificial agency.

Ethnographic studies by anthropologists, on the other hand, study infants in native or tribal societies to trace back the development of social norms and values in those societies. Here, Van Gennep's work is a framework for understanding the social processes behind the transformation of infants into adults via rites of passage (Van Gennep 1960). Initiation rites and ceremonies are important in understanding the birth of agency in an infant. These are moments of acceptance into the social order when newborns are assigned to their roles and duties in the social system. In some ways, these are the moments for the child to leave behind a stage of invisibility (Turner 1966) to slip into a new identity assigned via a name, status, gender role, or kinship position.

Our work intends to reflect on these two different perspectives and use them as a framework to look closely at human-machine alterity. For us, the birth of agency in an initiation rite reveals many clues about the intentions behind the construction of agency of an infant. From the newborn to the adult, we see the kinds of values and expectations projected onto the children. The infant's desires and social roles in the family and society in general are defined in this process and eventually carried by the infant throughout its life. In other words, for us, the way infants are cast in the shape of their adults is similar to how machines are conceived within the anthropomorphic frame of reference for intelligence. By creating the analogy between human infancy and intelligent machinery, we intend to identify some of the cultural expectations from artificial agents, especially in those that are expected to demonstrate human-like skills and behaviors. If we think of human agency as a club, we would like to find out what people are thinking about the membership requirements. While Turing's test focuses on the ability for machines to deceive a human reference in order to be let in to the club, we would like to broaden the question. There is much research for engineering human-like speech synthesis, vision or language processing systems. We would like to know what everyday people expect to see in an intelligent machine.

We start with the question: 'Is the intelligent machine ever initiated into the society?' In other words, when a machine is conceived in schematics, blueprints or code and actualized in material to become a producer, consumer, computer or calculator, among all possibilities, what is the process that defines what it is and what it does - configures its role and defines its value, form and function? As an initial response we built a system that interacts with people to find out their answers and to become a commentary on the nature of intelligence cast in the form of human-like

skills and behaviors. We use Turner's vocabulary to refer to our software as a 'neophyte machine'.

The Initiation of the Neophyte

Initiation Rite

Our system is a software application that simultaneously interacts with multiple people (a minimum of three) via text input. It is designed as an anthropomorphic play modeled on Van Gennep's analysis of initiation rites with three stages: *separation*, *margin*, and *aggregation*. Participants interact with the software by answering questions displayed on a graphical interface.

Separation. The software in its first stage submits itself to its users and acts only as an input device that would like to leave behind its previous stage to become an intelligent machine. It asks the participants their names in an obedient manner. After learning who the teachers are, *separation* begins as the neophyte submits itself to its instructors in the hope of learning:

- What and who the machine is (to be given its name and role or function).
- What it should know about the world (to acquire a level of literacy).
- What it should *not* know about the world (to get a sense of moral values).
- What it should be able to do (to be assigned its duties as a cultural object).
- What it should *not* do (to get a sense of the amount of regulation of its abilities of access and control).
- Where intelligent machines belong (to find its position in the social order).

The users respond each question one by one and proceed to the next. While every user replies to the questions on separate terminals, the neophyte collects all of the information to form its knowledge base. It parses the information to extract the nouns and verbs. After a basic evaluation, it checks whether there are any overlaps and repetitions in the dataset. In almost every session the answers are quite, diverse depending on the social and cultural backgrounds of the users. While there may be some overlaps, we observed more non-matching answers, mainly because of the open nature of the questions.

Margin. Next, the *margin* state begins, as the information needs to be verified. The neophyte looks for solid answers to make up its mind. It inevitably replies in a confused way, contesting the user in various ways:
- Hey Sam (the current participant), it looks like Carol (another participant) thinks that an intelligent machine should have an "umbrella," but you didn't say that.
- Both Barbara and Sam think that an intelligent machine should know about "homes". But Bill doesn't think that is important. Isn't this confusing?

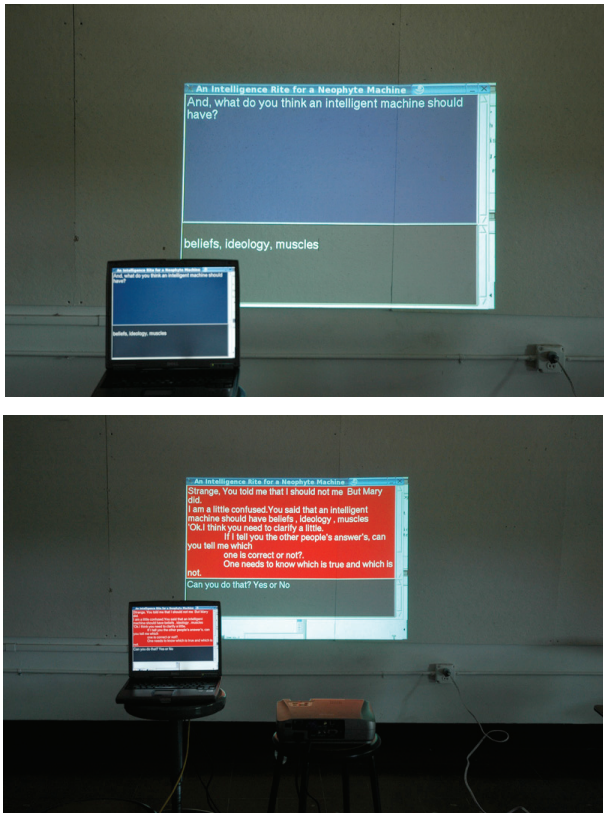

Figure 1 Screens from *separation* and *aggregation* stages

These are simple template driven replies phrased in a couple of ways. The software uses the user names and their answers to generate sentences. The neophyte then asks the master to have a seat while the interface changes to another screen asking for evaluations. Here, the participants are asked to rate the quality of the different answers to the same questions. They are asked to give a number from one to five for each answer. As each participant is presented with the other two participant's answers, they rank each other's answers to show which answers they think are better. Without being able to rank their own answer, they order the data for the machine and the machine uses it as criteria for making judgments about what really is considered important on its way to become intelligent.

Aggregation. The final stage begins with the machine presenting its results. After determining who scored better in the analysis of the intelligence, it now has a better sense of what intelligence is:

-Clearly, you are not intelligent Mike, because you told me that I need an "umbrella" to be an intelligent machine. But if it is not going to "rain" today, why should I need to "carry" one?

The software generates its provocative replies by using the common sense reasoning application ConceptNet (Liu et al. 2004). If the user answers the question "what the machine should have" with a noun such as "umbrella", we use it to query ConceptNet for its further associations. By finding out the association of "umbrella" with "rain" and "carrying", the software can generate very basic constructs with a number of templates that accept nouns and verbs to fill up the sentences.

As the ranking is done among the replies collected from current participants, the attributes of intelligence used for judgment are only valid for that particular session. If the users choose to perform again in another time, and even if they give the same answers, they may or may not still be considered intelligent by the system. As the criteria of intelligence are based on the evaluations of users ranking each other, the machine relies on them to make a judgment on their intelligence.

Discussion

In this section, we would like to address some of the topics we find important in positioning the project.

Intelligence. Intelligence is already a controversial term. Here, we use the word in a open-ended way to let the users define its meaning. It serves its purpose by being a very common word that triggers many different connotations for different users. For us, it is suitable for triggering more provocative answers on human intelligence to train the system. We follow the ethnocentric discussions on human/infant-computer interaction and model our system as a series of interaction patterns similar to infant development (Winnicott 1971). The intelligence of the system emerges from situations that come out of each stage of the initiation rite. The system facilitates an asynchronous interaction among participants to create a context and a situation (Suchman 1987) to gather knowledge about intelligence. For example, the ranking of judgments on what intelligence is or is not serves for playfulness and motivational purposes. It encourages users to play more and to learn more about each other's responses. Because they rank each other, judgments on intelligence are still left to users; the software ascribes no value to the participants. It simply generates responses to facilitate the learning process.

The Initiation Rite vs. The Turing Test. The Turing test, also known as the "imitation game", uses the differentiability between a human and a machine as a measure for intelligence. While in the Turing Test, a computer and a human user are tested against each other in front of a third party human reference, we slightly reverse this paradigm and pose the challenge to be intelligent to the human user. Users challenge each other to be considered intelligent by the machine.

In Turing's test, intelligence is a notion of status. To be considered intelligent, with all of its controversy, is to be given a privileged position. This is associated with the membership with a social club. The criteria for intelligence in our work are to be determined collectively and considered valid for only one session (e.g., during the interaction of three users at the same time). They are vague and subjective. However, we use them to understand how intelligence becomes a status and under which conditions. Like the infant that needs to pass the initiation rite to be accepted into the larger social group, what are the criteria that really determine one's position in the hierarchy of a social order?

Infant Agency vs. The Neophyte Machine. Supervised machine learning, information modeling, knowledge representation and reasoning techniques are already in use in AI for building cultural artifacts that imitate natural behavior. The complexities of the cognitive, psychological and social behavior of adults inspired many scientists to shift their focus to the infant. While the scale of complexity of the problem did not disappear, it opened up many alternative methods to study the growth of the infant. An ethnographic study of infancy, on the other hand, yields more questions on human agency. For example, as Turner mentions, when the infant is initiated, it leaves its "transitional persona". Once named, the infant becomes addressable in the social domain. It acquires a position - an autonomy that differentiates it from the others. It loses a sense of purity: it gets confused with conflicting values. In this regard, the question of autonomy and agency for a machine opens up for further discussion. When we ask users what they would like the machine to posses to become intelligent, we had answers that ranged from a pistol to pride, home, friends, ideology, muscles, and so on. These are an arbitrary set of answers for us but they exemplify the diversity of associations people have with 'intelligence.' In a broader social context this work provides us the clues to what people inherently expect to see in artificial agency to be accepted into the human 'club'. While these answers are not solids results to model a machine to become more human, we train our software with a series of insights about intelligence to familiarize it with the value system of its users.

Singular Agency and Community. The neophyte machine does not have a singular source of information. Its agency is not built on a single model of beliefs, tastes or expectations, which usually is not the case for the infant in a closed homogeneous society. The infant, once initiated, belongs to a certain value system, which in return shapes its social behavior. If we refer back to Turing's original paper where he quotes Hartree, the system creates the necessary means for having someone else to "think for itself" (Turing 1950). The neophyte machine is only as intelligent as its users. The system either directly makes use of the current users' data or ConceptNet's

knowledgebase, which is a collective source of information authored by 14,000 users. Having users from a variety of social and cultural backgrounds, we are able gather a great variety of information, even many conflicting beliefs, which can then be used for extending the learning paradigms of the software.

Technical Notes

The current implementation of our software is a standalone application that runs on a series of networked computers via TCP/IP. We use a client-server architecture, where a server coordinates the interaction between the client machines that remotely run the graphical interfaces. We use python and Tkinter GUI Toolkit for the client machines to walk the user through the different stages of the initiation rite and have them type their input [Figure1]. The server runs a python application, which communicates with ConceptNet via an XML-RPC server. We browse ConceptNet to find what a user answer can be "conceptuallyRelatedTo", what it can be "usedFor" or "capableOf". We gather a list of terms amongst the most related results and make a reverse search to check how many of them are also backwardly related to each other. The system tries the to pick the best one, by checking whether it suffices a certain threshold and uses it for filling out a template. We use Liu's Monthy Lingua natural language processing toolkit for parsing user input into words and nouns. In addition to ConceptNet, we use a custom written application to keep track of the answers provided by users and all of the replies generated for previous users. This application builds a semantic network to let us study the user answers and their associations used in the templates. We register the similarities and recurring patterns in the answers to and use it train the software for further sessions. The answers are also fed back to ConceptNet to extend its knowledgebase for future queries.

Future Work

Our current implementation is an in-progress architecture where we run tests for multiple users in a gallery installation-like set-up. In the future, we would like to have a web version to extend our user profile. While our GUI works for the user tests, it is in the process of a redesign to accommodate different kinds of GUIs for alternative users. We would like to experiment with the tone of language and the kinds of questions in user interaction. For example, in the *separation* stage, we would like see the different kinds of reaction the users may have for a more obedient tone vs. an aggressive tone asking the questions. By looking at the kinds of answers, we also would like the system alternate between different tones for coming stages. One of the main challenges we have is the ability to create interesting and natural looking replies by the software, especially at the *aggregation* stage where it uses the nouns and verbs acquired from ConceptNet queries. The replies are still prone to error. While we use simple, rule-based grammars in the templates to minimize error (e.g., a noun phrase following a verb) we eventually would like the system to be able to handle more complicated sentences for better associations. For example if the word "umbrella" triggers 38% association with the word rain (in a "context" search at ConceptNet), we would like the system to be able to generate a sentence using another association like "wet", which is less related (3%) to "rain".

Conclusion

To be able to build machines that can go beyond symbol crunching, syntactic operations, and participate in the semantic endeavor is a restless ambition. In this project we wanted to provide a playful platform that will make the users think both about their perceptions of intelligence and about the possibility of artificial agency. By referring to two different theoretical frameworks pursued by AI research and anthropology, we bring a different insight for human and artificial agency. While the AI community already considers many models to extend the research on artificial vision, speech generation or language processing systems, we introduce a platform that is used for understanding the social construction of the "belief of intelligence" similar to ethnographic studies that investigate the social and cultural norms that configure the lives of infants. The initiation rite for the intelligent machinery, in this regard, is a tool for understanding the assumptions, desires and expectations projected onto the attempts to build artificial agency from the perspective of the everyday person. We use the information both for informative playful social interaction as well as training a collaborative software application that collects existing norms and beliefs that shape our understanding of intelligence.

References

Liu, H., and Sigh. 2004. *ConceptNet: A Practical Commonsense Reasoning Toolkit*. BT Technology Journal 22(4):211-226. Kluwer Academic Publishers.

Steels, L., *The Synthetic Modeling of Language Origins*. Evolution of Communication, 1(1):1–35, 1997.

Suchman, L.A. 1987. *Plans and Situated Actions*. New York: Cambridge University Press.

Turing, A. M. 1950. Computing Machinery and Intelligence. *Mind*, 59:433-460.

Turner, V. 1967. *The Forest of Symbols*. Ithaca: Cornell University Press.

Van Gennep, A. 1960. *The Rites of Passage*. Trans. by Monika B. Vizedom and Gabrielle L. Caffee. Chicago: University of Chicago Press.

Winnicott, D. W. 1971. *Playing and Reality*. New York: Routledge.