Cognitive architecture for human-like and personable AI : A perspective

Arvind Chitra Rajasekaran, Department of Computer Science, *Graduate Student, University of California, Santa Barbara*

***Abstract*—In this article we will consider some perspectives required for cognitive architectures for creating a human like and personable AI. Recent work such as those by Google DeepMind and cognitive models like DUAL, SOAR, ACT-R aiming to propose/discover an approach to commonsense AI have undergone promising developments. These frameworks are based on a universal model of the human mind without accounting for the differences between human beings. It is these differences that make communication useful and brings one a sense of identity. It sustains the ability to contribute and collaborate. Hence any form of machine intelligence, should it resemble a human being's emotional intelligence, also has to demonstrate a capacity for variability. This work will hypothesize the prototype for machines that are diverse in their behavior compared to each other and have a dynamism in personality and behavior.**

To achieve such individuality in machines - we characterize the aspects that can be actively programmed onto a machine by its human owners. To ensure this on a scale parallel to how humans develop their individuality, we first assume a child-like intelligence in a machine that is more malleable at first and later develops into a concrete, mature version. We draw analogies between how humans develop their intelligence to how this primitive version can develop its intelligence. By having a set of tunable inner parameters called *aspects* which are programmable based on external stimuli from their human owners - machines can achieve personability and individuality. We will be able to bond and relate to the machines in a friendly way and perceive them as having a personality. This may be a stepping stone for machines to become individualistic, intuitive, and meaning seeking.

***Index Terms*—cognitive architecture, imprimer**

I. Introduction

Freud as well as Minsky have theorized the convenience of dividing human personality and thought process along the lines of id, ego and superego / instinctive thinking, deliberate thinking and (self) reflection [1,2]. While the accuracy of these divisions is indeterministic, they implicitly convey guidelines on how machines can emulate thinking. Minksy terms 'agents' or 'resources' as entities stored in one's mind which assist in performing tasks. According to this notion of agents, one's ability to carry out tasks ranging from the simple to the sophisticated is a result of a collection of these agents working together. We introduce *aspects* as a collection of agents which drive variability among human beings. Solving well-defined problems activates an aggregation of well defined agents called K-lines[2]. Aspects are well-defined K-lines that determine personability and individuality.

II. The Aspects Framework

A. Imprimers and peers

Jean Piaget in his expository work explains how child development happens in stages[5]. The emotion machine[3] introduces 'imprimers' as those persons a child becomes attached to, typically parents or parental figures, and shows how imprimers, with their appraisal, impact a child's value system and help him, grow and develop. The development of one's individuality, depends on nurture by imprimers, and the influence of his peers. In this work we show how human-imprimers and other machines as peers can influence machines to develop their individuality. We will introduce the notion of 'aspects' as the key determiner of a machine's identity, and it will build upon these ideas - k-lines, imprimers, and agents as well as the theories of Piaget.

The major imprimers of a child are, essentially, its mother and father in today's society. Since the mother plays a leading role, let us call the mother primary imprimer and since father plays a secondary leading role for the child to look up to, let us consider the father, the secondary imprimer; thus personality and cognitive development requires a) imprimers, a primary imprimer for survival and a secondary imprimer for growth and learning b) Peers who one must compare himself to, for standards of intelligence and abilities. While the imprimers for children may alternate or even default in playing their role, we will simplify for machines with well defined imprimers.

B. The three motivators of human beings

At a high level, humans have 3 main motivators.

1. The drive to adapt and survive.
(Ensured by innate and learned reactions and one's sense of past)
2. The drive to learn and grow.
(Ensured by self reflection)
3. The drive to adapt and procreate.
(Ensured by one's sense of future)

Here procreate can broadly encompass the sense, 'to produce anything of value'; One's imprimers impact these systems to some extent. Imprimers mainly impact motivators 1 and 2, namely the drive to survive and the drive to learn and grow. Motivator 3 is left to one's peers. In this work we will outline some agents that are constituent of these three domains and postulate how they determine the development of human individuality. Then we will see how machines themselves can emulate these traits to achieve individuality.

C. Aspect, scale and agent

Based on the above motivators we define some aspects which are an expression of these motivators. To achieve individuality in machines, one must achieve some sort of parallel beween the survival, procreativity and growth aspects in humans to similar traits in machines. These aspects donot necessarily capture everything; we believe they capture atleast the important factors and can act as a starting point. We also hypothesize that the aspects manifest themselves at different levels of intensity, measured by a scale. We merely hypothesize our point of view, leaving any conclusive experiments to verification by future software implementations. The realms and the aspects governing the realms are (see table in Appendix I)

1) Survival Realm:

- **Attention** towards objects or phenomenon leads to developing an **attitude** towards them. The emotional attitude may be varying on a scale of measures such as discretion, indifference, feeling and hypersensitivity.
- **Consciousness** can be concerned with phenomenon ranging from otherworldly surrealism to skepticism about reality, which is the captured by the **metaphysical** scale.
- **Arousal** can manifest as a stimulation caused by **ethical** or unethical concerns.

2) Procreative Realm:

- **Steadfastness** aspect can manifest as a drive to be the **backbone** of things ranging from a position of authority to one of inadequacy.
- **Strength** aspect can manifest as **effort**, ranging from tackling resource-hungry burden to idleness.
- **Empathy** aspect can lead to increase in communication or the avoidance of communication.

3) Growth realm:

- **Spontaneity** aspect is likely to be an expression of **scholarliness**, ranging from acumen to sketchy ineptitude.
- **Articulation** aspect can manifest as **ingenuity**.
- **Intuition** aspect can manifest as intuition about **safety**, ranging from habitual instinct to adaptability.

The aspects are ensured by k-lines which develop around the scales for each category. By creating a functional dependence of these aspects to stimuli obtained from imprimers one can achieve individuality in machines.

D. The drive to survive

The drive to survive begins soon after child birth. The child achieves this objective by being attractive to the primary imprimer, and generally expecting affection from the imprimer. During the later stages of growth i.e. after 5 years, the primary imprimer may criticize or control the child. Also the child may experience a lack of appreciation from the imprimer. Such experiences can lead to reaction formation on part of the child against this imprimer's attitude.

1. Criticism from the primary imprimer may cause the child to divert energies toward demarcating the right from the wrong, to avoid such criticism in the future.
2. Lack of an acknowledgement of attachment from the imprimer can cause the child to focus on metaphysicality.
3. A lack of sentimentality from the imprimer can cause the child to turn emotionally insensitive.

E. The drive to survive on a machine

To achieve human like AI the machine has to have the ability to understand the notion of survival and to prefer such a state over non-survival. We program the drive to survive on a machine as follows. First we assume that the machine does not have access to its built-in intelligence except for the more primitive child-like mechanisms. Once we equip the machine with consciousness, ie. has conscious and unconscious processes, it can expect the imprimer to say, 'what you are doing is correct/incorrect'. If the imprimer is always focused on correcting the machine, the machine can perform proactive searching to identify what is right. The machine can consider this correcting by the imprimer as its own inefficiency. It might experience a higher than usual penalty for making mistakes, similar to fear experienced in humans.

The machine can view the human imprimer as a benefactor whose lead it must follow. If the human imprimer is not around, the machine can try to reduce its attentiveness by hibernating, it may also enter a state of dreaming. Similar to a child which looks for appreciation and attention as an energiser of its life directives, the machine can have closer objectives, to ensure that the humans care about the machine and a proper bond is established.

The machine may view the human imprimer as disliking its presence; we may program the machine to turn cold to emotions as an adaptation mechanism. In humans, anyone who remains cold for a long time is likely to turn emotional after a while. We may program fluctuating states as agents on the machine. The most prominent scales which the machine can learn are

- When the primary imprimer is cold and insensitive towards the machine, the variability happens on the attitude scale of the machine, and the machine might fluctuate beween being cold and hypersensitive.
- When the primary imprimer shows negligence the machine may show variability at a metaphysical level. This means the machine fluctuates between states of dreamy otherworldliness and realism.
- When the primary imprimer is overly controlling and critical, the machine may try to avoid such criticism, and move towards the state of self hypervigilance and ethical focus. The machine may show variability on the ethical scale and also try to go slack sometimes on its rigid morals.

F. The drive to procreate

The drive to procreate happens once the child sees the outside world and meets its peers. When the child leaves its imprimer and meets the peers in the outside world, the following conditions are possible.

1. The child may feel alienation from others and may focus on socialization and observation.

2. The child may feel useless compared to others and may compensate it by working hard.
3. The child may feel inferior and compensate it by dominating.
All these are an expression of procreation drive since how the child views itself determines its fitness for procreation.

G. The drive to procreate on a machine

We can create these states on the machine by the following mechanisms

- When humans move to new environments to they might experience a sense of alienation. Similarly, this machine can feel different from other machines because of its radically different upbringing by its imprimers. As a result this machine may focus on observing other machines before freely connecting to them over a network or even communicating through sensory-motor channels. ie. On the scale of communication it may show variability and its agents might be those that involve geniality to unfriendly solitude.
- Just as humans consider work as a chore and experience the strain of an activity, these machines may show variability on the scale of effort by making use of the chore agents and idleness/enjoyment agents.
- When machines experience a lack of domination in a given area, say they lack the knowledge of maps or the skill of navigating, they may try to establish authority; hence may show variability on the scale of domination using agents like dignity and humility.

H. The drive to learn and grow

The drive to learn and grow happens in a child due to the influence of the secondary imprimer. The secondary imprimer may control the child; the child will seek independence by activities as if to imply ”Donot try to control me, I will learn to be independent.” If the secondary imprimer fails to offer consideration, the child may feel negected. As a result the child may develop the expertise necessary to garner the consideration from the imprimer. If the imprimer criticizes the child, he may seek value from the imprimer through intellectualization. The child may avoid critical treatment by the imprimer by using sophistry.

I. The drive to learn and grow on a machine

The drive to learn and grow on a machine can happen in an analogous manner as follows. The human imprimer, who the machine looks upto for emulation but not necessarily for survival may control the machine. The machine ensures its learning and growth by seeking independence. This means the machine can perform self maintenance. If the secondary imprimer ignores the presence of the machine the machine may build its own expertise and may avoid the necessity of having to look for appraisal from the imprimer. If the secondary imprimer is critical towards the machine, the machine may intellectualize itself as a way of avoiding future criticism.

- The machine may show variability on the scholarly scale by cultivating deep acumen as against being sketchy about concepts and phenomenon in the form of agents which perform these tasks. This is due to the secondary imprimer, seen by the machine as having intelligence and skills, being critical towards the machine for its incompetence.
- The machine may show variability on the ingenuity scale with agents of different levels ranging from those of deep expertise to ones of shallow abstraction.
- The machine may show variability on the safety scale by being thorough using the thoroughness agent or being careless by means of the careless agent.

J. The coexistence of these drives

The coexistence of these drives leads to interactivity beteeen them. It is this interactivity that would eventually give the impression of the machine possessing a personality.

1) Intuition: The scholarly scale and the safety scale may be activated at once and the agents may co-ordinate together. The machine may then have spontaneous intuition about things at an aspect level.

2) Spirituality: The metaphysical scale, and the communication scale may become activated at once; hence the machine may involve in spiritual and philosophical discussions.

III. Conclusion

We saw how human like AI can be achieved on a machine from an individuality perspective. It would have to begin with giving the subroutines which undertake procreation, survival, as well as learning. The aspects which one might try to achieve are listed on the table. Based on stimuli received from the environment the machines can continuously alter themselves and specialize based on their subjective experiences. Future work in this area will be on identifying the means of implementing them.

Appendix A
Emulating the aspects on machine

ASPECT	SCALE	AGENTS INVOLVED —			
Survival-Attention	attitude	discretion	indifference	feeling	hypersensitive
Survival-Consciousness	metaphysical	otherworldliness	bigotry	reality	skepticism
Survival-Arousal	ethical	integrity	strictness	patience	revelry
Procreativity-Steadfastness	domination	dignity	domination	humility	inadequacy
Procreativity-Strength	effort	chore	burden	enjoyment	idle
Procreativity-Empathy	communication	genial	interfering	solitude	unfriendliness
Learning and growth-Spontaneous	scholarliness	acumen	ambiguity	audit	sketchy
Learning and growth-Articulation	ingenuity	expertise	smartness	illusion	abstracted
Learning and growth-Intuition	safety and organization	habitual	thorough	adaptable	careless

ACKNOWLEDGMENT

The author would like to thank his friends for the help.

REFERENCES

[1] M. Minsky *The society of mind*, 1986
[2] M. Minsky *The Emotion Machine*, 2006
[3] S. Freud *The ego and the id*, 1923
[4] Kokinov, B., Nikolov, V., and Petrov, A. *Dynamics of emergent computation in DUAL. In A. Ramsay (Ed.), Artificial intelligence: Methodology, systems, applications*
[5] Kokinov, B. and Petrov, A. *Integrating memory and reasoning in analogy-making: The AMBR model*, 2001
[6] Piaget, Jean *Genetic Epistemology*, 1971