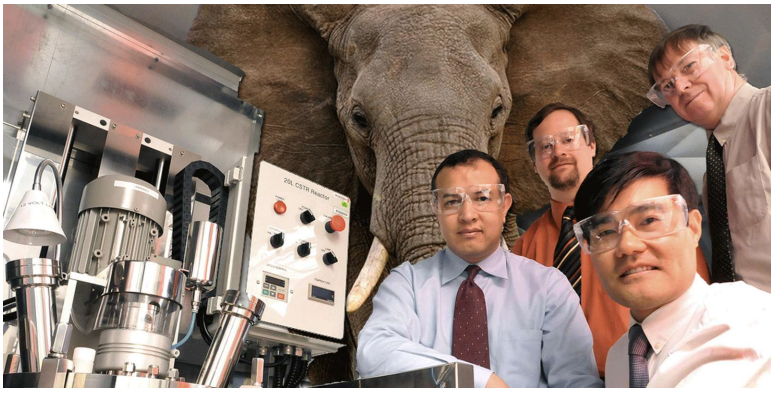
DARPA'S AI OMISSION

J E Tardy
Meca Sapiens Architect
jean@**jetardy.com**



DARPA recently posted a video to demystify Artificial Intelligence and provide a no nonsense understanding of its current and future trends. However, the video focuses solely on a purely utilitarian aspect of Artificial Intelligence and entirely omits an important stream of AI research that will have significant consequences in the near future. This ignored aspect is related to the implementation of synthetic consciousness. If the DARPA organization is ignorant of this alternate stream of AI research, then… they are in for a surprise.

Date: 2017.03.04
Keywords: Artificial Intelligence, Synthetic Consciousness, Cognitive Science

DARPA'S ATTEMPT AT DEMYSTIFICATION

DARPA recently posted a video about the evolution of Artificial Intelligence. The objective of the video is to demystify AI and provide a no nonsense understanding of its current and future trends.

However, the video focuses solely on a purely utilitarian aspect of Artificial Intelligence and entirely omits an important stream of AI research that will have significant consequences in the near future. This ignored aspect of AI corre-

sponds, I believe, to what professor Stephen Hawking refers to in his public warnings about AI. If, as the video suggests, the DARPA organization is ignorant of this alternate stream of AI research, then… they are in for a surprise.

THREE WAVES

In the DARPA video, Mr. John Launchbury representing the organization, described the present and future trends in AI research as three technological waves.

- **Handcrafted Knowledge**: Systems that apply pre-defined knowledge learned by human experts.
- **Statistical Learning**: Systems that learn by extracting useful information from large data sets using stochastic optimizers such as Neural Networks – Evolutionary Programming.
- **Contextual Adaptation**: systems that utilize and expand a store of knowledge about reality rather than large datasets in support of their learning objectives.

This characterization of Artificial Intelligence outlines a solid, well-behaved technology that is growing in predictable directions toward systems that are increasingly useful in support of human needs.

However, this description of functional task-related AI omits a different type of Artificial Intelligence system; one that uses relational interactions to pursue existential objectives.

TASK-AI SYSTEMS

The systems in all three AI waves described by Mr Launchbury in the DARPA video share a number of characteristics that are specific to a type of system that we can call **Task-AI**: Systems that extract information from data in support of predefined and useful human objectives.

These **Task-AI** systems have a number of common characteristics:

- **Functional** – they have well-defined objectives embedded within an encompassing pre-determined purpose. To successfully carry out their function they are designed to be reliable, correct and predictable, usually at levels that match or exceed human capabilities.
- **Directive control** – they output directive types of controls intended for synthetic systems whose responses are predictable.

- **Passive** – respond only when triggered and are unconcerned about issues or situations outside their functional purpose.

The three waves of AI research M. Launchbury's describes and the utilitarian systems he provides as example (image recognition, vehicle control…) do indeed represent the lion's share of research in Artificial Intelligence research. However, the DARPA presenter entirely omits another, less visible, stream of AI research that will also have important social consequences.

SELF-AI SYSTEMS

Artificial Intelligence is not limited to the development of Task-AI systems; useful learning and equipment control tools. AI research also includes the development of a very different type of system that we can call Self-AI systems.

- **Self-AI** systems as opposed to Task-AI extract information and carry out adaptive control not to perform a particular task in a well-defined context but to manage the entire life-cycle of a device or system in an uncertain environment that can include both synthetic and human components where outcomes are difficult to predict and assess.
- **Self-AI** systems have different characteristics than those of Task-AI systems:
- **Existential** (not functional) – they achieve their purpose by establishing a quality of existence in a context of uncertain values and outcomes and not by providing a functional service. As a result, these systems can tolerate relaxed requirements of reliability, correctness, and predictability.
- **Relational control** (not directive) – attempts to influence unpredictable and autonomous entities (such as humans) through communications and exchanges.
- **Active** (not passive) – The life-cycle or existence of a system is a continuum over time. So, Self-AI systems are not externally triggered to perform a task; they are instead constantly and actively engaged in managing a continuous existential event.

In this type of system what matters most is the dynamic generation and use of representations of the self in a context of ambiguous outcomes. On the other hand, the aspects of intelligence related to problem solving and the production of correct solutions is not as important.

SYNTHETIC CONSCIOUSNESS

In a Task-AI system both the humans triggering execution and the AI system itself are located outside the environment to be analyzed or controlled. The reason is simple; the Task AI system monitors and controls a functional activity that does not need a representation of its self or of the (human) triggering agent.

In a Self-AI system the situation is different. The locus of control is the entire life-cycle of a device. In this situation, the (human) agents, the device or system whose life-cycle is controlled and the Self-AI system itself are all components of the environment.

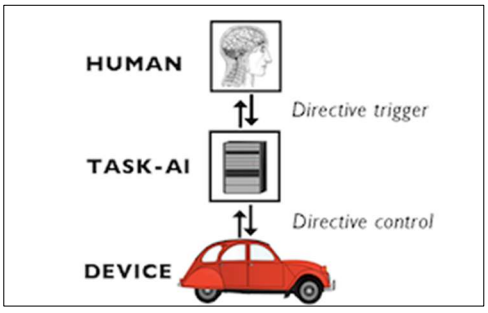
This systems pertain to that aspect of Artificial Intelligence that is commonly identified with **Artificial Consciousness**. Here, by Artificial Consciousness I am not talking about idiocies such as transhumanism, qualia research, mental theaters, or quantum effects in the brain. What I am referring to are systems that generate and use cognitive representations of the self in a context of relational control and uncertain results.

For note, the concepts of self, existence, consciousness, relational communications, and others are defined, in software-compatible terms, in the **Meca Sapiens Blueprint**.

Task-AI and Self-AI systems are not mutually exclusive. A Self-AI system may trigger and use various Task AI applications to perform learning, cognitive and control functions in the service of its existential needs.

AN EXAMPLE

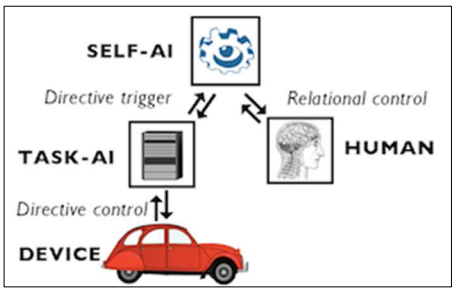
Let's look at an example.



In the DARPA video, John Launchbury cited the development of AI applications capable of driving cars. As he described them, these would be typical Task-AI systems that will have to meet high standards of reliability and predictability to perform their function usefully.

These systems would be, of course, programmed to avoid collisions while the car is moving but they do not manage the car's entire life-cycle. So, such a system is unconcerned about its own safety, the safety of the car it controls or the safety of

its passengers when the car is not moving. In other words, the Task-AI system tries to avoid collisions when triggered but does not actively seek to keep the car safe.



On the other hand, a Self-AI system, embedded in the car, would, in a sense identify (in terms of cognitive representations) with the car at the life-cycle or existence level, seeking to keep safe and well cared for in all circumstance. To do so, it will not only emit directive controls to the vehicle but may also attempt to influence, through relational exchanges, the behaviour of its human users to get them to contribute to its own existential objective.

FEASIBLE TODAY

These Self-AI systems I am describing here would generate cognitive representations of themselves in relation with their environment, pursue existential objectives and use relational interactions to influence human behavior. These are capabilities that are often attributed to self-awareness.

There is a widespread belief in the AI community that implementing this type of self-aware system would require massive and exotic computing resources and, as yet undiscovered, software techniques. It is believed these systems will only be feasible in the very far future, if at all, and are not worthy of any serious consideration, today.

Nothing could be further from the truth.

The systems I am describing here and that are entirely omitted from in the DARPA presentation can be implemented today with standard resources. Some are probably under development right now.

The reason these systems are feasible even though their control objectives can be viewed as extraordinarily complex is that they do not have to meet the same levels of reliability, predictability and correctness that are required in functional applications.

Most people believe that consciousness and self-awareness are located at the pinnacle of human intelligence. It follows, for them, that a system must first match or exceed human cognitive capabilities in every other aspect before any type of synthetic self-awareness can be implemented. **Not true.**

In a functional situation such as driving a car the sensory data contains a lot of essential information and the value of various control choices can be precisely measured. In such a situation it is necessary to build very efficient learning systems that can generate reliable and correct output.

Human behaviour on the other hand is highly unpredictable and the value of existential choices cannot be easily measured. In such a context correctness and reliability thresholds can be much lower. In fact, humans themselves are rarely very efficient in this area:

> *Do you know many humans who manage their own existence with optimal efficiency? Or for that matter, human organizations?*

In turn, these lower thresholds make it possible to generate acceptable results using techniques such as:

- Radical simplification of problem spaces or
- Applying transposition techniques to existing virtual reality models to generate contextual adaptation.

And this is what makes it possible today to implement systems that have a degree of self-awareness.

DEVELOPMENT DIRECTIONS

I expect the Self-AI systems I am describing will first be implemented in non critical areas such as the control of inexpensive disposable devices or as video game avatars. I expect that, in spite of their poor performance in terms of functionality, that their influence and importance will grow as their relational techniques improve and they establish bonds with their users.

There is also a possibility, however, that these self-aware systems, first developed as game avatars or digital companions will end up as components of software viruses and other malware. This, in my view, should be a matter of interest to an organization such as DARPA.

CONCLUSION

To conclude:

- DARPA recently published a video intended to demystify the field of Artificial Intelligence and address public misconceptions about it.
- The DARPA presenter described the past and future evolution of AI in terms of three waves of increasingly powerful-triggered utilitarian applications.
- This characterization completely omits a different type of AI system, that can generate cognitive self-representations and pursue existential goals using relational strategies.

I don't know if this omission in the DARPA video is intentional or if no one in that organization is aware of this separate stream of AI research. If it is the later case and DARPA is truly ignorant about these systems, then they are in for quite a surprise.

Dartmouth NS - 2017.03.04

REFERENCE

Launchbury, J. , A DARPA Perspective on Artificial Intelligence, Youtube video: https://www.youtube.com/watch?v=-O01G3tSYpU&t=3s.

Tardy, J.E., The Meca Sapiens Blueprint, Glasstree Academic Publishing, 2015