4. Interdisciplinary Conference on Electrics and Computer (INTCEC 2024)
11-13 June 2024, Chicago-USA

Brain Inspired Access Synthetic Consciousness Using A Neural Network Cluster

1st Oleksandr Jockusch
School of Computing
Southern Illinois University
Carbondale, IL, USA
oleksandr.jockusch@siu.edu

2nd Avery Leveque
School of Computing
Southern Illinois University
Carbondale, IL, USA
avery.leveque@siu.edu

3rd Henry Hexmoor
School of Computing
Southern Illinois University
Carbondale, IL, USA
hexmoor@siu.edu

***Abstract*—We are embarking on identifying rudiments of accountability for actions by the emerging AI models. A fundamental step on this path is to define how an AI model can be said to be conscious of its actions. We propose a novel model for synthetic consciousness and argue that this model is useful for AI accountability. With this proposed model, we aim to ascribe accountability for AI system-generated behavior that bears consequential impacts. The present framework is rooted in the neurology of the brain macro-structure, which we simulate using a cluster of artificial neural networks.**

***Index Terms*—neural network cluster, AI models, consciousness, accountability, neurology**

I. Introduction

Whereas phenomenal consciousness is a type of consciousness available to a living organism, *access consciousness* is a utilitarian variant that is characterized by explicit processes supporting consciousness (i.e., cognitive consciousness). We are embarking on developing access consciousness in part to narrowly identify accountability for actions performed by a specific AI-driven system. Such systems will be bestowed responsibility for their actions. Mobile robots as embodied AI systems are working alongside humans as mixed teams in various settings from hospitals to warehouses and even residential settings such as nursing homes. We must identify and hold to account parties (i.e., responsibility allocation for individuals) that perform consequential actions. A prerequisite for this is consciousness and actions that are intently generated. Hence, our goal is to develop forms of synthetic consciousness (SC). One form of this SC is derived from how the human brain lobes specialize with specific tasks and perform neural activities for synthesizing meaning (i.e., exteroception). *Exteroception* is the focus on processing data collected by the human sensory mechanisms from exogenous sources. In contrast, *interoception* is the focus on processing the internal needs of the organisms using the endogenously generated datum.

The Human life begins exclusively with the interoception bias and developmentally learns associations and processing that selectively balances interoception with exteroception. Consider that when cries and whimpers yield attention to needs, an infant learns to cry for attention. Certain brain lobes process raw sensory data to invoke precepts. We call the constellation of these sensory processing lobes as the *sense* generation component. In sum, the parietal lobe is separated from the occipital lobe by the parieto-occipital sulcus and is behind the central sulcus. It is responsible for processing much of the human sensory information. The parietal lobe also contains the somatosensory cortex. The Somatic (i.e., bodily) sensations arise from receptors positioned throughout the body that are responsible for detecting touch, *proprioception* (i.e. the position of the body in space), *nociception* (i.e. pain), and temperature. When such receptors detect one of these sensations, the information is sent to the thalamus and then to the primary somatosensory cortex. The primary somatosensory cortex is divided into 52 areas based on the delineations of the German neuroscientist Korbinian Brodmann [1].

The outputs of sensing are further processed in the brain by another neural bundle for invoking emotions and memories as well as encoding novel memories. The two-lobed Hippocampi are the predominant human brain counterparts of this apparatus. Specifically, episodic memories as well as recall of them take place within the hippocampi. We call this neural cluster the *think* component. E.g., Encountering a wild animal such as a gorilla in the zoo evokes a calm recall whereas encountering them in non-zoo environments generally will evoke fear and trepidation [1].

The output of the thinking component is fed into the neural component called the *act* that generates an action to be performed. Fig. 3 depicts our neural-inspired high-level components. Note the continual infinite loop with the triad of sense, think, and act. In **section 2** we will detail the motivations behind our system design, and introduce a proposed model of consciousness that is seen in biological organisms that can be applied to our SC model. **Section 3** will outline our proposed artificial neural network cluster (ANNC) that roughly parallels the functions of the brain lobes aligned with our model of consciousness. **Section 4** will illustrate the basic functions of the ANNC with a highly simplified model of a mouse brain that serves as our exemplar. **Section 5** will offer concluding remarks and the path forward.

II. Design Motivation

The idea for this endeavor emerged during the discussions on the evolution of biological organisms and their consciousness. These discussions later developed into deeper analysis of consciousness itself, and means of its emulation. We may apply the fundamentals of the biological evolution of organisms, especially the brain and central nervous system to the development of artificial intelligence, machine learning, and robotics.

In this section, we outline the two disparate concepts that coalesced as an inspiration for our proposed model - the structure and function of the brain and access synthetic consciousness loop.

A. Brain and The Central Nervous System

Natural evolution has resulted in the present-day hominids. The central nervous system (CNS) is a key component of hominids responsible for survival and prediction. An example of nervous system evolution is the emergence of the longest nerve known as the Vagus nerve, which is responsible for producing dorsal, fight or flight, and ventral states of user interface with the environment. Myriad forms of reasoning and planning are attributed to the human mind and extensively modeled with artificial intelligence algorithms. However, modeling is rarely focused on sections that mimic that of a CNS, whether that be hominids or any other organism. Our proposal attempts to build a model based on just that: the idea of a sectioned CNS, including a sort of **"brain"** with lobes.

Separation of cognitive tasks between the brain lobes help biological beings sense not only the world around them through *exteroception* but also collect information about themselves introspectively. For example, the *occipital lobe* facilitates perceiving the outside world. The auditory cortex within the temporal lobe does the same but for the auditory sense. The hypothalamus decrypts what is felt internally within the body such as tiredness and hunger. As part of the autonomous nervous system, it controls homeostasis subconsciously – as a whole, based on cues that may not rise beyond the subconsciousness for things like the heartbeat. Finally, regions like the prefrontal cortex, amygdala, and hippocampus take in all of these sensory experiences along with past experiences, i.e. memories, to create some sort of emotional response that will lead to eventual action.

This concept of modularity and division of tasks within the brain was in essence the inspiration for our system design. A system architecture with multiple interconnected modules which are responsible for different sub tasks, but are working in tandem to accomplish a bigger goal.

B. The Access Consciousness Loop

The second concept that served as a foundation of our system design is that of the access consciousness loop. Fig. 1 depicts our proposed model of consciousness as a continual feedback look for an organism (i.e., or a synthetic entity). The model is continual, meaning that it is cyclic. As long as the organism is alive (i.e., active), it cycles through basic steps. At the highest level, sensing leads to thinking, and that leads to acting. These three phases (*sense*, *think*, *act*) form a continual feedback loop. At a finer level, the loop is divided into eight steps, which break the major phases into more incremental tasks. Sensing is made up of collecting exteroceptive data from the environment (i.e., step 1), which triggers emotions, invoking relevant memories, or making new memories (i.e., step 2). Step 3 takes the triggers of emotion or memory (i.e., functions in the Hyppocampi) to determine an internal meaning for the organism. This can be considered a form of localization or establishing a context. Steps 3-5 largely mirror the functions in the Hyppocampi. Beyond establishing context, *awareness* is paramount. Determining the implications of the context is our 4th step. Once fully aware, the organism must make this awareness *personal* and find the impacts and implications for itself. Step 5 is the crux of our proposed *access consciousness*.

Once conscious, the organism must determine what it must do, which is the domain of the cerebral cortex. The cerebral cortex, also called gray matter, is the brain's outermost layer of nerve cell tissue. It has a wrinkled appearance from its many folds and grooves. The cerebral cortex plays a key role in memory, thinking, learning, reasoning, problem-solving, emotions, and consciousness. We call step 6 *desire formation*. We recall the continual agent model we called the *belief*, *desire*, and *intention* model (BDI) [2], [3], [4]. From "what to do" (i.e., step 6) we move on to the "how to do it" step (i.e., step 7). Performing the action determined from step 7 is our final step in the loop (i.e., step 8). After step 8, the loop returns to step 1, and so on. This model is inspired by the functions of biological organisms including humans.

It is also worth discussing the distinction between conscious and unconscious acts, as it is vital in establishing accountability for synthetically conscious AI systems. The endowment of the capacity for voluntary, conscious acts, must be preceded with powers of innate purpose. These purposes are the synthetic correlates to biological dives in living organisms. Furthermore, we must develop mechanisms to generate voluntary actions that support those purposes. For example, an enduring purpose for a robot can be to maintain and preserve a certain level of power. Nominally connected and voluntary, conscious actions are (a) report/notify low power status, (b) request power delivery/recharge, (c) relocate to a power source station. Consider when an animal produces alarm calls to get others to move away from a predator. With the act, it performs an *intentional*, conscious act to preserve safety. Acts that are conscious with sustained attention are known as *first-order intentional acts* [5].

These acts are contrasted with the zeroth order or unintended acts. The natural cause and effect of actions that are not conscious are unintended are known as zeroth order intentional acts. The producer of the first-order intentional acts is deemed accountable (i.e., responsible) for the consequences

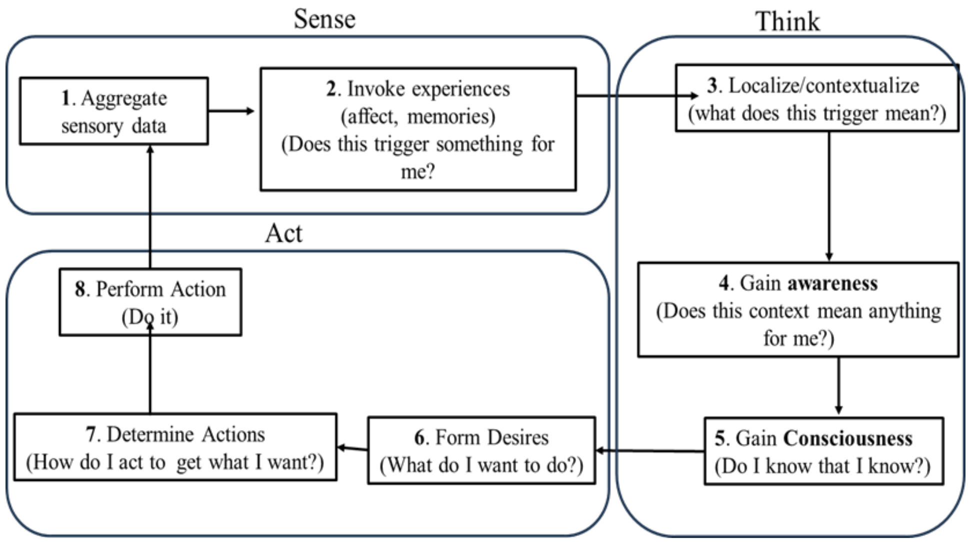

Fig. 1. Access consciousness loop

of those acts. There are more complicated (i.e., higher-order) intentional acts such as in the case of a mouse trap. The discussion of higher-order intentional acts and the degrees of subsequent accountability (i.e., degrees of liability and responsibility) are well beyond the scope of our current focus. By and large, first-order, premeditated actions are accountable, meaning that the performer of the action is eligible to be held to account and is responsible for results and ramifications. Fig. 2 illustrates the relationships among action types. There are a vast number of reasoning processes that produce action selection. For brevity, we consider planning and reasoning outside our current scope. Instead, our focus is on a single, atomic action chosen to be performed. Atomic actions are uncommon but serve to illustrate our main tenet surrounding consciousness.

A distinction is needed for actions performed by entities that can bear responsibility (i.e., humans and machines bestowed with legal status) versus entities that are devoid of legal responsibility (i.e., non-human animals). A predator may intentionally hunt and kill prey but there it is not blamed for the hunted prey. Whereas a misleading response from an AI assistant that leads to injury must be held liable and the system must account for the response. For entities with legal status, first-order intentional acts are *first-order accountable acts* herewith coined by us. To develop rigorous accountability by machines, we are striving to construct a model of synthetic consciousness as a first step. Beyond consciousness, subsequent steps will be required to conceive of *intentionality* and accountability.

Just as desirable for humans, we wish for actors (e.g., robots and humans) as well as disembodied actors (i.e., agents) to possess the traits of intentionality and accountability. Bovens [5] has put forward that the concept of accountability can be seen as both a virtue and a mechanism, introducing the idea that as long as someone or something is conscious, it can be held accountable. Therefore, understanding the role of accountability as part of consciousness is an important consideration when designing synthetically conscious AI systems, and we intend to delve deeper into this topic in our future research work.

III. Proposed System Description

Herein, we turn our attention to our proposed system capable of exhibiting *access synthetic consciousness*. At the core of our model is an Artificial Neural Network Cluster (ANNC). The ANNC architecture is inspired by the high-level structure and function of the brain, where different lobes are responsible for disparate cognitive functions. The proposed ANNC is composed of three blocks - *Sensory*, *Reflectory* (i.e., reasoning, planning, and problem-solving), and *Action*. The blocks correspond to three phases of the access consciousness cycle (*sense*, *think*, *act*) respectively. The sensory block contains three ANNs, while the reflectory and the action blocks both contain one ANN. Each ANN is responsible for a specific task or sub-task within the synthetic consciousness loop, vaguely paralleling cognitive function separation among the lobes within the brain. The motivation for combining multiple ANNs into a cluster is the fact that a singular ANN is not capable of performing all the tasks required for implementing the access consciousness loop. The use of multiple ANNs provides the ability to process heterogeneous data, as well as perform multiple distinct tasks, enabling synthetic consciousness to emerge. We present the details of the system description in the following subsections.

Hardware Requirements

Our proposed system is inherently software-focused, as we intend to maintain relative hardware independence. However, within the scope of this proposal, we are making several assumptions about the hardware that our system is going to be deployed with. The hardware system should include:

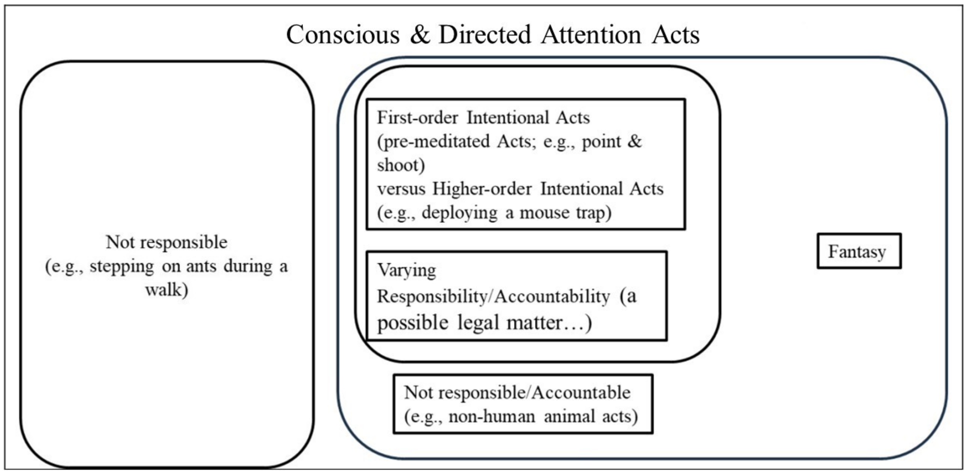

Fig. 2. Conscious and Direct Attention Acts

- depth perception capable cameras
- microphone(s)
- battery(s)
- an internal thermometer
- CPU

Data Representation within the ANNC

The mechanisms of representing the information exchanged within the ANNC are encoded. All the data exchanged between the ANNs is represented as *tensors*, as it is required with the application of ANNs. The data within a tensor indicates the presence or absence of a sensed object, emotion, or predicted action. For instance, if a specific object was predicted by the sensory ANNs as present, its corresponding value within the produced tensor would be set to ”1”; otherwise, the value would be set to ”0”.

Data Flow within the ANNC

The output tensors of the sensory ANN block are concatenated and passed on as input into the reflectory ANN block. Subsequently, the output tensor of the reflectory block is concatenated with the output tensors of the sensory block, and is passed on as input into the action ANN block, which in turn predicts the next action.

A. Sensory ANN Block

The three ANNs that constitute the sensory ANN block are responsible for the sense phase of the access synthetic consciousness cycle. More specifically, these ANNs simultaneously gather information about both the internal system state, as well as the external environment. For brevity, the typical anthropomorphic sensory data we are considering are visual, audio, and internal homeostatic system diagnostics data, with a separate ANN responsible for each type of sensory data. These can be extended to organic mechanisms for the organism as the occipital lobe, auditory cortex, and hypothalamus respectively. Following is the precise breakdown of each ANN in the sensory ANN block:

- The first ANN within the sensory ANN block is directly responsible for processing the visual data from the surrounding environment. For implementing this functionality, a Convolutional Neural Network (CNN) is our preferred choice due to its capacity to process spatial data. The visual data will be collected using Video Object Segmentation (VOS), which is a type of computer vision model that processes a video feed frame by frame and is trained to recognize objects within each frame. The objects recognized by the VOS CNN model represent the visual data collected during the sense phase of the access synthetic consciousness loop. As direct input, the VOS CNN will accept the pre-processed frames from the camera feed. As output, the VOS CNN will produce a tensor containing the information about which objects were detected in the current frame.
- The next ANN in the sensory ANN block is responsible for processing the audio data from the surrounding environment. This functionality is also likely to be implemented using a CNN. The machine learning technique best suited for this task is known as Audio Signal Segmentation (ASS). ASS CNN model is trained to recognize specific audio signals within an audio stream, which will give the ANNC the ability to recognize and differentiate between different sounds. The ASS CNN will take the feed of the microphones as input, and produce a tensor containing the information about which sounds were detected as output.
- The last ANN within the sensory block is responsible for processing the internal system diagnostics such as the battery level, the internal temperature, the CPU usage, etc. Based on the input data, this ANN will determine the overall state of the system as well as any potential concerns with the hardware or the software system components. This ANN serves as the function for replicating conscious organisms’ ability to sense and access their internal states. The ANN will take a tensor containing

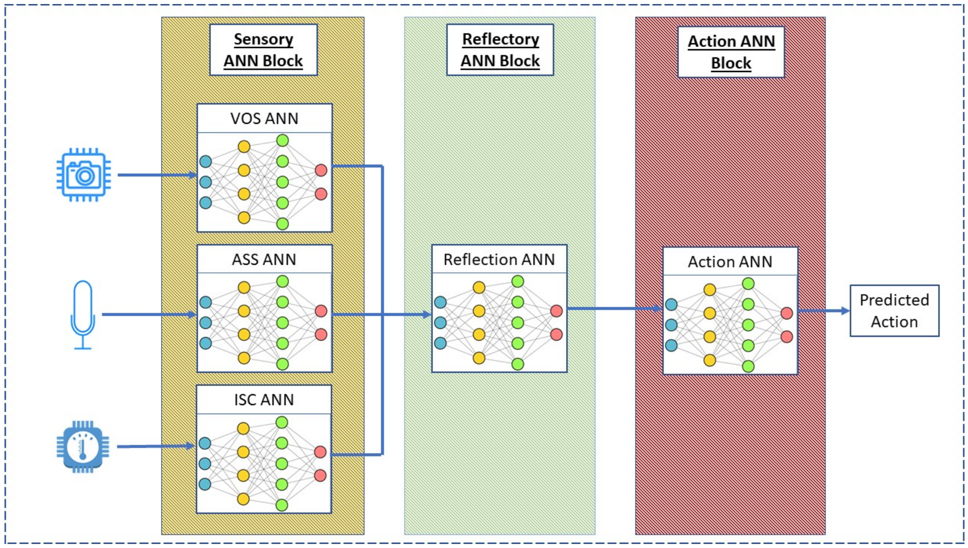

Fig. 3. Proposed ANNC Architecture

system diagnostic data as input, and produce a tensor containing the overall system state prediction as output.

B. Reflectory ANN Block

The reflectory block is made up of a single ANN, and serves to fulfill the *think* phase of the access synthetic consciousness cycle. The purpose of this block is to give ANNC the ability to contextualize and reflect on the gathered sensory information, gain awareness, and eventually access synthetic consciousness. The reflectory block takes in the data produced by the sensory block and synthesizes an appropriate emotional response. The output tensors of the sensory block ANNs are combined and passed on as input into the reflectory block ANN. The output tensor produced by the reflectory block indicates the predicted emotion/opinion/thought, which would be subsequently used in the next action prediction stage.

C. Action ANN Block

The action block contains one ANN and is responsible for the *act* phase of the access synthetic consciousness cycle. The purpose of this block is to give the ANNC the ability to form desires and decide on what actions to take, which is the last step in fulfilling the access synthetic consciousness loop. The action block takes in all the data previously collected in the ANNC and predicts the next action that is the most appropriate based on the received information. The input tensor of the action block contains the outputs of both the sensory and reflectory blocks. As a result, the action block ANN makes decisions based not only on the collected sensory data but also on the emotional response invoked by the sensory data, introducing more nuance to the decision-making process.

IV. Exemplar: Simulating a Simplified Consciousness of a Mouse

In this section we will illustrate the function of our proposed system by examining in detail a singe loop of the access synthetic consciousness cycle. For the sake of contextualization, we will simulate a simplified consciousness of a mouse. Note the formal depiction of the access synthetic consciousness cycle with ANNC outlined in the Algorithm 1.

A. Preconditions

Following are the preconditions we are considering in the scope of this example:

- The visual object classes that the system was trained to recognize are: mouse, cat, food
- The audio signal classes that the cluster was trained to recognize are: mouse noise, cat noise, person noise
- Internal system state classes that the cluster was trained to recognize: stable, low battery, overheating, CPU overload
- The emotional response classes that the cluster was trained to recognize: happiness, fear
- The action classes that the cluster was trained to predict: approach, escape

Note the order of the above mentioned classes, as it reflects the order the classes will be organized within the output tensors of their respective ANNs.

B. Access Synthetic Consciousness Loop Walkthrough

Following is the breakdown of a single access synthetic consciousness loop iteration using our proposed ANNC model:

- Step 1: Sense. During this step the visual, auditory, and internal sensory data is collected and encoded into tensors x, which are then appended to the tensor set R. Next, each tensor x is passed into its respective ANN, after which three output tensors z are produced. In our simulated example, we can assume that the visual perception CNN has detected a *cat* visual class, the audio signal segmentation CNN has detected a *cat noise* auditory class, and the internal state classification CNN has predicted a *stable* internal system state class. Fig. 4 depicts the output tensor of the visual perception CNN. Note that the second value in the tensor is equal to "1", which indicates that the cat visual object class was detected. All the other values in the tensor are equal to "0", which indicates that no other visual object classes were detected. Subsequently, the three output tensors are concatenated.

- Step 2: Think. During this step, the tensor containing the outputs of the sensory block is passed into the reflectory block ANN. The reflectory ANN outputs the tensor indicating the appropriate emotional response. Since the sensory block ANN has detected the presence of a nearby cat, the reflectory ANN will predict the *fear* emotional response as the most appropriate. Next, this output tensor will be concatenated with the output tensors of the sensory block. The resulting tensor will contain both the data captured by the sensory block and the emotional response predicted by the reflectory block.
- Step 3: Act. During this step the tensor containing the outputs of sensory and reflectory blocks is passed as input into the action block ANN. The action block then predicts the next action by taking into account both the collected sensory data and the predicted emotional response. Given that the sensory block detected the presence of a cat, and the reflectory block has predicted the *fear* emotional response, the action block will predict *escape* as the appropriate action. After the next action is predicted, the access synthetic consciousness loop start over.

Algorithm 1: Access Synthetic Consciousness Cycle Using Proposed ANNC Model

ANNC Initialized
for *access synthetic consciousness round* $i = 0, 1, 2 \ldots$ **do**
    Initialize blank tensor set R
    for *each input device* d **do**
        Collect sensory data from d into a tensor x
        $R_i \leftarrow$ Append the tensor x
    Initialize the sensory ANN block B_{sense}
    $\omega_s \leftarrow$ BlockUpdate (R, B_{sense})
    Initialize the sensory ANN block $B_{reflect}$
    $\omega_r \leftarrow$ BlockUpdate $(\omega_s, B_{reflect})$
    Initialize the sensory ANN block B_{act}
    $\omega_a \leftarrow$ BlockUpdate (ω_r, B_{act})
    $a \leftarrow \omega_a$
    Execute action a
BlockUpdate (R, B):
    Initialize blank tensor set ω_t
    for *each ANN* $a = 1, 2...$ *in* B **do**
        $z \leftarrow a(R_a)$
        $\omega_t \leftarrow$ Append tensor z

```
Tensor([[0 1 0]], shape=(1, 3)
```

Fig. 4. Output tensor of the visual perception CNN

V. Conclusions

We have proposed a generic artificial neural network framework that mimics brain macro-structure in a rudimentary sketch of an access synthetic consciousness process. With this, we hope to develop a measure of accountability for the burgeoning large-scale AI that is sweeping through all facets of our lives. Our model is rudimentary, anthropomorphic, and well-suited for a general-purpose autonomous mobile robot. The model is designed for ready scalability for modeling organic and man-made organisms. It is well suited to model embodied agents such as autonomous robots that have specialized sensory-motor capabilities beyond our generic description. Our model can be adapted to disembodied AI systems such as autonomous driverless terrestrial and aerial vehicles. We hope to herald a myriad of mature variants.

References

[1] D. Purves et al., Eds., *Neuroscience*, 6th ed. Suderland, MA: Sinauer Associates, an imprint of Oxford University Press, 2017.

[2] H. Hexmoor, and G. Beavers, "In search of simple and responsible agents," in *Workshop on Radical Agent Concepts*, Berlin, Heidelberg: Springer, 2002.

[3] H. Hexmoor, M. Huber, J. P. Müller, J. Pollock, and D. Steiner, "On the Evaluation of Agent Architectures," in *proceedings of the 6th international conference on Intelligent agents on agent theories, architectures, and languages, Orlando, Florida, USA, 1999*, Berlin, Heidelberg, New York: Springer Berlin, Heidelberg, 2000, pp. 106–116.

[4] H. Hexmoor, "Stages of autonomy determination," in *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 31, no. 4, pp. 509-517, Nov. 2001, doi: 10.1109/5326.983934.

[5] M. Bovens, "Two concepts of accountability: Accountability as a virtue and as a mechanism," *West European Politics*, vol. 33, no. 5, pp. 946–967, Aug. 2010, doi:10.1080/01402382.2010.486119.