Integrative Biological Simulation, Neuropsychology, and AI Safety

Gopal Sarma\textsuperscript{1,2} PhD, Adam Safron\textsuperscript{3} PhD, and Nick J. Hay\textsuperscript{4} PhD

1. Emory University School of Medicine, Atlanta, GA USA
2. The OpenWorm Foundation, Boston, MA USA
3. Northwestern University, Evanstan, IL USA
4. Vicarious AI, San Francisco, CA USA
• **Claim 1:** Simple organisms show complex behavior that continues to be difficult for modern AI systems. Neuronal simulations in virtual environments will allow these biological architectures to be used for AI research.
• **Claim 1**: Simple organisms show complex behavior that continues to be difficult for modern AI systems. Neuronal simulations in virtual environments will allow these biological architectures to be used for AI research.

• **Claim 2**: Value-alignment research may benefit from insights in neuropsychology and comparative neuroanatomy.
• **Claim 1:** Simple organisms show complex behavior that continues to be difficult for modern AI systems. Neuronal simulations in virtual environments will allow these biological architectures to be used for AI research.

• **Claim 2:** Value-alignment research may benefit from insights in neuropsychology and comparative neuroanatomy.

• **Claim 3:** Significant synergy may be achieved by coupling the two research programs described above.
Claim 1: Simple organisms show complex behavior that continues to be difficult for modern AI systems. Neuronal simulations in virtual environments will allow these biological architectures to be used for AI research.
Integrative Biological Simulations
Integrative Biological Simulations

- Computational platforms in which diverse, process-specific models, often operating at different scales, are combined into a global, composite model

Integrative Biological Simulations

• Computational platforms in which diverse, process-specific models, often operating at different scales, are combined into a global, composite model

• Why Develop These Platforms for the Life Sciences? Complexity of biological systems; complexity of the corresponding research community; reproducibility and research quality

Integrative Biological Simulation of Realistic Nervous Systems
Integrative Biological Simulation of Realistic Nervous Systems
Integrative Biological Simulation of Realistic Nervous Systems

- NEURON
- BluePyOpt
- NetPyNE
- Bionet
- Geppetto
- ChannelPedia
- NeuroMLDB
Claim 2: Value-alignment research may benefit from insights in neuropsychology and comparative neuroanatomy.
Neuropsychology and AI Safety
Neuropsychology and AI Safety

• View human values from the perspective of neuropsychological foundations

Neuropsychology and AI Safety

- View human values from the perspective of neuropsychological foundations
- Suggested decomposition of human values: 1) mammalian values 2) human cognition 3) several millennia of human social and cultural evolution

Neuropsychology and AI Safety

• View human values from the perspective of neuropsychological foundations

• Suggested decomposition of human values: 1) mammalian values 2) human cognition 3) several millennia of human social and cultural evolution

• Relevant disciplines include affective neuroscience, animal behavior, biological anthropology, comparative neuroanatomy, etc.

Neuropsychology and AI Safety

SEEKING

RAGE

FEAR

LUST

CARE

PANIC/GRIEF

PLAY

Claim 3: Significant synergy may be achieved by coupling the two research programs described above.
Integrative Biological Simulation

Neuropsychology and AI Safety

SEEKING
RAGE
FEAR
LUST
CARE
PANIC/GRIEF
PLAY

NEURON
BluePyOpt
NetPyNE
Bionet
Geppetto
ChannelPedia
NeuroMLDB
• **Claim 1:** Simple organisms show complex behavior that continues to be difficult for modern AI systems. Neuronal simulations in virtual environments will allow these biological architectures to be used for AI research.

• **Claim 2:** Value-alignment research may benefit from insights in neuropsychology and comparative neuroanatomy.

• **Claim 3:** Significant synergy may be achieved by coupling the two research programs described above.