

AAAI Spring Symposium Series 2014

Symposium 4: Implementing Selves with Safe Motivational Systems and Self-Improvement

Invited Speakers are one hour. Other speakers and session events are 30 minutes.

Monday, March 24 (Creating Self-Improving Selves)

9:00 am - 10:30 am Session
Opening, Logistics
Introductions
Mark Waser, Digital Wisdom Institute – “What does it mean to create a self?”

10:30 am - 11:00 am Coffee Break

11:00 am - 12:30 pm Session
Invited Speaker - Daniel Silver, Acadia University, “Lifelong Machine Learning and Reasoning”
Justin Brody, Goucher College, “Incorporating Elements of a Processual Self into Active Logic”

12:30 pm - 2:00 pm Lunch

2:00 pm - 3:30 pm Session
Boris Galitsky, Knowledge Trail Inc, “Finding Faults in Autistic and Software Active Inductive Learning System”
Andras Kornai, Budapest Institute of Technology, “Euclidean Automata”
Michael S. P. Miller, Piaget Modeler, “Serving Up Minds: JCB and PAM P2”

3:30 pm - 4:00 pm Coffee Break

4:00 pm - 5:30 pm Session
Sheldon O. Linker, Linker Systems, Inc. “Premise: A Language for Cognitive Systems”
Daniel Dewey, Future of Humanity Institute, “Reinforcement Learning and the Reward Engineering Principle”
Implementation Workshop – Self-Improving Selves

6:00 pm - 7:00 pm Reception

Tuesday, March 25 (Safe Motivational Systems)

9:00 am - 10:30 am Session
Rafal Rzepka, Hokkaido University, “Experience of Crowds as a Guarantee for Safe Artificial Self”
Jeanne Dietsch, Mobilerobots Inc, “AGI Ethics via Human Emotion: the Vital Link”
Richard Loosemore, Wells College, “The Maverick Nanny with a Dopamine Drip”

10:30 am - 11:00 am Coffee Break

11:00 am - 12:30 pm Session
Invited Speaker - Daniel Polani, University of Hertfordshire, “Empowerment: a Universal Utility for Intrinsic Motivation”
or “What to do when you do not know what to do?”
Morgan Waser, Virginia Commonwealth University, “A Human-Value-Driven Motivational System”

12:30 pm - 2:00 pm Lunch

2:00 pm - 3:30 pm Session
Invited Speaker – Steve Omohundro, Self-Aware Systems
“Ethics and Understanding & Managing the Unintended Consequences of Self-improvement”
Franco Cortese, IEET, “The Maximally Distributed Intelligence Explosion”

3:30 pm - 4:00 pm Coffee Break

4:00 pm - 5:30 pm Session
Deepak Justin Nath, “A Short Paper on Evaluation Schemes for Safe AGIs”
Mark Waser, Digital Wisdom Institute, “The Nuts & Bolts of Implementing a “Safe” Motivational System”
Implementation Workshop – “Safe Motivational Systems”

6:00 pm - 7:00 pm Plenary Session

Wednesday, March 26

9:00 am - 10:30 am Session
Invited Speaker – Pierre- Yves Oudeyer, Flowers Laboratory, Inria and Ensta ParisTech, France
“Developmental robotics: Lifelong learning and the morphogenesis of developmental structures”
By Invitation – Rafal Rzepka, Hokkaido University, “Pros and Cons of Borrowing Morality from P2P Civilization”

10:30 am - 11:00 am Coffee Break

11:00 am – 12:00 Session
By Invitation – Michal Ptaszynski, Kitami Institute of Technology, “Computational Phronesis as a Possible Path toward Machine Emotional Integrity”
Future Directions, Closing