

Lecture Notes in Artificial Intelligence

6830

Edited by R. Goebel, J. Siekmann, and W. Wahlster

Subseries of Lecture Notes in Computer Science

Jürgen Schmidhuber  
Kristinn R. Thórisson  
Moshe Looks (Eds.)

# Artificial General Intelligence

4th International Conference, AGI 2011  
Mountain View, CA, USA, August 3-6, 2011  
Proceedings

## Series Editors

Randy Goebel, University of Alberta, Edmonton, Canada  
Jörg Siekmann, University of Saarland, Saarbrücken, Germany  
Wolfgang Wahlster, DFKI and University of Saarland, Saarbrücken, Germany

## Volume Editors

Jürgen Schmidhuber  
University of Lugano & SUPSI, IDSIA  
Galleria 2, 6928 Manno, Switzerland  
E-mail: juergen@idsia.ch

Kristinn R. Thórisson  
Reykjavik University and Icelandic Institute for Intelligent Machines  
CADIA / School of Computer Science  
Menntavegi 1, 101 Reykjavik, Iceland  
E-mail: thorisson@ru.is

Moshe Looks  
Google Research  
1600 Amphitheatre Parkway, Mountain View, CA 94043, USA  
E-mail: madsience@google.com

ISSN 0302-9743 e-ISSN 1611-3349  
ISBN 978-3-642-22886-5 e-ISBN 978-3-642-22887-2  
DOI 10.1007/978-3-642-22887-2  
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2011933107

CR Subject Classification (1998): I.2, F.4.1, I.5, F.1-2

LNCS Sublibrary: SL 7 – Artificial Intelligence

© Springer-Verlag Berlin Heidelberg 2011

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

*Typesetting:* Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

# Preface

The original goal of the AI field was the construction of “thinking machines” – that is, computer systems with human-like general intelligence. Due to the difficulty of this task, for the last few decades the majority of AI researchers have focused on what has been called “narrow AI” – the production of AI systems displaying intelligence regarding specific, highly constrained tasks.

In recent years, however, more and more researchers have recognized the necessity – and feasibility – of returning to the original goals of the field. Increasingly, there is a call for a transition back to confronting the more difficult issues of “human level intelligence” and more broadly artificial general intelligence (AGI).

The AGI conferences are the only major conference series devoted wholly and specifically to the creation of AI systems possessing general intelligence at the human level and ultimately beyond.

Continuing the mission of the first three AGI conferences (most recently AGI-10, held at the University of Lugano, Switzerland), in August 2011, AGI-11 was held at the Google headquarters in Mountain View, California. AGI-11 gathered an international group of leading academic and industry researchers involved in scientific and engineering work aimed directly toward the goal of artificial general intelligence.

Keynote speeches were delivered by Ernst Dickmanns, the pioneer of self-driving cars, Peter Norvig, co-author of the highest-cited AI textbook, Zhongzhi Shi, and Aaron Sloman. The special session on neuroscience and AGI included a keynote speech delivered by Ed Boyden, a co-founder of optogenetics. Of the 103 submissions, 28 were accepted as full papers for this volume (27%), and 26 as short papers. Enjoy!

August 2011

Jürgen Schmidhuber  
Kristinn R. Thórisson  
Moshe Looks

# Conference Organization

## Organizing Committee

Moshe Looks (Conference Chair)	Google, USA
Ben Goertzel	Novamente, USA
Marcus Hutter	Australian National University, Australia
Sarah Bull	NICTA, Australia
David Orban	Humanity+, USA
John Laird	University of Michigan, USA
Randal Koene	Halcyon Molecular, USA
Bas Steunebrink	IDSIA, Switzerland
Brandon Rohrer	Sandia National Laboratories, USA
Blerim Emruli	Lulea University of Technology, Sweden
Stephen Reed	Texai, USA

## Program Chairs

Juergen Schmidhuber	IDSIA, Switzerland
Kristinn R. Thórisson	Reykjavik University, Iceland

## Program Committee

Sam Adams	IBM, USA
Sebastian Bader	Rostock University, Germany
Anirban Bandyopadhyay	National Institute for Materials Science, USA
Eric Baum	Baum Research Enterprises
Anslem Blumer	Tufts University, USA
James Bonaiuto	California Institute of Technology, USA
Joanna Bryson	University of Bath, UK
Antonio Chella	University of Palermo, Italy
Haris Dindo	University of Palermo, Italy
Yoni Donner	Google, USA
Marco Dorigo	Université Libre de Bruxelles, Belgium
Wlodzislaw Duch	Nicolaus Copernicus University, Poland
Matteo Gagliolo	Vrije Universiteit Brussel, Belgium
Deon Garrett	Icelandic Institute for Intelligent Machines, Iceland
Nil Geisweiller	Novamente LLC, USA
Faustino Gomez	IDSIA, Switzerland
J. Storrs Hall	Institute for Molecular Manufacturing, USA

Demis Hassabis	Gatsby Unit, UK
Pascal Hitzler	Wright State University, USA
Marcus Hutter	Australian National University, Australia
Matt Ikle	Adams State College, USA
Kamilla Jóhannsdóttir	University of Akureyri, Iceland
Benjamin Johnston	University of Technology, Sydney, Australia
Emanuel Kitzelmann	University of Bamberg, Germany
Andras Kornai	Boston University, USA
Jan Koutnik	IDSIA, Switzerland
Kai-Uwe Kuehnberger	University of Osnabrueck, Germany
Christian Lebiere	Carnegie-Mellon University, USA
Shane Legg	Gatsby Unit, UK
Thor List	Communicative Machines Ltd., UK
Andras Lorincz	Eötvös Loránd University, Hungary
Matt Luciw	IDSIA, Switzerland
Jiang Min	Xiamen University, China
Victor Ng-Thow-Hing	Honda Research Institute, USA
Eric Nivel	Reykjavik University, Iceland
Leo Pape	IDSIA, Switzerland
Jan Poland	ABB Corporate Research, Switzerland
Mark Ring	IDSIA, Switzerland and USA
Paul Rosenbloom	University of Southern California, USA
Sebastian Rudolph	Karlsruhe Institute of Technology, Germany
Daniil Ryabko	INRIA Lille, France
Rafal Rzepka	Hokkaido University, Japan
Alexei Samsonovich	George Mason University, USA
Ricardo Sanz	Technical University of Madrid, Spain
Tom Schaul	IDSIA, Switzerland
Matthias Scheutz	Indiana University, USA
Ute Schmid	University of Bamberg, Germany
Stuart Shapiro	SUNY Buffalo, USA
Lokendra Shastri	SETLabs, Infosys Technologies Limited, Germany
Zhongzhi Shi	Chinese Academy of Sciences, China
John Sowa	VivoMind, USA
Bas Steunebrink	IDSIA, Switzerland
Julian Togelius	IT University of Copenhagen, Denmark
Marc Toussaint	Technical University of Berlin, Germany
Lyle Ungar	University of Pennsylvania, USA
Wendell Wallach	Yale University, USA
Pei Wang	Temple University, USA
Marco Wiering	University of Groningen, The Netherlands
Mary-Anne Williams	University of Technology, Sydney, Australia
David Wolpert	NASA Ames, USA
Yixin Zhong	Beijing University of Posts and Telecommunications, China

## Local Organizing Committee

Peter Norvig	Google, USA
Bruce Klein	Singularity University, USA
Rod Furlan	Entrepreneur

## Demo Track Committee

Itamar Arel (Chair)	University of Tennessee, USA
---------------------	------------------------------

## Steering Committee

Ben Goertzel	Novamente, USA
Marcus Hutter	Australian National University, Australia

# Table of Contents

## Long Papers

Self-Modification and Mortality in Artificial Agents .....	1
<i>Laurent Orseau and Mark Ring</i>	
Delusion, Survival, and Intelligent Agents .....	11
<i>Mark Ring and Laurent Orseau</i>	
Coherence Progress: A Measure of Interestingness Based on Fixed Compressors .....	21
<i>Tom Schaul, Leo Pape, Tobias Glasmachers, Vincent Graziano, and Jürgen Schmidhuber</i>	
Sequential Constant Size Compressors for Reinforcement Learning .....	31
<i>Linus Gisslén, Matt Luciw, Vincent Graziano, and Jürgen Schmidhuber</i>	
Planning to Be Surprised: Optimal Bayesian Exploration in Dynamic Environments .....	41
<i>Yi Sun, Faustino Gomez, and Jürgen Schmidhuber</i>	
Optimal Direct Policy Search .....	52
<i>Tobias Glasmachers and Jürgen Schmidhuber</i>	
Nonlinear-Dynamical Attention Allocation via Information Geometry ...	62
<i>Matthew Ikle and Ben Goertzel</i>	
An Information Theoretic Representation of Agent Dynamics as Set Intersections .....	72
<i>Samuel Epstein and Margrit Betke</i>	
On More Realistic Environment Distributions for Defining, Evaluating and Developing Intelligence .....	82
<i>José Hernández-Orallo, David L. Dowe, Sergio España-Cubillo, M. Victoria Hernández-Lloreda, and Javier Insa-Cabrera</i>	
Structural Emergence in Partially Ordered Sets Is the Key to Intelligence .....	92
<i>Sergio Pissanetzky</i>	
Integrating Perception and Cognition for AGI .....	102
<i>Unmesh Kurup, Christian Lebiere, and Anthony Stentz</i>	



Systematically Grounding Language through Vision in a Deep, Recurrent Neural Network .....	112
<i>Derek D. Monner and James A. Reggia</i>	
Comparing Humans and AI Agents.....	122
<i>Javier Insa-Cabrera, David L. Dowe, Sergio España-Cubillo, M. Victoria Hernández-Lloreda, and José Hernández-Orallo</i>	
The LIDA Framework as a General Tool for AGI .....	133
<i>Javier Snaider, Ryan McCall, and Stan Franklin</i>	
From Memory to Problem Solving: Mechanism Reuse in a Graphical Cognitive Architecture .....	143
<i>Paul S. Rosenbloom</i>	
Rational Universal Benevolence: Simpler, Safer, and Wiser Than “Friendly AI” .....	153
<i>Mark Waser</i>	
The Collection of Physical Knowledge and Its Application in Intelligent Systems .....	163
<i>Benjamin Johnston</i>	
Rationality and General Intelligence .....	174
<i>Helmar Gust, Ulf Krumnack, Maricarmen Martínez, Ahmed Abdel-Fattah, Martin Schmidt, and Kai-Uwe Kühnberger</i>	
Wagging the Dog: Human <i>vs.</i> Machine Inference of Causality in Visual Sequences .....	184
<i>Florin Popescu</i>	
Learning Problem Solving Skills from Demonstration: An Architectural Approach .....	194
<i>Haris Dindo, Antonio Chella, Giuseppe La Tona, Monica Vitali, Eric Nivel, and Kristinn R. Thórisson</i>	
Compression and Intelligence: Social Environments and Communication .....	204
<i>David L. Dowe, José Hernández-Orallo, and Paramjit K. Das</i>	
OpenPsi: Realizing Dörner’s “Psi” Cognitive Model in the OpenCog Integrative AGI Architecture .....	212
<i>Zhenhua Cai, Ben Goertzel, and Nil Geisweiller</i>	
Extending Cognitive Architectures with Semantic Resources .....	222
<i>Alessandro Oltramari and Christian Lebiere</i>	
A Motivational System for Cognitive AI .....	232
<i>Joscha Bach</i>	

On Fast Deep Nets for AGI Vision . . . . .	243
<i>Jürgen Schmidhuber, Dan Cireşan, Ueli Meier, Jonathan Masci, and Alex Graves</i>	
Considerations for a Neuroscience-Inspired Approach to the Design of Artificial Intelligent Systems . . . . .	247
<i>Serge Thill</i>	
Brain Anatomy and Artificial Intelligence . . . . .	255
<i>L. Andrew Coward</i>	
<b>Short Papers</b>	
Information, Utility and Bounded Rationality . . . . .	269
<i>Pedro Alejandro Ortega and Daniel Alexander Braun</i>	
A Family of Gödel Machine Implementations . . . . .	275
<i>Bas R. Steunebrink and Jürgen Schmidhuber</i>	
Reinforcement Learning and the Bayesian Control Rule . . . . .	281
<i>Pedro Alejandro Ortega, Daniel Alexander Braun, and Simon Godsill</i>	
Societies of Intelligent Agents . . . . .	286
<i>Bill Hibbard</i>	
Towards a General Vision System Based on Symbol-Relation Grammars and Bayesian Networks . . . . .	291
<i>Elias Ruiz, Augusto Melendez, and L. Enrique Sucar</i>	
Reasoning in Non-Axiomatic Logic: A Case Study in Medical Diagnosis . . . . .	297
<i>Pei Wang and Seemal Awan</i>	
Measuring Agent Intelligence via Hierarchies of Environments . . . . .	303
<i>Bill Hibbard</i>	
Learning What to Value . . . . .	309
<i>Daniel Dewey</i>	
Learning the States: A Brain Inspired Neural Model . . . . .	315
<i>András Lőrincz</i>	
AGI Architecture Measures Human Parameters and Optimizes Human Performance . . . . .	321
<i>András Lőrincz and Dániel Takács</i>	
Multigame Playing by Means of UCT Enhanced with Automatically Generated Evaluation Functions . . . . .	327
<i>Karol Walędzik and Jacek Mańdziuk</i>	

From Sensorimotor Graphs to Rules: An Agent Learns from a Stream of Experience . . . . .	333
<i>Marius Raab, Mark Wernsdorfer, Emanuel Kitzelmann, and Ute Schmid</i>	
Three Hypotheses about the Geometry of Mind . . . . .	340
<i>Ben Goertzel and Matthew Ikle</i>	
Imprecise Probability as a Linking Mechanism between Deep Learning, Symbolic Cognition and Local Feature Detection in Vision Processing . . . . .	346
<i>Ben Goertzel</i>	
Generalization of Figure-Ground Segmentation from Binocular to Monocular Vision in an Embodied Biological Brain Model . . . . .	351
<i>Brian Mingus, Trent Kriete, Seth Herd, Dean Wyatte, Kenneth Latimer, and Randy O'Reilly</i>	
The Illusion of Internal Joy . . . . .	357
<i>Claude Touzet</i>	
Philosophically Inspired Concept Acquisition for Artificial General Intelligence . . . . .	363
<i>Iris Oved and Ian Fasel</i>	
Machine Lifelong Learning: Challenges and Benefits for Artificial General Intelligence . . . . .	370
<i>Daniel L. Silver</i>	
A Demonstration of Combining Spatial and Temporal Perception . . . . .	376
<i>Jianglong Nan and Fintan Costello</i>	
Towards Heuristic Algorithmic Memory . . . . .	382
<i>Eray Özkural</i>	
Complex Value Systems in Friendly AI . . . . .	388
<i>Eliezer Yudkowsky</i>	
Real-World Limits to Algorithmic Intelligence . . . . .	394
<i>Leo Pape and Arthur Kok</i>	
AGI and Neuroscience: Open Sourcing the Brain . . . . .	401
<i>Randal A. Koene</i>	
What Makes a Brain Smart? Reservoir Computing as an Approach for General Intelligence . . . . .	407
<i>Janelle Szary, Bryan Kerster, and Christopher Kello</i>	
<b>Author Index . . . . .</b>	<b>415</b>