

### LNAI Series Editors

Randy Goebel

*University of Alberta, Edmonton, Canada*

Yuzuru Tanaka

*Hokkaido University, Sapporo, Japan*

Wolfgang Wahlster

*DFKI and Saarland University, Saarbrücken, Germany*

### LNAI Founding Series Editor

Joerg Siekmann

*DFKI and Saarland University, Saarbrücken, Germany*

More information about this series at <http://www.springer.com/series/1244>

Tom Everitt · Ben Goertzel  
Alexey Potapov (Eds.)

# Artificial General Intelligence

10th International Conference, AGI 2017  
Melbourne, VIC, Australia, August 15–18, 2017  
Proceedings

*Editors*

Tom Everitt   
Australian National University  
Canberra, ACT  
Australia

Alexey Potapov   
St. Petersburg State University  
St. Petersburg  
Russia

Ben Goertzel  
OpenCog Foundation  
Hong Kong  
China

ISSN 0302-9743                      ISSN 1611-3349 (electronic)  
Lecture Notes in Artificial Intelligence  
ISBN 978-3-319-63702-0              ISBN 978-3-319-63703-7 (eBook)  
DOI 10.1007/978-3-319-63703-7

Library of Congress Control Number: 2017947039

LNCS Sublibrary: SL7 – Artificial Intelligence

© Springer International Publishing AG 2017

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, 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.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Printed on acid-free paper

This Springer imprint is published by Springer Nature  
The registered company is Springer International Publishing AG  
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

# Preface

The original goal of artificial intelligence (AI) was to build machines with human-level intelligence. As the field evolved, efforts became scattered across a wide range of “narrow” AI domains. The goal of the Artificial General Intelligence (AGI) community is to refocus on the original goal of human-level intelligence, and to explore the space of possible intelligences. Sometimes this means tying together narrow AI technologies into more complex systems and cognitive architectures. But equally often, AGI research involves finding new ways of looking at intelligence, including new algorithms, mathematical frameworks, and conceptualizations. A third branch of research covers the societal impact of AGI, and how to ensure its safe applicability.

This, the tenth AGI conference, took place during August 15–18 in Melbourne, Australia, against the backdrop of many exciting developments in traditional AI and machine learning. It is therefore only appropriate that the conference was hosted back-to-back with some major traditional AI and machine learning conferences: ICML and UAI in Sydney, and IJCAI also in Melbourne.

We received 35 high-quality papers, spanning a wide range of AGI topics. Out of these submissions, 21 papers (60%) were accepted for oral presentation. An additional six papers were accepted for poster presentation. Keynotes, tutorials, and workshops provided additional perspectives. In the keynotes, Christian Calude explored practical and theoretical aspects of incomputability, Marcus Hutter advertised universal artificial intelligence, Peter Cheeseman discussed recursively self-improving AI, and Elkhonon Goldberg connected biological insights about the brain with AI architectures. In the tutorials, Alexey Popatov suggested possible cross-fertilizations between AGI approaches, and Ben Goertzel envisioned a future unification. Naotsugu Tsuchiya gave a tutorial on AGI and consciousness. Finally, part of the last day was devoted to a workshop on understanding.

We wish to extend a deep thanks to the Program Committee for performing the essential task of quality control of the submissions. All papers received constructive feedback and an impartial evaluation from at least two independent reviewers. Only a handful of papers received less than three reviews. We also wish to thank the local committee for organizing the conference, and thank our generous sponsors, the OpenCog Foundation and Hanson Robotics.

June 2017

Tom Everitt  
Alexey Potapov  
Ben Goertzel

# Organization

## Program Committee

Hadi Afshar	Australian National University, Australia
Joscha Bach	MIT Media Lab, USA
Tarek Richard Besold	University of Bremen, Germany
Jordi Bieger	Reykjavik University, Iceland
Dietmar Bruckner	Bernecker + Rainer
Cristiano Castelfranchi	Institute of Cognitive Sciences and Technologies
Antonio Chella	Università di Palermo, Italy
Mayank Daswani	Australian National University, Australia
Tom Everitt	Australian National University, Australia
Stan Franklin	Institute for Intelligent Systems, University of Memphis, USA
Arthur Franz	Independent researcher
Nil Geisweiller	Novamente LLC
Ben Goertzel	AGI Society
Jose Hernandez-Orallo	Universitat Politecnica de Valencia, Spain
Bill Hibbard	University of Wisconsin - Madison, USA
Marcus Hutter	Australian National University, Australia
Matt Iklé	Adams State University, USA
Benjamin Johnston	University of Technology Sydney, Australia
Garret Katz	University of Maryland College Park, USA
Kevin Korb	Monash, Australia
Kai-Uwe Kuehnberger	University of Osnabrück, Institute of Cognitive Science, Germany
Ramana Kumar	Datat61, CSIRO, and UNSW, Australia
John Licato	Indiana University/Purdue University - Fort Wayne, USA
Sean Markan	Eudelic Systems LLC
Maricarmen Martinez	Universidad de los Andes, Colombia
Amedeo Napoli	LORIA Nancy, France
Eric Nivel	CADIA, Reykjavik University, Iceland
Abdul Rahim Nizamani	Dfind IT
Eray Özkural	Gök Us Araştırma ve Geliştirme Ltd.
Sergei Obiedkov	National Research University Higher School of Economics
Laurent Orseau	Google Deepmind
Günther Palm	Ulm University, Germany
Maxim Peterson	ITMO University
Alexey Potapov	AIDEUS

Nico Potyka	Universität Osnabrück, IKW, Germany
Paul S. Rosenbloom	University of Southern California, USA
Rafal Rzepka	Hokkaido University, Japan
Oleg Scherbakov	ITMO university
Ute Schmid	University of Bamberg, Germany
Leslie Smith	University of Stirling, UK
Javier Snaider	Google
Bas Steunebrink	IDSIA
Claes Strannegård	Chalmers University of Technology, Sweden
Kristinn Thórisson	Reykjavik University, Iceland
Volkan Ustun	USC Institute for Creative Technologies
Viktoras Veitas	Global Brain Institute, VUB
Mario Verdicchio	Università degli Studi di Bergamo, Italy
Pei Wang	Temple University, USA
Roman Yampolskiy	University of Louisville, USA
Byoung-Tak Zhang	Seoul National University, South Korea

## **Additional Reviewer**

Majeed, Sultan

# Contents

## Architectures

From Abstract Agents Models to Real-World AGI Architectures: Bridging the Gap . . . . .	3
<i>Ben Goertzel</i>	
A Formal Model of Cognitive Synergy . . . . .	13
<i>Ben Goertzel</i>	
Generic Animats . . . . .	23
<i>Claes Strannegård, Nils Svängård, Joscha Bach, and Bas Steunebrink</i>	
Self-awareness and Self-control in NARS. . . . .	33
<i>Pei Wang, Xiang Li, and Patrick Hammer</i>	
DSO Cognitive Architecture: Unified Reasoning with Integrative Memory Using Global Workspace Theory . . . . .	44
<i>Khin Hua Ng, Zhiyuan Du, and Gee Wah Ng</i>	

## Mathematical Foundations

A General (Category Theory) Principle for General Intelligence: Duality (Adjointness). . . . .	57
<i>Steven Phillips</i>	
Abstract Representations and Generalized Frequent Pattern Discovery . . . . .	67
<i>Eray Özkural</i>	
On Hierarchical Compression and Power Laws in Nature. . . . .	77
<i>Arthur Franz</i>	
From First-Order Logic to Assertional Logic . . . . .	87
<i>Yi Zhou</i>	

## Algorithms

Genetic Algorithms with DNN-Based Trainable Crossover as an Example of Partial Specialization of General Search . . . . .	101
<i>Alexey Potapov and Sergey Rodionov</i>	



Deductive and Analogical Reasoning on a Semantically Embedded Knowledge Graph . . . . .	112
<i>Douglas Summers-Stay</i>	
Computational Neuroscience Offers Hints for More General Machine Learning . . . . .	123
<i>David Rawlinson and Gideon Kowadlo</i>	
Generating Single Subject Activity Videos as a Sequence of Actions Using 3D Convolutional Generative Adversarial Networks . . . . .	133
<i>Ahmad Arinaldi and Mohamad Ivan Fanany</i>	
One-Shot Ontogenetic Learning in Biomedical Datastreams . . . . .	143
<i>John Kalantari and Michael A. Mackey</i>	
The MaRz Algorithm: Towards an Artificial General Episodic Learner . . . . .	154
<i>Christian Rodriguez, Giselle Marston, William Goolkasian, Ashley Rosenberg, and Andrew Nuxoll</i>	
<b>Safety</b>	
A Game-Theoretic Analysis of the Off-Switch Game. . . . .	167
<i>Tobias Wängberg, Mikael Böörs, Elliot Catt, Tom Everitt, and Marcus Hutter</i>	
What People Say? Web-Based Casuistry for Artificial Morality Experiments . . . . .	178
<i>Rafal Rzepka and Kenji Araki</i>	
Malevolent Cyborgization . . . . .	188
<i>Nadisha-Marie Aliman</i>	
<b>Understanding</b>	
Understanding and Common Sense: Two Sides of the Same Coin? . . . . .	201
<i>Kristinn R. Thórisson and David Kremelberg</i>	
The Pedagogical Pentagon: A Conceptual Framework for Artificial Pedagogy . . . . .	212
<i>Jordi Bieger, Kristinn R. Thórisson, and Bas R. Steunebrink</i>	
<b>Human Cognition</b>	
An Information-Theoretic Predictive Model for the Accuracy of AI Agents Adapted from Psychometrics . . . . .	225
<i>Nader Chmait, David L. Dowe, Yuan-Fang Li, and David G. Green</i>	

Bandit Models of Human Behavior: Reward Processing in Mental Disorders. . . . .	237
<i>Djallel Bouneffouf, Irina Rish, and Guillermo A. Cecchi</i>	
Analyzing Human Decision Making Process with Intention Estimation Using Cooperative Pattern Task . . . . .	249
<i>Kota Itoda, Norifumi Watanabe, and Yoshiyasu Takefuji</i>	
Pursuing Fundamental Advances in Human Reasoning . . . . .	259
<i>Timothy van Gelder and Richard de Rozario</i>	
<b>Philosophy</b>	
<i>A Priori</i> Modeling of Information and Intelligence . . . . .	265
<i>Marcus Abundis</i>	
<b>Author Index</b> . . . . .	275