

Cognitive Technologies

Managing Editors: D.M. Gabbay J. Siekmann

Editorial Board: A. Bundy J.G. Carbonell
M. Pinkal H. Uszkoreit M.M. Veloso W. Wahlster
M.J. Wooldridge

More information about this series at <http://www.springer.com/series/5216>

Robert Trappl

Editor

A Construction Manual for Robots' Ethical Systems

Requirements, Methods, Implementations



Springer

Editors

Robert Trapp
Austrian Research Institute for Artificial
Intelligence (OFAI) of the Austrian Society
for Cybernetic Studies (OSGK)
Vienna, Austria

Center for Brain Research,
Medical University of Vienna
Vienna, Austria

Managing Editors

Prof. Dov M. Gabbay
Augustus De Morgan Professor of Logic
Department of Computer Science
King's College London
Strand, London, UK

Prof. Dr. Jörg Siekmann
Forschungsbereich Deduktions- und
Multiagentensysteme, DFKI
Saarbrücken, Germany

ISSN 1611-2482

ISSN 2197-6635 (electronic)

Cognitive Technologies

ISBN 978-3-319-21547-1

ISBN 978-3-319-21548-8 (eBook)

DOI 10.1007/978-3-319-21548-8

Library of Congress Control Number: 2015957251

Springer Cham Heidelberg New York Dordrecht London

© Springer International Publishing Switzerland 2015

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.

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media
(www.springer.com)

Preface

In 1906, the Austrian writer Alexander Roda Roda, born Sándor Friedrich Rosenfeld, wrote the story “Das Justizklavier” (The Justice Piano). In some state in the Maghreb—Roda Roda does not locate it precisely—an inventor asks for an audience with the potentate. The potentate, interested in what the man might offer him, receives him. The inventor opens a big trunk and inside there is a small piano. The potentate is curious about this piano, and the inventor explains: “You have big expenses for judges, lawyers, and attorneys, and problems with people who complain about the length of trials. This is my invention, the Justice Piano, that will remove all these problems.”

“You see that my piano has white and black keys, like any piano. But these keys are inscribed, the black ones with the possible crimes of the defendant, e.g., ‘Burglary’, ‘Leg Fracture’, ‘Adultery’. [In German the words rhyme: ‘Einbruch’, ‘Beinbruch’, ‘Ehebruch’.] The white keys are for extenuating causes, e.g., ‘Minor’, ‘Without Previous Convictions’, ‘Drunk’. You simply press the white and black keys and without any delay the piano prints out the verdict.”

At first, the potentate is enthusiastic but then starts pondering and finally says: “A piano normally has at least two pedals, one for forte and one for piano. Couldn’t you attach two pedals, one with the inscription ‘Opposition’ and the other with ‘Government’?”

Things have become far more complex in the meantime, and a simple rule-based expert system would not suffice to help a robot in making ethical decisions. Many papers and even a few books have been published on how to enable robots to make such decisions, a few even claiming that such decisions should not be made by robots. But if it is accepted that robots will have to make decisions which are ethical, then how should a designer of the robot’s software proceed? What are the prerequisites of ethical systems, what methods are available, and are there already applications?

In an attempt to answer these questions, the Austrian Research Institute for Artificial Intelligence (OFAI) identified potential contributors, mostly from their previous publications, invited them to submit position papers, and then invited them

to a 2-day workshop at the OFAI in Vienna to present and exchange their ideas. The workshop formed the basis for most of the chapters of this book.

The contributors are affiliated with various universities, among them Eindhoven University of Technology; the University of Hartford; Rensselaer Polytechnic Institute; Université Pierre et Marie Curie, Paris; and Universidade Nova de Lisboa.

A book like this would not have been possible without the commitment of many persons. First, I thank the authors who took great pains to enhance their original position papers into book chapters by including new material and by considering the comments in and outside the discussions.

Second, I want to thank my colleagues at the OFAI who have been of great help, especially Karin Vorsteher for her great efforts in proofreading and formatting and for many other activities for which there is not enough space to list.

Third, I thank the series editors, Jörg Siekmann and Dov Gabbay, for including this book in the Cognitive Technologies book series, and the Springer Computer Science editor Ronan Nugent for his support in the publication process.

Finally, I thank the Austrian taxpayers whose money allowed us to develop the workshop, pay for the travel and hotel expenses of the participants, and then prepare this book. We received this money through the Austrian Federal Ministry for Transport, Innovation and Technology, with Doris Bures as then Federal Minister, now President of the Austrian Parliament, and the very supportive officers, Ingolf Schaedler, Michael Wiesmueller, and Karl Supa, to whom I offer my sincere gratitude for trusting that I would finally present a useful product.

I hope you enjoy studying this book.

May 2015

Robert Trapp

Contents

1	Robots' Ethical Systems: From Asimov's Laws to Principlism, from Assistive Robots to Self-Driving Cars.....	1
	Robert Trapp	
Part I Requirements		
2	Robot: Multiuse Tool and Ethical Agent.....	11
	Brigitte Krenn	
3	Towards Human–Robot Interaction Ethics.....	31
	Sabine Payr	
4	Shall I Show You Some Other Shirts Too? The Psychology and Ethics of Persuasive Robots	63
	Jaap Ham and Andreas Spahn	
Part II Methods		
5	Ethical Regulation of Robots Must Be Embedded in Their Operating Systems.....	85
	Naveen Sundar Govindarajulu and Selmer Bringsjord	
6	Non-monotonic Resolution of Conflicts for Ethical Reasoning	101
	Jean-Gabriel Ganascia	
7	Grafting Norms onto the BDI Agent Model	119
	Mihnea Tufiş and Jean-Gabriel Ganascia	
Part III Implementations		
8	Constrained Incrementalist Moral Decision Making for a Biologically Inspired Cognitive Architecture.....	137
	Tamas Madl and Stan Franklin	

9 Case-Supported Principle-Based Behavior Paradigm..... 155
Michael Anderson and Susan Leigh Anderson

**10 The Potential of Logic Programming as a Computational
Tool to Model Morality** 169
Ari Saptawijaya and Luís Moniz Pereira