# An Anthropology of Robots and Al

Annihilation Anxiety and Machines

Kathleen Richardson

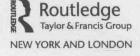


# An Anthropology of Robots and AI

Annihilation A

chines

Kathleen Richaruson



First published 2015 by Routledge 711 Third Avenue, New York, NY 10017

and by Routledge 2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

Routledge is an imprint of the Taylor & Francis Group, an informa business

© 2015 Taylor & Francis

The right of Kathleen Richardson to be identified as author of this work has been asserted in accordance with sections 77 and 78 of the Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this book may be reprinted or reproduced or utilised in any form or by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying and recording, or in any information storage or retrieval system, without permission in writing from the publishers.

Trademark notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Library of Congress Cataloging-in-Publication Data Library of Congress Control Number: 2015930211

ISBN: 978-1-138-83174-2 (hbk) ISBN: 978-1-315-73642-6 (ebk)

Typeset in Sabon by Apex CoVantage, LLC

## An Anthropology of Robots and AI

This book explores the making of robots in labs at the Massachusetts Institute of Technology (MIT). It examines the cultural ideas that go into the making of robots, and the role of fiction in co-constructing the technological practices of the robotic scientists. The book engages with debates in anthropological theorizing regarding the way that robots are reimagined as intelligent, autonomous and social, and woven into lived social realities. Richardson charts the move away from the "worker" robot of the 1920s to the "social" one of the 2000s, as robots are reimagined as companions, friends and therapeutic agents.

Kathleen Richardson is Senior Research Fellow in the Ethics of Robotics in the School for Computer Science and Informatics, Faculty of Technology, De Montfort University, Leicester, UK.

### Routledge Studies in Anthropology

1 Student Mobility and Narrative in Europe
The New Strangers

The New Strangers Elizabeth Murphy-Lejeune

- 2 The Question of the Gift Essays across Disciplines Edited by Mark Osteen
- 3 Decolonising Indigenous Rights Edited by Adolfo de Oliveira
- 4 Traveling Spirits
  Migrants, Markets and Mobilities
  Edited by Gertrud Hüwelmeier
  and Kristine Krause
- 5 Anthropologists, Indigenous Scholars and the Research Endeavour Seeking Bridges Towards Mutual Respect Edited by Joy Hendry and Laara Fitznor
- 6 Confronting Capital
  Critique and Engagement in
  Anthropology
  Edited by Pauline Gardiner
  Barber, Belinda Leach and
  Winnie Lem
- 7 Adolescent Identity
  Evolutionary, Cultural and
  Developmental Perspectives
  Edited by Bonnie L. Hewlett

- 8 The Social Life of Climate Change Models Anticipating Nature Edited by Kirsten Hastrup and Martin Skrydstrup
- 9 Islam, Development, and Urban Women's Reproductive Practices Cortney Hughes Rinker
- 10 Senses and Citizenships
  Embodying Political Life
  Edited by Susanna Trnka,
  Christine Dureau, and Julie Park
- 11 Environmental Anthropology
  Future Directions
  Edited by Helen Kopnina and
  Eleanor Shoreman-Ouimet
- 12 Times of Security
  Ethnographies of Fear, Protest and the Future
  Edited by Martin Holbraad and Morten Axel Pedersen
- 13 Climate Change and Tradition in a Small Island State The Rising Tide Peter Rudiak-Gould
- 14 Anthropology and Nature Edited by Kirsten Hastrup
- 15 Animism and the Question of Life Istvan Praet

- 16 Anthropology in the Making Research in Health and Development Laurent Vidal
- 17 Negotiating Territoriality
  Spatial Dialogues Between State
  and Tradition
  Edited by Allan Charles Dawson,
  Laura Zanotti and Ismael Vaccaro
- 18 HIV/AIDS and the Social Consequences of Untamed Biomedicine Anthropological Complicities Graham Fordham

- 19 Environmentalism, Ethical Trade, and Commodification Technologies of Value and the Forest Stewardship Council in Chile Adam Henne
- 20 An Anthropology of Robots and AI Annihilation Anxiety and Machines Kathleen Richardson

In loving memory of my brother Mark

## Acknowledgments

First and foremost, I would like to thank Rodney Brooks, Jamie Rollins, Una-May O'Reilly (and the Le-Baron family who welcomed me into their home), Martin Martin, Brian Adams, Paul Fitzpatrick, Eduardo Torres-Jara, Jessica Howe, Charlie Kemp, Jessica Banks, Aaron Edsinger, Theresa Langston, Anne Lawthers (aka Ladybug), Annika Pfluger and Ron Wiken. I am very grateful to all the members of the lab for allowing me to come and see their fascinating work. I would like to say a special thank you to Lijin Aryananda, whose ongoing help and support I continue to receive as a very personal friend and as a professional robotic scientist. Also Max Berniker, Rebecca Bureau and Minshu Son—these were really the best of friends. Also I would like to thank all the users of the Stata Center who shared their thoughts with me about the space. I was so deeply moved by the generosity of the staff, students and faculty at MIT and made many lifelong friends.

There were many others who supported my academic studies and helped me along the way as friends, family and supporters, and these include: Stephen & Karen Jones, Michael Sun, Andy Aryananda and Popo Wiryanti. I could not have completed my studies without the crucial financial support of the following: The Jirehouse Foundation, The Economic Social Research Council and The British Academy Postdoctoral Fellowship.

I would like to thank the editors at *Etnofoor* for their kind permission in allowing me to reproduce some of the arguments in Chapter 5 from: Richardson, K 2010, 'Disabling as mimesis and alterity: making human-oid robots at the Massachusetts Institute of Technology', *Etnofoor*, vol. 22, no. 1, pp. 75–90.

I would like to thank two mentors from the anthropology department at Cambridge, Marilyn Strathern and Nikolai Ssorin-Chiakov. A special thanks to my friend Vita Peacock who is my important confidant, and new friends who offered ideas on different chapters: Floyd Codlin, Marek Sinason and Robert Oates. I would also like to thank the editors at Routledge for their support.

Finally, I would like to make a special tribute to CB who has helped me to understand the true meaning of love as a mutual and creative relationship between I and Thou.

# Contents

	Acknowledgments	xi
	Introduction: Annihilation Anxiety and Machines	1
1	Revolutionary Robots	24
2	Out of Body Minds	39
3	Social Robots	60
4	The Gender of the Geek	77
5	The Dissociated Robot	92
6	Fantasy and Robots	112
	Conclusion: Loving the Attachment Wounded Robot	130
	Index	133

## Introduction

## Annihilation Anxiety and Machines

In an extreme view, the world can be seen as only connections, nothing else.

Tim Berners-Lee, Weaving the Web, 1999, p. 14.

The Terminator movies (1984–2003) show examples of robots that are super-advanced intelligent machines intent on destroying humanity to assure their supremacy. The Terminator is significant to begin this narrative, as it is one of the most popular fictions of a robot and it carries a central theme about human destruction. Whether you look to the past of robots or the present, this enduring theme of destruction returns. I respect that there are many other kinds of robots to consider such as robot companions, robot lovers, therapeutic robots, domestic robots and others, and we will explore these different imaginings of the robot in what follows, but for now, we will focus on the theme of human annihilation by robots.

The Terminator film caused something of a stir when first released in 1984 and was seen by millions of people around the world in the first year of broadcast. It features a high-profile Hollywood action actor, Arnold Schwarzenegger, who has a very unique stature; he is known for his toned muscled physique and stands at 1.88 meters or over 6 feet 2". Schwarzenegger's speech is marked by his strong Austrian-intoned English, and his speech and language are jokingly referred to as mechanical and formalistic. Some have rudely suggested he makes the "perfect" robot! While this is not the case, it is true that we take our cultural and technological models of robots from fictions. Multiple tides flow from fictions to living practices of technoscience.

A quick summary of the plot of the first *Terminator* is necessary. Set in the future of 2029 (not far away now), a super-advanced cyborg is sent to 1984 to kill Sarah O'Connor. In this dystopian future of 2029, super-intelligent machines rule the Earth, and authority over the remaining humans is maintained by killer robots. The Terminator T-800 Model 101 is sent back in time and must destroy Sarah O'Connor. Sarah is the mother of the future leader of the human rebellion. The machines figure if they can stop her child from being born, they can save themselves problems later. In this tale that

twists and turns, and folds together the future, present and past, human and nonhuman into its narrative texture, the film represents an iconic Euro-American portraval of robots as destrovers.

This popular fiction of robots has something important to tell us about the cultural theme of destruction, and more frightening themes followed in each subsequent film: Terminator 2: Judgment Day (1991), Terminator 3: Rise of the Machines (2003), and Terminator Salvation (2004). And the Terminator story is not over yet: in July 2015, Terminator: Genesis is scheduled for release, and speculating from the title, involves a hint at a rebirth. Only time will tell what the next installment of this robot saga has in store!

Hollywood filmmakers may receive some reward for shaping the cultural imagination of robots, but it was not them, but another more esoteric and radical avant-garde playwright to whom we must make our first tribute in

recognizing the robot as a cultural entity, and a destructive one.

The first robots emerged as characters 1920s play, R.U.R. (Rossum's Universal Robots), written by Czech playwright Karel Čapek. The play is unique. It is the first to coin the term "robot" and features the first cultural representations of robots. R.U.R. is the first play of modern fiction to bring about the end of humanity as a narrative plot of complete human annihilation (Reilly 2011). This being the first work of modern fiction to do this is significant, as prior to this, only in religious tales such as the New Testament's Book of Revelation is human annihilation a central feature when the apocalypse comes.

The robot—as first given life in a text and through theatrical performance by its creator Capek in R.U.R.—is a device to explore the fears of terminus in human existence brought about by mechanization, political ideologies and high modernism, and it speaks to the theme of humanity's end. Set in the tumultuous political era of the 1920s, Čapek took the idea of the factory worker one step further by inventing the robot. He created a laboring entity to work with limited subjectivity, a functionally competent laboring device. The term robot is from the Slavic term for work ("robota"), but Čapek, inspired by his artist brother Josef, drew on another meaning of the term relating to the "robota economy", an agricultural system where peasants work extra, providing for their landowners needs before their own. Robot is Czech for 'compulsory service', akin to Slav "robota", meaning 'servitude, hardship' (Merriam-Webster 1971, p. 1964).

As we reflect back on the robot in the 1920s play and the contemporary fiction of it, there is a recurring message:

### BEWARE YOUR END, HUMANITY!

In which case, we must take seriously the fear of the end of the human that is circulated in robot narratives.

In this book, I will preface and interlace each chapter with tales from robotic fictions because I want to argue that robotic fictions are taken into the lived realities of robotic practices and transferred into the making of robots, returning back into those fictions. This book is a reading and appreciation of these fictions by observing the making of robots in labs at the Massachusetts Institute of Technology (MIT). MIT is a world-renowned science and technological institution, repeatedly in the top three of the world's top research institutions (QS World Rankings 2014). MIT has a presence in popular culture that has formalized its mystique. In the 1950s classic, *The Day the Earth Stood Still* (1951), an alien spacecraft lands in Washington, guarded by a robotic life form. In the panic that ensues, it is MIT scientists that the US government calls on to help "rationalize" the situation, decipher the mystery of the alien visitor and calm the American public. In American culture at least, MIT scientists stand for impersonal rationality and arguably masculine authority in the fields of science and technology.

By the time I began thinking of my fieldwork in the early 2000s, humanoid robot labs only numbered a few around the world, notably in cities in Japan, such as Tokyo and Waseda. MIT's robot lab was one of the first in the US to begin a program of making humanoid robots, funded by a generous Defense Advanced Research Projects Agency (DARPA) grant. Robots and violence are frequent bedfellows, even when the robots produced by military funding seem to have no direct application for a military purpose,

such as building a robot child.

What is in a name anyway? In labs at MIT, I realized lab titles were fragile, coming and going depending on the grant or new focus of the research director. As a visiting researcher to robot labs in the US and the UK, the lab name is really an umbrella term, often for a multitude of research activities, some humanoid in focus and others not. The MIT robotics lab was set in the Artificial Intelligence Laboratory but shared the same physical space as MIT Computer Science. In 2003 these two departments merged to become The Computer Science and Artificial Intelligence Laboratory (CSAIL). All the CSAIL researchers were relocated to a new campus building, designed by architect Frank Gehry, known for his radical, geometrically distorted designs (Gilbert-Rolfe and Gehry 2002).

In keeping with issues of anthropological commitments to confidentiality, I have given the people and the robots in the lab pseudonyms. Some of these pseudonyms I have playfully taken from Čapek's play, R.U.R. As the robotic scientists I work with produce artifacts such as memos, scholarly dissertations, books, papers and robots, I have only referred to public material if such activities do not conflict with my initial commitment to honor the relationships with my interlocutors, many of whom are still my friends. I experienced considerable generosity from the lab group and researchers at MIT. I found their work and their lifeworlds extraordinary, and I hope some of that uniqueness is reflected in this book.

The robotics lab had taken a humanoid turn in early 2000s after its group director had been inspired to build humanlike robots after watching the film 2001: A Space Odyssey, made in 1968 but set in the year 2001.

#### 4 Introduction

The lab pioneered the first sociable robot (an oxymoronic term no doubt): A robot designed for social interaction with the intended aim of developing to such an extent that its future kin would be sophisticated enough to be a companion to humans.

This book then is about the theories and technologies that go into the making of robots, as well the people who make them and how their stories

and narratives feed into the machines they create.

Haraway's observation that the boundary between fiction and reality is thoroughly breached by technoscience is, of course, accurate (1991). But what I will attempt to show in these pages is that the Real is continually asserting itself in the making of robots, and there is a sphere outside cultural constructions that has its own separate properties. The Real is the boundary. Robotics in its own ways is confronted by its own realities. When constructing models of the mechanical human, theory and practice become intertwined in distinctive, sometimes unpredictable ways. In these labs, the robotic scientists continually referenced robotic fictions when producing robots, and the robots were repeatedly meeting the constraints of the Real: the physical, social and cultural environments that acted as containers. The Real and the fictional played off against each other in unusual ways, most notably in how the theme of robot destruction was addressed by these researchers. The cultural image of the threatening robot informed the making of the robots in the lab. The following information was provided on a robotics lab website at MIT:

Q: Are you ever worried that your robot might get 'too intelligent' or 'too powerful'?

A: No—we have programmed the robot to spare our lives in the event that it ever attempts to organize its brethren in a bloody revolution against the human race.

(MIT Humanoid Robotics Group n.d.)

Here the theme of destruction is taken up and diffused in a light-hearted way, but robots and artificial intelligence (AI) threats are present and receive more than a passing dismissal as will become apparent in the pages that follow.

#### ANNIHILATION ANXIETY: TO REDUCE TO NOTHING

The last few decades of anthropological theorizing have been beset by a number of theoretical problems that have resisted the dualistic analytical consequence of Cartesian dualisms and the ways these constructions have played themselves out in the construction of what life is (Latour 1993, Haraway 1991). One may say that anthropology as a discipline has suffered (and overcome) a kind of separation anxiety—about how to describe, resolve and

explain dichotomous relations including those between: persons and things (Gell 1998; Strathern 1988), humans and machines (Haraway 1991; Haraway 2003; Suchman 2006; Hicks 2002; Rabinow 2011), humans and animals (Haraway 2003; Haraway 1991; Ingold 2012), the body and the mind (Csordas 1999; Featherston & Burrows 1995), humans and nonhumans (Latour 1993; Latour 2005), fact and fiction (Haraway 1991; Graham 2002), and public and private spaces (Buchli and Lucas 2001; Buchli 1997).

If anthropology is said to have dealt with and overcome separation anxiety, why is the theme of human terminus brought about by machines a persistent and recurring theme in contemporary Euro-American cultural life? Latour (1993) takes this one step further and proposes that underscoring the fear of machines is a result of asymmetrical humanism (separation anxiety):

How could the anthropos be threatened by machines? It has them, it has put itself into them, it has divided up its own members among their members, it has built its own body with them. How could it be threatened by objects? They have all been quasi-subjects circulating within the collective they traced. It is made of them as much as they are made of it.

(1993, p. 138)

For Latour, the fear of the machine is an outcome of artificially separated categories, and this is reflected in the fear of objects (robots, viruses, supercomputers or meteors) that possess autonomy and can come back and haunt humanity as detached other.

Could the fear of the machines really be an outcome of 'asymmetrical humanism', as Latour proposes? I want to suggest that fear of robots and machines is the outcome of symmetrical anti-humanism, where humans and nonhumans are placed on a par, and the human is ascribed no distinctive quality over other agents—where human agents are reduced to nothing. This is presented as an anthropological emphasis on process in the absence of ontological difference. The robot has historically been a way to talk about dehumanization and the elevation of the nonhuman. The first meanings of the robot were primarily about dehumanization, and hence Čapek's robots were human, made of flesh, blood, bones and veins, but assembled on a mechanical production line with a scientific formula (2004, p. 13). It was other artists in the 1920s that took the robot character from the play and turned it into a machine. We can look to the robot in its historical sense and its contemporary manifestations in labs and in fictions to explore these points further. I frame this recurring fear, in contrast to separation anxiety as annihilation anxiety.

Annihilation anxieties are produced by an analytical position that rejects ontological separations, combined with radical anti-essentialism—when humans and nonhumans become comparable. The dystopian horror presented in *R.U.R.* and *The Terminator* films relate to a fear of terminus,

#### 6 Introduction

but the fear of the robot uprising is an existential fear about the end of the human (though with \$8 billion in forecasted funding for military robots by 2016 (ABIResearch 2011), the physical threat of destruction is not so fictional). The robot is a way to reflect on the violence of World War I and the unprecedented destruction of human life mediated by machines. The end of the human then is intimately related to violence: death is the ultimate end of the human.

What does annihilation mean? Annihilation is one of those terms encompassing multiple meanings, and I call upon all of those meanings in proposing an analytical framework to make sense of robots and AI systems. On the one hand, annihilation means the 'act of annihilating' or 'state of being annihilated' (Webster's Third New International 1971, p. 87). We are already familiar with Euro-American narratives of technological revenge in the form of Frankenstein's monster from the nineteenth-century classic tale by Mary Shelly (1969), or The Matrix Trilogy (1999-2003), where humans are imagined as batteries for AI systems. Annihilation also means 'cessation of being: NOTHINGNESS' (Webster's Third New International 1971, p. 87). Annihilation is derived from the verb "annihilate" ('ending'), and the Latin "annihilates", past participle of "annihilare": 1. 'to cause to be of no effect', 2. 'to look upon as nothing', 3. 'to reduce to nothing' (Webster's Third New International 1971, p. 87). These meanings open up another way of reflecting on endings and nothingness. Central to this discourse on robots is to highlight the reduction of the human to nothing, as a nondistinct agent in anthropological theorizing. 'To reduce to nothing' is also about the erasing of differences between humans and nonhumans. As anthropological theorizing takes an 'ontological turn' shaped by 'actornetworks', 'assemblages', 'meshwork', and 'companion species', so too are the human and nonhuman interconnected, even enmeshed with each other (Latour 2005, Rabinow 2011, Ingold 2012, Haraway 2003). Aside from the meanings the term annihilation possesses in popular language, it has meanings in physics, too, which are worth considering: '. . . the process whereby an electron and a positron unite and consequently lose their identity as particles transforming themselves into short gamma rays' (Webster's Third New International 1971, p. 87). In this sense, annihilation means something more than the mere disappearance and end of phenomena: a stage of merging occurs before one thing is created from these two forms. Out of nothing does come something-at least in theoretical physics.

In Buddhist philosophies too, annihilation of the ego is the highest state of being a human can attain. Japanese roboticist Masahiro Mori, theorist of the uncanny valley, writes, 'human beings have self or ego, but machines have none at all. Does this lack cause machines to do crazy, irresponsible things? Not at all. It is people, with their egos who are constantly being led by selfish desires to commit unspeakable deeds. The root of man's lack of freedom (insofar as he actually lacks it) is his egocentrism. In this sense, the ego-less machine leads a less hampered existence' (Mori 1999, p. 49; my

emphasis). In the Buddha and the Robot, Mori writes of his Buddhist vision of science, technology and robots. In Buddhist philosophies, the relations between different kinds of things are seen as interrelated. 'As I consider questions of this sort, I am reminded of the Buddhist axiom that "nothing has an ego". This means that nothing exists in isolation; everything is linked with everything else' (Mori 1999, p. 28).

Cartesian dualism, which proposed the mind as transcendent and the body as immanent, did capture something about the nature of ontological difference. In rejecting Cartesian dualism, anti-dualist categories have emerged ('cyborgs', 'meshworks', 'actor-networks', and 'assemblages'), but such styles propose a multiplicity without proposing any ontological difference of the different entities. Cartesian dualism has not been resolved-it has been side-stepped into a form of merging. In rejecting the ontological difference that was captured in the theorizing of dualism, a form of the "I/ego" is also threatened: the "I" as a human subject and different from other entities. Robotic scientists and AI theorists bring these issues to the fore in the way they create artificial beings (Helmriech 1998).

#### **EVERYTHING IS CONNECTED**

"Everything is connected" is a phrase we hear repeatedly: from chaos theory, when a butterfly flaps its wings, creating havoc elsewhere (Gleick 1994), to globalization (Erikson 2003) that emphasizes global flows of markets, labor, goods, services and capital. Let us consider Marilyn Strathern's (2014) points in relation to this statement:

Indeed, the more so-called 'bounded' notions of society and culture are held up to criticism, along with the systems and structures that were once their scaffold, the more relations, relationships, the relational, relationality, are evoked as prime movers (of sociality) in themselves. Quite aside from identifying relations in structures, systems of classification, co-variation, and so forth, the concept is equally forcefully applied to any new object of knowledge, emergent configuration, or coconstruction, and not only in a passive sense (everything is connected), but in the active sense of the observer making phenomena appear, illuminating them, by the concept.

(p. 5; my emphasis)

In the 'active sense of making phenomena appear', Strathern highlights the construction of connections between everything. Strathern (2014, p. 10) takes up these points and develops philosopher John Locke's ideas of association when what becomes connected to something else is dependent on the types of associations that are crafted. Locke's theory of associations was also a theme that interested cybernetics pioneer Norbert Weiner. In Weiner's classic text Cybernetics: or, Control and Communication in the Animal and the Machine, he outlines a theory of cybernetic systems as organic, mechanical and tied together by control and communication systems (1961). Wiener draws on Locke's themes, exploring 'the possibility of assigning a neural mechanism to Locke's theory of the association of ideas' (Wiener 1961, p. 156). Weiner uses the example of recognition of the face of a man to explore this issue: 'how do we recognize the identity of the features of a man, whether we see him in profile, in three-quarters face, or in full face? (1961, p. 156). Weiner, like Strathern, was interested in the parts of the person, and how ever-diminishing parts could still stand in for the whole of the person (Strathern 1988).

The feminist model of the cyborg developed by Donna Haraway needs to be honored in this history of humans and machines, and though written over 20 years ago, the cyborg reveals something distinctive when contrasted with different cultural imaginations of robots. The cyborg is an analytical device with which to assess the breakdown of organism and machine as distinctive categories:

Although the cyborg image originated in space and science fiction to refer to forms of life that are part human and part machine, it is by no means confined to the world of technology. Rather, cyborg anthropology calls attention more generally to the cultural production of human distinctiveness by examining ethnographically the boundaries between humans and machines and our vision of the differences that constitute those boundaries.

(Downey, William & Dumit 1995, pp. 264-265)

The cyborg was appropriated by Haraway as a polemical tool to critique social relations, and in this sense, it is similar to the robot. Whereas the robot as imagined by its creator Čapek expressed the fear of boundary transgressions between human and nonhuman, Haraway's cyborg takes them as given, and she pushes the boundary transgressions further in her political work:

I want to signal three crucial boundary breakdowns that make the following political-fiction (political-scientific) analysis possible. By the late twentieth century in United States scientific culture, the boundary between human and animal is thoroughly breached. . . . The second distinction is between animal-human (organism) and machine. . . . The third distinction is a sub-set of the second: the boundary between physical and non-physical is very imprecise for us'.

(1991, pp. 151-153)

For Haraway, the cosmologies that constitute modernism have been called into question via new technologies and feminist theorizing. The