

Reading List

Vår-2017 INFO381

Articles

"The Nature, Importance, and Difficulty of Machine Ethics" by James H. Moor
http://www.psy.vanderbilt.edu/courses/hon182/The_Nature_Importance_and_Difficulty_of_Machine_Ethics.pdf

"Machine morality: bottom-up and top-down approaches for modelling human moral faculties"
Wendell Wallach, Colin Allen and Iva Smit
<http://link.springer.com/article/10.1007/s00146-007-0099-0>

"The social dilemma of autonomous vehicles," by Jean-François Bonnefon, Azim Shariff, and Iyad Rahwan, Science, 2016
<http://science.sciencemag.org/content/352/6293/1573>

"Towards an ethical robot: Internal models, consequences and ethical action selection."
Winfield, A. F., Blum, C. and Liu, W. (2014) In: Mistry, M., Leonardis, Aleš, Witkowski, M. and Melhuish, C., eds. Advances in Autonomous Robotics Systems: Proceedings of the 15th Annual Conference, TAROS 2014, Birmingham, UK, 1-3 September 2014., pp. 85-96
<http://eprints.uwe.ac.uk/23487>

"Formal verification of ethical choices in autonomous systems" by Louise Dennis, Michael Fisher, Marija Slavkovic, Matt Webster
<http://www.sciencedirect.com/science/article/pii/S0921889015003000>

"ETHEL: Toward a Principled Ethical Eldercare System." M. Anderson and S. Anderson
<http://www.aaai.org/Papers/Symposia/Fall/2008/FS-08-02/FS08-02-002.pdf>

"Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings"
Bolukbasi et al.
<https://arxiv.org/pdf/1607.06520.pdf>

"Granny and the robots: ethical issues in robot care for the elderly" Sharkey and Sharkey
<http://link.springer.com/article/10.1007/s10676-010-9234-6>

"Governing lethal behavior: embedding ethics in a hybrid deliberative/reactive robot architecture." R. Arkin
<http://dl.acm.org/citation.cfm?doid=1349822.1349839>

"Governing Lethal Behavior: Embedding Ethics in a Hybrid Deliberative/Reactive Robot Architecture - Part 2: Formalization for Ethical Control." R.Arkin
<http://ebooks.iospress.nl/publication/4069>

"Unethical Research: How to Create a Malevolent Artificial Intelligence" by Roman Yampolsky and Federico Pistono
<https://arxiv.org/abs/1605.02817>

Chapters 1,7, 18,19,20,21,26,27 from a)
Chapters 4,6,7 and 12 from b)

Books

- a) Artificial Intelligence: A Modern Approach Paperback – 3rd edition by Stuart Russell, Peter Norvig
- b) Beginning Ethics: An Introduction to Moral Philosophy by Lewis Vaughn

Lecture notes are also part of the literature list.