

Silo-Busting Building a Transdisciplinary Community for Beneficial AI



Wendell Wallach
Asilomar, January 5, 2016

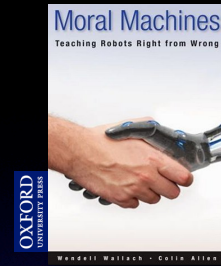


“Alas, poor Yorick! I knew him.”

- Moral Philosophy and Applied Ethics
 - Machine Ethics or Machine Morality
 - Engineering Ethics

- Managing Complex Adaptive Systems
 - Testing
 - Forensics
 - Resilience Engineering
 - Joint Cognitive Systems

- Science and Technology Policy
 - Soft governance
 - Agile or adaptive governance



- Robotics
 - Social Robotics
 - Affective Computing
- Philosophy
 - Naturalistic Philosophy
 - Philosophy and Computing
 - Deontic Logic

Value Alignment



Stuart Russell



The Hastings Center/FLI Project

Control and Responsible
Innovation in the Development
of Autonomous Systems

Three Brainstorming Workshops

- April 2016 – Exposure.
 - Excluded Law, LAWS, & Tech Unemployment
 - Representatives from: ONR, WEF & IEEE
- Oct/Nov 2016 – Value Alignment/Moral Machines.
- April 2018 – Specific practical near-term challenges that lay foundations for solving the control problem.

Challenges/Tensions

- Languages/Terms/Misunderstandings
 - The Languages of Ethics
- Near Term vs. AGI/Existential Risks
- Social Theorists v. Physical Scientists/Engineers
 - Ease/Difficulty of the Challenge
 - Naysaying v. ways forward

~~naysaying~~

Verboten!

Measuring Results



Open a Highway to



New Model for the Governance of Emerging Technologies



Gary Marchant



Issues Manager

- Coordinate activities of stakeholder
- Comprehensive monitoring
 - Flag issues/gaps
 - Find solutions within the robust set of available mechanisms
 - Mandated to avoid regulation where possible
 - Favor soft governance
- Nimble/Flexible/Adaptive/Lean

Existing Policy Mechanisms

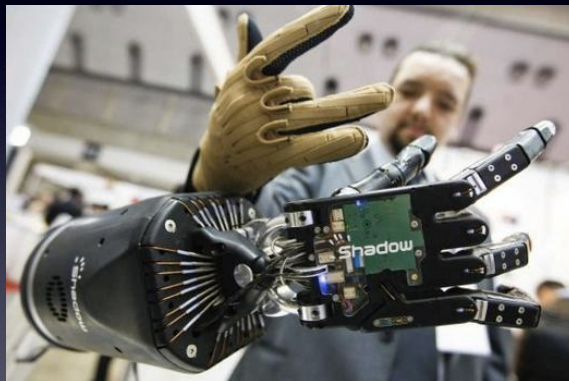
- **Government**

- Financing of Research
- Laws & Regulations
 - Govt. agencies
 - Oversight
- Research Ethics
 - Human Subject
 - Animal
 - Biosafety

- **Soft Governance**

- Industry Standards
- Codes of Conduct
- Certification
- Laboratory Practices and Procedures
- Insurance Policy
- Statement of principles
- Social norms

Technological Solutions



Pilot Projects



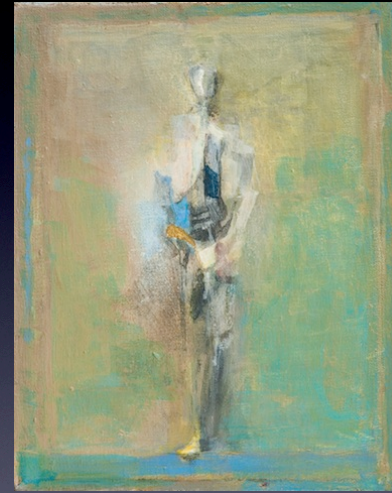
- AI/Robotics
- Synthetic Biology

International Coordination or Harmonization



**Building Global
Infrastructure to Ensure
that AI and Robotics are
Beneficial**

Thank You!



Email address: wendell.wallach@yale.edu