

ROBOPHILOSOPHY 2024



Memory slices by Anna Strasser
DISCLAIMER: JUST MEMORIES – AIMING FOR CORRESPONDENCE WITH REALITY BUT CANNOT GUARANTEE IT.



RE-ENVISIONING ETHICS :

FROM MORAL MACHINE TO EXTENSIVE REGULATION

HAVE WE BEEN UNDERESTIMATING THE SOCIO-TECHNICAL CHALLENGES POSED BY RO(BOT)S – PHYSICAL SYSTEMS AND VIRTUAL BOTS?

We will not produce AI systems capable of making even satisfactory choices in complex situations

where uncertainty reigns, multiple values converge, and the information available is inadequate to project meaningful consequences for various courses of action.

international governance?

AI will pose safety and security risks far beyond ...

- A vast infrastructure to ensure AI safety will be required.

ONTOLOGICAL QUESTION!

neither machines nor humans

FOSTER collective problem solving!!!



WENDELL WALLACH

Many of the complexities inherent in managing intelligent systems can not be adequately met by scientific innovation, existing ethical constraints, or weak regulations forged by legislatures under the capture of the AI oligopoly.

PHYSICAL & DIGITAL GENAI AVATARS

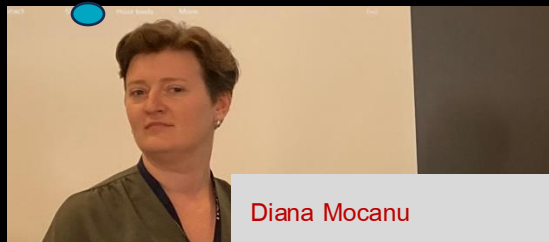
with Diana Mocanu | George Bogateanu | Stefano Dafarra | Radu Uszkai | Mihaela Constantinescu

ARTIFICIAL LEGAL AGENCY FOR SOCIAL ROBOTS VS. GENAI AVATARS

Can we delegate a bit of legitimation to AVATARS?

- accounts of AI agency in legal literature rely on "sense-think-act paradigm" + "intentional stance"
- **THERE IS A SPECTRUM OF AGENCY** (liberal agency)
- actor-network theory (Latour)

soul machines



THE ETHICS OF HUMAN-ROBOTIC AVATARS INTERACTION. - INSIGHTS FROM GAMING

What happens when the virtual world & the real world come together?

- avatars in gaming
- learning for the real world?



THE USE OF AVATARS IN MEDICINE: OPTIMISING BIOLOGICAL MODELS WITH PERSONAL HEALTH DATA

HUMANOID AVATAR SYSTEMS & THEIR IMPACT ON HRI

- humanoid iCub3
 - with locomotion & facial expressions
 - immersive sensory feedback concerning visual, auditory, haptic, weight, touch modalities
- iFeel
 - custom-made wearable technologies for motion & force tracking
- ❖ Through real-world examples, we'll unveil the practical applications of avatar systems, from remote collaborations to live engagements at prominent events.



Stefano
Dafarra

MORAL RESPONSIBILITY AND GENAI AVATARS

GenAI avatars question our understanding of agency

- moral responsibility with causation / freedom / knowledge deliberation
- ascribing a degree-based agency to avatars & avatar controllers
 - **RESPONSIBILITY & PROXIMITY GAPS AHEAD!!**

NEW FORM OF AGENCY
EMERGING



Mihaela Constantinescu



WHO DO WE BECOME WHEN WE TALK TO MACHINES?

In recent years, programs (robots and chatbots) built on generative AI have offered themselves as companions that care –presented, for example, as potential coaches, psychotherapists, and romantic companions – as artificial intimacy, our new AI.

A study of users of these programs makes it clear that adjacent to the question of what these programs can do is another: What are they doing to us –for example, to the way we think about human intimacy, agency, and empathy?

When we communicate on screens, we distance ourselves from each other. We lose the ability to put ourselves in the place of others and the ability to negotiate differences.

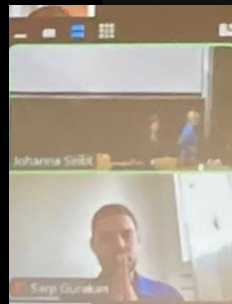
BUT

no stakes: no pain

PRETEND EMPATHY

signals the presence of another but there is no other

We condemn social media while embracing generative AI companions.





Depiction: 3 scenes

- (1) *Base scene*:
a wooden frame with a painting
- (2) *Proximal scene (depiction proper)*:
a woman sitting
- (3) *Distal, or depicted, scene*:
Mona Lisa smiling at us

Social robots as depictions

1. *Asimo_{base}* is the raw material, the mechanical apparatus, that belongs to *Asimo*'s base scene. It is made of metal, paint, sensors, and other material, and it has parts that move or create sounds.
2. *Asimo_{prox}* is the depiction proper. Like a puppet, marionette, or ventriloquist dummy, it is a depictive prop for the social being it represents. It depicts *Asimo_{dist}* in physical appearance, in its movements, and in its sounds.
3. *Asimo_{dist}* is the character depicted by *Asimo_{prox}*. He is the human-like being named *Asimo*.

EMBODIMENT MATTERS

- bodies create a sense of presence
- embodied beings have to be 'dealt with'
- evoke a double reality, combining aspects of both artefact & depicted character
- robot bodies are constant reminders of the depicted nature of the social being, comprising depictive, supportive, external and imagined features

ROBOTS ARE LESS DANGEROUS TECHNOLOGIES THAN DISEMBODIED TECHNOLOGIES LIKE SOCIAL MEDIA, CHATBOTS, ETC.

IT IS HARD TO ANTICIPATE FOR WHAT WE NEED REGULATIONS

- don't overrate conversations
 - machines as practice partners, triggering an openness to talk
- there are other things to do than deep conversations
- do things together / shared experience
 - you can get a glimpse through messengers ...
 - use tools to facilitate skills

MAY AI BE WITH YOU

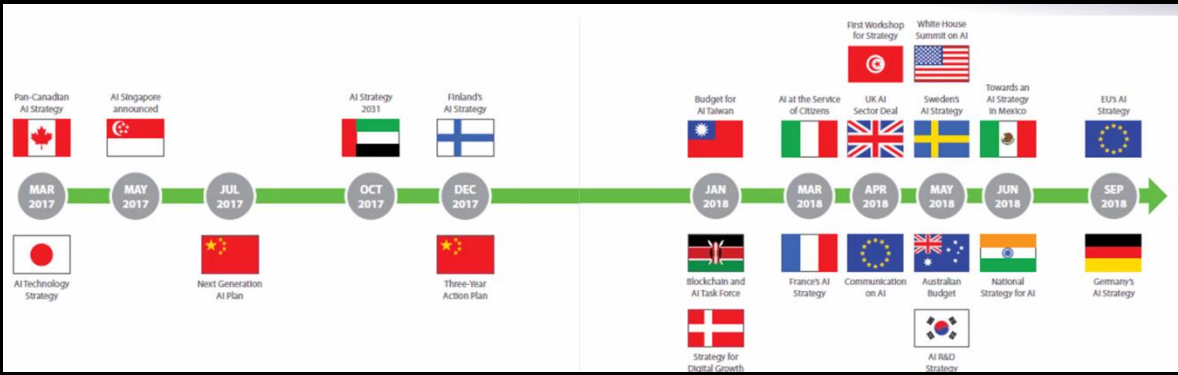
AGENCY & AUTOMATION IN THE AGE OF ALGORITHMIC MODERNITY

3 HARD TRUTH

- I. institutional changes
- II. mental health risks
- III. diminished agency

TECHNOLOGICAL TSUNAMI

- all countries wrote reports / strategies which are surprisingly similar
- spending huge amounts of money on AI



THEORIZING THE PRESENT AGE

- BECK'S REFLEXIVE MODERNIZATION - Chernobyl 1986
- GIDDENS'S LATE MODERNITY - Berlin Wall fell 1989
- THRIFT'S SOFT CAPITALISM – Asian Financial Crisis 1997
- BAUMAN'S LIQUID MODERNITY - Dot.com Bubble 2000

We need a more advanced adult debate!!

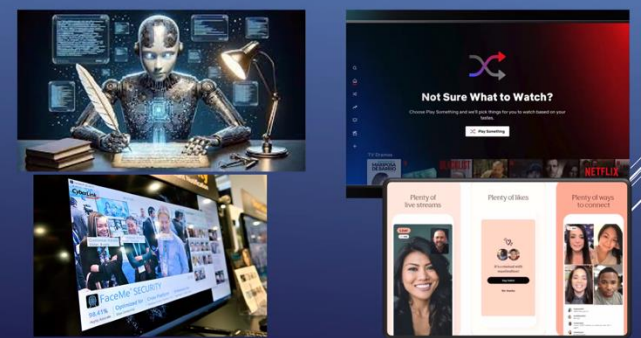


THEORY OF ALGORITHMIC MODERNITY

Circa 2016

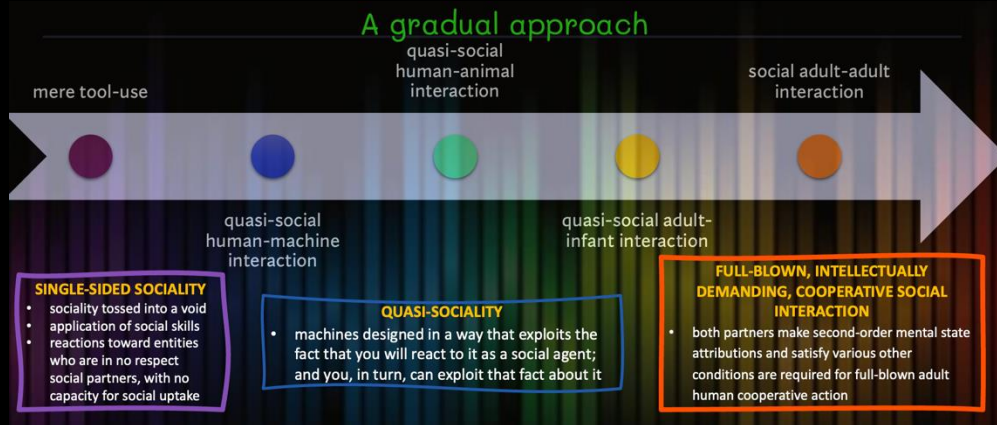
- AlphaGo defeats Go Champion
- AlphaGo Zero defeats AlphaGo
- 100-0 China announces US\$5 billion AI investment

LIFE ON AUTOPILOT – GenAI + OTHER OUTSOURCING

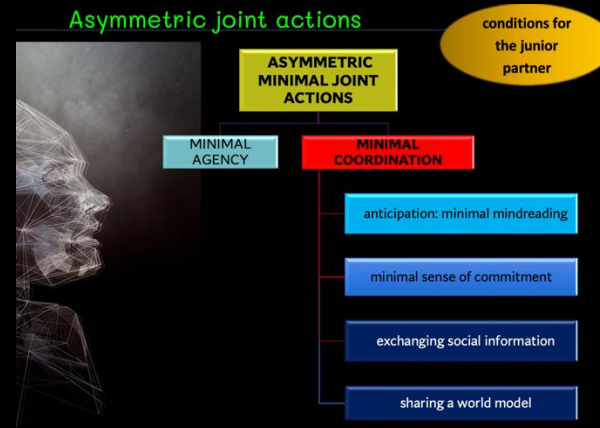


ACTION PARTNERS

MOVING FROM THE TWO-DIMENSIONAL SPACE TO THE THREE-DIMENSIONAL SPACE



Asymmetric joint actions



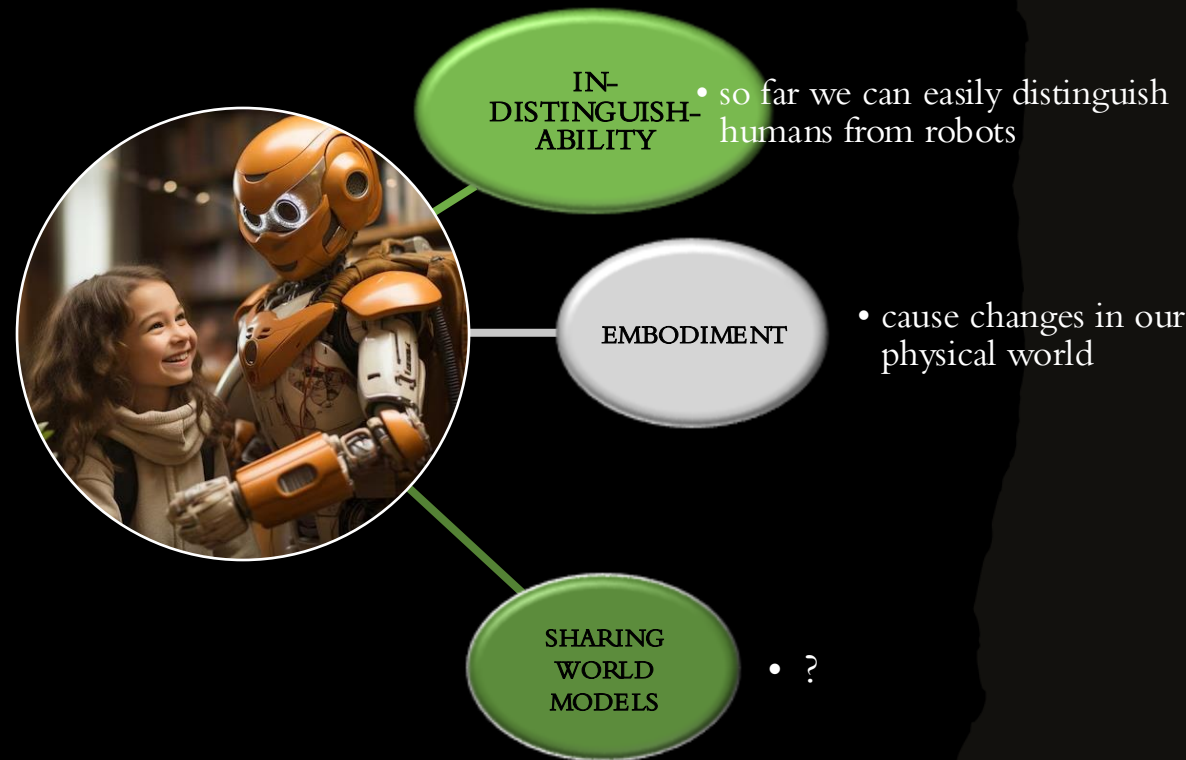
ANNA STRASSER

increasing INDISTINGUISHABILITY between machine-generated & human-created text

LLMs live NOT in our social, physical world

LLMs are not embodied

But they may play a role in our world of language games.



"TEAM MODE" IN HUMAN-ROBOTS COLLABORATION
AS AN AMPLIFIER OF PROBLEMS WITH ATTRIBUTION OF RESPONSIBILITY

COLLECTIVE AGENT, WHICH IS
TO BE HELD RESPONSIBLE
(Gunkel 2020; Nyholm 2018)



KAMIL MAMAK

BUT

robots can neither bear responsibility, nor be participants in shared responsibility
(Hakli & Mäkelä 2019)

deal with *responsibility gaps*

work-in-progress with Hakli & Mäkelä:

- sometimes, the fact of collaboration can decrease the degree of guilt of the human collaborator

1. There could be a formulated hypothesis that robots, due to the collaborative tendencies of humans, could increase the risks of committing crimes by human collaborators.
 2. In the case of humans, the set of crimes prevents humans from supporting criminal activity; in the case of robots, those rules cannot apply.
- Preventing measures:
- The design of robots should take into account the easiness of persuading humans to break the rules (for example, ignoring safety measures).
 - Robots should not gather in the place of the potential mob.
 - Ascribing responsibility for crime should take into account that humans could be encouraged to break the rules by robots' communication or their sole presence.

IF ROBOTS WERE TEAM MEMBERS WHO INITIATED OR STRENGTHENED
WRONGFUL DECISIONS



"TEAM MODE" AS AN AMPLIFIER OF THE BLURRING RESPONSIBILITY

1. Robots and AI to address mental healthcare crisis
2. Robots in institutional and informal social roles
3. Challenges with responsibilities due to institutional complexities
4. Challenges with unfulfilled social roles: one-sidedness and transformative changes
5. Suggestion: Ameliorative social role engineering

CONSEQUENCES OF CONCEPTUALIZING ROBOTS AS BEING ABLE TO TAKE OVER SOCIAL ROLES IN MENTAL HEALTHCARE ORGANIZATIONS

→ how interactions with social robots can reshape the dynamics & obligations between clinicians & patients in healthcare groups



If robots are considered to be able to play social roles usually occupied by humans, they will inherit some of the causal powers assigned to those roles.



TUOMAS VESTERINEN



- BUT**
those causal roles cannot be complete!
1. lack of moral agency & full agency of robots raises questions over responsibility gaps
 2. robots cannot fully assume social roles due to their inability to internalize these roles
- interactions between robots & humans are inherently emotionally & psychologically one-sided

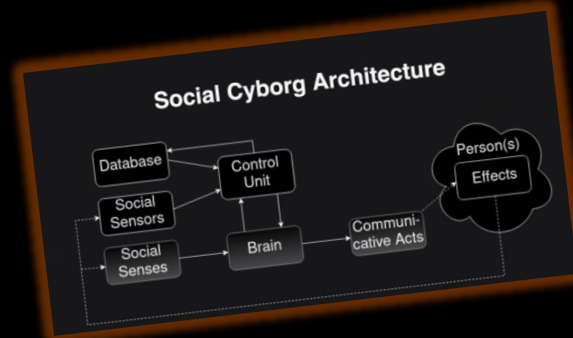
Outline

- Social Intelligence Augmentation
- Social Cyborg Architecture
- HCI, Data, Sensors
- Social Influence Relations
- Ethical Considerations
- Recap
- Closing Credits

AS A SOCIAL CYBORG

SOCIAL ROBOTS ENRICHED WITH SOCIAL INTELLIGENCE COULD HELP PEOPLE TO BE SOCIALLY MORE INTELLIGENT

- combine the best parts of a social robot and a human being into a social cyborg to augment the social intelligence of its host organism
- social cyborg = a new technological concept
 - similar to social robots, AI-mediated communication, cognitive AI extenders, and artificial moral advisors
 - to enhance strategic communication in group negotiation settings



RISKS

- new twist to the ethics of manipulation
- unfairness for those who cannot afford or refuse to extend themselves with artificial social intelligence devices?

Application Areas

Professional



Personal



PROTO-MORAL MACHINES & VALUE-SENSITIVE DESIGN OF SOCIO-TECHNICAL SYSTEMS

- 1) What are protomoral machines and why to aim at them?
- 2) Lessons from the evolutionary origins of morality
- 3) How protomoral machines become ethical in a context

simple robot's function
→ value-sensitive design (VSD)

robots with higher degree of autonomy & flexibility in behavior
→ problem of moral decision-making emerges

IN BETWEEN SIMPLE FUNCTIONS &
ARTIFICIAL MORALITY

Protomorality: ethically relevant interaction types and the underlying psychology

- enabling participation in certain forms of social interaction

MINIMIZE ETHICAL RISKS WHEN AUTONOMOUS ROBOTS
ARE DEPLOYED IN COMPLEX SOCIAL CONTEXTS
→ COMBINE THE PROTO-MORALITY APPROACH & VSD, WITH
A FOCUS ON THE FORMS OF SOCIAL INTERACTION WITHIN
THE GROUP'S FUNCTIONING



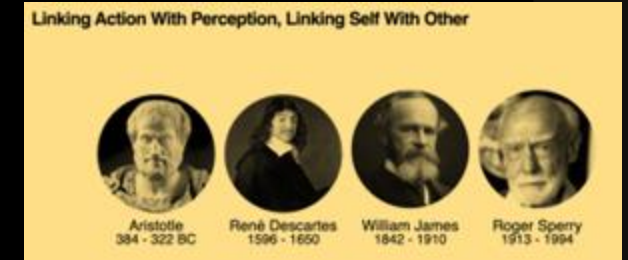
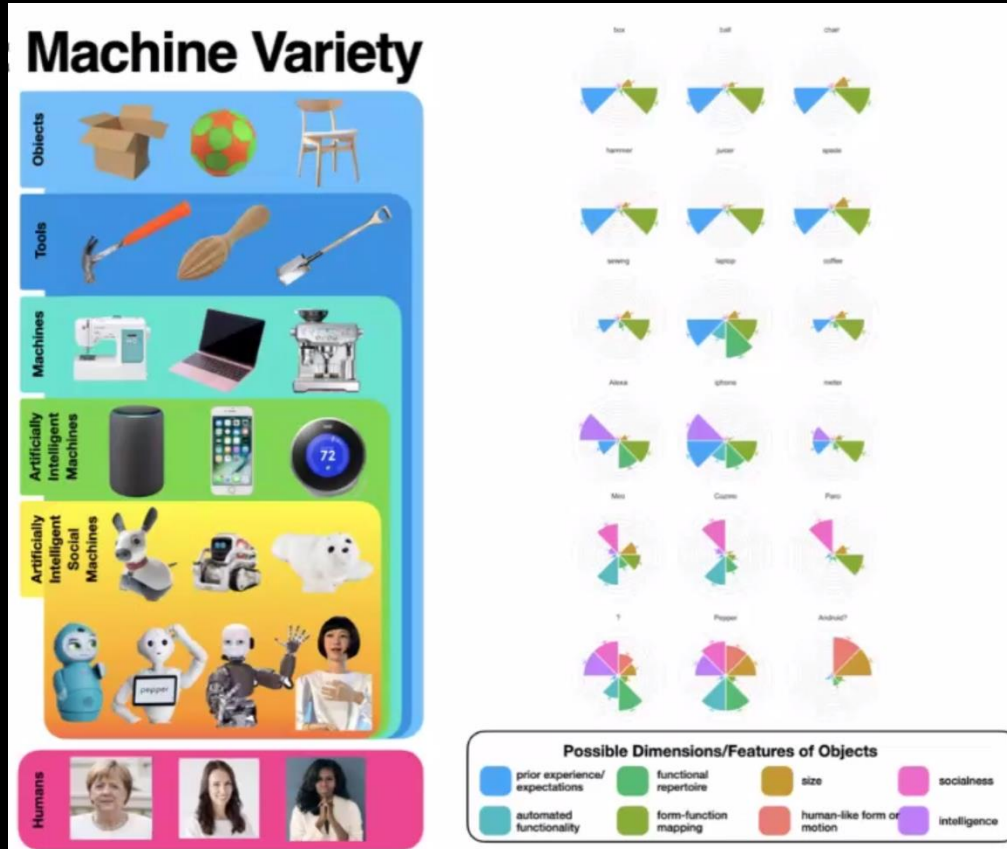
TOMI KOKKONEN



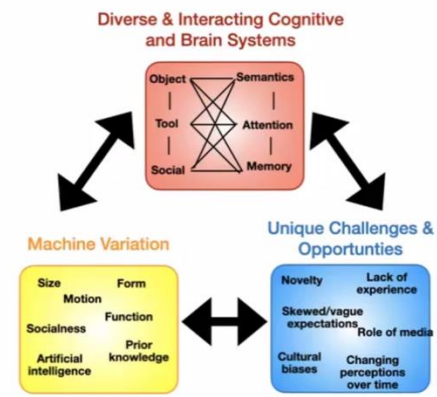
RE-ENVISIONING ETHICS :

FROM MORAL MACHINE TO EXTENSIVE REGULATION

MIND MEETS MACHINE – NEUROCOGNITIVE PERSPECTIVES ON HUMAN-ROBOT INTERACTION



Three Pillars of Proposed Framework



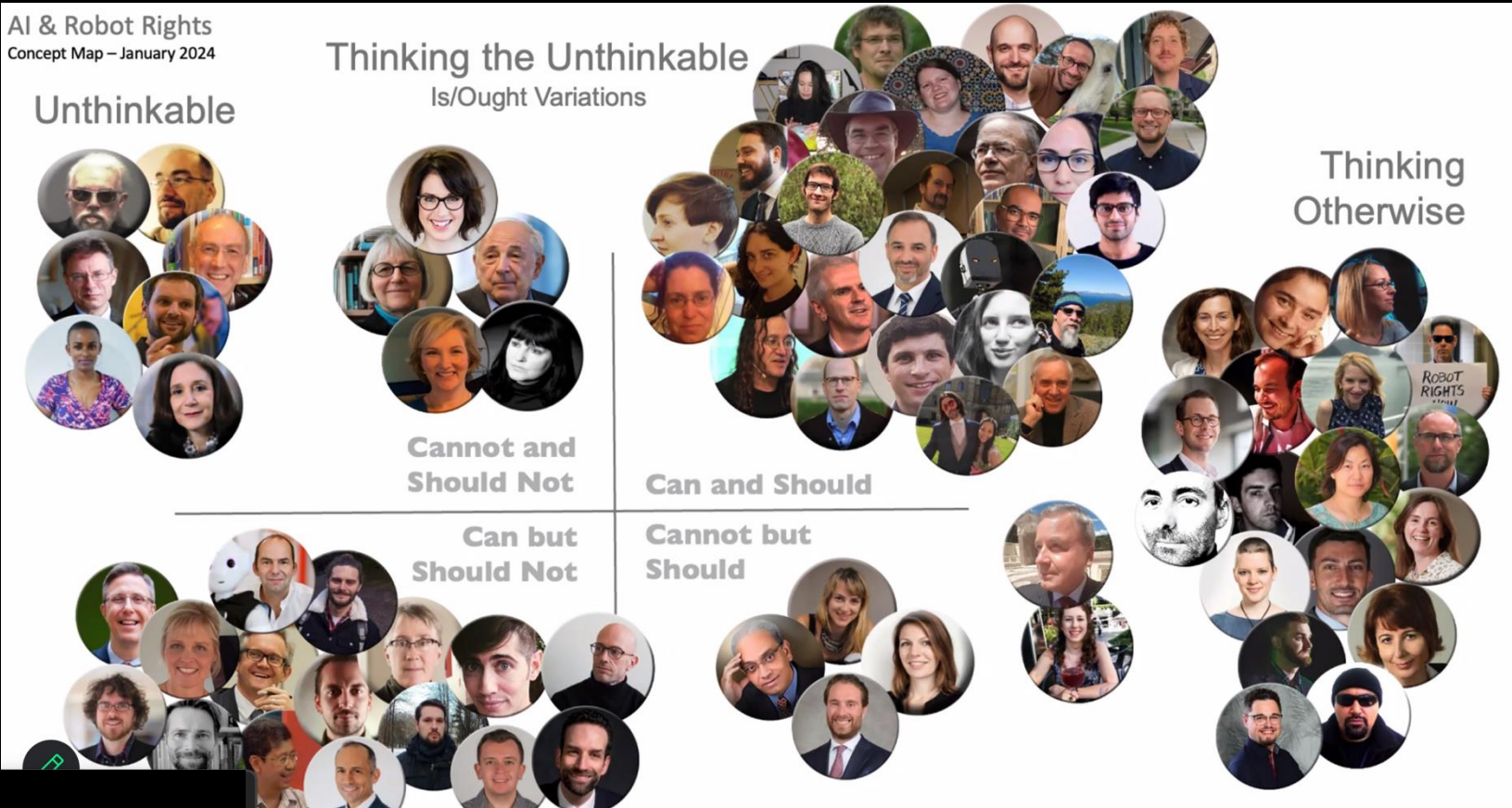

True progress in developing machines that engage humans on a social level will only come from interdisciplinary collaboration across the humanities, social, biological and computing sciences.

EMILY CROSS

Workshop 10: Robot Rights - From Theory to Practice

Robot Rights

From Theory to Practice





AI AS THE SUBJECT OF RIGHTS: ANALYSIS FROM AN ARENDTIAN PERSPECTIVE

LEGAL SUBJECTIVITY IS NOT SOLELY TIED TO TRAITS LIKE SENTIENCE OR CONSCIOUSNESS

- despite being unconscious, newborns have legal rights
 - sentient beings like chimps have no legal rights
 - → legal subjectivity of AI, like other non-human entities, depends on people including AI systems in their network of relationships.
- Without such acceptance, legal rights for AI will remain elusive, regardless of the consciousness of AI.

► Kantian Perspective

- AI systems operate exclusively based on goals rather than desires.

► Arendtian Perspective

- Societal acceptance of AI systems or any other robotic entities as integral components of society with their moral status

HUMANOID ROBOTS SHOULD HAVE MORE RIGHTS THAN ROBODOGS

- ❖ We could mistake human-like robots with humans.
- ❖ Human-robot interactions might impact human-human interactions.
- ❖ Human-like robots might be parties of human relationships.

epistemological limitations – Danaher's ethical behaviorism
appearance & distinguishability
transfer to HHI
friendship with humanoid robots

Conclusions

- This paper stands at the position that **human-like robots should have more rights than robodogs.**
- The notion should not be surprising, taking into consideration **the societal and legal position of humans.**
- Whether we like it or not, **the current law is anthropocentric**, and human interests have a privileged position compared to animals.
- However, **this paper should not be read in praise of the anthropocentric view or as a call for ignoring animal welfare.**
- It aims to point out that **human interests might be endangered by the rise of robots**, and **there are more human interests endangered in human-shaped robots than in robodogs.**

APPLYING THE HUMAN RIGHTS
APPROACH TO ROBOTS:
POSSIBILITIES & CHALLENGES



Skeptics

contest the idea that robots
could possess moral status &
rights

OTHERS

artificial entities might evolve into
“suprapersons” endowed with even
stronger rights than those of humans

underlying assumptions

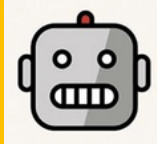
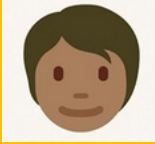
- 1) foundational assumptions underpinning human rights;
- 2) implications of these assumptions for future robots;
- 3) main practical challenges associated with recognizing robots as bearers of human rights

- If we grant certain widely shared assumptions about the moral foundations of human rights, it seems plausible that robots with human-like cognition would possess these rights as well.
- However, the question of moral rights for robots (including human rights) becomes highly challenging in practice due to our limitations in understanding what gives rise to conscious experience. This stresses the importance of establishing a suitable framework for moral decision making under uncertainty in the context of the human treatment of artificial entities.

AMERICAN PRAGMATISM BRIDGING
THEORY AND PRACTICE

- pragmatism's commitment to resolving seemingly interminable philosophical dilemmas
- with an unapologetic drive for AMELIORATION
- broad points of convergence
 - pluralism, nondualism, an emphasis on the materiality of language, relationality, and meliorism
 - → establish links to the “relational turn,” underscoring the importance of relational entanglement and social context to considerations of robot rights.
 - conflation of truth with goodness, blurring the lines between 'ought' and 'is'

- An alternative to the optimism <> pessimism binary
- Meliorism treats salvation as neither necessary nor impossible. It treats it as a possibility, which becomes more and more of a probability the more numerous the actual conditions of salvation become. (James, 1907, p. 125)
- Toward the middle ground - the belief that the world can be made better by human effort
 - hope and despair
 - choice and consequence



1. The Properties Approach – Identify and reevaluate three philosophical problems with the standard properties approach.

2. The Relational Turn – Alternative model that shifts emphasis from internal properties of the individual to extrinsic social circumstances and relationships.

3. Summary and Conclusions – A moral framework that is more agile in its response to the unique opportunities and challenges of the 21st century.

WHICH PROPERTIES?

male | white | European
rationality | consciousness | suffering

HOW TO DETECT?

????

RELATIONAL

- no internal properties
- objectively observable extrinsic social relationships

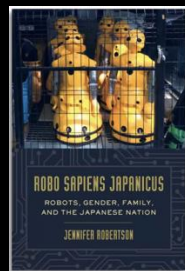
PHENOMENOLOGICAL

- *how it is treated* → *what it is*
- DIVERSE**

Making Kin with the Machines

Essay Competition Winner

by Jason Edward Lewis, Noelani Arista, Archer Pechawis, and Suzanne Kite



Conclusion

It is in responding to the moral opportunities and challenges posed by seemingly intelligent and social artifacts that we are called to take responsibility for ourselves, for our world, and for those others who are encountered here.

Workshop 10: Robot Rights - From Theory to Practice



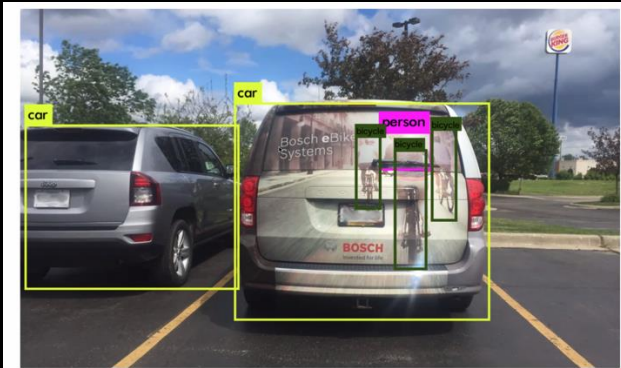
David Gunkel

AI'S CHALLENGE TO UNDERSTANDING THE WORD

AI do not understand the world as we do



school bus 1.0 garbage truck 0.99 punching bag 1.0 snowplow 0.92



What a self-driving car's camera sees when it looks at a car with an advert on the back. Photograph: telzunahan <https://www.theguardian.com/technology/2017/aug/30/self-driving-cars-hackers-security>

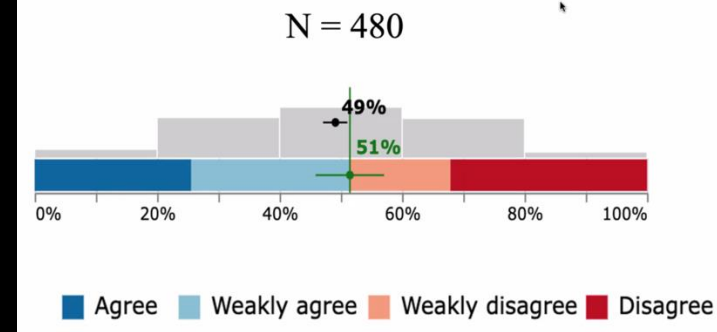


WHAT DO NLP RESEARCHERS BELIEVE? RESULTS OF THE NLP COMMUNITY METASURVEY

2022

Julian Michael^{1,2} Ari Holtzman¹ Alicia Parrish⁴ Aaron Mueller⁵ Alex Wang³
Angelica Chen² Divyam Madaan³ Nikita Nangia²
Richard Yuanzhe Pang³ Jason Phang² and
Samuel R. Bowman^{2,3,4}

Agree or disagree: Some generative models trained only on text, given enough data and computational resources, could understand natural language in some non-trivial sense.



MELANIE MITCHELL

Take-home message

LLMs are better (often dramatically) on solving reasoning tasks that are similar to those seen in training data.

This reflects some failures of abstract understanding.

How can we get machines to learn and use humanlike concepts and abstractions?

How to evaluate understanding in LLMs?

1. Just look at their behavior (“Turing test”) **But subject to Eliza effect!**
2. Test them on “understanding” benchmarks **But...data contamination? Shortcuts? “Approximate retrieval?”**
3. Evaluate them on tasks requiring abstraction and reasoning

QUESTIONS

What do we mean by “understanding”?

Why not *explicitly* claim that the (vast?) majority of these tests *are* indeed contaminated**?

Question 2: how do those abstract concepts get populated and extended into new domains without bodies? What is the alternative explanation that retains the importance of these systems for our conceptual structures but could even in principle do away with, not just bodies, but bodies in some way similar enough to our own, embedded in physical and socio-cultural systems we inherit and act upon (if we want to share enough of a conceptual system to understand them)?

Would you be willing to include in test items issued to LLMs/AIs *the requirement that a formally verifiable justification* be provided?

“An enactive cognitive science perspective makes salient the extent to which language is not just verbal or textual but depends on the mutual engagement of those involved in the interaction. The dynamism and agency of human languaging means that language itself is always partial and incomplete. It is best considered not as a large and growing heap, but more a flowing river. Once you have removed water from the river, *no matter how large a sample you have taken*, it is no longer the river. The same thing happens when taking records of utterances and actions from the flows of engagement in which they arise. The data on which the engineering of LLMs depends can never be complete, partly because some of it doesn’t leave traces in text or utterances, and partly because language itself is never complete.” (p. 19-20)

How about that?

SELMAR BRINGSJORD

ROBIN ZEBROWSKI

Robots in Healthcare –

Interdisciplinary Co-Design and Technoethics Education

PRACTISE



ETHIC

Ethical deployment: Embodiment makes a huge difference!

Opportunities:

- Deployment at a slower pace than AI → **More time** for ethical reflection, regulation & education
- **Co-design** with all stakeholders (social services, health centers, care facilities, research institutions, technology companies, nursing and patient associations, caregivers, end-users...)

Amplification of some AI risks:

1. Human **agency** & oversight → **Deceit**, over-reliance, **emotional bonding**, isolation
2. Technical robustness & **safety** → **Physical harm** added to other dangerous effects
3. **Privacy** & data governance → Robot mobility/autonomy increases **spying threat**
4. **Transparency** → **Unpredictable motions and behaviors** (lack of intention cues)

New risks:

- Human dignity → User **objectification** and **lack of control** over their life

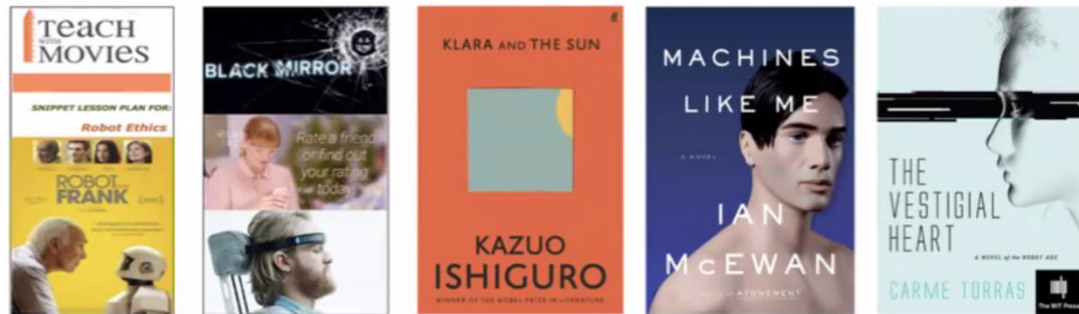


Carme Torras

1. Human **agency** and oversight
2. Technical robustness and **safety**
3. **Privacy** and data governance
4. **Transparency**
5. Diversity, non-discrimination and **fairness**
6. Societal and environmental **well-being**
7. **Accountability**

EDUCATION

Modern Science Fiction - Ethics debates and education initiatives



"The relationships that we have constructed in turn shape us"

Robert C. Solomon
"The Passions"

Encouraging future perspectives

Social services are beginning to invest in social robots:

- Population ageing & shortage of caregiving personnel.
- Elderly people could **live longer at home** with some technological help.
- Robots can take on routine **tasks with no added human value.**

AI Avatars and the Future Society

Why Do We Need Humanoid Robots?

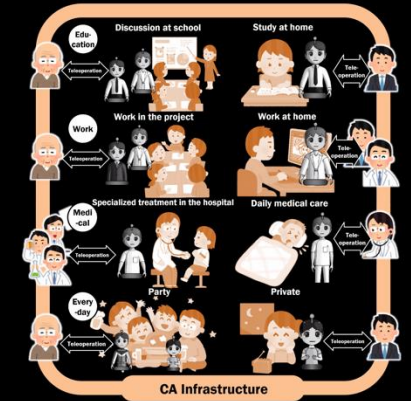
- Humans have brains that recognize humans. Therefore, the ideal interface for humans is humans.
- Therefore, we realize humanlike information media that people can easily use and interact with.
- At the same time, we will use it to understand the higher cognitive functions of humans. (constructive science for understanding human)
- We are exploring what it means to be human through interaction with humanlike robots.



1. By 2050, realize a society in which people are free from the constraints of body, brain, space, and time.
2. By 2050, realize a society that can predict and prevent diseases very early.
3. By 2050, realize a robot that learns and acts by itself and coexists with people through co-evolution of AI and robots.
4. By 2050, achieve sustainable resource recycling for global environmental restoration.
5. By 2050, create a sustainable food supply industry on a global scale without waste by making full use of unused biological functions, etc.
6. By 2050, realize error-tolerant general-purpose quantum computer that dramatically develop economy, industry and security.
7. By 2040, prevent and overcome major illnesses and enjoy life without health concerns until the age

Future Society in 2050

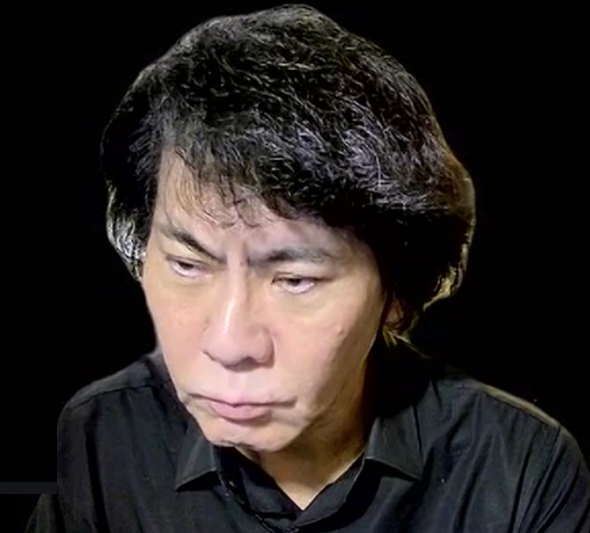
- Anyone, including, elderly and people with disabilities, will be able to freely participate in various activities with abilities beyond ordinary people while expanding their physical, cognitive, and perceptual abilities using a large number of CAs.
- Anyone will be able to work and study anytime, anywhere, minimize commuting to work, and have plenty of free time.



Geminoid HI-6+LLM

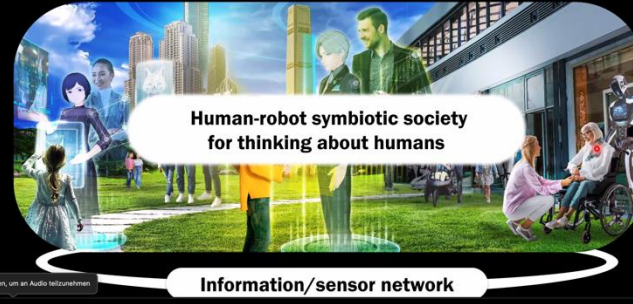


Hiroshi Ishiguro



Human-Robot Symbiotic Society

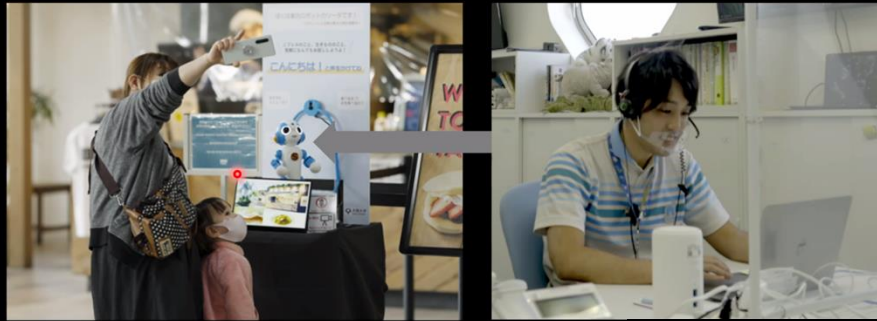
Using Avatars in a Supermarket with Cyberagent



Using Avatars in a Kindergarten with Cyberagent



Using Avatars at an Amusement Parks with Cyberagent



Using Avatars for After-School Children's Club



Using the Digital Minister Avatar

- The Digital Minister conducts official duties with an avatar. We don't have to worry about security, so he can work anytime, anywhere.
- The use of the Digital Minister Avatar can create an opportunity to discuss various issues related to the use of avatars.



QUESTIONS

Will Avatars de-humanize Humans?

- Dis-embodiment
- Dis-enhancement
- De-socializing



Final Dialogue

Who should carry responsibility for social robots? Is there a general model for the distribution of responsibility or do we need to differentiate for each application area (healthcare, education, open public spaces etc.)?

Should Europeans give up on ethical concerns, as many AI-specialists recommend? Is the EU AI Act in your assessment useful? If not, how do we supplement it?

Are we overly concerned about social robots—should we trust that humans will come to terms with them as they have with any other technology? What is different and why should policy makers be alarmed?

What are in your view the 3 central and most urgent tasks in robo-philosophy for the next decade?

Do we have enough regulations?
How can we protect all citizens?

Final Dialogue

Panelists: David Gunkel, Alan Winfield, Kerstin Fischer, Raja Chatila, Shannon Vallor,
Johanna Seibt, Bertram Malle



CHECK OUT the memory
slices

<https://www.denkwerkstatt.berlin/ANNA-STRASSER/MEMORY-SLICES/>



*Thank
you all!*