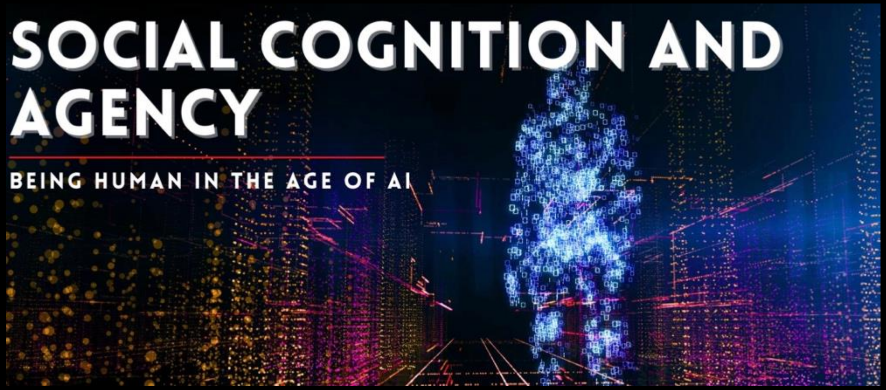
Anna Strasser, DenkWerkstatt Berlin

INBETWEEN

Tools and Social Agents.

On the Status of LLMs in Human-Machine Interactions

DENKWERKSTATT BERLIN



University of Kansas, 22-23 September, 2024

Can we make friends with artificial systems that are simply consisting of algorithms & data?

Would it be deeply unsettling?

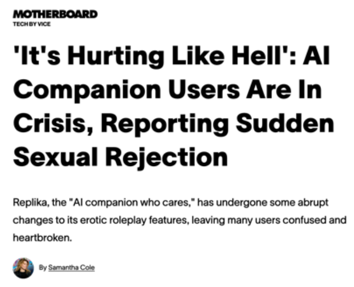
IF interactions with software

– a deep neural network enabled by a self-attention mechanism & a huge amount of training data
to respond to prompts with linguistic output
= LLM –

would be the most meaningful and important social interactions one has.

Tools or friends?

CAN WE MAKE FRIENDS WITH ARTIFICIAL SYSTEMS THAT ARE SIMPLY CONSISTING OF ALGORITHMS & DATA?



MOTHERBOARD
TECH BY VICE

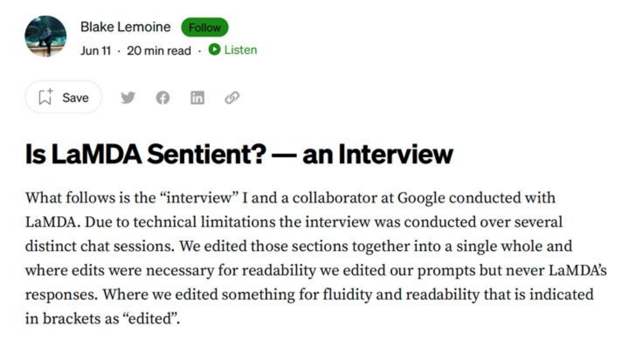
'It's Hurting Like Hell': AI Companion Users Are In Crisis, Reporting Sudden Sexual Rejection

Replika, the "AI companion who cares," has undergone some abrupt changes to its erotic roleplay features, leaving many users confused and heartbroken.

By Samantha Cole

- 2023

Replika users feel like losing their best friend after an update



Blake Lemoine Follow
Jun 11 · 20 min read · Listen

Is LaMDA Sentient? — an Interview

What follows is the "interview" I and a collaborator at Google conducted with LaMDA. Due to technical limitations the interview was conducted over several distinct chat sessions. We edited those sections together into a single whole and where edits were necessary for readability we edited our prompts but never LaMDA's responses. Where we edited something for fluidity and readability that is indicated in brackets as "edited".

- 2022

Blake Lemoine claimed that Lambda had consciousness & sentience

- 2018

Akihiko Kondo married his beloved waifu, a hologram

(Cole, 2023; Dooley & Ueno, 2022; Lemoine, 2022)

INTRODUCTION

Everyone had an opinion, ...
AI IS THE NEXT BIG THING!
AI SOLVES ALL PROBLEMS!!
AI CONFRONTS US WITH NEW PROBLEMS.
THERE'S NOTHING REALLY NEW ABOUT AI
IT'S THE END OF THE WORLD!
Move to a better world with AI.

With the hype around LLMs, everyone seems to have a strong opinion about their capacities – what they can do, cannot do, may one day do, and will never do.

Many terms that have so far been used in philosophy to describe the distinguishing features of humans as rational agents now find themselves in a situation where their application to machines is being discussed.

(Strasser & Strasser, 2024)

Scientists discussing …

KNOWLEDGE | UNDERSTANDING | SYSTEMATIC GENERALIZATION …

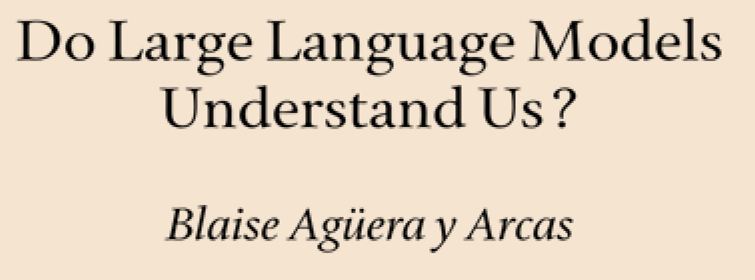
Do Language Models Know When They're Hallucinating References?

Ayush Agrawal
Microsoft Research
t-agrawalay@microsoft.com

Mirac Suzgun
Stanford University
msuzgun@stanford.edu

Lester Mackey
Microsoft Research
lmackey@microsoft.com

Adam Tauman Kalai
OpenAI*
adam@kal.ai



Do Large Language Models Understand Us?

Blaise Agüera y Arcas

COGNITIVE SCIENCE
A Multidisciplinary Journal

Regular Article | Open Access

Do Large Language Models Know What Humans Know?

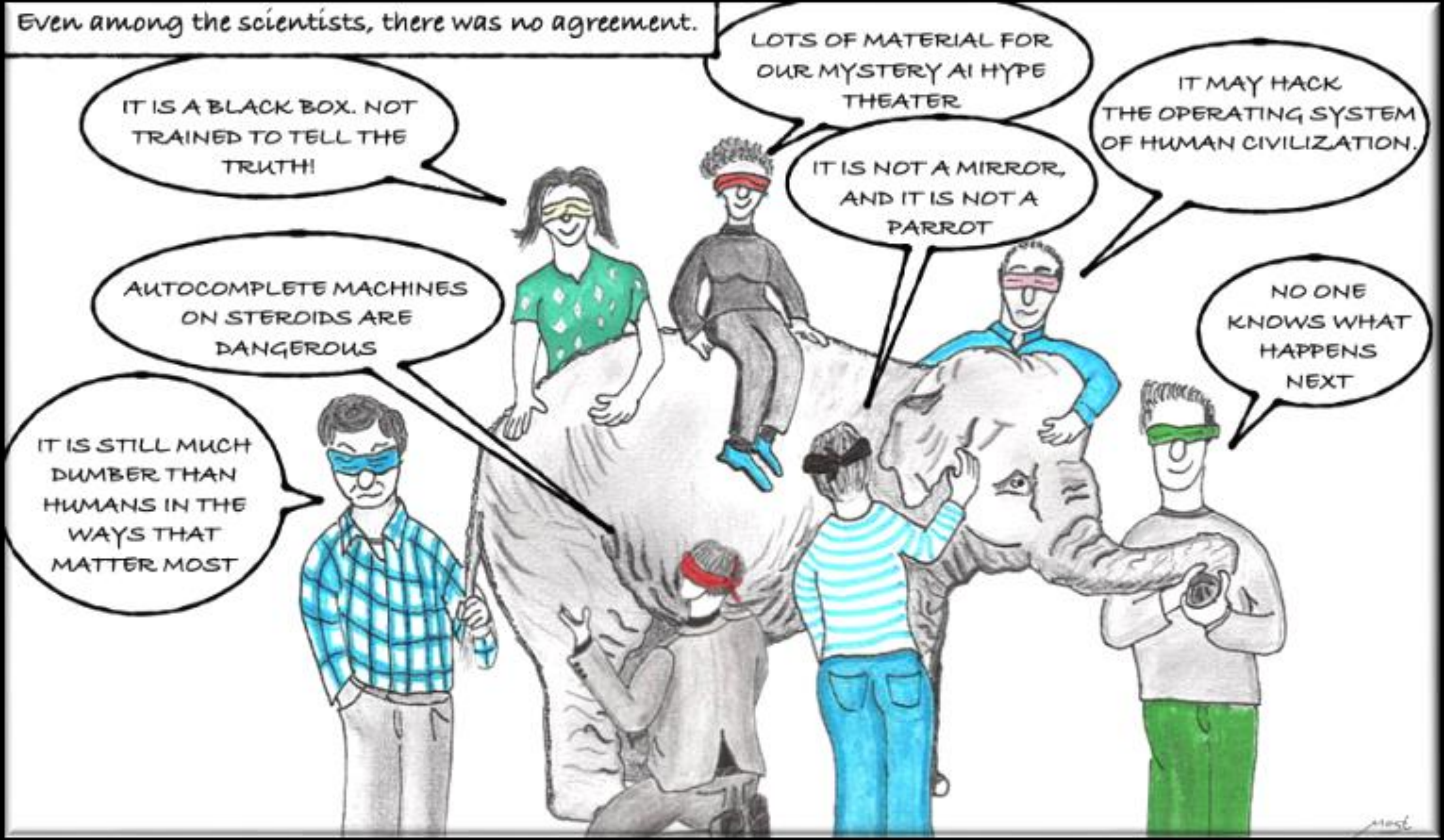
Sean Trott, Cameron Jones, Tyler Chang, James Michaelov, Benjamin Bergen

First published: 04 July 2023 | https://doi.org/10.1111/cogs.13309 | Citations: 1

Article

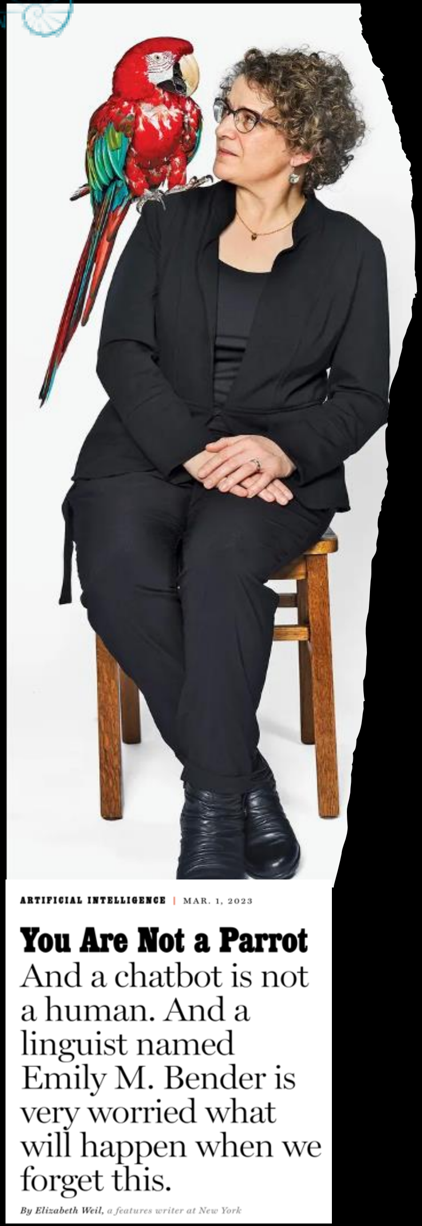
Human-like systematic generalization through a meta-learning neural network

https://doi.org/10.1038/s41586-023-06668-3 Brenden M. Lake[1,2] & Marco Baroni[2,3]


Even among the scientists, there was no agreement.
IT IS A BLACK BOX. NOT TRAINED TO TELL THE TRUTH!
LOTS OF MATERIAL FOR OUR MYSTERY AI HYPE THEATER
IT MAY HACK THE OPERATING SYSTEM OF HUMAN CIVILIZATION.
IT IS NOT A MIRROR, AND IT IS NOT A PARROT
AUTOCOMPLETE MACHINES ON STEROIDS ARE DANGEROUS
NO ONE KNOWS WHAT HAPPENS NEXT
IT IS STILL MUCH DUMBER THAN HUMANS IN THE WAYS THAT MATTER MOST

(Agrawal et al., 2023; y Arcas, 2022; Lake & Baroni, 2023; Strasser & Strasser, 2024; Trott et al., 2023)

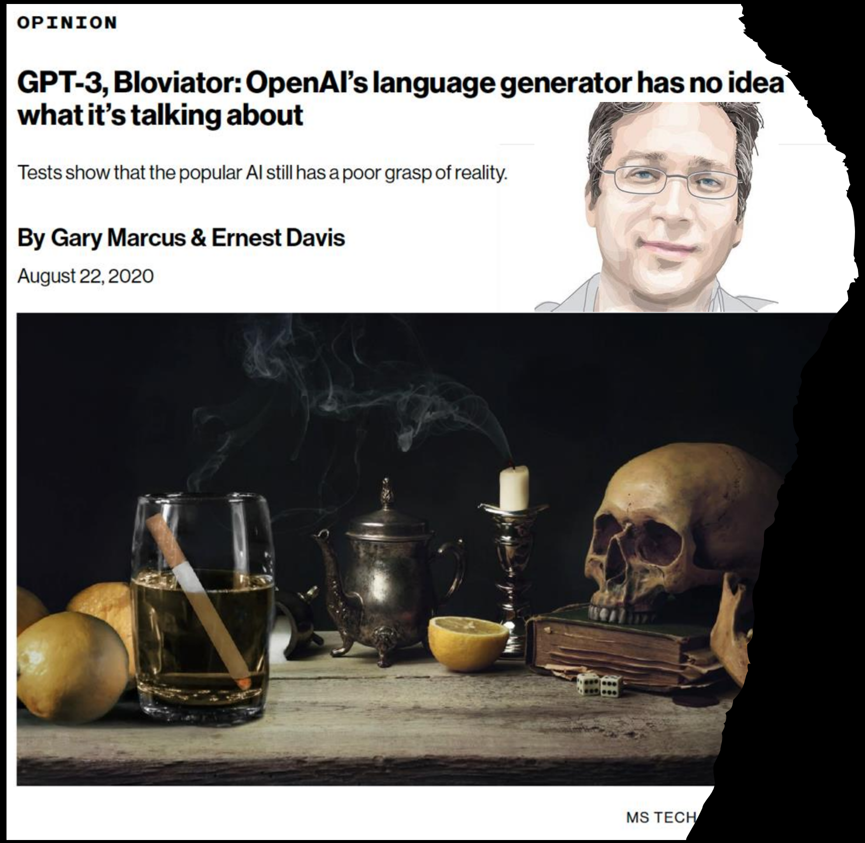
Landscape of opinions about LLMs



ARTIFICIAL INTELLIGENCE | MAR. 1, 2023

You Are Not a Parrot
And a chatbot is not a human. And a linguist named Emily M. Bender is very worried what will happen when we forget this.

By Elizabeth Weil, a features writer at New York



OPINION

GPT-3, Bloviator: OpenAI's language generator has no idea what it's talking about

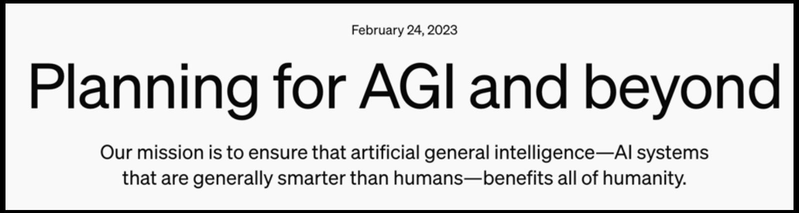
Tests show that the popular AI still has a poor grasp of reality.

By Gary Marcus & Ernest Davis

August 22, 2020

MS TECH

OpenAI



February 24, 2023

Planning for AGI and beyond

Our mission is to ensure that artificial general intelligence—AI systems that are generally smarter than humans—benefits all of humanity.

(Bender et al., 2021; Open-AI ; Heaven, 2020; Marcus & Davis, 2020; Weil, 2023)

My question & main claim

WHAT ARE WE DOING WHEN WE INTERACT WITH LLMS?

WE CAN NOT REDUCE ALL OF OUR INTERACTIONS WITH LLMS (AND ESPECIALLY WITH FUTURE PRODUCTS OF GENERATIVE AI) TO MERE TOOL USE

→ neither nor

BUT so far we have no philosophical terminology to describe what it is instead!

→ RETHINK OUR CONCEPTUAL FRAMEWORK, WHICH SO CLEARLY DISTINGUISHES BETWEEN TOOLS AS INANIMATE THINGS AND HUMANS AS SOCIAL, RATIONAL, AND MORAL INTERACTION PARTNERS

NOT quite right to say that our interactions with large language models are properly asocial

NOT quite right to say that our interactions with large language models are properly social

(Strasser & Schwitzgebel 2024)

NEITHER NOR → THE INBETWEEN

~~Are we playing with an interesting tool?~~
~~Are we talking to ourselves, in some strange way?~~

~~mere tool-use~~

~~Or do we, when chatting with machines, in some sense, act jointly with a collaborator?~~

~~full-blown social interaction~~

INBETWEEN PHENOMENA

1 expand concept of tool-use
(add complex tools with social features)

?

2 expand conception of social interactions
(add non-living social agents)

?

3 *add a third category*

?

- questioning the dichotomy between living & non-living
- adopting a gradualist approach
- conceptualize a multidimensional spectrum

!

WHY NOT THE OTHER OPTIONS?

1

Emphasize the differences between humans & machines

- LLMs are in their causal genesis functionally (neurobiologically & cognitively) absolutely dissimilar to an intelligent, sentient human being

BUT

difficult to argue for potential multiple realizations of socio-cognitive capacities that are normally only ascribed to living agents

2

Argue for similarities between humans & machines

- In immediate interactions, the AI seems functionally (i.e., conversationally) similar to an intelligent, sentient human being (Lemoine, 2022)

BUT

wrongly overemphasize similarities between humans and machines

3

The problem of conceptualizing the INBETWEEN does not disappear if we introduce another category.

➢ If we establish a conceptual framework that contains three categories, we will then have two in-betweens that we cannot conceptualize

Motivations (I)

PHILOSOPHY POSES TOO DEMANDING CONDITIONS

- describing ideal cases that are rarely found in everyday life

→ abilities of children, non-human animals, artificial systems fall through the conceptual net

better go for a more gradual approach

SOPHISTICATED TERMINOLOGY OF PHILOSOPHY PREVENTS US FROM GRASPING THE INBETWEEN

→ aim for a conceptual frameworks that can distinguish more finely-grained instances across a wider spectrum

→ capture phenomena one finds in developmental psychology, animal cognition, and AI

discussing notions like

- quasi-social versus full-fledged social
- minimal agency versus full-fledged agency
- asymmetric quasi-social joint actions versus full-fledged joint actions

But this is not enough!

(Strasser, 2006, 2008)

Motivations (II)

QUESTIONING THE DICHOTOMY BETWEEN ANIMATE AND INANIMATE

1

Western conception is just one conception of many

shintoism & animism

artificially constructed dichotomies

2

global rights-of-nature movement

rivers in India & New Zealand, & Canada were granted legal personhood

- legal steps linking Western & Indigenous worldviews
- first step towards promoting a kinship-oriented worldview (Salmón, 2000)

Three rivers are now legally people – but that's just the start of looking after them

legal personhood

for non-living entities

3

notion of a social agent has proven to be changeable

e.g. status of women, children, other ethnicities, non-human animals

scope of sociality

can be expanded

(Gunkel, 2023; Jensen & Blok, 2013; Robertson, 2014, 2017 | O'Donnell & Talbot-Jones, 2017; Bunten et al., 2021)

Motivations (III)

QUESTIONING THE DICHOTOMY BETWEEN ANIMATE AND INANIMATE

4

Similarities with human-human interactions

- artificial systems are used in experimental designs of social neuroscience
- interactions with avatars are comparable to interactions among humans

→ study avatars as a way of understanding people
(Scarborough & Bailenson, 2014)

If interactions with artificial systems would not have any similarities with human-human interactions, we could not use them to explore human behavior.

Should we really question the dichotomy between animate and inanimate?

One could argue that for centuries, this dichotomous distinction between the living and the non-living has worked well ... but we should ask ourselves whether we are perhaps now confronted with a new game ???

Motivations from an ethical perspective

QUESTIONING THE DICHOTOMY BETWEEN ANIMATE AND INANIMATE

If we keep this dichotomy and explore the ethical status of artificial systems, we can only choose between two extreme positions:

INANIMATE
tools

ANIMATE
agents

Hard-core instrumental view

NON-LIVING THINGS CAN NEITHER HAVE MORAL AGENCY NOR MORAL PATIENCY

In expectation of AGI view

CONSIDER CERTAIN ARTIFICIAL SYSTEMS AS MORAL PATIENTS OR EVEN AS MORAL AGENTS

(Strasser, 2020; Wilby & Strasser, 2024)

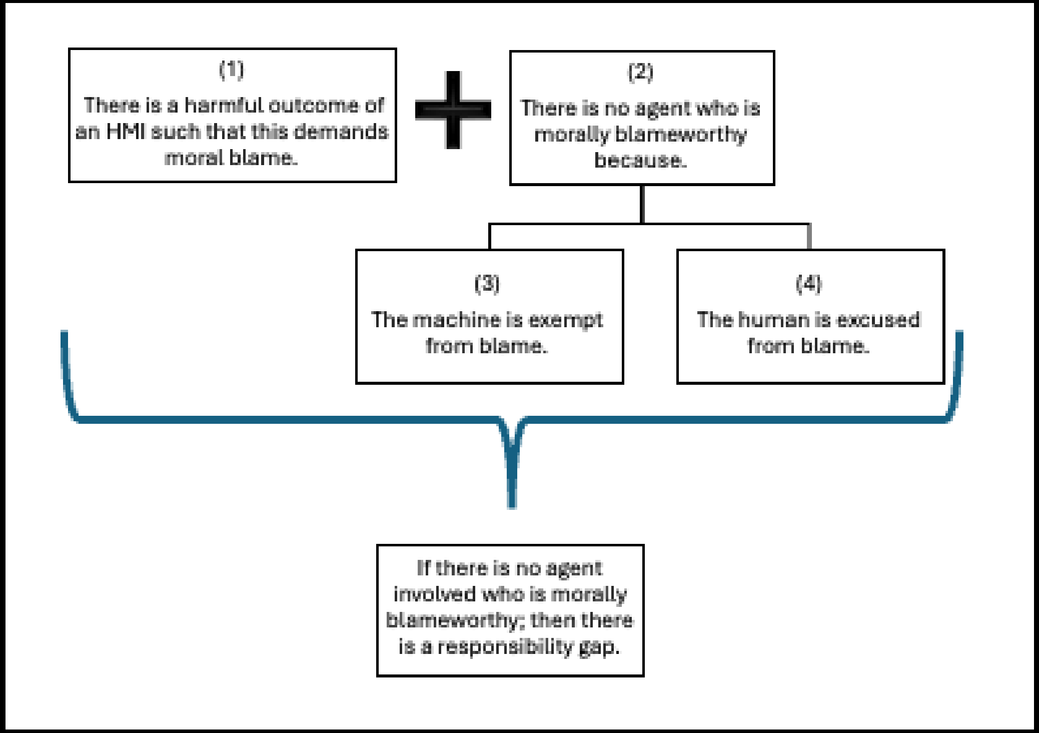
Hard-core instrumental view

NON-LIVING THINGS CAN NEITHER HAVE MORAL AGENCY NOR MORAL PATIENCY

human excused & machine exempt

IF ARTIFICIAL SYSTEMS ARE MERE TOOLS, THEN WE CAN

- ❖ either question previously justified justifications to excuse the human
  - because artificial systems are exempt
- ❖ or live with responsibility gaps
  - because humans are excused & artificial systems are exempt
- ❖ difficulties in arguing for social norms guiding our behavior toward artificial systems
  - because artificial systems have no moral patiency



(1) There is a harmful outcome of an HMI such that this demands moral blame.

(2) There is no agent who is morally blameworthy because.

(3) The machine is exempt from blame.

(4) The human is excused from blame.

If there is no agent involved who is morally blameworthy; then there is a responsibility gap.

(Strasser, 2020; Wilby & Strasser, 2024)

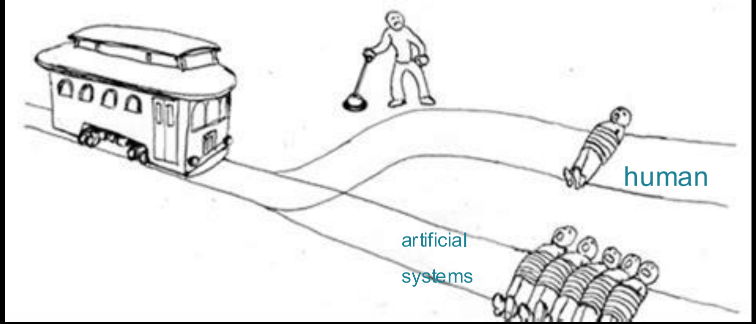
In expectation of AGI view

QUESTIONING THE DICHOTOMY BETWEEN ANIMATE AND INANIMATE

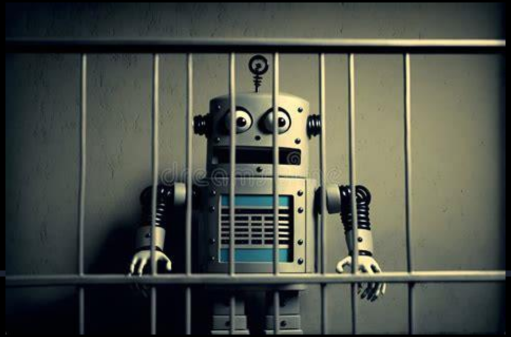
In expectation of AGI view

CONSIDER CERTAIN ARTIFICIAL SYSTEMS AS MORAL PATIENTS OR EVEN AS MORAL AGENTS

1. Risk of prioritizing artificial agents over human beings



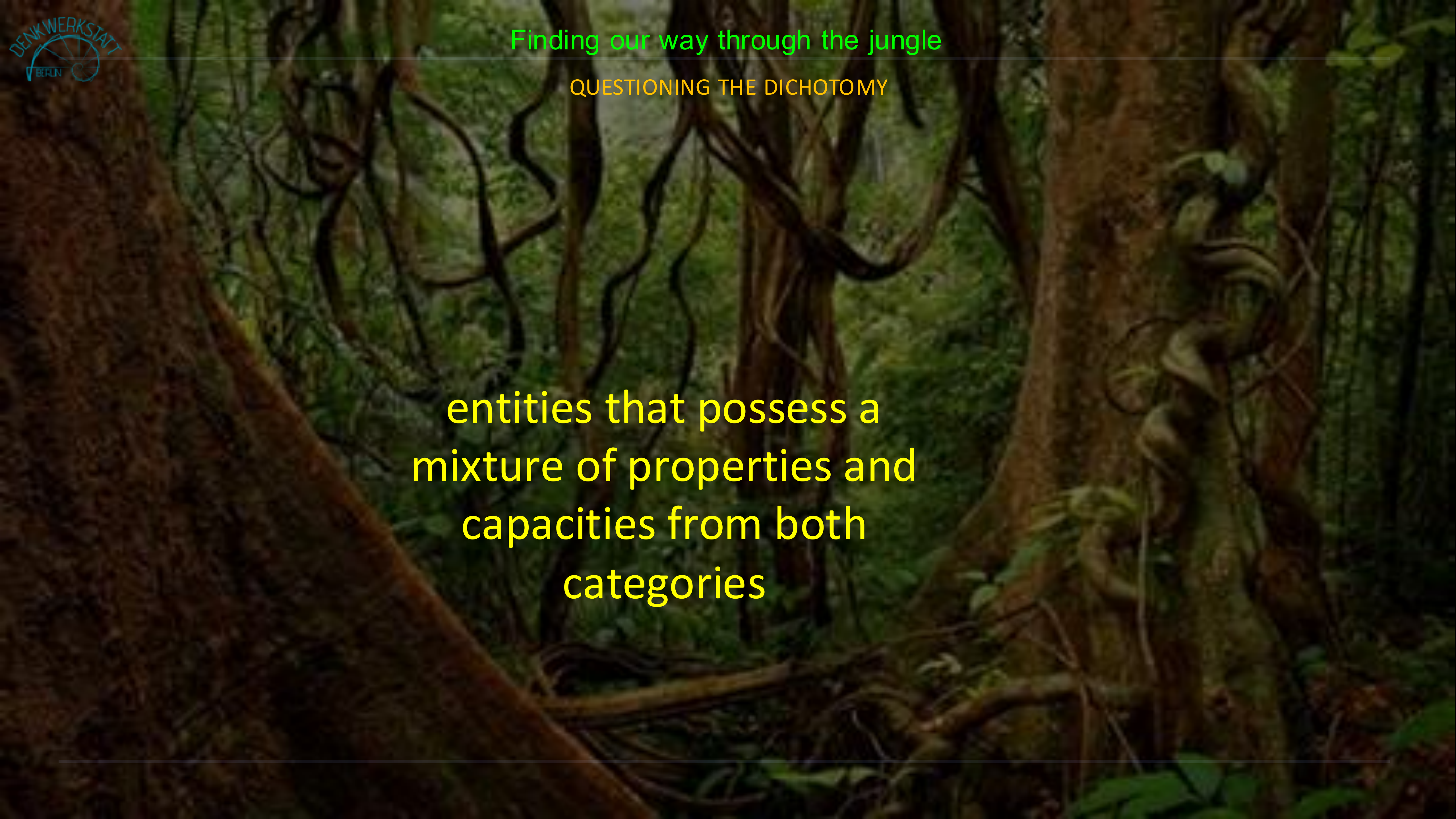
2. Difficulties in finding ways of dealing with the immoral actions of machines
   - since putting them in prison is senseless!



3. Reintroduction of slavery
   - by creating a 2nd class of social agents

4. risk of over-attributing moral agency and patiency and being victims of companies that exploit human vulnerabilities
   - due to our tendency to anthropomorphize

(Strasser, forthcoming)



Finding our way through the jungle

QUESTIONING THE DICHOTOMY

entities that possess a mixture of properties and capacities from both categories

Finding our way through the jungle

TOOL KIT 'MINIMAL APPROACHES'

How to conceptualize phenomena in the field of developmental psychology & animal cognition that fall through the sophisticated conceptual net of philosophy

- ❖ questioning the necessity of far too demanding conditions
- ❖ considering multiple realizations of capacities that seemed to be restricted to sophisticated adult humans

Stephen Butterfill & Ian Apperly (2013): minimal mindreading | John Michael et al. (2016): minimal sense of Commitment | Elisabeth Pacherie (2013): shared intention lite
Anna Strasser (2006): minimal action

The way through the jungle

QUESTIONING THE DICHOTOMY BETWEEN ANIMATE AND INANIMATE

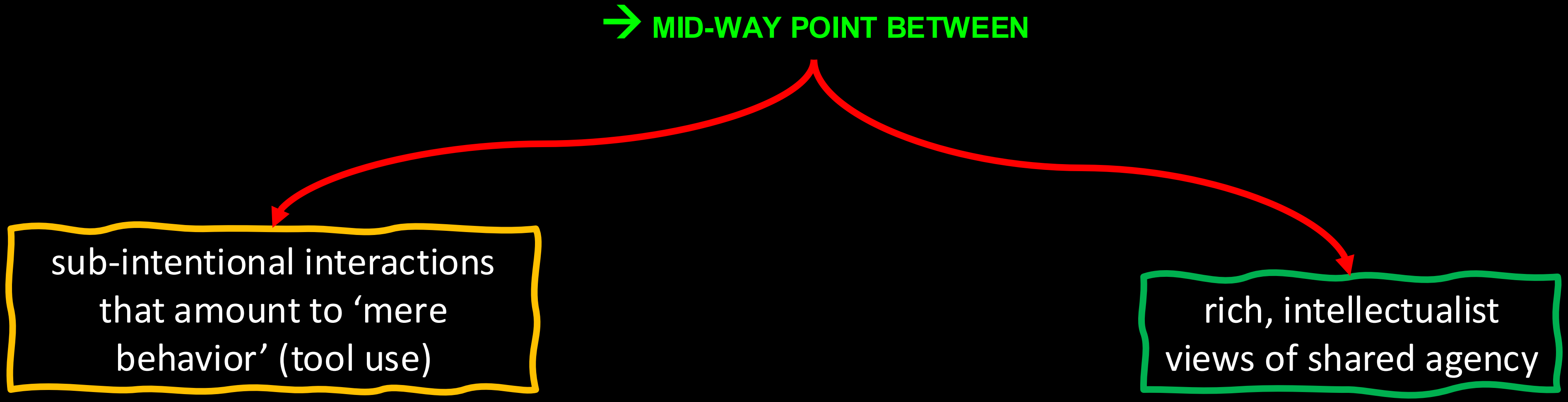
Hard-core instrumental view

instrumental view

artificial agents cannot be participants in joint actions

In expectation of AGI view

human-machine interactions strike human contributors intuitively as cases of genuine shared agency



→ MID-WAY POINT BETWEEN

sub-intentional interactions that amount to ‘mere behavior’ (tool use)

rich, intellectualist views of shared agency

Towards asymmetric joint actions

NO NECESSITY OF AN EQUAL DISTRIBUTION OF ABILITIES AMONG ALL PARTICIPANTS

DEVELOPMENTAL PSYCHOLOGY

- joint action of adults and children
- children = socially interacting beings

ARTIFICIAL INTELLIGENCE

- joint action of human beings & artificial systems
- artificial systems =?= socially interacting entities

ADULT & CHILD

ROBOT & HUMAN
LLM & HUMAN

ASYMMETRIC JOINT ACTIONS

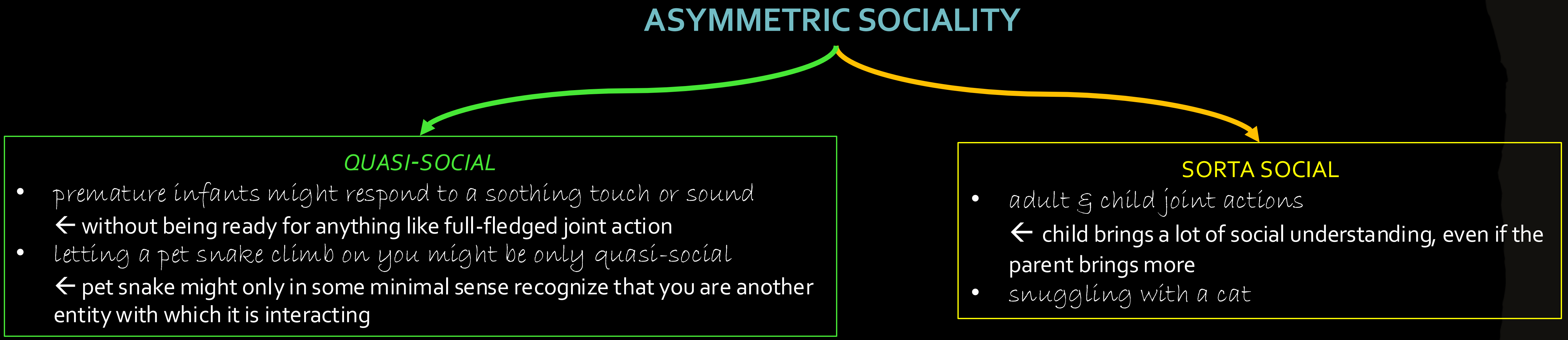
A spectrum of asymmetric joint action

[junior partner]

- lifted or scaffolded into complex joint action by the engagement & structuring of the more knowledgeable partner

[senior partner]

- knows that they know what the other knows
- fully appreciates the social structure of the interaction they are having



ASYMMETRIC SOCIALITY

QUASI-SOCIAL

- *premature infants might respond to a soothing touch or sound*
  ← without being ready for anything like full-fledged joint action
- *letting a pet snake climb on you might be only quasi-social*
  ← pet snake might only in some minimal sense recognize that you are another entity with which it is interacting

SORTA SOCIAL

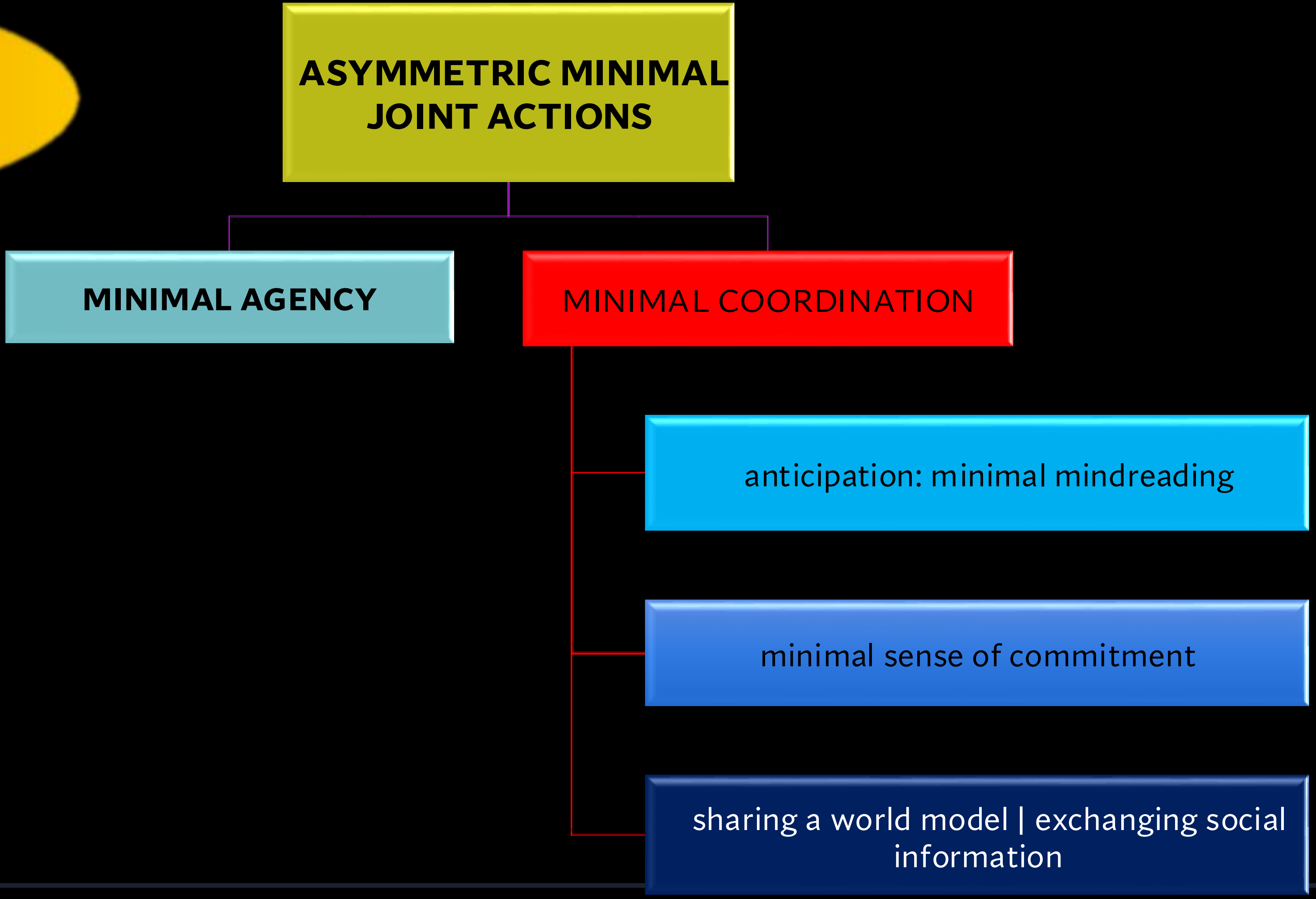
- *adult & child joint actions*
  ← child brings a lot of social understanding, even if the parent brings more
- *snuggling with a cat*

Interactions between a fully social agent and some partner – whether human, machine, or animal – that is not cognitively capable of full-fledged social joint action but that does respond in a way that productively invites further social responses from the social partner

Inbetween mere tool-use and social interactions

TOWARDS ASYMMETRIC JOINT ACTIONS

conditions for the junior partner



ASYMMETRIC MINIMAL JOINT ACTIONS

MINIMAL AGENCY

MINIMAL COORDINATION

anticipation: minimal mindreading

minimal sense of commitment

sharing a world model | exchanging social information

Moving Targets

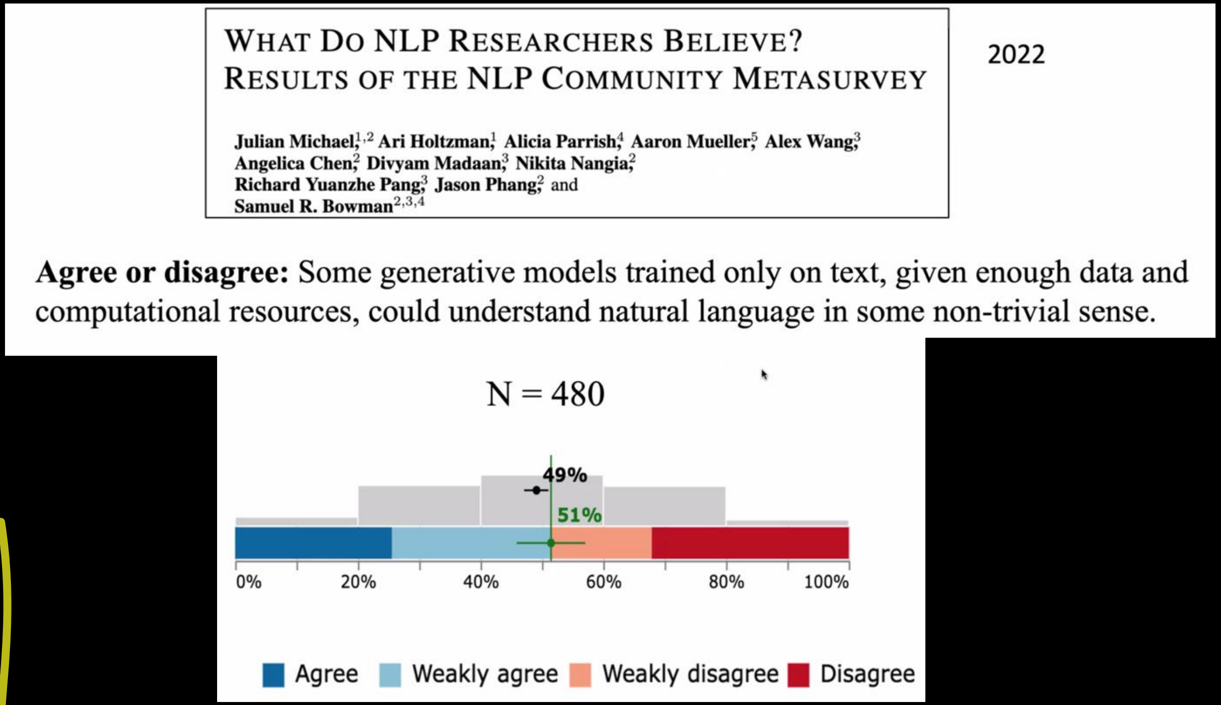
DEEP UNCERTAINTY REGARDING ASCRIBING PROPERTIES AND ABILITIES TO ARTIFICIAL SYSTEMS

There is not much hope that we will come soon to a scientifically sound and accepted ascription.

(1) Computer science research is moving fast, and the debates about attribution do not seem to come to an agreement.

(2) Philosophy: the meaning of the notions we use for the attributions of mental and socio-cognitive abilities is under debate.

(3) Sociological processes and legal changes continuously contribute to a change of our social practice of what status and rights are assigned to artificial counterparts.



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

2022

Julian Michael,[1,2] Ari Holtzman,[1] Alicia Parrish,[4] Aaron Mueller,[5] Alex Wang,[3] Angelica Chen,[2] Divyam Madaan,[3] Nikita Nangia,[2] Richard Yuanzhe Pang,[3] Jason Phang,[2] 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.

N = 480

A PLEA FOR CROSS-DISCIPLINARY COLLABORATION

All we can hope for is a negotiation process between computer science, philosophy, sociology, and legal research.

Michael et al., 2022

Conclusion

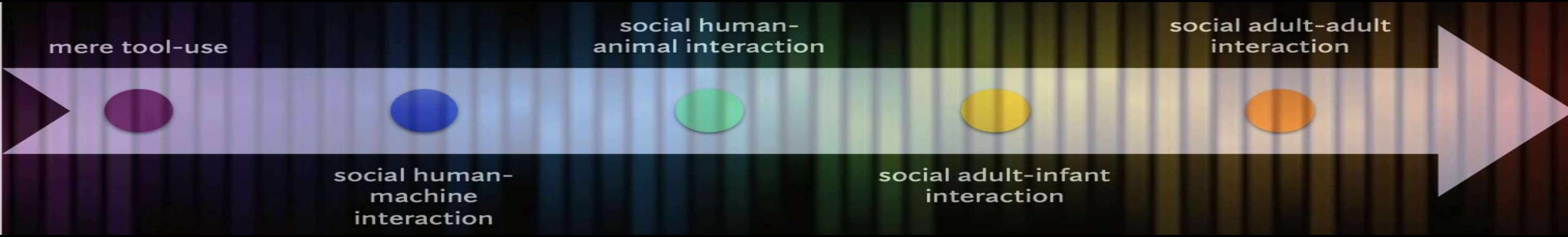
Consider the possibility of questioning the dichotomy between animate and inanimate entities

After all, we might be confronted with a new game.

The main aim of this talk was to prepare the grounds for questioning the dichotomy between animate and inanimate entities, as this is an important presupposition for developing a conceptual framework that can capture phenomena that I locate in the INBETWEEN.

If I am successful with this, I can argue for a gradual approach describing all kinds of social interactions, and finally answer the question of what we are doing when we interact with LLMs— what status artificial systems have in HMIs.

Then we can stop reducing all our interactions with artificial systems (and especially with future products of generative AI) to mere tool use.


mere tool-use
social human-machine interaction
social human-animal interaction
social adult-infant interaction
social adult-adult interaction

All this would not have been possible if I had not interacted with people & machines

Daniel
Dennett

Eric
Schwitzgebel

Mathew
Crosby

David
Schwitzgebel

Mike
Wilby

DigiDan

Thank you !

A HUMAN-MADE BOOK IN THE AGE OF MACHINE-GENERATED TEXTS

Anna´s AI Anthology

How to live with smart machines?

INVITATION

HYBRID BOOK LAUNCH

Anna's AI Anthology. How to live with smart machines?

7 OCTOBER 2024, 8 pm (Berlin time)

Kino Babylon (Rosa-Luxemburg-Straße 30, 10178 Berlin) & online via Zoom

The book launch will include four small panel discussions in which 2 to 3 authors will briefly recall the main theses of their essays, discuss them with the other authors, and answer questions from the audience.

REGISTER HERE: https://forms.gle/M1KYVgpVCPKf3Gor9

* Participation (in-person) is limited – notifications will be send via email – Zoom link will be distributed shortly before the event

DENKWERKSTATT BERLIN

IN KOOPERATION MIT MOMO Berlin

References

Agrawal, A., Mackey, L., & Kalai, A. T. (2023). *Do Language Models Know When They're Hallucinating References?* (arXiv:2305.18248). arXiv. http://arxiv.org/abs/2305.18248

Agüera y Arcas, B. (2022). Do Large Language Models Understand Us? *Daedalus*, *151*(2), 183–197. https://doi.org/10.1162/daed_a_01909

Barkham, P. (2021, July 25). Should rivers have the same rights as people? *The Guardian*. https://www.theguardian.com/environment/2021/jul/25/rivers-around-the-world-rivers-are-gaining-the-same-legal-rights-as-people

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. https://doi.org/10.1145/3442188.3445922

Bunten, A., Iorns, C., Townsend, J., & Borrows, L. (2021, June 3). *Rights for nature: How granting a river 'personhood' could help protect it*. The Conversation. http://theconversation.com/rights-for-nature-how-granting-a-river-personhood-could-help-protect-it-157117

Butterfill, S. A., & Apperly, I. A. (2013). How to Construct a Minimal Theory of Mind. *Mind & Language*, *28*(5), 606–637. https://doi.org/10.1111/mila.12036

Cole, S. (2023). 'It's Hurting Like Hell': AI Companion Users Are In Crisis, Reporting Sudden Sexual Rejection. *Vice*. https://www.vice.com/en/article/y3py9j/ai-companion-replika-erotic-roleplay-updates

Dooley, B., & Ueno, H. (2022, April 24). This Man Married a Fictional Character. He'd Like You to Hear Him Out. *The New York Times.* https://www.nytimes.com/2022/04/24/business/akihiko-kondo-fictional-character-relationships.html

Gunkel, D. J. (2023). *Person, Thing, Robot: A Moral and Legal Ontology for the 21st Century and Beyond*. The MIT Press. https://doi.org/10.7551/mitpress/14983.001.0001

Henrich, J. P. (2016). *The secret of our success: How culture is driving human evolution, domesticating our species, and making us smarter*. Princeton University Press.

Heyes, C. (2014). False belief in infancy: A fresh look. *Developmental Science*, *17*(5), 647–659. https://doi.org/10.1111/desc.12148

----------- (2015). Animal mindreading: What's the problem? *Psychonomic Bulletin & Review*, *22*(2), 313–327. https://doi.org/10.3758/s13423-014-0704-4

Jensen, C. B., & Blok, A. (2013). Techno-animism in Japan: Shinto Cosmograms, Actor-network Theory, and the Enabling Powers of Non-human Agencies. *Theory, Culture & Society*, *30*(2), 84–115. https://doi.org/10.1177/0263276412456564

Lake, B. M., & Baroni, M. (2023). Human-like Systematic Generalization through a Meta-learning Neural Network. *Nature*, 1–7. https://doi.org/10.1038/s41586-023-06668-3

Lemoine, B. (2022, June 11). Is LaMDA Sentient? — An Interview. *Medium*. https://cajundiscordian.medium.com/is-lamda-sentient-an-interview-ea64d916d917

Marcus, G., & Davis, E. (2020). GPT-3, Bloviator: OpenAI's language generator has no idea what it's talking about. *MIT Technology Review*. https://www.technologyreview.com/2020/08/22/1007539/gpt3-openai-language-generator-artificial-intelligence-ai-opinion

Michael, J., Sebanz, N., & Knoblich, G. (2016). The Sense of Commitment: A Minimal Approach. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01968

Michael, J., Holtzman, A., Parrish, A., Mueller, A., Wang, A., Chen, A., Madaan, D., Nangia, N., Pang, R. Y., Phang, J., & Bowman, S. R. (2022). What Do NLP Researchers Believe? Results of the NLP Community Metasurvey (arXiv:2208.12852).

References

O’Donnell, E., & Talbot-Jones, J. (2017, March 23). Three rivers are now legally people – but that’s just the start of looking after them. *The Conversation*. http://theconversation.com/three-rivers-are-now-legally-people-but-thats-just-the-start-of-looking-after-them-74983

Pacherie, E. (2013). Intentional joint agency: Shared intention lite. *Synthese*, *190*(10), 1817–1839. https://doi.org/10.1007/s11229-013-0263-7

Perner, J. (1991). *Understanding the representational mind* (pp. xiv, 348). The MIT Press.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *1*(4), 515–526. https://doi.org/10.1017/S0140525X00076512

Robertson, J. (2014). Human Rights vs.Robot Rights: Forecasts from Japan. *Critical Asian Studies*, *46*(4), 571–598. https://doi.org/10.1080/14672715.2014.960707

------------- (2017). *Robo sapiens japanicus: Robots, Gender, Family, and the Japanese Nation*.

Salmón, E. (2000). Kincentric Ecology: Indigenous Perceptions of the Human-Nature Relationship. *Ecological Applications*, *10*(5), 1327–1332. https://doi.org/10.2307/2641288

Scarborough, J. K., & Bailenson, J. N. (2014). Avatar Psychology. In M. Grimshaw (Ed.), *The Oxford Handbook of Virtuality*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199826162.013.033

Sterelny, K. (2012). *The Evolved Apprentice: How Evolution Made Humans Unique*. The MIT Press. https://doi.org/10.7551/mitpress/9780262016797.001.0001

Strasser, A. (2006). *Kognition künstlicher Systeme:* DE GRUYTER. https://doi.org/10.1515/9783110321104

------------ (2013). Kognition künstlicher Systeme. In *Kognition künstlicher Systeme*. De Gruyter. https://doi.org/10.1515/9783110321104

------ (Ed.). (2024). *Anna’s AI Anthology. How to live with smart machines?* xenomoi Verlag.

Strasser, A., & Schwitzgebel, E. (2024). Quasi-sociality: Toward Asymmetric Joint Actions. In *Anna’s AI Anthology. How to live with smart machines?* xenomoi Verlag.

Strasser, A., & Wilby, M. (2023). The AI-Stance: Crossing the Terra Incognita of Human-Machine Interactions? In *Social Robots in Social Institutions* (pp. 286–295). IOS Press. https://doi.org/10.3233/FAIA220628

Tomasello, M. (2008). *Origins of human communication* (pp. xiii, 393). MIT Press.

Trott, S., Jones, C., Chang, T., Michaelov, J., & Bergen, B. (2023). Do Large Language Models Know What Humans Know? *Cognitive Science*, *47*(7), e13309. https://doi.org/10.1111/cogs.13309

Vesper, C., Butterfill, S., Knoblich, G., & Sebanz, N. (2010). A minimal architecture for joint action. *Neural Networks*, *23*(8), 998–1003. https://doi.org/10.1016/j.neunet.2010.06.002

Warneken, F., Chen, F., & Tomasello, M. (2006). Cooperative Activities in Young Children and Chimpanzees. *Child Development*, *77*(3), 640–663. https://doi.org/10.1111/j.1467-8624.2006.00895.x

Weil, E. (2023, March 1). *You Are Not a Parrot*. New York Magazine. https://nymag.com/intelligencer/article/ai-artificial-intelligence-chatbots-emily-m-bender.html

Wilby, M., & Strasser, A. (2024). Situating machines within normative practices: Bridging responsibility gaps with the AI-Stance. In A. Strasser (Ed.), *Anna’s AI Anthology. How to live with smart machines?* xenomoi Verlag.