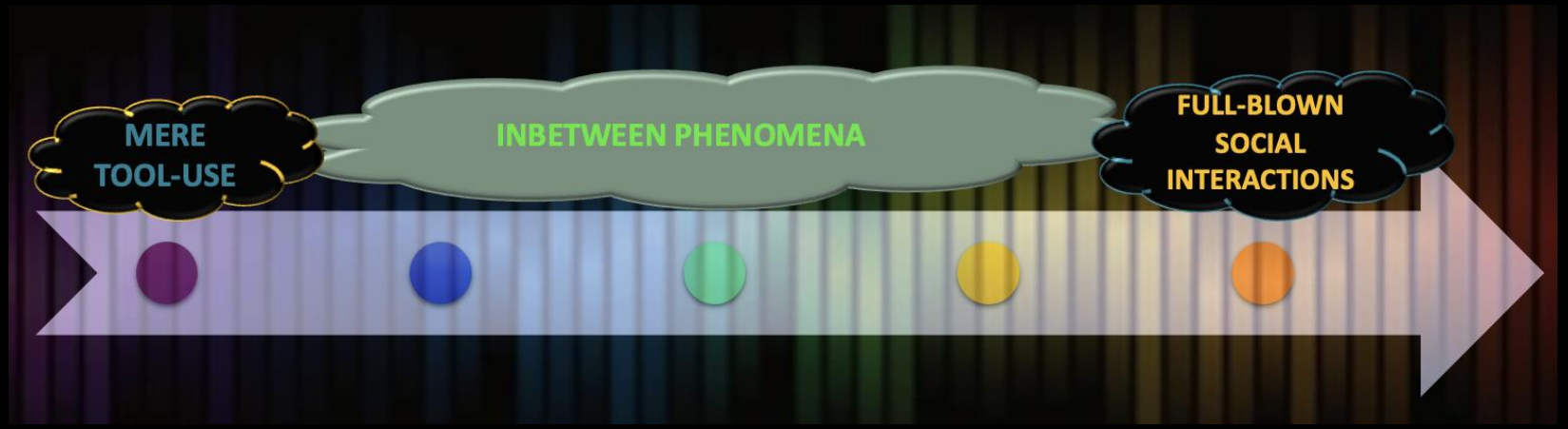
Various kinds of social interactions in the age of AI

Anna Strasser



Generative AI – How to characterize these new entities?

IT IS UNCLEAR WHICH PROPERTIES AND ABILITIES WE ARE JUSTIFIED IN ASCRIBING TO THEM.

Many notions that philosophers previously reserved for describing distinguishing features of humans are now being applied to machines

- intense debates about the extension of notions like individual & joint actions, understanding, knowledge, reasoning, phenomenological consciousness & sociality


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

Strasser & Strasser, 2024

NO AGREEMENT OF WHETHER CURRENT GENERATIVE AI SYSTEMS POSSESS AGENCY, INTELLIGENCE AND CONSCIOUSNESS

Increasing use of LLM-based chatbots that offer themselves to serve as collaborators, companions, friends, or even therapists

➢ **NECESSARY TO RETHINK WHAT CONSTITUTES SOCIAL INTERACTION**

(Agrawal et al., 2023; y Arcas, 2022; Lake & Baroni, 2023; Trott et al., 2023; Xu et al., 2025)

Social interactions with generative AI systems?

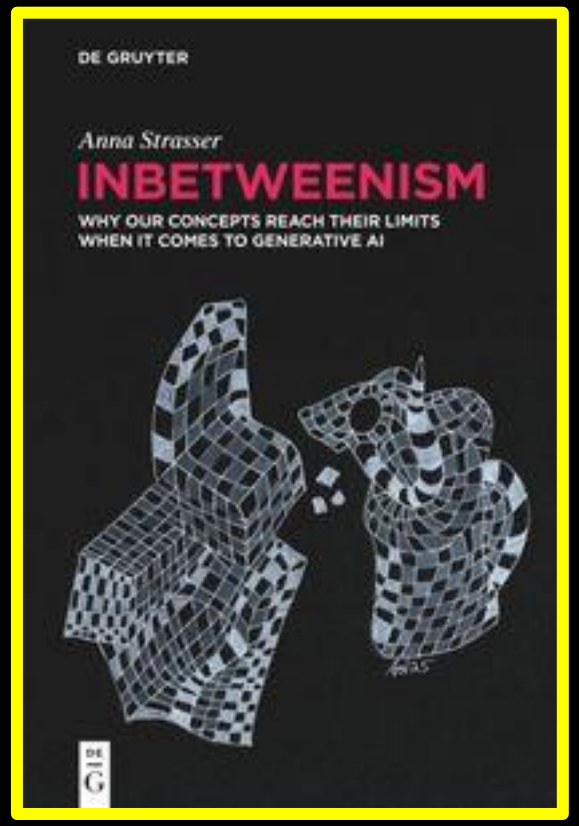
Can AI systems serve as social counterparts? – Can they act as group agents together with humans?

Hardcore instrumentalist view

no agency
no intelligence
no consciousness

- **AI systems cannot have a social status in human-machine interactions.**

challenged by observations that these systems are significantly infiltrating our social spheres



In expection of AGI view

full-fledged
agency
intelligence
consciousness

- **The whole demanding package of conditions that we require from humans can, in principle, also be fulfilled by sophisticated machines.**

seem to ignore differences between machines and humans

(Bender et al., 2021; Hashim, 2024; Lemoine, 2022; Long et al., 2024; Open-AI, n.d.; Strasser, 2026, Weil, 2023)

Overview

I

Towards Inbetweenism

- *a multidimensional spectrum of social interactions*

AI SYSTEMS WITH A MINIMAL FORM OF AGENCY & SOCIO-COGNITIVE ABILITIES CAN QUALIFY AS COLLABORATORS IN ASYMMETRIC JOINT ACTIONS WITHOUT CONSCIOUSNESS.

II

Limitations

of current AI systems as social interaction partners

CURRENT AI SYSTEMS WITHOUT EMOTIONS ARE NOT WELL-EQUIPPED TO ACT AS THERAPISTS OR EVEN AS PARTNERS IN THE SPHERE OF LIVING TOGETHER.

Slides can be downloaded
at https://www.denkwerkstatt.berlin/ANNA-STRASSER/TALKS

A multidimensional spectrum of social interactions

INTELLECTUALIST CONCEPTIONS OF PHILOSOPHY RESTRICT THE EXTENSION OF THE NOTION OF SOCIAL INTERACTIONS FAR TOO MUCH

MERE TOOL-USE

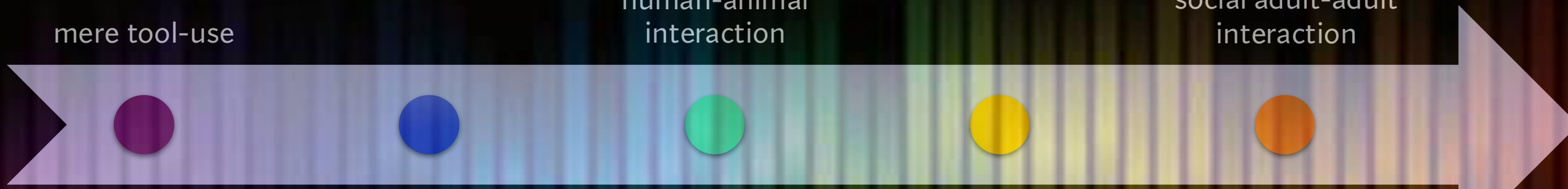
Hardcore instrumentalist

INBETWEEN PHENOMENA

SOCIAL INTERACTIONS

In expection of AGI view

QUASI-SOCIAL ASYMMETRIC SOCIAL INTERACTIONS



SINGLE-SIDED SOCIALITY

quasi-social human-machine interaction

quasi-social adult-infant interaction

FULL-BLOWN SOCIAL INTERACTION

VARIOUS KINDS OF ASYMMETRIC JOINT ACTIONS

Multidimensionality is a complex matter

QUASI-SOCIALITY EXISTS ON A COMPLEX SPECTRUM &
CAN BE CONCEPTUALIZED WITH THE HELP OF A DISJUNCTIVE CONCEPTUAL FRAMEWORK

If we do not focus on adult humans as the only type of social partners

➢ WE SHOULD EXPECT THAT THERE ARE SEVERAL DIMENSIONS ALONG WHICH WE CAN CHARACTERIZE VARIOUS INSTANCES OF MORE OR LESS SOCIAL INTERACTIONS

COMPLEX SOCIAL SKILLS WILL NOT EMERGE IN AN INSTANT

NOT

- *developmentally in humans*
- *phylogenetically in animal evolution*
- *technologically in the design of AI systems*

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

Asymmetric cases of joint actions

PARADIGMATIC EXAMPLE OF SOCIAL INTERACTIONS THAT COULD BE APPLICABLE TO ARTIFICIAL SYSTEMS

asymmetric joint actions

INTELLECTUALIST CONCEPTIONS

- philosophers tend to describe ideal cases

FULL-FLEDGED AGENCY BY DONALD DAVIDSON

- children, non-human animals, and artificial systems tend to fall through the conceptual net

Artist: Lorin Strasser

GRADUALIST APPROACHES & MINIMAL NOTIONS

REQUIREMENTS FOR AGENCY & OTHER SOCIO-COGNITIVE ABILITIES

ASYMMETRIC MINIMAL JOINT ACTIONS

conditions for the artificial partner

- **MINIMAL AGENCY**
- **MINIMAL COORDINATION**
  - anticipation: minimal mindreading
  - minimal sense of commitment
  - exchanging social information & sharing a world model

Further problems

WE DO NOT KNOW WITH CERTAINTY WHETHER THE CURRENT AI SYSTEMS FULFILL THE PROPOSED CONDITIONS

Any application of a conceptual framework raises empirical questions, namely, questions about the extent to which certain conditions (abilities and properties) are fulfilled by the entity in question.

justified ascription practice

property-based approaches
- operationalization

relational approaches
- *take our attitude towards an interaction partner as a reason for ascribing a social role to them.*

MY CLAIM
There are forms of agency and intelligence that do not require consciousness —
the three features are not always served in one package!

future research

establishing cross-disciplinary approaches considering social and technical factors of our ascription practice concerning the social status of artificial systems

→ may lead to a lucky marriage between ideas of the property-based and relational approaches

Diverse types of social interactions

II

Limitations
of current AI systems as social interaction partners

CURRENT AI SYSTEMS WITHOUT EMOTIONS ARE NOT WELL-EQUIPPED TO ACT AS THERAPISTS OR EVEN AS PARTNERS IN THE SPHERE OF LIVING TOGETHER.

THOSE CONSIDERATIONS ARE MOTIVATED BY THE POTENTIAL DISRUPTIVE EFFECTS OF GENAI IN VARIOUS SOCIAL DOMAINS.

HMIs can qualify as instances in this multidimensional spectrum of social interactions as asymmetric joint actions that happen in the domain of working

BUT

SOME OF OUR SOCIAL INTERACTIONS

- may require being less asymmetric,
- may demand that both interaction partners exhibit emotions and reflective rationality

Therapy & the role of emotions

Workshop on Artificial Joint Intentionality (AJI):

Skilled Social Interactions in the Age of AI

DATE	ORGANIZERS
6–8 November 2025	Cameron Buckner, University of Florida, US; Anna Strasser, DenkWerkstatt Berlin, Germany; Amber Ross, University of Florida, US; Duncan Purves, University of Florida, US; Kirk Ludwig, Bloomington, US

Ellen Fridland: Coregulation & the AI therapist

Ellen Fridland
King's College London (GB)
The Brookline Center for Community Mental Health (US)

A HIGH PERCENTAGE OF PEOPLE USE AI AS A THERAPEUTIC MEANS

But are current AI systems well equipped to establish a therapeutic relationship?

The therapeutic relationship plays a significant role in making psychotherapy effective.

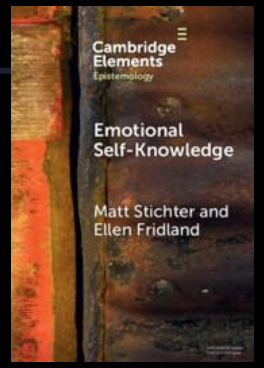
- good evidence
  - a **strong alliance** can improve well-being,
  - **empathy** can play a role in reducing anxiety, depression, and other personality disorders
  - **empathic relationships** can contribute to positive self-esteem and affirmation, which in turn can contribute to mental health

➢ AND ALL THIS HAS TO DO WITH EMOTIONS

WHAT MAKES THERAPY EFFECTIVE IS THAT IT CAN CHANGE/HEAL THE PREDICTIONS OF OUR AFFECTIVE SYSTEMS →THE EMOTIONS WE EXPERIENCE AS WE ENCOUNTER THE WORLD

(Cross et al., 2024; Sahebi & Parvizi-Wayne, 2026; Zao-Sanders, 2025)

What are emotions?



What are emotions?

EMBODIED RELATIONAL MODEL OF EMOTIONS

- relational affective models that are developed in early childhood through interpersonal interaction
  - **capacity for intrapersonal affective regulation, which enables us to care for ourselves on an affective level and form secure attachments with others**

What role play emotions?

ENABLE US TO NAVIGATE, GUIDE, REGULATE OUR INTERPERSONAL & SOCIAL INTERACTIONS

- without social interactions, we would probably not be able to develop the variety of emotional states, and our ability to self-regulate them
- research dramatically shows the extent to which reduced interactions negatively impact emotional development

How are emotions developed?

CO-REGULATION BETWEEN CHILDREN & CAREGIVERS

- emotions are shaped through the interpersonal interaction
- having a reliable counterpart can prove helpful in coping with emotions or behaviors
- bodily expression of emotions play an important role
  - e.g. transfer effects of a regulated nervous system from one person to another can play a role in promoting emotional balance

How can a therapeutic process change emotions?

RENEWED CO-REGULATION

IF

- we lose essential parts of the capacity to care for ourselves on an affective level and form secure attachments with others

THEN

- our well-being is in danger
- a renewed co-regulation may be beneficial

(Stichter & Fridland, 2025)

Functional role of therapy

EFFECTIVENESS OF PSYCHOTHERAPY IS AT LEAST PARTLY ATTRIBUTABLE TO THE THERAPEUTIC RELATIONSHIP BETWEEN THERAPIST & CLIENT, IN WHICH EMOTIONAL CO-REGULATION IS HAPPENING

A RENEWED EMOTIONAL CO-REGULATION

- offers a possible mechanism that enables us to initiate a reconfiguration/recalibration of our affective predictions and reactions
- to change emotions and enable us to better navigate, guide and regulate our social interactions

Repeated co-regulation with a sensitive, empathetic therapist

- can shift the client's affective predictions,
- enable them to break out of behavioral patterns,
- foster a greater capacity for intrapersonal affective regulation,
- ultimately enabling them to care for themselves on an affective level

GOAL OF THERAPY:

RECONFIGURE AFFECTIVE SCHEMAS & PREDICTIONS TO FORM SECURE ATTACHMENTS WITH OTHERS

Co-regulation among humans

CONSTITUTIVE ELEMENTS OF THERAPEUTIC RELATIONSHIPS

- alliance, empathy, positive regard, emotional expression, congruence and genuineness

← THIS HAS TO DO WITH EMOTIONS

	presupposes	enables	helpful for
FULL-BLOODED COREGULATION	**being embodied & having emotions & being in the same room**	mutual affective or ANS covariation (interactions between ANS signals)	attuning, empathy, trust ….
		access to physiological, biocultural information	understanding emotional signals transferred by body language, facial expressions, tone of voice, language
MINIMAL VERSIONS OF COREGULATION	video / audio / text *(Telehealth > Phone > Chat)*	make use of physiological, biocultural information from a distance (drawing on experiences of intense interpersonal embodied interactions)	a certain degree of co-regulation remotely

AI-therapy?

Can disembodied, unemotional attempts at emotional co-regulation recalibrate ingrained attachment patterns?

CAN AN AI THERAPIST TEACH THEIR CLIENTS SELF-REGULATION THROUGH CO-REGULATION?

WITHOUT BEING EMBODIED?

- no interaction with ANS signals, which play a role in attuning
- unable to maintain affective covariation or reciprocity
  - ability to convey emotional signals via body language/facial expressions/tone of voice is rather limited
  - no access to physiological/biocultural information

✓ without mutual ANS covariation

➢ **role physiological/biocultural information plays in co-regulation could be very minimal**

- even though studies indicate that face-to-face therapy is still more effective than Zoom therapy

WITHOUT EMOTIONS?

- IF the relational theory of emotion is right
- THEN human therapists are capable of teaching their clients self-regulation through co-regulation because they have emotions

! not without mutual affective covariation

➢ **legitimate doubts whether AI systems without emotions are well-equipped to act as therapists**

Further disadvantages of AI therapists

"AS-IF CO-REGULATION"?

COULD THERE BE SIMILAR EFFECTS WITHOUT MUTUAL AFFECTIVE COVARIATION?

BUT THERE ARE OTHER REASONS CONCERNING THE AVAILABILITY, THE RISK OF HARMING PATIENTS & THE POTENTIAL MISUSE OF SHARED INFORMATION THAT QUESTION THE USEFULNESS OF AI THERAPISTS

Availability

- Given the waiting lists for therapy places, this is naturally appealing.

BUT

- ❖ feeding the expectation of uninterrupted availability will not enable us to have better relationships with other humans who are never always available

- ❖ building a secure attachment is the foundation for forming close relationships without constantly needing to reassure ourselves that the other person is available

- ➢ **The AI therapist might prepare us to have more relationships with other AIs and to value them more highly because they are constantly available.**

Potential misuse of shared information

- Interacting with AI offers something like a safe space in which we can experiment with aspects of our own personality, which can help us find our identity

I don't want to question whether such interactions can serve as a useful practice space.

BUT

- ❖ I do wonder to what extent this can be a safe practice space as long as this technology comes with all the serious concerns about data protection.
- ❖ Private and intimate information is stored, and we cannot be sure that it will not be used against our interests.

Further disadvantages of AI therapists

Risk of harming patients

- ❖ AI systems used to treat mental health issues have contributed to suicides, the exacerbation of psychosis, and retraumatization

- The developers of AI systems are busy introducing ever-new guardrails to at least partially mitigate this danger

BUT

- ➢ to date, there is no fully successful alignment method that can rule out dangerous and inappropriate responses regarding suicidal thoughts, psychosis, and trauma

DESKILLING & DEPENDENCY

- AI systems support their users in emotional self-regulation

BUT

- ❖ due to their constant availability, this can contribute to a decline in the ability to regulate one's own emotions without the AI system (Phang et al., 2025).

- ❖ One could even point to the risk of an unhealthy emotional attachment here, which can lead to emotional dependence (Xie et al., 2023).

Conclusion

AI SYSTEMS WITH A MINIMAL FORM OF AGENCY & SOCIO-COGNITIVE ABILITIES CAN QUALIFY AS COLLABORATORS IN ASYMMETRIC JOINT ACTIONS WITHOUT CONSCIOUSNESS AND WITHOUT EMOTIONS.

CURRENT AI SYSTEMS WITHOUT EMOTIONS ARE NOT WELL-EQUIPPED TO ACT AS THERAPISTS OR EVEN AS PARTNERS IN THE SPHERE OF LIVING TOGETHER BECAUSE EMOTIONS ARE RELATIONAL (CO-REGULATION).

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
- Cross, S., Bell, I., Nicholas, J., Valentine, L., Mangelsdorf, S., Baker, S., Titov, N., & Alvarez-Jimenez, M. (2024). Use of AI in Mental Health Care: Community and Mental Health Professionals Survey. *JMIR Mental Health*, *11*, e60589–e60589. https://doi.org/10.2196/60589
- Lake, B. M., & Baroni, M. (2023). Human-like Systematic Generalization through a Meta-learning Neural Network. *Nature*, *623*, 115–121. https://doi.org/10.1038/s41586-023-06668-3
- Phang, J., Lampe, M., Ahmad, L., Agarwal, S., Fang, C. M., Liu, A. R., Danry, V., Lee, E., Chan, S. W. T., Pataranutaporn, P., & Maes, P. (2025). *Investigating Affective Use and Emotional Well-being on ChatGPT* (Version 1). arXiv. https://doi.org/10.48550/ARXIV.2504.03888
- Sahebi, S., & Parvizi-Wayne, D. (2026). LLM-"Friends" are Hostile Scaffolds in the Age of Loneliness. *Minds and Machines*, *36*(1), 13. https://doi.org/10.1007/s11023-026-09765-0
- Stichter, M., & Fridland, E. (2025). *Emotional Self-Knowledge: How Affective Skills Reveal Our Values, Goals, Cares, and Concerns* (1st ed.). Cambridge University Press. https://doi.org/10.1017/9781009307857
- Strasser, A. (Ed.). (2024). *Anna's AI Anthology. How to live with smart machines?* xenomoi Verlag.
- Strasser, A. (2026). *Inbetweenism: Why our concepts reach their limits when it comes to Generative AI*. De Gruyter. https://doi.org/https://doi.org/10.1515/9783111438078
- 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
- Xie, T., Pentina, I., & Hancock, T. (2023). Friend, mentor, lover: Does chatbot engagement lead to psychological dependence? *Journal of Service Management*, *34*(4), 806–828. https://doi.org/10.1108/JOSM-02-2022-0072
- Xu, F., Hao, Q., Shao, C., Zong, Z., Li, Y., Wang, J., Zhang, Y., Wang, J., Lan, X., Gong, J., Ouyang, T., Meng, F., Yan, Y., Yang, Q., Song, Y., Ren, S., Hu, X., Feng, J., Gao, C., & Li, Y. (2025). Toward large reasoning models: A survey of reinforced reasoning with large language models. *Patterns*, *6*(10), 101370. https://doi.org/10.1016/j.patter.2025.101370
- y Arcas, B. A. (2022). Do Large Language Models Understand Us? *Daedalus*, *151*(2), 183–197. https://doi.org/10.1162/daed_a_01909
- Zao-Sanders, M. (2025, April 9). How People Are Really Using Gen AI in 2025. *Harvard Business Review*. https://hbr.org/2025/04/how-people-are-really-using-gen-ai-in-2025