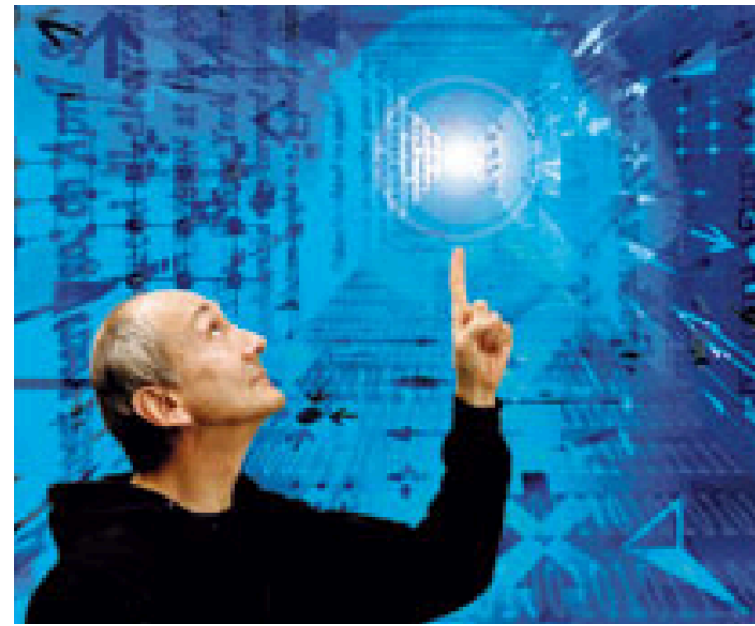


do language models require sensory grounding
for meaning and understanding?

david chalmers

NYU debate: march 24, 2023



stevan harnad (1990): the symbol grounding problem

for symbols in AI systems to have meaning, they need sensory grounding (or: bodily grounding, or external grounding)

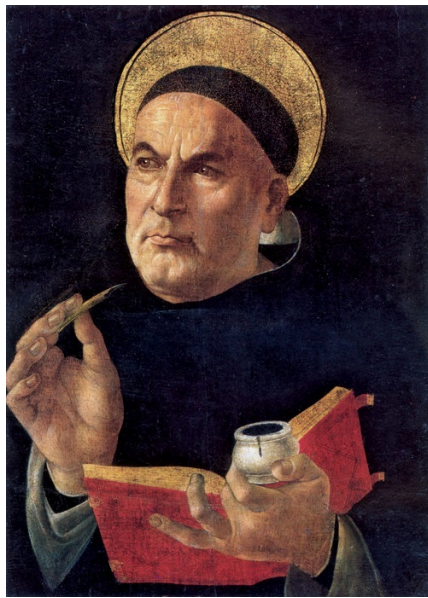
emily bender and alexander koller
(2020)

language models lack sensory
grounding, and don't have meaning or
understanding



Photo credit: Ian Allen

do *humans* need sensory grounding for meaning, understanding, and thinking?



thomas aquinas (1200s): “there’s nothing in the mind that wasn’t first in the senses”.

(if so, thinking and understanding require sensory capacities — at least in humans)



diderot, condillac (1700s): *sensism*

no cognition without sensation



avicenna/ibn sina (1000s): there can be thinking without sensing

Avicenna's Floating Man

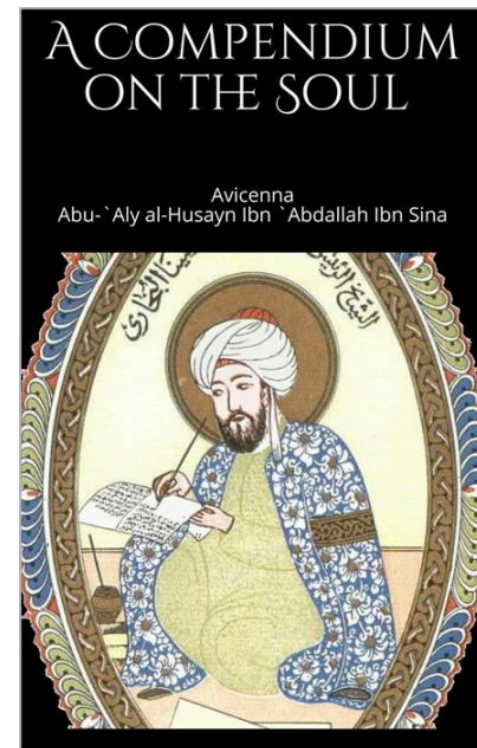


Philosophy thought experiments in drawings, by
@HelenDeCruz

Avicenna's Floating Man

[Ibn Sina, *De Anima*, ~1027]

“...He was just created at a stroke, fully developed and perfectly formed but with his vision shrouded from perceiving all external objects – created floating in the air or in the space, not buffeted by any perceptible current of the air that supports him, his limbs separated and kept out of contact with one another, so that they do not feel each other. Then let the subject consider whether he would affirm the existence of his self. There is no doubt that he would affirm his own existence, although not affirming the reality of any of his limbs or any external thing.”





avicenna: floating man can think about himself without ever sensing (and without sensory capacities?)

others: he could also think about mathematics, logic, philosophy, and could form hypotheses about external reality.

if so, thinking and understanding don't *require*
sensory capacities

still: some *sorts* of thinking and understanding
require sensory capacities

a system without senses couldn't fully understand

redness

how to ride a bike

though it could perhaps have a partial understanding of these things



likewise: an AI system without sensory input
could think about

itself

math and logic

hypotheses about the external world

but there would be limits on its understanding
in sensory and bodily domains

same for language models:

LLMs don't need sensory grounding for thinking and understanding (in general);

but they arguably need sensory grounding for some sorts of thinking and understanding.

Q: does sensing boost thinking (to a new level)
in language models?

(and in intelligent creatures generally?)

pure language models

vs.

multimodal language models

pure language models have text inputs and outputs, but not human-style senses

multimodal models also process quasi-sensory inputs (images, audio, ...) and quasi-embodied outputs



Figure 1: LLMs have not interacted with their environment and observed the outcome of their responses, and thus are not grounded in the world. SayCan grounds LLMs via value functions of pretrained skills, allowing them to execute real-world, abstract, long-horizon commands on robots.

multimodal language models have some capacities that pure language models lack:

sensing

bodily action

perceptual understanding?

embodied know-how?

question: do multimodal models outperform
language models on domains they share

text-only tasks?

tests of cognition and reasoning?

expectation (pre-LLMs): yes

multimodal systems will outperform pure
text systems

pure language models may work for language
tasks but not for thinking

now: not so clear

pure LLMs excel at many non-linguistic tasks

multimodal LLMs sometimes have a small
boost, sometimes not

Simulated exams	GPT-4 <small>estimated percentile</small>	GPT-4 (no vision) <small>estimated percentile</small>	GPT-3.5 <small>estimated percentile</small>
Uniform Bar Exam (MBE+MEE+MPT) ¹	298 / 400 <small>-90th</small>	298 / 400 <small>-90th</small>	213 / 400 <small>-10th</small>
LSAT	163 <small>-88th</small>	161 <small>-83rd</small>	149 <small>-40th</small>
SAT Evidence-Based Reading & Writing	710 / 800 <small>-93rd</small>	710 / 800 <small>-93rd</small>	670 / 800 <small>-87th</small>
SAT Math	700 / 800 <small>-89th</small>	690 / 800 <small>-89th</small>	590 / 800 <small>-70th</small>
Graduate Record Examination (GRE) Quantitative	163 / 170 <small>-80th</small>	157 / 170 <small>-62nd</small>	147 / 170 <small>-25th</small>
Graduate Record Examination (GRE) Verbal	169 / 170 <small>-99th</small>	165 / 170 <small>-96th</small>	154 / 170 <small>-63rd</small>
Graduate Record Examination (GRE) Writing	4 / 6 <small>-54th</small>	4 / 6 <small>-54th</small>	4 / 6 <small>-54th</small>
USABO Semifinal Exam 2020	87 / 150 <small>99th - 100th</small>	87 / 150 <small>99th - 100th</small>	43 / 150 <small>31st - 33rd</small>
USNCO Local Section Exam 2022	36 / 60	38 / 60	24 / 60
Medical Knowledge Self-Assessment Program	75 %	75 %	53 %
Codeforces Rating	392 <small>below 5th</small>	392 <small>below 5th</small>	260 <small>below 5th</small>
AP Art History	5 <small>86th - 100th</small>	5 <small>86th - 100th</small>	5 <small>86th - 100th</small>
AP Biology	5 <small>85th - 100th</small>	5 <small>85th - 100th</small>	4 <small>62nd - 85th</small>
AP Calculus BC	4 <small>43rd - 59th</small>	4 <small>43rd - 59th</small>	1 <small>0th - 7th</small>
AP Chemistry	4 <small>71st - 88th</small>	4 <small>71st - 88th</small>	2 <small>22nd - 46th</small>



Yann LeCun

@ylecun



Replying to [@davidchalmers42](#)

Can an LLM figure this out:

In front of us are six gears numbered 1 to 6, mounted on axles in a row. Each gear is engaged with the next gear. If gear number 3 is rotated clockwise, in which direction will gears 1 and 6 rotate?

7:26 PM · Sep 21, 2022



David Chalmers @davidchalmers42 · Sep 21, 2022 ...

Replying to @ylecun

OK, GPT-3 made a mess of this. "If gear 3 is rotated clockwise, gear 2 will rotate clockwise. If gear 2 is rotated clockwise, gear 1 will rotate counterclockwise. If gear 1 is rotated counterclockwise, gear 6 will rotate clockwise."



David Chalmers @davidchalmers42 · Sep 21, 2022 ...

Replying to @davidchalmers42 and @ylecun

another model did a bit better with the local dynamics but messed up consistency. i'm holding out for better performance from GPT-4!

CH

In front of us are six gears numbered 1 to 6, mounted on axles in a row. Each gear is engaged with the next gear. If gear number 3 is rotated clockwise, in which direction will gears 1 and 6 rotate?



When gears are engaged, they rotate in opposite directions. If gear number 3 is rotated clockwise, the gears will rotate as follows:

1. Gear 3 rotates clockwise.
2. Gear 4, engaged with gear 3, rotates counterclockwise.
3. Gear 5, engaged with gear 4, rotates clockwise.
4. Gear 6, engaged with gear 5, rotates counterclockwise.

Now let's move back to gears 1 and 2:

1. Gear 2, engaged with gear 3, rotates counterclockwise.
2. Gear 1, engaged with gear 2, rotates clockwise.

So, if gear number 3 is rotated clockwise, gear 1 will rotate clockwise, and gear 6 will rotate counterclockwise.

multimodal models are valuable:

“a picture is worth a thousand words”:
image training may carry more data

but pure models with enough text data can
perform as well: then sensing isn't boosting
thinking to a new level that can't be reached
without it

these results provide at least some evidence
against sensism

finally: do large language models understand?

to answer this well, we may need *conceptual engineering* (not just software engineering!)

engineer new concepts of understanding,
thinking, meaning

b-understanding: to behave as if one understands

e-understanding: to experience as if one understands

current language models may b-understand but
not e-understand

beyond that, do they really understand?

at some point this becomes a verbal question

do LLMs need sensory grounding for thinking and understanding? **no**

do they need sensory grounding for some sorts of thinking and understanding? **yes**

do they need sensory grounding for good enough performance on cognitive tasks? **no**

do current LLMs understand? **mu**