

Advanced NLP

Lecture 11: What Constitutes “Intelligence”? On Moving Limits in Nonconscious Computation

Dr. Gabriel Stanovsky

June 25, 2023



- LLMs achieve **impressive performance** on various tasks

- LLMs achieve **impressive performance** on various tasks
 - Math, world knowledge, common sense, and more

- LLMs achieve **impressive performance** on various tasks
 - Math, world knowledge, common sense, and more
- Often with only **in-context supervision**

- LLMs achieve **impressive performance** on various tasks
 - Math, world knowledge, common sense, and more
- Often with only **in-context supervision**
 - **Caveat:** unclear if they see signal during supervision

- LLMs achieve **impressive performance** on various tasks
 - Math, world knowledge, common sense, and more
- Often with only **in-context supervision**
 - **Caveat:** unclear if they see signal during supervision
- Raises a debate: Do they **really** understand?



Abeba Birhane

@Abebab

...

gentle reminder: all large language models are good at
predicting the next word in a sequence based on
previous words they've seen. that's all. there's no
understanding of meaning whatsoever

9:47 AM · Nov 30, 2022



Abeba Birhane
@Abebab

gentle reminder: all large language models are good at
is predict the next word in a sequence based on
previous words they've
understanding of mean

9:47 AM · Nov 30, 2022

...



@emilymbender@dair-community.social on ... @emilymb... · Dec 1

Yes, exactly this. I wish we didn't need to keep reminding people, and
@Abebab is commendable for being gentle about it!

For the long form of this argument, see Bender & @alkoller 2020:
aclanthology.org/2020.acl-main....



Abeba Birhane
@Abebab

gentle reminder: all large language models are good at
is predict the next word in
previous words they've seen
understanding of meaning

9:47 AM · Nov 30, 2022



@emilymbender@dair-community.social on ... @emilymb... · Dec 1
Yes, exactly this. I wish we didn't need to keep reminding people, and
[@Abebab](#) is commendable for being gentle about it!

For the long form of this argument, see Bender & [@alkoller](#) 2020:
aclanthology.org/2020.acl-main...



((((J()() 'yoav)))) 🦜 @yoavgo · Dec 3

dismissing these models as "just silly stochastic parrots" is an extremely
an anti-curious and anti-scientific position, in my view.

 (((((yoav))))))! @yoavgo · Dec 3

dismissing these models as "just silly stochastic parrots" is an extremely
an anti-curious and anti-scientific position, in my view.



Abeba Birhane @Abebab · Dec 3

healthy scepticism and criticism not is anti-scientific. science without
healthy scepticism is dogma

 (((J)(J) 'yoav)))) @yoavgo · Dec 3

disn
an a



Abeba Birhane @Abebab · Dec 3

healthy skepticism and criticism not in anti-scientific science without

healthy skepticism and criticism not in anti-scientific science without

 (((J)(J) 'yoav)))) @yoavgo · Dec 3

Replying to @Abebab

skepticism is not sufficient for a falsifiable theory.



@okeefekat@dair-community.social @okeefekat · Dec 3

Replying to @BlancheMinerva @yoavgo and @Abebab

You do realize there is a whole field of actual science studying cognition, yes? And that @Abebab's PhD is in cognitive science?

...

It's not a quick gotcha question for Twitter.



 (((J)(J) 'yoav)))) @yoavgo · Dec 3

...

Replying to @okeefekat @BlancheMinerva and @Abebab

but that's the core point. i haven't seen a concrete proposal (from @Abebab or @emilymbender or others) for a definition of "understanding" that we can test a model against (with the caveat that I am more familiar with emily's writing than abeba's). only philosophical arguments.



2



2



17



Tip

(((J)(J) 'yoav))) @yoavgo · Dec 3

disn
an a



Abeba Birhane @Abebab · Dec 3

(((J)(J) 'yoav))) @yoavgo · Dec 3

@okeefekat@dair-community.social @okeefekat · Dec 3
Replying to @BlancheMinerva @yoavgo and @Abebab

(((J)(J) 'yoav))) @yoavgo · Dec 3

Replying to @okeefekat, @BlancheMinerva and @Abebab
but that's the core point, I haven't seen a concrete proposal (from
@Abebab or @emilymbender or others) for a definition of "understanding"
that we can test a model against (with the caveat that I am more familiar
with emily's writing than abeba's). only philosophical arguments.

0 2 t 2 17 Tip



@emilymbender@dair-community.social on ... @emilymb... · Dec 4 ...

Replying to @BlancheMinerva @yoavgo and @Abebab

Why in the world is the burden of proof --- i.e. the development of
definitions and tests --- somehow supposed to be on those of us who are
calling BS on claims of "understanding" rather than on those making the
claims?

 (((J)(J) 'yoav))) @yoavgo · Dec 3

disn
an a
 **Abeba Birhane** @Abebab · Dec 3

 (((J)(J) 'yoav))) @yoavgo · Dec 3

 @okeefekat@dair-community.social @okeefekat · Dec 3
Replying to @BlancheMinerva @yoavgo and @Abebab

 (((J)(J) 'yoav))) @yoavgo · Dec 3
Replying to @okeefekat @BlancheMinerva and @Abebab

 @emilymbender@dair-community.social on ... @emilymb... · Dec 4 ...
Replying to @BlancheMinerva @yoavgo and @Abebab

Why in the world is the burden of proof --- i.e. the development of definitions and tests --- somehow supposed to be on those of us who are calling BS on claims of "understanding" rather than on those making the claims?

 **Stella Rose Biderman** @BlancheMinerva · Dec 4

Replying to @emilymbender @yoavgo and @Abebab

The burden of proof is not on you. All we are doing is asking for feedback about what kinds of evidence would be compelling.



2



4



Tip

Our take in the discussion: Moving of the Goalposts

- AI research describes Intelligence as having **conscious experience**

Our take in the discussion: Moving of the Goalposts

- AI research describes Intelligence as having **conscious experience**
 - Usually without saying so explicitly

Our take in the discussion: Moving of the Goalposts

- AI research describes Intelligence as having **conscious experience**
 - Usually without saying so explicitly
 - We'll survey literature to show this

Our take in the discussion: Moving of the Goalposts

- AI research describes Intelligence as having **conscious experience**
 - Usually without saying so explicitly
 - We'll survey literature to show this
- However, **objectively** measuring consciousness is famously hard

Our take in the discussion: Moving of the Goalposts

- AI research describes Intelligence as having **conscious experience**
 - Usually without saying so explicitly
 - We'll survey literature to show this
- However, **objectively** measuring consciousness is famously hard
 - **The mind-body problem™** — frustrating research for centuries

Our take in the discussion: Moving of the Goalposts

- AI research describes Intelligence as having **conscious experience**
 - Usually without saying so explicitly
 - We'll survey literature to show this
- However, **objectively** measuring consciousness is famously hard
 - **The mind-body problem™** — frustrating research for centuries
- Instead, AI focuses an **proxy** intelligence benchmark (e.g., chess)

Our take in the discussion: Moving of the Goalposts

- AI research describes Intelligence as having **conscious experience**
 - Usually without saying so explicitly
 - We'll survey literature to show this
- However, **objectively** measuring consciousness is famously hard
 - **The mind-body problem™** — frustrating research for centuries
- Instead, AI focuses an **proxy** intelligence benchmark (e.g., chess)
 - Implicitly assuming models **must be** conscious to solve it

Our take in the discussion: Moving of the Goalposts

- AI research describes Intelligence as having **conscious experience**
 - Usually without saying so explicitly
 - We'll survey literature to show this
- However, **objectively** measuring consciousness is famously hard
 - **The mind-body problem™** — frustrating research for centuries
- Instead, AI focuses an **proxy** intelligence benchmark (e.g., chess)
 - Implicitly assuming models **must be** conscious to solve it
 - When ML models “solve” the task, a new one is chosen

Our take in the discussion: Moving of the Goalposts

- AI research describes Intelligence as having **conscious experience**
 - Usually without saying so explicitly
 - We'll survey literature to show this
- However, **objectively** measuring consciousness is famously hard
 - **The mind-body problem™** — frustrating research for centuries
- Instead, AI focuses an **proxy** intelligence benchmark (e.g., chess)
 - Implicitly assuming models **must be** conscious to solve it
 - When ML models “solve” the task, a new one is chosen
- **This isn't productive**

Our take in the discussion: Moving of the Goalposts

- AI research describes Intelligence as having **conscious experience**
 - Usually without saying so explicitly
 - We'll survey literature to show this
- However, **objectively** measuring consciousness is famously hard
 - **The mind-body problem™** — frustrating research for centuries
- Instead, AI focuses an **proxy** intelligence benchmark (e.g., chess)
 - Implicitly assuming models **must be** conscious to solve it
 - When ML models “solve” the task, a new one is chosen
- **This isn't productive**
 - It's not informing us about what constitutes intelligence

Our take in the discussion: Moving of the Goalposts

- AI research describes Intelligence as having **conscious experience**
 - Usually without saying so explicitly
 - We'll survey literature to show this
- However, **objectively** measuring consciousness is famously hard
 - **The mind-body problem™** — frustrating research for centuries
- Instead, AI focuses an **proxy** intelligence benchmark (e.g., chess)
 - Implicitly assuming models **must be** conscious to solve it
 - When ML models “solve” the task, a new one is chosen
- **This isn't productive**
 - It's not informing us about what constitutes intelligence
 - Or how to build “real” intelligent systems

Today

- 1 For foundational AI, Intelligence = Having consciousness
- 2 Moving the Goalposts
- 3 Moving Forward
- 4 Conclusions

Outline

- 1 For foundational AI, Intelligence = Having consciousness
- 2 Moving the Goalposts
- 3 Moving Forward
- 4 Conclusions

Conscious Experience

- **Immediate** to the experiencer through senses

Conscious Experience

- **Immediate** to the experiencer through senses
 - E.g., feeling sadness, the experience of seeing the color **red**

Conscious Experience

- **Immediate** to the experiencer through senses
 - E.g., feeling sadness, the experience of seeing the color **red**
 - Sometimes termed **Qualia**

Conscious Experience

- **Immediate** to the experiencer through senses
 - E.g., feeling sadness, the experience of seeing the color **red**
 - Sometimes termed **Qualia**
- Cannot be currently **objectively** measured by an external observer

Conscious Experience

- **Immediate** to the experiencer through senses
 - E.g., feeling sadness, the experience of seeing the color **red**
 - Sometimes termed **Qualia**
- Cannot be currently **objectively** measured by an external observer
- Instead mediated through **abstractions**

Conscious Experience

- **Immediate** to the experiencer through senses
 - E.g., feeling sadness, the experience of seeing the color **red**
 - Sometimes termed **Qualia**
- Cannot be currently **objectively** measured by an external observer
- Instead mediated through **abstractions**
 - “I feel like a trainwreck”

Conscious Experience

- **Immediate** to the experiencer through senses
 - E.g., feeling sadness, the experience of seeing the color **red**
 - Sometimes termed **Qualia**
- Cannot be currently **objectively** measured by an external observer
- Instead mediated through **abstractions**
 - “I feel like a trainwreck”
 - “This color reminds of sunsets in my childhood”

Conscious Experience

- **Immediate** to the experiencer through senses
 - E.g., feeling sadness, the experience of seeing the color **red**
 - Sometimes termed **Qualia**
- Cannot be currently **objectively** measured by an external observer
- Instead mediated through **abstractions**
 - “I feel like a trainwreck”
 - “This color reminds of sunsets in my childhood”
- Being conscious is a **binary, singular** trait

Conscious Experience

- **Immediate** to the experiencer through senses
 - E.g., feeling sadness, the experience of seeing the color **red**
 - Sometimes termed **Qualia**
- Cannot be currently **objectively** measured by an external observer
- Instead mediated through **abstractions**
 - “I feel like a trainwreck”
 - “This color reminds of sunsets in my childhood”
- Being conscious is a **binary, singular** trait
 - You can't be 90% conscious

Conscious Experience

- **Immediate** to the experiencer through senses
 - E.g., feeling sadness, the experience of seeing the color **red**
 - Sometimes termed **Qualia**
- Cannot be currently **objectively** measured by an external observer
- Instead mediated through **abstractions**
 - “I feel like a trainwreck”
 - “This color reminds of sunsets in my childhood”
- Being conscious is a **binary, singular** trait
 - You can't be 90% conscious
 - You can't be conscious in math but not in analyzing sentiment

Being conscious is a **binary, singular** trait. You can't be 90% conscious,
You can't be conscious in math but not in analyzing sentiment

Being conscious is a **binary, singular** trait. You can't be 90% conscious,
You can't be conscious in math but not in analyzing sentiment

- I think this matches AI's search for a singular binary "intelligence"

Being conscious is a **binary, singular** trait. You can't be 90% conscious,
You can't be conscious in math but not in analyzing sentiment

- I think this matches AI's search for a singular binary "intelligence"
- Next we'll read through some quotes from **seminal work**

Being conscious is a **binary, singular** trait. You can't be 90% conscious,
You can't be conscious in math but not in analyzing sentiment

- I think this matches AI's search for a singular binary "intelligence"
- Next we'll read through some quotes from **seminal work**
- Keep the definition of **conscious experience** in mind

Turing (1950)

Not until a machine can write a sonnet or compose a concerto because of **thoughts and emotions felt**, and not by the chance fall of symbols, could we agree that machine equals brain—that is, not only write it but **know that it had written it**. No mechanism could feel (and not merely artificially signal, an easy contrivance) pleasure at its successes, grief when its valves fuse, be warmed by flattery, be made miserable by its mistakes, be charmed by sex, be angry or depressed when it cannot get what it wants.

Searle (2010)

I demonstrated years ago with the so-called Chinese Room Argument that the implementation of the computer program is not by itself sufficient for **consciousness** or intentionality.

- **The Chinese Room** questions if a computer can be truly intelligent by imagining a non-Chinese speaker using a rule-book to manipulate Chinese symbols, seemingly having comprehension without real intelligence.
- In the above quote Searle acknowledges that in his view intelligence = having consciousness

Bender et al. (2021)

Our human understanding of coherence derives from our ability to recognize interlocutors' **beliefs and intentions** within context. That is, human language use takes place between individuals who share common ground and are mutually **aware** of that sharing (and its extent), who have communicative intents which they use language to convey, and who **model each others' mental states as they communicate.**

O'Gieblyn, (2021)

As AI continues to blow past us in benchmark after benchmark of higher cognition **we quell our anxiety by insisting that what distinguishes true consciousness is emotions, perception, the ability to experience and feel.** The qualities, in other words, which we share with animals.

Outline

- 1 For foundational AI, Intelligence = Having consciousness
- 2 Moving the Goalposts
- 3 Moving Forward
- 4 Conclusions

So why aren't we measuring consciousness in models?

- Objective experiences

So why aren't we measuring consciousness in models?

- **Objective experiences**
 - E.g., chemical or electrical activity in our brain, weights in ANN

So why aren't we measuring consciousness in models?

- **Objective experiences**
 - E.g., chemical or electrical activity in our brain, weights in ANN
 - Results don't depend on who's measuring

So why aren't we measuring consciousness in models?

- **Objective experiences**
 - E.g., chemical or electrical activity in our brain, weights in ANN
 - Results don't depend on who's measuring
- **Subjective experiences**

So why aren't we measuring consciousness in models?

- **Objective experiences**
 - E.g., chemical or electrical activity in our brain, weights in ANN
 - Results don't depend on who's measuring
- **Subjective experiences**
 - E.g., private thoughts, desires

So why aren't we measuring consciousness in models?

- **Objective experiences**
 - E.g., chemical or electrical activity in our brain, weights in ANN
 - Results don't depend on who's measuring
- **Subjective experiences**
 - E.g., private thoughts, desires
 - Not communicated, immediately available

So why aren't we measuring consciousness in models?

- **Objective experiences**
 - E.g., chemical or electrical activity in our brain, weights in ANN
 - Results don't depend on who's measuring
- **Subjective experiences**
 - E.g., private thoughts, desires
 - Not communicated, immediately available
 - Can't be measured objectively (kg, ml)

So why aren't we measuring consciousness in models?

- **Objective experiences**
 - E.g., chemical or electrical activity in our brain, weights in ANN
 - Results don't depend on who's measuring
- **Subjective experiences**
 - E.g., private thoughts, desires
 - Not communicated, immediately available
 - Can't be measured objectively (kg, ml)
- **The mind-body problem:**
Q: How to objectively measure subjective experience?

So why aren't we measuring consciousness in models?

- **Objective experiences**
 - E.g., chemical or electrical activity in our brain, weights in ANN
 - Results don't depend on who's measuring
- **Subjective experiences**
 - E.g., private thoughts, desires
 - Not communicated, immediately available
 - Can't be measured objectively (kg, ml)
- **The mind-body problem:**

Q: How to objectively measure subjective experience?
- Has been **frustrating** researchers for centuries

Proxy Tasks

- Measuring consciousness is hard, AI focuses on proxy tasks (chess)

Proxy Tasks

- Measuring consciousness is hard, AI focuses on proxy tasks (chess)
 - Implicit assumption: models **must be** conscious to solve them

Proxy Tasks

- Measuring consciousness is hard, AI focuses on proxy tasks (chess)
 - Implicit assumption: models **must be** conscious to solve them
- However, in doing so, moved to objective multifaceted measurements

Proxy Tasks

- Measuring consciousness is hard, AI focuses on proxy tasks (chess)
 - Implicit assumption: models **must be** conscious to solve them
- However, in doing so, moved to objective multifaceted measurements
 - You can be 90% accurate in chess moves

Proxy Tasks

- Measuring consciousness is hard, AI focuses on proxy tasks (chess)
 - Implicit assumption: models **must be** conscious to solve them
- However, in doing so, moved to objective multifaceted measurements
 - You can be 90% accurate in chess moves
 - You can be terrible at chess but still be conscious

Proxy Tasks

- Measuring consciousness is hard, AI focuses on proxy tasks (chess)
 - Implicit assumption: models **must be** conscious to solve them
- However, in doing so, moved to objective multifaceted measurements
 - You can be 90% accurate in chess moves
 - You can be terrible at chess but still be conscious
 - Turns out you can be great at chess but not very conscious

Proxy Tasks

- Measuring consciousness is hard, AI focuses on proxy tasks (chess)
 - Implicit assumption: models **must be** conscious to solve them
- However, in doing so, moved to objective multifaceted measurements
 - You can be 90% accurate in chess moves
 - You can be terrible at chess but still be conscious
 - Turns out you can be great at chess but not very conscious
- So AI moves to another task

Chess



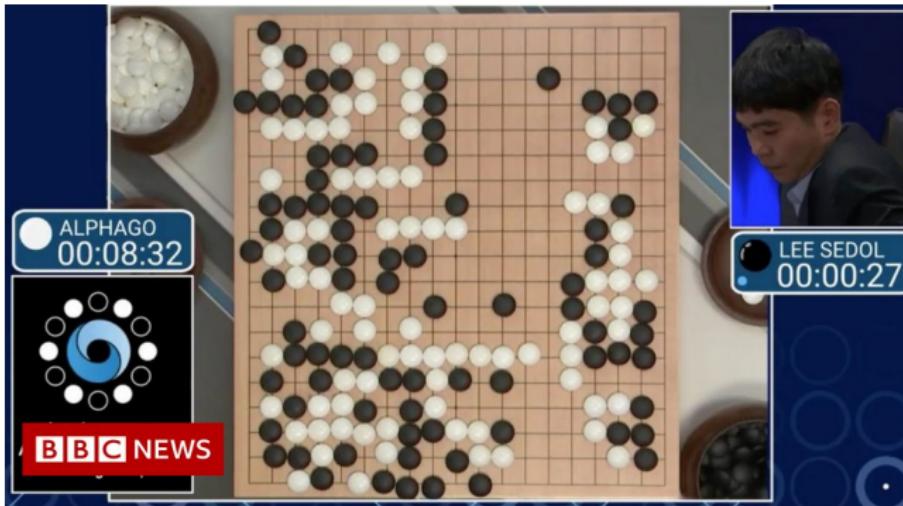
- Magnus Carlsen ELO: 2880

Chess



- Magnus Carlsen ELO: 2880
- Stockfish ELO: 3546

Go



- Watch the *AlphaGo* movie (2017)

Now it's happening with ChatGPT



<https://www.youtube.com/watch?v=8oxHS5qemAE>

Is ChatGPT “really” understanding?



<https://www.youtube.com/watch?v=8oxHS5qemAE>

Is ChatGPT “really” understanding?



Is ChatGPT “really” understanding?



<https://www.youtube.com/watch?v=8oxHS5qemAE>

Is ChatGPT “really” understanding?



<https://www.youtube.com/watch?v=8oxHS5qemAE>

AI Research is at an Impasse

- Decades of **goalposts moving** improve AI functional understanding

AI Research is at an Impasse

- Decades of **goalposts moving** improve AI functional understanding
 - E.g., all of the tasks we discussed in the course

AI Research is at an Impasse

- Decades of **goalposts moving** improve AI functional understanding
 - E.g., all of the tasks we discussed in the course
- Very little progress on “**True Intelligence**”

AI Research is at an Impasse

- Decades of **goalposts moving** improve AI functional understanding
 - E.g., all of the tasks we discussed in the course
- Very little progress on “**True Intelligence**”
 - Understanding what it constitutes

AI Research is at an Impasse

- Decades of **goalposts moving** improve AI functional understanding
 - E.g., all of the tasks we discussed in the course
- Very little progress on “**True Intelligence**”
 - Understanding what it constitutes
 - How to measure it

AI Research is at an Impasse

- Decades of **goalposts moving** improve AI functional understanding
 - E.g., all of the tasks we discussed in the course
- Very little progress on “**True Intelligence**”
 - Understanding what it constitutes
 - How to measure it
 - How to build machines which posses it

AI Research is at an Impasse

- Decades of **goalposts moving** improve AI functional understanding
 - E.g., all of the tasks we discussed in the course
- Very little progress on “**True Intelligence**”
 - Understanding what it constitutes
 - How to measure it
 - How to build machines which posses it

Turing (1950)

Not until a machine can write a sonnet or compose a concerto because of **thoughts and emotions felt**, and not by the chance fall of symbols, could we agree that machine equals brain—that is, not only write it but **know that it had written it**.

Outline

- 1 For foundational AI, Intelligence = Having consciousness
- 2 Moving the Goalposts
- 3 Moving Forward
- 4 Conclusions

Disentangling consciousness & functional understanding

- Revisit the original motivation for AI

Disentangling consciousness & functional understanding

- Revisit the original motivation for AI
- Explicitly tackle the question of conscious experience in machines

Disentangling consciousness & functional understanding

- Revisit the original motivation for AI
- Explicitly tackle the question of conscious experience in machines
 - Leveraging multi-disciplinary research

Disentangling consciousness & functional understanding

- Revisit the original motivation for AI
- Explicitly tackle the question of conscious experience in machines
 - Leveraging multi-disciplinary research
- No proxy (coding, driving, writing poetry) would entail consciousness

Biological Markers of Consciousness

- Theories inferring consciousness from the brain w/o subjective reports

Biological Markers of Consciousness

- Theories inferring consciousness from the brain w/o subjective reports
 - Via temporal and spatial dynamics unique to conscious

Biological Markers of Consciousness

- Theories inferring consciousness from the brain w/o subjective reports
 - Via temporal and spatial dynamics unique to conscious
- Consciousness associated with rapid broadcast to large brain areas

Biological Markers of Consciousness

- Theories inferring consciousness from the brain w/o subjective reports
 - Via temporal and spatial dynamics unique to conscious
- Consciousness associated with rapid broadcast to large brain areas
- Find analogs to spatio-temporal propagation in ANN

Biological Markers of Consciousness

- Theories inferring consciousness from the brain w/o subjective reports
 - Via temporal and spatial dynamics unique to conscious
- Consciousness associated with rapid broadcast to large brain areas
- Find analogs to spatio-temporal propagation in ANN
 - E.g., spiking neural networks

The Function of Consciousness

- Vast research on the scope of nonconscious human cognition

The Function of Consciousness

- Vast research on the scope of nonconscious human cognition
 - E.g., Fast vs. slow thinking

The Function of Consciousness

- Vast research on the scope of nonconscious human cognition
 - E.g., Fast vs. slow thinking
- NLP can contribute to this research

The Function of Consciousness

- Vast research on the scope of nonconscious human cognition
 - E.g., Fast vs. slow thinking
- NLP can contribute to this research
 - By examining what can be addressed by statistical models

The Function of Consciousness

- Vast research on the scope of nonconscious human cognition
 - E.g., Fast vs. slow thinking
- NLP can contribute to this research
 - By examining what can be addressed by statistical models
- ... and benefit from it

The Function of Consciousness

- Vast research on the scope of nonconscious human cognition
 - E.g., Fast vs. slow thinking
- NLP can contribute to this research
 - By examining what can be addressed by statistical models
- ... and benefit from it
 - By focusing on tasks which require consciousness

Outline

- 1 For foundational AI, Intelligence = Having consciousness
- 2 Moving the Goalposts
- 3 Moving Forward
- 4 Conclusions

- AI research describes Intelligence as having **conscious experience**

- AI research describes Intelligence as having **conscious experience**
 - Usually without saying so explicitly

- AI research describes Intelligence as having **conscious experience**
 - Usually without saying so explicitly
- However, **objectively** measuring consciousness is famously hard

- AI research describes Intelligence as having **conscious experience**
 - Usually without saying so explicitly
- However, **objectively** measuring consciousness is famously hard
 - **The mind-body problem™** — frustrating research for centuries

- AI research describes Intelligence as having **conscious experience**
 - Usually without saying so explicitly
- However, **objectively** measuring consciousness is famously hard
 - **The mind-body problem™** — frustrating research for centuries
- Instead, AI focuses an **objective** intelligence benchmark (e.g., chess)

- AI research describes Intelligence as having **conscious experience**
 - Usually without saying so explicitly
- However, **objectively** measuring consciousness is famously hard
 - **The mind-body problem™** — frustrating research for centuries
- Instead, AI focuses an **objective** intelligence benchmark (e.g., chess)
 - Implicitly assuming models **must be** conscious to solve it

- AI research describes Intelligence as having **conscious experience**
 - Usually without saying so explicitly
- However, **objectively** measuring consciousness is famously hard
 - **The mind-body problem™** — frustrating research for centuries
- Instead, AI focuses an **objective** intelligence benchmark (e.g., chess)
 - Implicitly assuming models **must be** conscious to solve it
 - When ML models “solve” the task, a new one is chosen

- AI research describes Intelligence as having **conscious experience**
 - Usually without saying so explicitly
- However, **objectively** measuring consciousness is famously hard
 - **The mind-body problem™** — frustrating research for centuries
- Instead, AI focuses an **objective** intelligence benchmark (e.g., chess)
 - Implicitly assuming models **must be** conscious to solve it
 - When ML models “solve” the task, a new one is chosen
- **This isn't productive**

- AI research describes Intelligence as having **conscious experience**
 - Usually without saying so explicitly
- However, **objectively** measuring consciousness is famously hard
 - **The mind-body problem™** — frustrating research for centuries
- Instead, AI focuses an **objective** intelligence benchmark (e.g., chess)
 - Implicitly assuming models **must be** conscious to solve it
 - When ML models “solve” the task, a new one is chosen
- **This isn't productive**
 - It's not informing us about what constitutes intelligence

- AI research describes Intelligence as having **conscious experience**
 - Usually without saying so explicitly
- However, **objectively** measuring consciousness is famously hard
 - **The mind-body problem™** — frustrating research for centuries
- Instead, AI focuses an **objective** intelligence benchmark (e.g., chess)
 - Implicitly assuming models **must be** conscious to solve it
 - When ML models “solve” the task, a new one is chosen
- **This isn't productive**
 - It's not informing us about what constitutes intelligence
 - Or how to build “real” intelligent systems

- AI research describes Intelligence as having **conscious experience**
 - Usually without saying so explicitly
- However, **objectively** measuring consciousness is famously hard
 - **The mind-body problem™** — frustrating research for centuries
- Instead, AI focuses an **objective** intelligence benchmark (e.g., chess)
 - Implicitly assuming models **must be** conscious to solve it
 - When ML models “solve” the task, a new one is chosen
- **This isn't productive**
 - It's not informing us about what constitutes intelligence
 - Or how to build “real” intelligent systems
- **Multidisciplinary research is needed to make progress**