Muhammad Houzair Koussa

Software engineer @ Stratalis
Maintainer @ thephilosophicalcode.com
Member @ frontend.mu

thephilosophicalcode.com

Open-source article series exploring the intersection of philosophy & code

Repo: github.com/houzyk/thephilosophicalcode

frontend.mu

Mauritian tech community based around front-end development

Monthly meet-ups, networking and open-source opportunities

Repo: github.com/Front-End-Coders-Mauritius

Thinking about thinking machines

Philosophically analyse the question “can machines think?”

Agenda

Not gonna answer the question “can machines think?”

- It’s a loaded question with ambiguous terms
- It’s easy to talk pass each other

So, let’s focus on clearing some of that ambiguity

Particularly, we’ll:

1. Deconstruct the question
2. Build a “framework” for understanding our question

Agenda

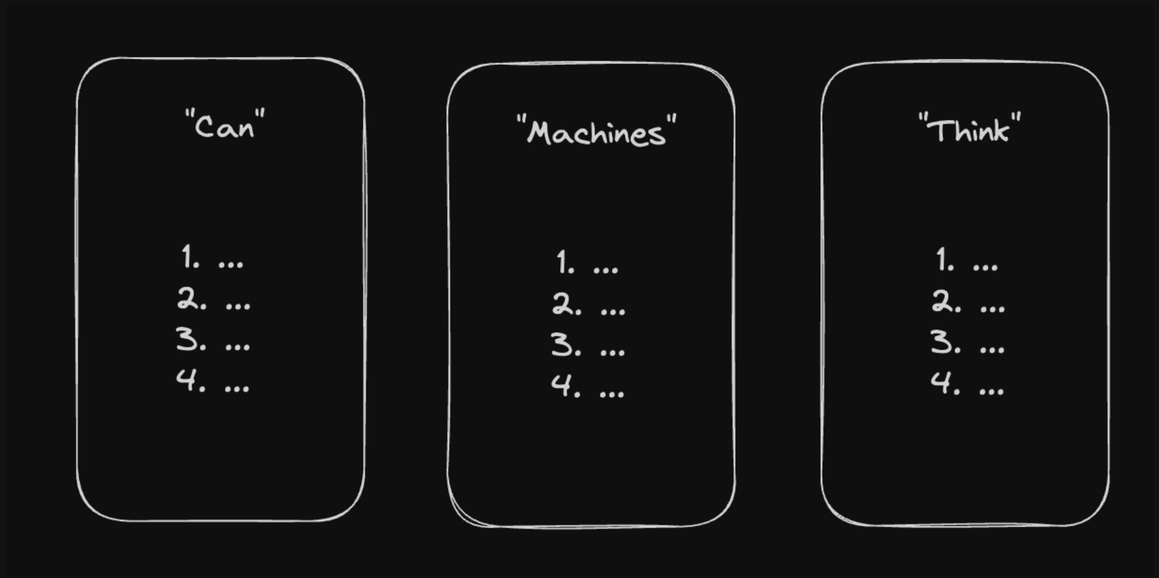
1. Context - deconstruction method & framework
2. Deconstruction
3. Wrap up
4. Q & A

1. Deconstruction method & framework

We have 3 ambiguous terms - “can”, “machines” & “think”

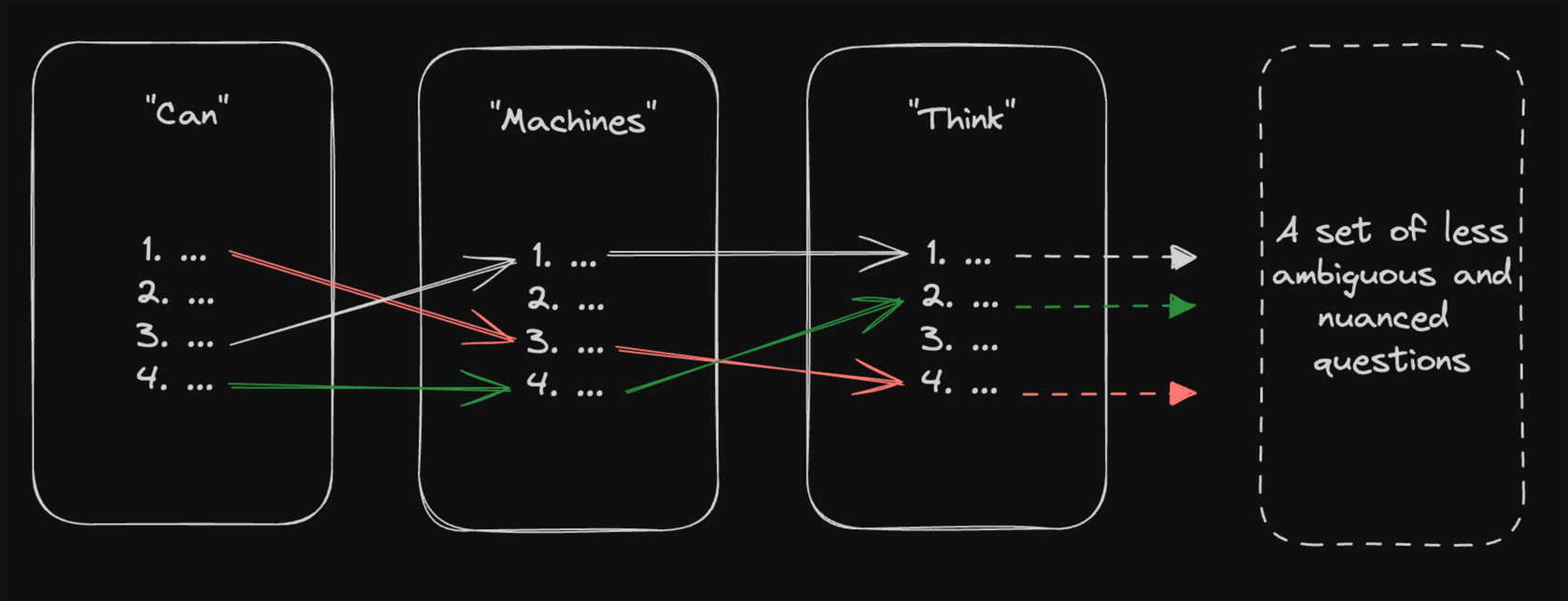
For deconstruction, we’ll:

1. Analyse each term as a concept in itself
2. Come up with a set of less ambiguous phrasings for each term



1. Deconstruction method & framework

With these phrasings, we can build a framework through permutations



All permutations = a set of less ambiguous and nuanced questions

Easier to further analyse, contextualise & (hopefully) answer our original question “can machines think?”

2. Deconstruction - “can”

Intuitively, “can” is about possibilities

In the context of thinking machines, we may see these possibilities as:

A. Existence

B. Knowledge

C. Ability

D. Ethics (not really but it matters)

2. Deconstruction - “can” as existence

“Can” as existence = the possible existence of thinking machines

Two ways to understand “possible existence”:

A. In some possible world quite distinct from our own

So, we’re asking “is it possible that machines think?”

B. In our actual world. It’s about asking if thinking machines may exist given our current laws of nature and history

So, we’re asking “is it actually possible that machines think?”

2. Deconstruction - “can” as knowledge

The (possible) existence of X, does not entail knowledge *about and of* X (such as X’s inner workings, properties & relations)

Even if thinking machines exist, that does not entail that we can know anything about them

“Can” as knowledge = the possibility to know anything about thinking machines

Particularly, we’re asking “can we know if machines think?

Big picture - thinking machines as a solvable epistemic/scientific question

2. Deconstruction - “can” as ability

Knowledge about and of X, does not entail the immediate ability to build X

“Can” as ability = the possibility to build thinking machines

Particularly, we’re asking “can we build machines that think?

Big picture - thinking machines as a solvable engineering problem

2. Deconstruction - “can” as ethics

It’s common philosophical practice to separate “can” & “should”

If something can happen, this doesn’t mean that it should happen

However, it’s not common practice to separate the “can” and the “should”

Talk about the *possibility* of thinking machines is often mixed with talk about whether we *should* build them

So, it’s fair to deconstruct “can” as ethical & ask “should we build machines that think?”

2. Deconstruction - “can”

“Can” as	Less ambiguous phrasing
Existence in some possible world	Is it possible that machines think?
Existence in the actual world	Is it actually possible that machines think?
Knowledge	Can we know if machines think?
Ability	Can we build machines that think?
Ethics	Should we build machines that think?

2. Deconstruction - “machines”

In the context of thinking machines, we may deconstruct “machines” as:

A. Computers

B. Networks

C. Non-natural objects

2. Deconstruction - “machines” as computers

Computers in their most general sense - Turing Machines

So, we’re asking “can Turing Machines think?”

There’s a general idea that Turing Machines only manipulate symbols

Symbolic manipulation may not imply thought - The Chinese Room Argument

Imagine a room with a guy and a Chinese dictionary. He does not understand Chinese

We feed in Chinese text through the door. He gives back perfect English translations

2. Deconstruction - “machines” as networks

Connectivism - just like the brain, a thinking machine would be a network instead of a single machine

So, our question becomes “can networks think?”

Not everyone subscribes to the idea of “just like the brain”

It’s not necessary that thought arises *only* out of structures that mimic natural brains

2. Deconstruction - “machines” as non-natural objects

Fallback - deconstruct “machines” as just any object

If materialism is true, we are thinking machines - that’s a bit trivial

(Materialism - reality (especially “the soul”) is wholly derivable from the physical)

So, it feels more natural to think of these objects as non-natural

However, “non-naturalness” is a philosophical can of worms

For lack of better terms, we’ll simply deconstruct into “can non-natural objects think?”

2. Deconstruction - “machines”

“Machines” as	Less ambiguous phrasing
Computers	Can Turing Machines think?
Networks	Can networks think?
Non-natural objects	Can non-natural objects think?

2. Deconstruction - “think”

The concept of thought is highly ambiguous on its own

Consider “inward thought” and “outward thought”. It’s not easy to distinguish between the two

We see others as having “thought” because we see them behaving as such

I think that you think because you act as if you think like I think!

So, we’ll implicitly understand “thought” as both inward and outward

2. Deconstruction - “think”

Despite of our shortcomings, we’ll deconstructing “think” as:

A. Rationality

B. Natural mental states

C. Consciousness

D. Non-anthropomorphic thought

…

2. Deconstruction - “think” as rationality

A model of rationality - ability in reaching one’s goal effectively

Rationality does not necessarily imply mental states

Not all intelligent actors have mental states

Not all thinking beings act rationally

So, we may deconstruct our question as “can machines be rational?” or “can machines act rationally?”

2. Deconstruction - “think” as natural mental states

Another can of worms - natural mental states are highly diverse

Natural mental states may include:

A. Emotion

B. Perception

C. Reaction

…

We should bear that diversity in mind as we deconstruct our question into “can machines have mental states?” or “can machines act as having mental states?”

2. Deconstruction - “think” as consciousness

Natural mental states do not necessarily entail consciousness - The Philosophical Zombie

Imagine someone who acts, lives and breaths exactly like us but is all blank *inside*

Intuitively, they would not posses that deeply personal thing which we call phenomenological self-awareness (the quintessential human experience)

2. Deconstruction - “think” as non-anthropomorphic thought

Anthropomorphic value judgement - the attribution of human values onto the non-human

It’s important to recognise that saying that *consciousness is special* is potentially an anthropomorphic value judgement

Non-anthropomorphic thought - it’s possible that there is a form of thought unlike ours

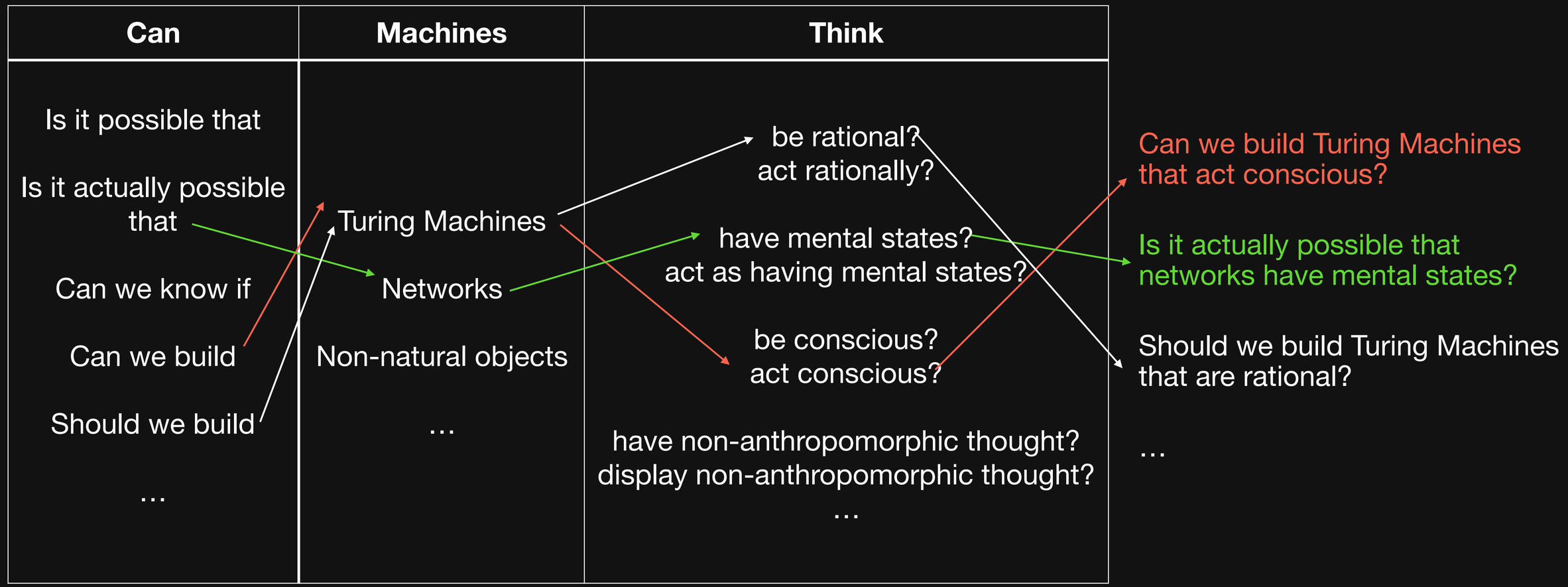
That form of thought is not necessarily weaker or more powerful

2. Deconstruction - “think”

“Think” as	Not so less ambiguous phrasing
Rationality	Can machines be rational? Can machines act rationally?
Natural mental states	Can machines have mental states? Can machines act as having mental states?
Consciousness	Can machines be conscious? Can machines act conscious?
Non-anthropomorphic thought	Can machines have non-anthropomorphic thought? Can machines display non-anthropomorphic thought?

3. Wrap up

At least 120 less ambiguous and nuanced questions:



3. Wrap up

We have more questions than we started with and no answers

But we can now further analyse, contextualise and (hopefully) find answers

- Which questions are valid?
- What kind of methods should we use in answering them?
- Are they just yes/no questions?
- What’s the relationship among and between them like?

- Do these questions show us any kind of limitations/quirks?

Thank you!

Slides - github.com/houzyk/talks

thephilosophicalcode.com
frontend.mu