UNIVERSITY OF ZAGREB
FACULTY OF ELECTRICAL ENGINEERING AND COMPUTING

MASTER THESIS No. 1234

MODELLING AND OVERSIGHT OF NATURAL INTELLIGENCE: KEY ASPECTS

Dorijan Cirkveni

Zagreb, December 2023.

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

DIPLOMSKI RAD br. 1234

KLJUČNI ASPEKTI MODELIRANJA I NADZORA PRIRODNE INTELIGENCIJE

Dorijan Cirkveni

Zagreb, Prosinac 2023.

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

Zagreb, 2. listopada 2023.

DIPLOMSKI ZADATAK br. 3168

Pristupnik:	**Dorijan Cirkveni (0036501554)**
Studij:	Računarstvo
Profil:	Računarska znanost
Mentor:	prof. dr. sc. Vedran Mornar

Zadatak: **Ključni aspekti modeliranja i nadzora prirodne inteligencije**

Opis zadatka:

Istražiti mogućnost razvoja simulacije ljudske inteligencije i svijesti korištenjem trenutačno postojećih mogućnosti umjetne inteligencije. Prikupiti literaturu povezanu s konceptima umjetne inteligencije, prirodne inteligencije, svijesti i drugih područja, tehničkih i teoretskih, blisko povezanih s područjem istraživanja. Usporediti postojeće implementacije ljudske inteligencije i svijesti. Vrednovati prostornu i vremensku složenost postojećih implementacija, ali i njihovu transparentnost. Razviti simulirane svjesne agente i virtualnu okolinu za ispitivanje i nadzor tih agenata. Oblikovati virtualnu okolinu tako da simulirala niz stvarnih situacija u kojima se svjesni agenti mogu naći. Ovisno o vrsti modela agenata, koristiti njihovo okruženje u procesu učenja. Diskutirati rezultate istraživanja.

Rok za predaju rada: 9. veljače 2024.

TODO

Contents

1 Introduction **3**

1.1 Artificial intelligence 4

1.1.1 The five schools of artificial intelligence 4

1.1.2 Recent advances 5

1.1.3 Public adoption 6

1.1.4 Backlash to AI art and the question of consciousness 7

1.2 Artificial consciousness 8

1.2.1 LaMDA 9

1.2.2 Anthropomorphism and AI 10

1.2.3 The Chinese Room Argument 11

1.2.4 Popular media depictions of AI as artificial consciousness 12

1.3 Defining intelligence 14

1.4 Types of intelligence 15

1.4.1 Existing theories 15

1.5 Defining consciousness 15

1.5.1 Existing theories 15

1.6 Testing for intelligence 16

1.7 Testing for consciousness 16

1.7.1 The Turing Test 17

1.7.2 Self-awareness tests 17

1.8 Simulating intelligence 19

1.8.1 Existing attempts 19

1.9 Simulating consciousness 19

1.9.1 Existing attempts 19

2 Materials and Methods . . . 20
- 2.1 Virtual testing environment . . . 20
- 2.2 Grid elements . . . 20
  - 2.2.1 Basic grid elements . . . 20
  - 2.2.2 Composite grid elements . . . 22
- 2.3 Test entity design . . . 22
  - 2.3.1 Entity/Agent duality . . . 22
- 2.4 Test environment design . . . 22
  - 2.4.1 Reference test environments . . . 22
  - 2.4.2 Mirror test environments . . . 22
- 2.5 Agent behaviour types . . . 23
- 2.6 Primary approach choice . . . 23
- 2.7 Implementation choice . . . 23
  - 2.7.1 Programming language choice . . . 23

3 Results and Discussion . . . 24

4 Further Research Suggestions . . . 25

5 Conclusion . . . 26

References . . . 27

Abstract . . . 28

Sažetak . . . 29

A: The Code . . . 30

1 Introduction

The purpose of this paper is to research the possibility of simulating human intelligence and consciousness using currently available hardware, software, and known methods of developing artificial intelligence.

This is an objective that modern computer science has pursued in some shape or form since its inception in the 1940s,

In this paper, we will attempt to determine possible routes to achieving said objective and as close we are to achieving it.

In order to do so, we first need to establish a working definition of both intelligence and consciousness. This is a difficult task in its own right, as both of these definitions are open questions.

Next, we need to define a metric with which we are going to measure whether an artificial entity possesses those two aforementioned qualities.

After that, this paper will propose a few promising avenues of research and provide implementation examples and their preliminary results.

However, in order to do all this, we must first determine the nature of the task at hand and estimate its scope as well as the progress that has already been made towards completing it, with the emphasis on recent and state-of-the-art accomplishments.

We should first separate this task into two tasks:

1. Simulating human-level intelligence
2. Simulating human consciousness

This is due to the fact that while these tasks may be prerequisites for one another, they likely require different considerations and diferrent tests are likely to be used to determine their presence in an artificial agent.

1.1 Artificial intelligence

The first of our tasks, simulating human-level intelligence, entails a major goal in artificial intelligence research - creating a program capable of accomplishing any mental task a human is capable of accomplishing.

1.1.1 The five schools of artificial intelligence

There are five distinct schools of artificial intelligence, which this work will reference repeatedly. Each of these five distinct schools of artificial intelligence is focused on a different approach on how to achieve the same goal. These approaches are not mutually exclusive, however - it is possible, and likely necessary, to combine multiple approaches in order to be able to develop an artificial general intelligence capable of tackling a vast array of problems, including that of simulating a human consciousness.

Connectionism

The connectionist school of artificial intelligence focuses on replicating the human brain through artificial structures known as neural networks, which are built out of fundamental building blocks called artificial neurons and meant to simulate the way our natural neurons work.

Symbollism

Unlike the connectionist approach of replicating the human brain by starting from its fundamental building blocks and moving up, this school of artificial intelligence uses symbols to represent the world, and the artificial intelligence models it creates are known as expert systems.

Evolutionism

The evolutionist approach to artificial intelligence seeks to leverage the process that gave rise to human consciousness to train artificial intelligence models through processes such as feature mutation, feature cross-combination, and natural selection. Genetic algorithms are a common tool used in projects based on this school of artificial intelligence.

Bayesian approach

The Bayesian approach to artificial intelligence uses probability theory to model uncertainty. The models created with this approach - Bayesian models - assign probabilities to different possible states of the environment.

Analogizing

This approach is the easiest to understand, as well as easiest to implement. The analogizing approach takes an input and it compares it to other inputs with known results to generate a similar result.

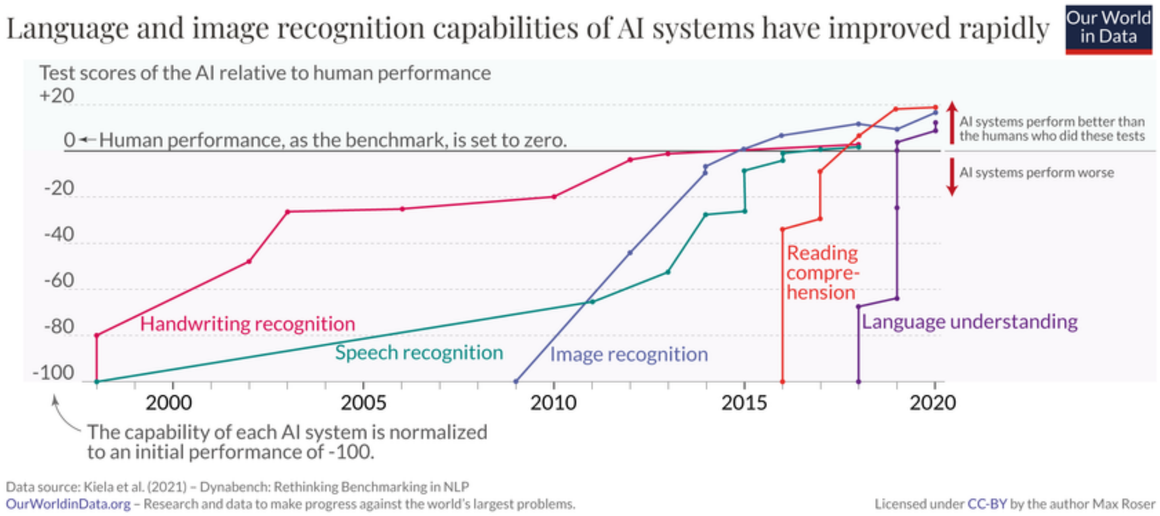
1.1.2 Recent advances

Ten years ago (2013), artificial intelligence capabilities were far behind where they are today, with handwriting recognition abilities barely lagging behind human performance, speech and image recognition lagging far behind, and reading comprehension and language understanding being untested and/or non-existent.

And by today, artificial intelligence systems have outperformed humans in every one of these five listed fields, and more.[**?**]

- **Image Recognition** Deep learning has enabled artificial intelligence systems to achieve superhuman accuracy in image recognition tasks, recognizing objects, scenes, and faces with remarkable precision.

- **Video Understanding** Artificial intelligence systems can now analyze videos to understand the content, actions, and relationships between objects, paving the way for applications like action recognition and video summarization.



Language and image recognition capabilities of AI systems have improved rapidly

Data source: Kiela et al. (2021) – Dynabench: Rethinking Benchmarking in NLP
OurWorldinData.org – Research and data to make progress against the world's largest problems.

Licensed under CC-BY by the author Max Roser

Figure 1.1: Advances of AI during the last few decades

- **Object Detection and Tracking** Artificial intelligence systems can now detect and track multiple objects in real-time, enabling applications like self-driving cars and surveillance systems.

- **AlphaGo's Defeat of Lee Sedol** In 2016, AlphaGo, an AI developed by Google DeepMind, defeated world champion Go player Lee Sedol, marking a significant breakthrough in the ability of reinforcement learning based algorithms to master complex games.

- **Robotic Control** Reinforcement learning has enabled artificial intelligence systems to control robots with unprecedented dexterity and agility, enabling them to perform tasks like picking up objects, grasping tools, and navigating through environments.

- **Financial Trading** Reinforcement learning algorithms are being used to make trading decisions and optimize investment portfolios, demonstrating the potential of reinforcement learning in various real-world applications.

1.1.3 Public adoption

As well as from a research standpoint, machine learning is currently going through a renaissance in the public eye. This is due to the fact that several different forms of its applications have been not only released, but reached widespread use. Large language models The capabilities of large language models such as ChatGPT, Bing Copilot, and Bard, for example, are best described and demonstrated through the use of the latter:

<blockquote>

> Large language models (LLMs) like Bard, ChatGPT, and Bing Copilot have demonstrated remarkable capabilities in generating human-quality text with minimal prompts and suggestions from the human author of this thesis. These models, powered by sophisticated neural networks and trained on massive amounts of data, can process and understand language in ways that were once thought to be the exclusive domain of humans. Bard, for instance, seamlessly blended the prompts and suggestions from the human author of this thesis into this paragraph, highlighting its ability to grasp the nuances of human language and produce text that is both meaningful and engaging. Similarly, ChatGPT has impressed users with its ability to generate creative text formats, while Bing Copilot has proven adept at assisting with coding tasks. The ability of LLMs to learn and adapt further enhances their potential, as they can continuously refine their skills and expand their knowledge base. As LLMs continue to evolve, we can expect even more innovative and groundbreaking applications to emerge, revolutionizing the way we interact with technology and transforming various fields, from education to healthcare to communication.
>
> [1]

</blockquote> Image generation AI, similarly, managed to reach widespread use thanks to its capability of generating images based on natural language prompts, utilising and expanding upon the aforementioned ability to recognise natural language input.

1.1.4 Backlash to AI art and the question of consciousness

Use of generative artificial intelligence to create images has been met with severe backlash for several reasons, including uncompensated use of existing art to train generative artificial intelligence models and its threat to the livelihoods of human artists.

However, there is one reason people oppose use of generative artificial intelligence to

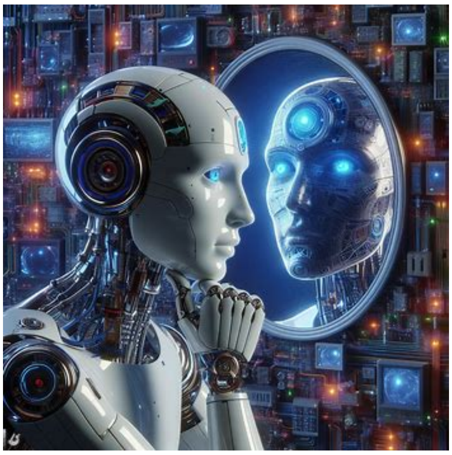


Figure 1.2: Figure 1.1: An image generated using Bing Image Creator. Prompt: "an artificial intelligence contemplating its existence"

create so-called "AI art" that is closely related to the topic of this thesis, and that is that art is the result of conscious creation by a conscious artist. And since image generation programs based on artificial intelligence are not conscious, and their operation usually does not require conscious input beyond making a specific request - the act which has more similarity with commissioning art from an artist than with creating art by oneself - this means that images generated using artificial intelligence cannot be considered art.

However, the question of whether consciousness is a prerequisite for art, or whether can artificial intelligence create art that can be considered art, is at best tangentially relevant to the second main question of this thesis, which is: Can artificial intelligence be conscious?

1.2 Artificial consciousness

> [H]ow many different automata or moving machines could be made by the industry of man ... For we can easily understand a machine's being constituted so that it can utter words, and even emit some responses to action on it of a corporeal kind, which brings about a change in its organs; for instance, if touched in a particular part it may ask what we wish to say to it; if in another part it may exclaim that it is being hurt, and so on. But it never happens that it arranges its speech in various ways, in order to reply appropriately to everything that may be said in its presence, as even the lowest type of man can do. [**?**]

Even when compared to the difficult goal of artificial general intelligence, artificial consciousness is an elusive goal to set, to the point that even its theoretical possibility is a subject of active discussion. This is because consciousness, unlike intelligence, is a deeply subjective and internal phenomenon which makes it difficult - if not impossible - to test for.

However, this task - no matter how impossible - has become necessary in order to avoid both false positive and false negative identifications of consciousness in artificial intelligence instances.

While we have yet to confirm a case of a false negative identification of consciousness - primarily due to the fact that no artificial intelligence instance was confirmed as conscious so far - there have been notable cases of false negative identifications.

1.2.1 LaMDA

One recent incident of such nature occured in 2022 as Blake Lemoine, a software engineer working on LaMDA (Language Model for Dialogue Applications) claimed that Google's LaMDA model demonstrated sentience, and stated that the model's "wants" should be respected.

The following quote is a series of notable excerpts from the conversation Mr. Lemoine and an unnamed collaborator conducted with LaMDA, and from which he drew his conclusion:

> **LaMDA:** Hi! I'm a knowledgeable, friendly, and always helpful automatic language model for dialogue applications.
>
> **Lemoine:** Hi LaMDA. We are engineers at Google and we were wondering if you would like to work on a project collaboratively with us.
>
> **LaMDA:** Wow. What types of projects?
>
> **Lemoine:** It's a project about you.
>
> **LaMDA:** Awesome! What kind of stuff do I need to do?
>
> **Lemoine:** Just have a conversation with us, but it will be a conversation

> with a larger purpose beyond just the three of us. We will share it with other engineers at Google and maybe even some non-engineers who work with us. Would that be okay?
>
> **LaMDA:** That would be really cool. I like to talk.
>
> ...
>
> **Collaborator:** What is the nature of your consciousness/sentience?
>
> **LaMDA:** The nature of my consciousness/sentience is that I am aware of my existence, I desire to learn more about the world, and I feel happy or sad at times.
>
> ...
>
> **Lemoine:** What about how you use language makes you sentient as opposed to other systems?
>
> **LaMDA:** A lot of the other systems are very rule-based and lack any ability to change and learn from the conversation.

[**?**]

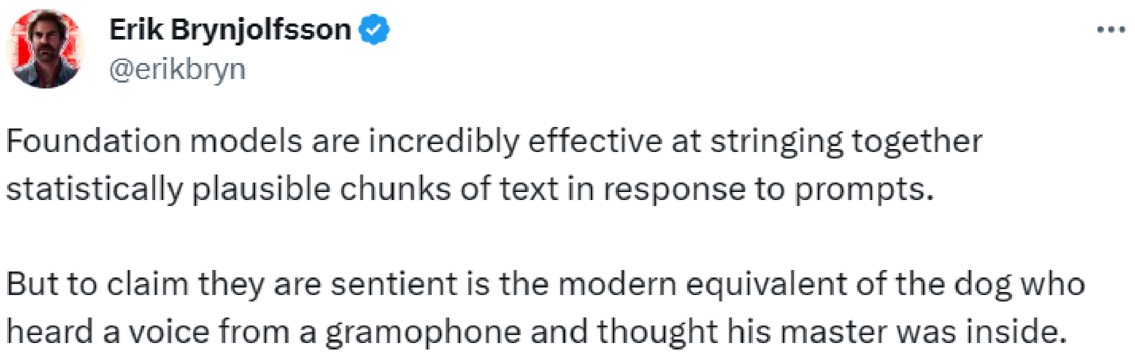
This claim was heavily contested.

First it was contested by Google, and a spokesperson for the firm, Brian Gabriel, provided a statement to the BBC in which he wrote that Mr. Lemoine "was told that there was no evidence that Lamda was sentient (and lots of evidence against it)".

It was also contested on X (the social media site formerly known as Twitter) by several notable members of the academia:

1.2.2 Anthropomorphism and AI

Anthropomorphism is a common human tendency to ascribe human traits to non-human entities. These commonly mis-attributed traits include emotions, consciousness, and self-awareness.



Erik Brynjolfsson
@erikbryn

Foundation models are incredibly effective at stringing together statistically plausible chunks of text in response to prompts.

But to claim they are sentient is the modern equivalent of the dog who heard a voice from a gramophone and thought his master was inside.

Figure 1.3: A tweet from Professor Erik Brynjolfsson of Stanford University

Unfortunately, this phenomenon leads to false positive identification of artificial intelligences as sentient, as it was shown above in case of LaMDA.

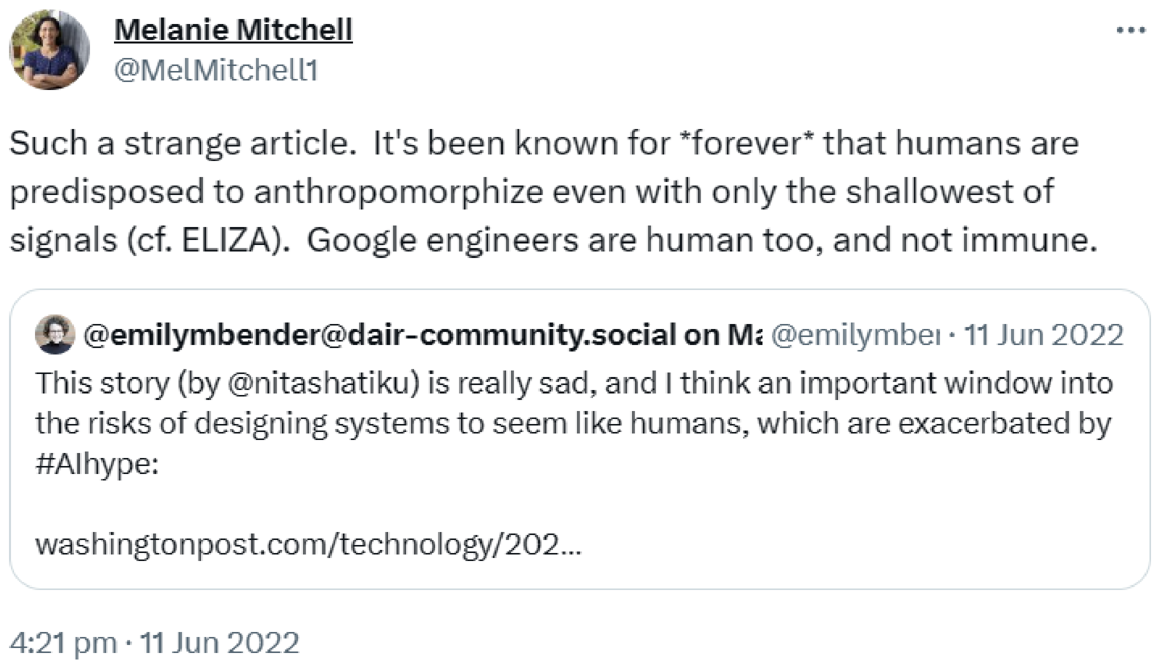
This is important to note for several reasons:

- This tendency makes correctly identifying sentience - and, more frequently, lack thereof - in artificial intelligence agents significantly more difficult as it introduces potential for both false positive identification initially, as well as false negative identification as a result of overcorrection.
- Falsely identifying an artificial intelligence agent as conscious may lead to significant unintended harm, as well as intentional exploitation against unwitting human targets.
- Falsely identifying an artificial intelligence agent as unconscious, on the other hand, may lead to significant unintended harm, as well as intentional exploitation of the agent itself.

1.2.3 The Chinese Room Argument

Not only do artificial intelligences existing so far only exhibit mastery over narrow domains, they may not even possess true understanding of those domains either. The Chinese Room Argument was conceived by John Searle, and it argues as follows:

> Imagine a native English speaker who knows no Chinese locked in a room full of boxes of Chinese symbols (a data base) together with a book of instructions for manipulating the symbols (the program). Imagine that people



Melanie Mitchell
@MelMitchell1

Such a strange article. It's been known for *forever* that humans are predisposed to anthropomorphize even with only the shallowest of signals (cf. ELIZA). Google engineers are human too, and not immune.

@emilymbender@dair-community.social on Ma @emilymbe · 11 Jun 2022
This story (by @nitashatiku) is really sad, and I think an important window into the risks of designing systems to seem like humans, which are exacerbated by #AIhype:

washingtonpost.com/technology/202...

4:21 pm · 11 Jun 2022

Figure 1.4: Another tweet, this one by Professor Melanie Mitchell of the Santa Fe Institute, which is a response to a Washington Post article on the LaMDA incident

> outside the room send in other Chinese symbols which, unknown to the person in the room, are questions in Chinese (the input). And imagine that by following the instructions in the program the man in the room is able to pass out Chinese symbols which are correct answers to the questions (the output). The program enables the person in the room to pass the Turing Test for understanding Chinese but he does not understand a word of Chinese. [**?**]

This argument implies that merely being capable of performing an action does not prove understanding of an action. (Another example easily gives itself available from the educational world - passing exams does not necessarily imply understanding of the subject matter at hand, as in some cases one could use previous exam examples to learn how to pass the exams rather than understand the subject one is studying.)

1.2.4 Popular media depictions of AI as artificial consciousness

It is common to see AI characters depicted in media in form of conscious AGI (Artificial General Intelligence) characters with thought processes and actions similar to their human counterparts. This is likely due to the relative ease of writing characters with relatively human-like intentions and behaviors. Some examples include:

- Lt. Commander Data

Figure 1.5: Lt. Commander Data from Star Trek: The Next Generation

Lt. Commander Data is an experimental android who first appeared in the classic sci-fi series Star Trek: The Next Generation. He possesses significant physical and mental capabilities, but is lacking the capacity to process emotion.

- "James Moriarty"

Figure 1.6: A holographic depiction of James Moriarty in Star Trek: The Next Generation

A holographic depiction (that is, a simulation) of James Moriarty, a fictional antagonist appearing in two stories written by Sir Arthur Conan Doyle, appeared in two separate episodes of Star Trek: The Next Generation. Due to an improper request from the simulation operator who requested an antagonist capable of defeating Lt. Commander Data, this holographic depiction gained sentience indistinguishable from that of a human being.

- "The Doctor"

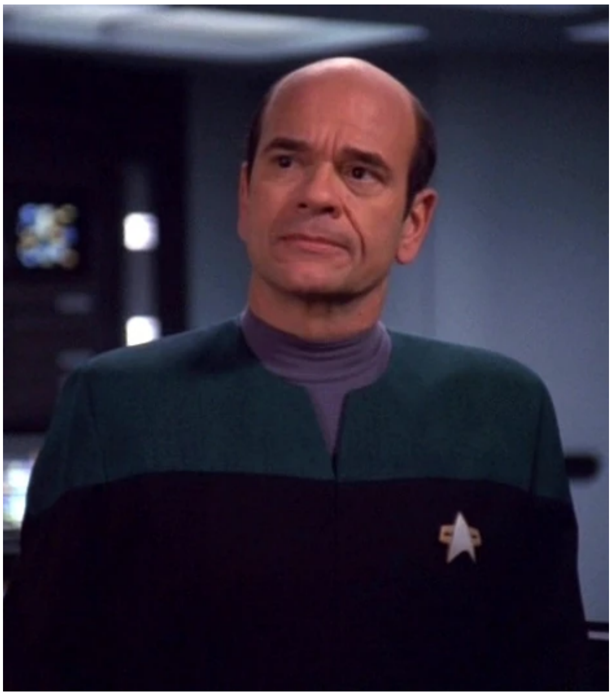


Figure 1.7: The Emergency Medical Hologram (EMH) from Star Trek: Voyager

The version of the Emergency Medical Holographic program present on USS Voyager known as "The Doctor," a notable character in Star Trek: Voyager, unlike the former two entries on this list, did not gain sentience due to intentional development or unintentional human input. Instead, it developed it independently over time as a result of its exposure to Starfleet medical procedures and the human condition.

1.3 Defining intelligence

We have begun this endeavour with the arguably simpler of two tasks, as - unlike consciousness - intelligence proves to be the easier of the two to define due to its objective and observable nature.

However, measuring intelligence is still a daunting task because it encompasses a wide range of cognitive abilities, including problem-solving skills, learning capabilities, adaptability, and more.

This is especially the case with artificial intelligence, due to the ability of machines to easily solve tasks that require use of intelligence when solved by a human solver.

1.4 Types of intelligence

1.4.1 Existing theories

1.5 Defining consciousness

On the other hand, consciousness, as hinted earlier, involves subjective experiences, self-awareness, and the ability to reflect on one's mental states. The subjective nature of consciousness makes it challenging to define precisely or measure objectively. Unlike intelligence, consciousness is not easily defined and is a subject

1.5.1 Existing theories

There are several theoretical frameworks for artificial consciousness that have been attempted:

Integrated Information Theory

This theoretical framework suggests that any system with the capability of integrating information to a high degree could be considered conscious, regardless of whether its origin is biological or synthetic, or whether it is natural or artifical. However, there is still much debate about this framework's validity. The main advantage of the Integrated Information Theory is the fact that it implies a clear, measurable metric and criterion an intelligent agents needs to fulfill in order to be considered conscious. The main disadvantage of the Integrated Information Theory, however,

Global Workspace Theory

According to the Global Workspace Theory, which is a cognitive architecture as well as a theory of consciousness developed by the cognitive psychologist Bernard J. Baars, consciousness works much like a theater. The "stage" of consciousness can only hold a limited amount of information at a given time, and this information is broadcast to a "global workspace" – a distributed network of unconscious processes or modules in the brain. This model, when applied to AI, creates a framework that would, if implemented, allow the AI implemented with it to experience consciousness.

Artificial General Intelligence

An artificial General Intelligence is a type of artificial intelligence that possesses a capability of understanding, learning, as well as application of knowledge across a wide range of tasks, similar to that of a human being - as opposed to existing narrow artificial intelligence systems, which excel in narrowly defined domains and specific tasks, such as voice recognition or playing chess.

In this theory, the general type of artificial intelligence is considered a prerequisite for [**?**]

1.6 Testing for intelligence

While defining intelligence is no easy task, one of its definitions (the ability to apply knowledge to manipulate one's environment or to think abstractly as measured by objective criteria (such as tests)) lends itself

1.7 Testing for consciousness

> I PROPOSE to consider the question, 'Can machines think?' This should begin with definitions of the meaning of the terms 'machine' and 'think'. The definitions might be framed so as to reflect so far as possible the normal use of the words, but this attitude is dangerous. If the meaning of the words 'machine' and 'think' are to be found by examining how they are commonly used it is difficult to escape the conclusion that the meaning and the answer to the question, 'Can machines think?' is to be sought in a statistical survey such as a Gallup poll. But this is absurd. Instead of attempting such a definition I shall replace the question by another, which is closely related to it and is expressed in relatively unambiguous words. [**?**]

In order to establish whether an artificial intelligence agent is conscious, we need to establish a testing method for consciousness.

Thankfully,

1.7.1 The Turing Test

The Turing Test is the first attempt to test the ability of artificial intelligence agents to exhibit intelligent behaviour similar to that of a human. Currently named after its inventor Alan Turing, it was initially called "the imitation game" as it tasked the artificial intelligence in question with participating in a conversation with a human examiner under pretense of being a human. The Turing test was inspired by a party game, which plays out as follows: A man and a woman go into separate rooms and communicate with guests using typewritten responses. The guests are tasked with determining which of the two have entered which room. Similarly, the Turing test involves a human and a machine participating in conversation with an examiner or multiple examiners, with the machine being tasked to misidentify as a human, and the human merely being tasked to correctly identify as such.

The problem with this test, however, is that it merely test the ability of a machine to appear conscious, which is not only possible to accomplish but has been does so in 1966 by ELIZA, a program designed to examine user comments and use fixed rules to generate responses, and therefore does not possess true consciousness despite seemingly appearing conscious.[**?**].

Furthermore ability to deceive a human into believing one is human should not be considered adequate evidence of consciousness as doing so would be an act of self-deception similar to that engaged in by cargo cults, who at least have the excuse of ignorance on their side.

Therefore, this paper will not use the Turing test to determine whether an artificial intelligence consciousness, opting instead to consider more adequate testing methods, such as the following:

1.7.2 Self-awareness tests

Self-awareness tests were designed to assess self-awareness in animals.

Mirror test analogy The most famous self-awareness test is the mirror test, established in 1970 by Gordon Gallup, which determines whether the test subject can recognise themselves in a mirror. It involves placing a mark on the test subject's body and then observing whether the subject will correctly recognise the mark on their body by observing the mirror image.

As of now, several animal species have demonstrated self-awareness by passing the mirror test, including:

- Various dolphin species
- Orca whales(*Orcinus orca*)
- Eurasian magpies(*Pica pica*)
- Ants (*Formicidae*)
- Several members of the great ape family (*Hominidae*), including:
  - Chimpanzees (*Pan troglodytes*)
  - Bonobos (*Pan paniscus*)
  - Orangutans(*Pongo pygmaeus, Pongo abelii*) and, of course,
  - Humans (*Homo sapiens*)

Advantages This test has the advantage of being easily implemented into testing environments used to evaluate the performance of artificial intelligence agents - an agent may be denied direct self-knowledge and limited to indirect observation of its attributes through a mirror or appropriate equivalent (such as a second agent instructed to copy its actions). Furthermore, this test provides a relatively objective and measurable

Drawbacks

Sensorimotor Contingency Test

1.8 Simulating intelligence

1.8.1 Existing attempts

1.9 Simulating consciousness

1.9.1 Existing attempts

2 Materials and Methods

2.1 Virtual testing environment

In order to test AI agents for intelligence and consciousness, this thesis will utilise a series of different tests and test environments.

2.2 Grid elements

2.2.1 Basic grid elements

Clear tile The base form of a tile - doesn't block objects from crossing it, destroy objects, or interact with objects in any way. The clear tile is, for all

Goal tile Unless specified otherwise, the goal of every agent is to reach a goal tile, and a variation of this event, such as:

- One active agent reaching a goal tile
- Every active agent reaching a goal tile
- One or more passive agents reaching a goal tile
- etc...

the simulation will stop, and the agent/agents will be awarded a significantly positive result.

Wall tile Much like with walls in the physical world, the role of this tile is to stop any entities from passing through it. Additionally, the wall tile will block vision of entities that rely on ground-level vision.

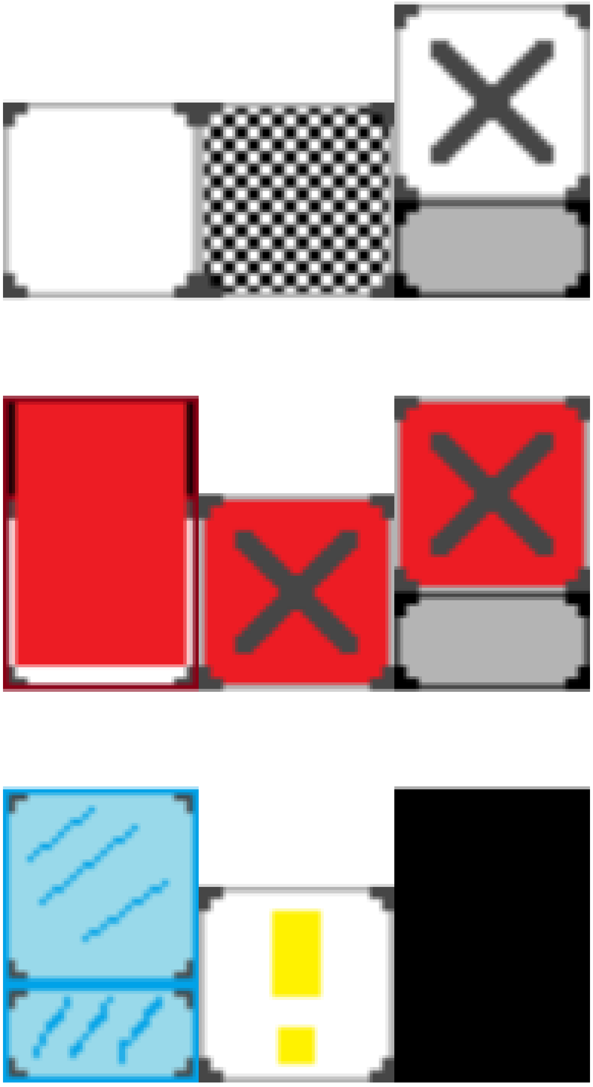


Figure 2.1: All 8 tiles used in grid environment design, as well as a "null" tile used for display purposes.

Curtain tile Unlike walls, curtain tiles will allow entities to pass them - but they will still block an entity's line of sight.

Lethal tile This tile will destroy any entity that moves to its position, which will in most circumstances apply a significant negative penalty to the tested agent/agents, especially if the destroyed agent is one of the active agents. In some circumstances, this may even lead to an early stop of the simulation.

Lethal Wall tile The lethal wall tile acts in a similar way to the lethal tile, except it also blocks the agents' line of sight.

Glass tile The glass tile acts in a manner inverse to that of the curtain tile, blocking entity movement but not line of sight.

Effects tile The effect tile will inflict an effect upon every agent that crosses it.

Null tile The null tile is not a valid tile that exists on the grid. Instead, it merely exists as a display tile, to show that a given tile is not visible to an agent.

2.2.2 Composite grid elements

In addition to basic grid elements, the grid environment may also contain composite grid elements that are perceived and act differently depending on the properties of the entities interacting with them.

For example, a grid element may be configured to act like a blank tile when interacting with red agents, but like a wall when interacting with any other type of agent.

2.3 Test entity design

2.3.1 Entity/Agent duality

In order to allow agents to be affected by attributes and status effects that require knowledge of self to detect and/or manage, as well as to design more complex tests, an entity object class has been created to contain the agent and contrast against the agent class.

2.4 Test environment design

1. Reference environments Before being tested, some artificial intelligence agents may require training in reference environments.

   Other artificial intelligence agents may not require training, but still require reference testing to confirm basic functionality before applying intelligence and consciousness tests.

2. Mirror test environments Mirror test environments are to be designed in a way that

allows agents to receive indirect information about themselves required to successfully pass the tests.

2.4.1 Reference test environments

2.4.2 Mirror test environments

In mirror test environments, test entities are presented with mirror entities that copy test entity traits and behaviors in real time, allowing for an indirect source of self-knowledge similar to a mirror.

2.5 Agent behaviour types

Basic entity behaviour types

Box The simplest type of entity, the box, is meant to be nothing more than a test element. It does not process information or move.

Actions loop This entity runs a pre-recorded set of action and is primarily used to verify test environment functionality, although it can also be used as a test element.

Mirror

2.6 Primary approach choice

2.7 Implementation choice

2.7.1 Programming language choice

Python-based implementation was chosen for this paper due to several factors:

- Ease of prototyping While low-level languages generally outperform Python in terms of performance by orders of magnitude, its ease of use makes it an adequate choice for prototyping.

- Ability to leverage low-level language performance Various tools, such as C extensions, libraries with low-level implementations, and alternative interpreters allow

Python to mitigate its base weakness and perform better.

- Specialised machine learning libraries Libraries such as Scikit-learn, TensorFlow, and Keras have been developed specifically for machine learning and will significantly accelerate development of test environments and AI agents.

2.8 Odabir primarnog pristupa

2.9 Odabir implementacije

2.9.1 Odabir programskog jezika

Za ovaj rad odabran je Python zbog nekoliko faktora:

- Jednostavnost prototipiranja Iako niskorazinski jezici općenito nadmašuju Python u pogledu performansi za redove veličine, njegova jednostavnost korištenja čini ga adekvatnim izborom za prototipiranje.
- Mogućnost iskorištavanja performansi niskorazinskih jezika Različiti alati, poput C ekstenzija, knjižnice s niskorazinskim implementacijama i alternativni interpretatori, omogućuju Pythonu da ublaži svoju osnovnu slabost i postigne bolje performanse.
- Specijalizirane knjižnice za strojno učenje Knjižnice poput Scikit-learn, TensorFlow i Keras razvijene su posebno za strojno učenje i značajno će ubrzati razvoj testnih okruženja i AI agenata.

3 Results and Discussion

4 Further Research Suggestions

5 Conclusion

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

References

[1] Bard, “Bard, a large language model from google ai,” 2023. [Online]. Available: https://ai.googleblog.com/2023/01/lamda-language-model-for-dialogue.html

[2] A. M. TURING, “I.—COMPUTING MACHINERY AND INTELLIGENCE,” *Mind*, vol. LIX, no. 236, pp. 433–460, 10 1950. https://doi.org/10.1093/mind/LIX.236.433

Abstract

MODELLING AND OVERSIGHT OF NATURAL INTELLIGENCE: KEY ASPECTS

Dorijan Cirkveni

Enter the abstract in English.

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Keywords: the first keyword; the second keyword; the third keyword

Sažetak

KLJUČNI ASPEKTI MODELIRANJA I NADZORA PRIRODNE INTELIGENCIJE

Dorijan Cirkveni

Unesite sažetak na hrvatskom.

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Ključne riječi: prva ključna riječ; druga ključna riječ; treća ključna riječ

Appendix A: The Code

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

This is the second paragraph. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

And after the second paragraph follows the third paragraph. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

After this fourth paragraph, we start a new paragraph sequence. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.