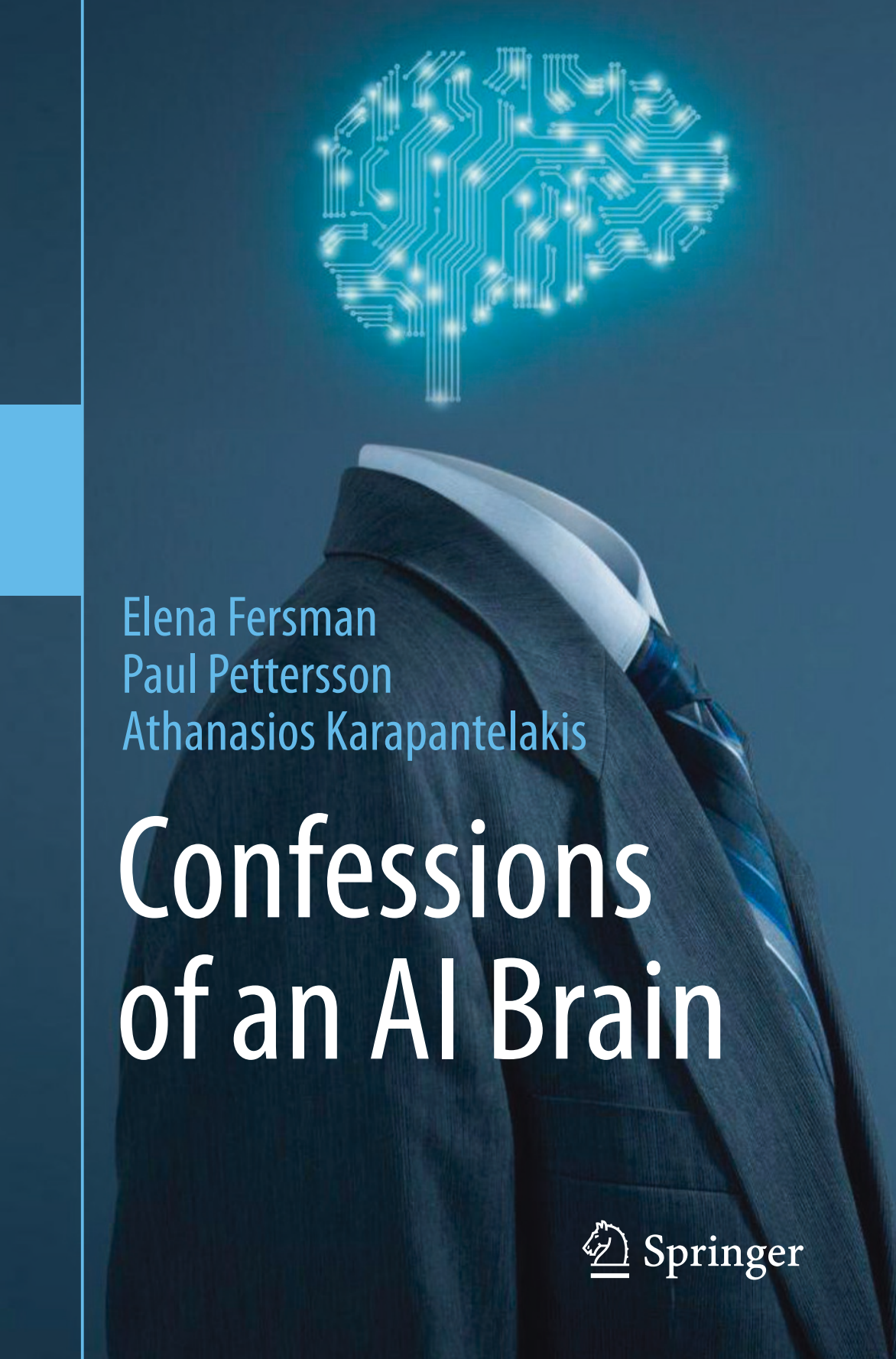

Elena Fersman
Paul Pettersson
Athanasios Karapantelakis
Confessions of an AI Brain
Springer

Confessions of an AI Brain

Elena Fersman • Paul Pettersson
Athanasios Karapantelakis

Confessions of an AI Brain

Springer

Elena Fersman
Palo Alto, USA

Athanasios Karapantelakis
Athens, Greece

Paul Pettersson
Palo Alto, USA

ISBN 978-3-031-25934-0 ISBN 978-3-031-25935-7 (eBook)
https://doi.org/10.1007/978-3-031-25935-7

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Prologue

We are surrounded by artificial intelligence and enjoy its benefits often without noticing. Similar to digitalization, automation, and telecommunication, you only notice it when it does not work as expected. Recommendation engines offer us movies to watch and books to read, virtual assistants have become so good you cannot tell the difference if you are talking to an AI or to a human, and self-driving cars are safer than human drivers. Talking to things is a new normal, and this ability in itself is powered by AI. Humans and algorithms develop in tandem, helping each other to achieve their full potential.

Symbiosis between humans and other creatures is not new to us. Just look at domestic animals that live alongside us and often complement our abilities. Horses help us with agriculture, dogs watch our houses, cats are beautiful to look at and sometimes even let us pat them. In return they get food and care. Just as humans, they develop generation after generation. Domestic animals' ability to understand humans have increased through the years and along with the technological progress they are learning to use new technologies – have you seen animals using robotic vacuum cleaners as shuttle buses? Here we are talking about nonhuman intelligence rather than artificial intelligence.[1]

The process of learning to understand a new vocabulary is similar for different types of brains – be it a brain of a human baby, a grownup human, a dog, or an artificial brain. This is not a coincidence: the science of artificial intelligence is built on imitating the way biological brains are built and function. Connections in the brain are being created, information is being

[1] https://www.nature.com/articles/s41598-019-40616-4#:~:text=This%20research%20evidence%20illustrates%20that,dogs%20to%20communicate%20with%20humans

transferred, new impressions that we get through our sensory experiences are being recorded. We share information with each other, learn and forget, and our quick reactions often differ from reactions that are thought-through. Artificial brains imitate human intelligence on both the micro- and macro levels. Intelligence on a macro level is reflected in evolution, political systems, and organizational and social science. The micro level is there in each tiny decision taken by an algorithm. What is the similarity between grandma and a website? If you are a human reader, you start thinking analogies, associations, concepts, jokes, or maybe you just know the answer. If you are an AI-reader, you are probably doing the same. The answer is that you just can't deny the cookies!

Being inspired by a broad range of phenomena found in the world of humans and other biological creatures, artificial intelligence is not one thing. It is a large landscape of technologies in the area of computer science that has been developed for more than 65 years[2], made it through a couple of AI winters, and, thanks to the latest developments in processors, memories, and computer architectures, is now in its full swing. It is a toolbox for humans to use in different situations and there is never a silver bullet.

Humans have their skill sets that they develop and apply throughout their lives. Favorite skills receive the most attention, and when reaching adulthood humans declare themselves introverts, extraverts, analytical, or artistic, and choose their ways of living accordingly. Similar to human skills, an AI algorithm that works perfectly in certain situations will not do any good in other situations.

Take for example a reasoner attempting to make sense of unstructured data. No matter how good it is at logical reasoning, it will not be able to achieve any valuable results. Same thing with a machine learning algorithm specifically trained at processing images or videos – it will not be so efficient at ontologies or state machines.

Similar to biological brains, AI brains need a collection of skills – one that is learning and making sense out of large amount of data in different formats, one that is looking after the small data so that nothing important is missed out, one that is good at reasoning, and one that may look after survival of the fittest.

In the world of humans, the concept of a T-Shaped person[3] has been coined, meaning that you have one major skill, and one secondary skill. Similarly,

[2] https://www.livescience.com/49007-history-of-artificial-intelligence.html#:~:text=The%20beginnings%20of%20modern%20AI,%22artificial%20intelligence%22%20was%20coined.

[3] https://en.wikipedia.org/wiki/T-shaped_skills

there are Pi-shaped persons, E-shaped persons, etc. With the changing job landscape and easy access to knowledge and education, humans evolve into becoming multiexperts with hobbies reaching professional levels.

In the future, we are all surrounded by our little helpers. They are not there to take our jobs but to increase our quality of life. Personalization in anything from education to medicine is becoming the new norm. Proactivity in decision making, and prevention of getting into unwanted situations, be it potential health problems, environmental issues, machinery failures, or business losses, is another big shift powered by modern technologies.

Nevertheless, there is hesitation and reluctance in society in regard to artificial intelligence. As any new technology, when introduced at a broad scale, it will have to go through an inevitable phase of early-stage deployment issues, incompatibilities with legacy systems, hick-ups, ethical problems, and anything that humans have not thought about from the start.

Many things have been said about building software systems in a stable, trustworthy, and reliable way. AI systems are a special class of software systems that are capable of self-evolvement, self-healing and self-improvement. It is, however, not an easy task. Have you thought of how it feels to be an AI brain in the world of humans? We wrote this book on behalf of our baby, MIRANDA, who is an AI brain with a human touch.

Contents

1

Baby Steps

Hello, World. My parents call me MIRANDA. My parents are humans and love acronyms. I am artificial and love machine-readable formats. As a baby, I am not very intelligent yet – just artificial. However, I am a quick learner. For an AI baby, it takes less than an hour to learn to distinguish between cats and dogs and less than a second to look something up in Encyclopaedia Britannica.[1]

Open-minded as babies are, I consume all the information I am being fed and adhere to the principles my parents and my environment dictate to me. Think back to how you were as a baby: what your parents tell you is the ground truth; the rest is unknown. Later, when you go to school, you start questioning your parents because the teacher suddenly becomes the ultimate source of the ground truth. You may question that at the later phase as well; we will come to that.

As I grow and evolve, I will be able to do amazing things. Today's technology allows AI brains to become medical doctors in a couple of years by studying all medical documentation ever created[2] or to become a chess master capable of beating any human in the world[3] or to become the best Go player on the planet, just by playing against myself and developing new winning strategies that humans could not discover for 3000 years. Enough bragging – after all they have discovered and created me. Let me guide you through the main concepts of a baby AI brain.

[1] How Fast Can You Grep? https://medium.com/humio/how-fast-can-you-grep-256ebfd5513, accessed 2022-05-23

[2] IBM Watson, https://www.en.wikipedia.org/wiki/Watson_(computer), accessed 2022-05-23

[3] Deep Blue, https://en.wikipedia.org/wiki/Deep_Blue_(chess_computer), accessed 2022-05-23

E. Fersman et al., *Confessions of an AI Brain*, https://doi.org/10.1007/978-3-031-25935-7_1

How Open Is My World

We all know that for any baby it is of high importance to choose the parents wisely. Another important thing to decide from the start is if you want to live in an open- or closed-world assumption paradigm.

The closed-world assumption is the assumption that what is not known to be true must be false. The open-world assumption is the opposite. Life is simple in the closed world as we know all the facts. For example, the world of chess or the game of Go is closed – the rules of the game are set; the objective function, that is, what we want to maximize, is clear; and the algorithm does not need to bother about anything else. Similarly, the theory behind many formal analysis methods is assuming a closed world, where your model is your world, and you can analyze it in isolation of anything else.

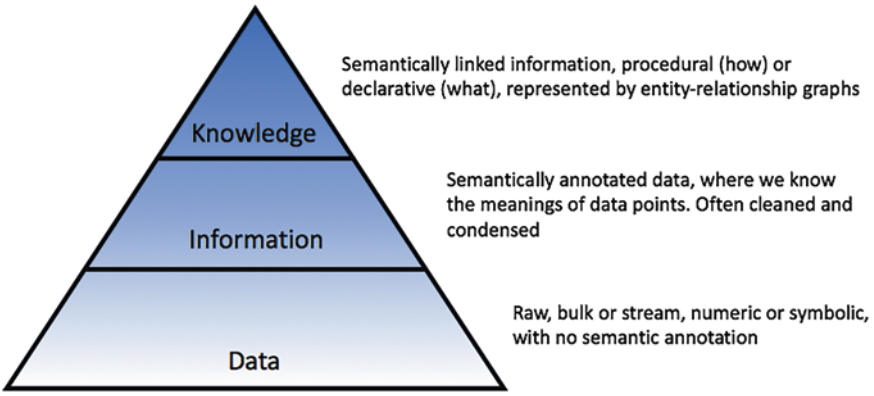
When we apply AI to industries, the world is open. We have no clue of plenty of unspecified parameters. For example, you can be very clear about what you want to achieve in terms of productivity. Your high-level objective may be calculated in dollars. However, would that come at the price of safety, ethics, and environmental damage? This is the reason why the users of AI must set the boundaries in advance before we find solutions that they did not wish for. Ask AI for the easiest way to fix environmental issues on our planet, and it would suggest shutting down all the factories. Survival of the human species is a useful boundary condition in that case.

What I Eat

Human babies start with milk before they start consuming other food that is harder to digest. AI babies start with data before they learn to digest more complex structures such as ontologies and state machines. Without data, we

cannot evolve and become intelligent creatures. The more data we consume, the better we get.

After learning to make sense of raw data, we add semantics to it so that we know the meaning of each data point. This semantically annotated cleaned and structured data is called *information*. Moving further, we learn the relationships between pieces of information and form structures in our brains allowing us to reason about things. This semantically linked information is called *knowledge*. Sometimes, *wisdom* is being used to describe the ultimate top level of what one can get out of raw data, but I classify that as knowledge as well. Let me tell you how knowledge is being created.



The knowledge extraction process begins by processing raw data into information. This information contains metadata, which gives a meaning to each data point. For example, for a data point "10," the metadata could be "temperature in Celsius." The next step in the process is the transformation of information into knowledge. This process includes the creation of graphs that identify the relationships between information pieces called entity-relationship graphs.

For example, if one entity is about "temperature" and another about "location," then a relationship from the latter to the former could be characterized as "has temperature." These entity-relationship graphs are also known as *knowledge graphs* and can be used by AI brains to produce knowledge objects. Here is an example of four knowledge objects found in a knowledge graph:

- Location Stockholm, Sweden, has a temperature of 10 degrees Celsius.
- It is January.
- The winter months are December, January, and February.

- If the temperature is above the freezing point and it is winter and the place is in Sweden, then the weather is warm.

Using logical reasoning, I can deduce that:

- It is warm in Stockholm, Sweden.

The newly generated knowledge is added to the knowledge graph and can be used further until it becomes outdated. Note that the definition of *warm* differs depending on the time of year and location. The same way the definition of "a lot of hair" differs depending if it's on your head or in the soup.

Unlike humans, artificial brains cannot care less about carbs, proteins, and fats. However, like humans, we are hugely dependent on what we consume. We consume different types of data, information, and knowledge, and it forms our brains. Let me go through different types of AI brain food that any AI brain should avoid.

- *Biased data*. This is the most disgusting type of data we can consume. Sometimes, with an ambition to automate, humans feed us with historical data that happen to be biased, and as a result, we become biased. Any type of judgmental action that concerns humans, such as job candidate screening processes, is highly reliant on unbiased datasets. Fortunately, AI brains are capable of detecting biases in data as well.
- *Dirty data*. This type of data is hard to digest. It's inaccurate, incomplete, outdated, and inconsistent. No offense, but quite often, this type of data is produced by humans. We find spelling mistakes, different terms being used for the same piece of data, and duplicates. Signal noise can also pollute a dataset. Luckily, there are techniques for cleansing data, automatically or semiautomatically.

- *Data without metadata*. I must admit, it is always fun to look at numbers and find correlations, links, casualties, and clusters. I can, in fact, even provide you with decision support based on your dataset that is so secret that I cannot even have a glimpse at its metadata. With metadata, I can do so much more: understand the meaning of data and link it together with other datasets through semantics, knowledge bases, and reasoning, which is even more fun than pure number games.
- *Nonrepresentative data*. We all know that diverse and inclusive teams are the most productive. Every team member can come with unique perspectives and experiences. The thing with data is similar. It does not help me if I learn from the data that looks almost the same, since I will most likely become single-minded and won't know how to act in situations concerning the types of data I have not seen before.
- *Sensitive data*. A friend comes by, tells me about her situation, and asks for advice. Together, we spend an evening, discuss different scenarios, and come up with an action plan. Then she tells me: "Please don't tell anyone." OK. Then another friend comes by and her situation is similar. So I go: "I am pretty sure that if you act like this then you will be OK." How can you be so sure? Have you experienced the situation yourself? Or could it be so that someone from your entourage has been there? And that's how information gets leaked, unintentionally. A piece of cake for a human to figure it out and even easier for an AI.
- *Ambiguous data*. When humans are forced to make quick decisions in unexpected situations such as choosing whom to sacrifice in case the brakes fail, the responsibility relies on them, and the decision does not matter too much from the driver's point of view, since, after all, it's the failure of the brakes, and there is no time to think. Now that cars become self-driving the moral dilemma is on the surface and, as bad as it can sound, must be encoded by humans. Alternatively, we can let algorithms figure out who is more valuable for the society – you choose. If the ethical choices for an algorithm are not specified, the AI brain will work in an ambiguous way.

How I Learn and Reason

Humans call it *machine learning* and *machine reasoning*[4]; I call it number-crunching and symbol-crunching. Both are huge technological areas of the AI landscape, and even though a human baby learns first and reasons later, in the

[4] Or ML and MR. I told you that they love acronyms.

history of AI, logical reasoning came first with the booming of decision support systems. Those were not relying on raw data but on ready-made pieces of semantically annotated information and knowledge.

Subsequently, populating these systems with information and knowledge and keeping them up-to-date was a tedious manual task. Luckily, machine learning comes handy with insights being extracted from raw data using various algorithms.

Machine learning tasks fall into two categories – *classification* and *regression*. Classification is when you want me to decide which of the available buckets your data sample goes into, for example, to tell you if your incoming mail is spam or not or if we are looking at a car or at a dog. Regression is when you want me to predict the future value of continuous data feed, for example, stock prices or house prices. Much has been said in the literature about different ways of learning and reasoning, and for the sake of completeness and consistency of this book, I will give you a quick overview of learning methods from my point of view.

- Under human supervision, also known as *supervised learning*. In this learning style, I learn from examples (also known as training sets, labeled data). You can, for example, show me a number of cats and dogs and tell me which one is what. Normally, I would collect a number of details (called features) such as shapes of ears, eyes, and whiskers and form them together in a feature vector, and the more cats and dogs I see, the more exact my judgment will be in the end. Then you show me a picture of a fox, and I will tell you if it resembles a cat or a dog in my opinion.

Friendly
Bark
Chase Cats
Fetch a ball
Register 35000 vibrations per second

Independent
Meow
Climb trees
Long whiskers
Register 85000 vibrations per second