# Full Self-Driving Skynet, and Other Artificial Intelligence Myths

### The realities of decision making with deep machine learning models

Brad Flaugher

January 5, 2023

We have now accumulated sufficient evidence to see that whatever language the central nervous system is using, it is characterized by less logical and arithmetical depth than what we are normally used to.
– John von Neumann [1]

# Preface

This book is a work in progress, I hope it helps demystify the world of deep learning as I understand it.

Humans won't be able to control superintelligent AI, talk about that here[2]

Talk about Bostrom and GPAI here, and Erdi's answer to that. [3] [4]

Talk about the alignment problem and Ethical freakouts about AI. Talk about the big 3 from [5] [6]

Funding and startups, everybody is doing it, I'm trying to make sense of it

*Brad Flaugher*

# Contents

# List of Figures

# List of Tables

# List of Listings

# Playing chess in 1997 | 1

## 1.1 "Textbook" AI in 1997

Dr. Elaine Rich's textbook on Artificial Intelligence, published in the 1980s, was a groundbreaking work that helped to establish many of the foundational concepts and techniques in the field of AI. However, the rapid advancements in AI over the past few decades have led to many of the chapters in this textbook becoming obsolete.

One of the main reasons for this is the prevalence of deep learning, big data, and large-scale statistical models in modern AI. These techniques have largely replaced the symbolic, rule-based approach to AI that was emphasized in the textbook, making many of the chapters on knowledge representation and expert systems less relevant.

Additionally, the explosion of data and the availability of powerful computing resources have made it possible to apply machine learning techniques at a scale that was previously unimaginable. This has led to the development of highly effective machine learning models that can handle complex tasks such as image and speech recognition with a high degree of accuracy, making many of the chapters on simpler machine learning techniques such as decision trees[1] and linear regression less relevant. [7] [2]

We'll discuss this history and a few examples from the "early days" of AI to help us understand where we are headed. We'll start with machine translation, then discuss chess and finally neural networks, which will be the focus of the rest of this book.

## 1.2 Teaching computers to translate

Noam Chomsky is a linguist and philosopher who has made significant contributions to the field of linguistics with his theory of universal grammar. Chomsky believes that all human languages share a common underlying structure, and that this structure is innate to humans. He proposes that this innate structure is the result of a "language acquisition device" present in the human brain, which allows us to learn and produce language. Chomsky also argues that the structure of language is largely independent of its content, and that the ability to produce and understand language is a fundamental aspect of human nature. His theory has been influential in the field of linguistics and has sparked much debate and research on the nature of language and its relationship to the human mind.

For English speakers or anyone who has learned English as a second language you'll have many examples of special cases, irregular verbs, bad english and former street slang that became good and proper over time. For

[1] Although mathematically, Neural Networks are Decision Trees

[7]: Rich et al. (2009), *Artificial Intelligence*

[2] the book is now in its third edition and unlikely to be updated as Dr. Rich as retired utexas.edu

programmers this is a nightmare, how can we codify human knowledge in a timely fashion? If we tried to write the rules of the english language in code (which many have tried to do) the rules themselves might change before we were finished writing them.

Explicitly translating languages through code is a difficult task because it requires a thorough understanding of the grammar, vocabulary, and syntax of both languages, as well as the nuances and subtleties of their respective cultures[3]. Simply coding rules for how to translate words or phrases from one language to another is not sufficient, as there are often multiple valid translations for a given phrase depending on the context in which it is used.

A more effective approach to translation is to use statistical techniques that rely on a large corpus of translated data, such as Canadian laws[4]. This type of data-driven approach involves training a machine learning model on a large dataset of translations, allowing it to learn the patterns and relationships between the languages. The model can then use this knowledge to make educated translations of new phrases or sentences, taking into account the context in which they are used.

While this approach is not perfect, it has proven to be highly effective in machine translation and can produce accurate translations even for languages that are very different from each other. The use of a large dataset of translations also allows the model to learn from the mistakes and variations present in real-world translations, further improving its accuracy.

## 1.3 Codified human knowledge

When we "teach" a computer to perform a task by explicitly writing down all of the rules of that task, we are really codifying human understanding.[5] When we codify human understanding we write down every rule that we know explicitly. For small tasks we can do this with 100 percent accuracy, and only minor headache on the part of the sofware developer.

For example, let's write a boring function to tell you the number of days for a given month.

```
def days_in_month(year, month):
  if month in [1, 3, 5, 7, 8, 10, 12]:
    return 31
  elif month in [4, 6, 9, 11]:
    return 30
  elif month == 2:
    if (year % 4 == 0 and year % 100 != 0) or year % 400 == 0:
      return 29
    else:
      return 28
  else:
    return "Invalid month"
```

[3] For programmers this is a nightmare, how can we codify human knowledge in a timely fashion? If we tried to write the rules of the English language in code (which many have tried to do) the rules themselves might change before we were finished writing them.

[4] They're in French AND English, which is useful data that we can use to correllate phrases and transform English to French and vice-versa.

[5] Programming this way makes some software development totally boring, I almost switched my major in college to math after considering what a life would look like manually writing rules for handling "edge cases" for the rest of my natural life.

Writing code can be a tedious and repetitive task, especially when it comes to debugging and testing. It can be especially frustrating when you're working on a large project and you're trying to track down a specific bug that's causing the program to crash. Testing code can also be boring, as it often involves running the same tests over and over again to ensure that the code is working correctly.

Additionally, writing code can be boring because it requires a lot of concentration and focus. It can be easy to get lost in the details and lose track of time, especially if you're working on a complex problem. It can also be challenging to come up with creative solutions to problems, and it can be frustrating when your code doesn't work as expected.

While writing and testing code can be rewarding and fulfilling, it can also be a tedious and boring process. It requires a lot of patience, persistence, and attention to detail, and it can be easy to get frustrated and lose motivation. However, with practice and perseverance, it is possible to overcome these challenges and find enjoyment in the process of writing and testing code.

AI has traditionally operated by explicitly codifying human knowledge into machine-readable formats by doing the boring job of coding. This approach, which I'm calling "codified human knowledge" relies on humans to carefully structure and organize information in a way that can be understood by the AI system. The AI system then uses this structured knowledge to make decisions and perform tasks.

However, recent advances in AI have largely ignored the knowledge representation problem and instead have focused on using statistical techniques and neural networks to automatically learn patterns and relationships in data. This approach, known as "deep learning," involves training large neural networks on vast amounts of data, allowing the AI system to make educated classifications and transformations of data without explicit human guidance.

Deep learning has proven to be highly effective in a variety of applications, such as image and speech recognition, and has contributed to the rapid progress we have seen in AI in recent years. However, the reliance on large amounts of data and the lack of transparency in these models can make it difficult to understand how they are making decisions, which can be a concern in certain applications (hence the title of this book).

## 1.4 Deep Blue's brute force victory

Deep Blue was a revolutionary computer developed by IBM that was specifically designed to play chess at the highest level. It was programmed with a vast database of chess knowledge and was able to analyze millions of positions per second.

Garry Kasparov was the reigning world chess champion at the time, and he was considered to be one of the greatest players in history. He had never lost a match to a computer before, and he was confident that he would be able to defeat Deep Blue.

However, things did not go as Kasparov had expected. Deep Blue was able to analyze the positions on the board with incredible speed and accuracy, and it was able to come up with highly sophisticated strategies that Kasparov had never seen before.

Despite Kasparov's best efforts, he was no match for the sheer brute force of Deep Blue's computational power. In the end, Deep Blue emerged victorious, defeating Kasparov in a historic match that changed the world of chess forever.

Deep Blue was a turning point in the development of AI, but Deep Blue's methods (namely calculating every possible outcome of a Chess game to determine the best move) was not suitable for many of the world's problems. It turns out that Chess is fun, but the world is not like chess. The "real" future of AI was being developed elsewhere, using statistics and a toy model of the brain to solve a very practical problem for banks.

## 1.5 Meanwhile at the bank

It was the early 1990s and Yann LeCun was a researcher at Bell Labs in New Jersey. At the time, the process of reading and processing checks was a tedious and time-consuming task that was done manually by bank employees. LeCun saw the potential for using artificial intelligence to automate this process, and he began experimenting with using convolutional neural networks (CNNs) to recognize patterns in images of checks.

At the time, CNNs were a relatively new type of neural network that had been developed in the 1980s for image recognition tasks. They were inspired by the structure of the human visual system, and were able to process images in a way that was similar to how the human brain does.

LeCun's work was groundbreaking, and he was able to achieve impressive results using CNNs to process checks. By 1993, he had developed a system that was able to read and process checks with a high degree of accuracy, significantly reducing the amount of time and effort that was required to process checks manually.

LeCun's work on using CNNs for check processing was a major milestone in the field of artificial intelligence, and it laid the foundation for the development of many other applications of CNNs in the years that followed. Today, CNNs are widely used in a variety of applications, including facial recognition, image classification, and natural language processing. [6]

## 1.6 Programmer intelligence, data intelligence and artificial intelligence

I think it's useful to separate the actors in the AI problem-space into three groups. The data, the programmer and the machine learning (or AI) together they make the programs that we use every day, and for the rest of

[6] check out Yann LeCun domonstrating a convolutional neural network in 1993 at youtube.com

this book I'll try and separate the discussion of the smarts of each to help us better understand the world. [7]

Programmer intelligence refers to the ability of a human programmer to design, write, and debug computer programs. This type of intelligence involves problem-solving skills, logical thinking, and the ability to learn and adapt to new programming languages and technologies. Most books don't talk about programmer intelligence and use the even more vague word "Algorithm" to describe both the programmer's output and the AI that might contibute to decision making, which can lead to misunderstanding.

Artificial intelligence is a terrible term. It generally refers to the ability of a machine or computer system to perform tasks that would normally require human intelligence, such as learning, problem-solving, and decision-making. Artificial intelligence systems can be trained to perform a wide range of tasks, from simple tasks like recognizing patterns in data to more complex tasks like understanding and generating natural language. Becuase the term is so broad I'll avoid it, and instead talk about machine learning and deep learning instead. [8]

Data intelligence [9] refers to the ability to extract meaningful insights and knowledge from large datasets. This involves using statistical and analytical methods to discover patterns and trends in data, and using this information to inform business decisions or solve problems. Data intelligence requires a combination of programming skills and statistical and analytical expertise.

Overall, while all three types of intelligence are important in the field of computer science, they involve different skill sets and focus on different aspects of problem-solving and decision-making. Programmer intelligence is essential for designing and implementing computer programs, machine learning is focused on statistically mimicing human-like output in machines, and data intelligence involves using data to inform decision-making and solve problems.

By framing intelligence in this way we can chip away at the AI myths that abound and think about what is really happening. Programmers use data to make AI, there are many places where things can go awry, and many layers of misunderstanding that can get baked into AI products.

[7] I'll also repeatedly encourage you to use the words "algorithm" and "artificial intelligence" sparingly. They'll confuse your thinking, I promise.

[8] In case I am confusing you, I think that Deep Blue was made with "Programmer Intelligence" or "Codified Human Knowledge" where Yann LeCun's Check OCR is "Machine Learning" with some "Data Intelligence" baked in.

[9] I'm making this phrase up, my point is that there is information in the data that might not be successfully extracted by a given AI (or a human-programmer).

# Self-driving with statistics 2

Hardware got amazing, we gave up teaching the way we teach ourselves and let the data do the work

We leveraged huge statistical models to regress our way to success

We used building blocks of regression and neurons to train huge models

These models are statistical and deterministic, but ultimately chaotic black boxes..

TODO talk about these books [8] [9] [10] [11]

Talk about ChatGPT, deterministic vs probabalistic and Thomas Hobbes [1]

**Figure 2.1:** The Mona Lisa. https://commons.wikimedia.org/wiki/File:Mona_Lisa,_by_Leonardo_da_Vinci,_from_C2RMF_retouched.jpg

# A Shakespearean comedy of numbers 3

It’s all numbers man!!!!

## 3.1 Knowlede Representation, or not

This is some text and a link to Hey if you want to site something on the side use[2]

"AI Scientists disagree as to whether these language networks posess true knowledge or are just mimicking humans by remembering the statistics of millions of words. I don’t beloive any kind of deep learning network will achieve the goal of AGI if the network doesn’t model the world the way the brain does. Deep learning networks work well, but not because they solved the knowledge representation problem. They work well because they avoided it completely, relying on statistics and lots of data instead. How deep learning networks work is clever, their performance impressive, and they are commercially valuable. I am only pointing out that they don’t possess knowledge and, therefore, are not on the path to having the ability of a five-year-old child." [12]

```
cd myproject
docker run tensorflow
#profit!
```

`tex.stackexchange.org` for help.

[2]: Andreu et al. (2021), *Humans won’t be able to control a superintelligent AI, according to a study*

# Derivative artworks of the future? 4

GPT-3, BERT and Bloom

Link some cool shit here, Draw Owl!

Who owns this shit anyway? Copilot and FSF plus lawsuits

Prove it, asshole!

I can make a "model" that behaves like a database. Just memorizes shit

# The data is the hardest part 5

You are essentially programming with data, so if your data sucks so will your prediction, you also really can't generalize, only correllate.

are you predicting the right thing? Are you really predicting how valuable the company is or just whether it'll be the next meme stock?

"I hope for some sort of peace—but I fear that machines are ahead of morals by some centuries and when morals catch up there'll be no reason for any of it." Harry Truman, 1945 [13]

Representation, "fixing the training set" [5], or the Impossibility of Fairness from a model.

TODO talk about these books [14] [15] [6] [5]

"The second requirement of goal-misalignment risk is that an intelligent machine can commandeer the Earth's resources to pursue its goals, or in other ways prevent us from stopping it... We have similar concerns with humans. This is why no singer person or entity can control the entire internet and why we require multiple people to launch a nuclear missile. Intelligent machines will not develop misaligned goals unless we go to great lengths to endow them with that ability. Even if they did, no machine can commandeer the world's resources unless we let it. We don't let a single human, or even a small number of humans, control the world's resources. We need to be similarly careful with machines." [12]

```
cd myproject
docker run tensorflow
#profit!
```

tex.stackexchange.org for help.

[13]: McCullough (1992), *Truman*

[5]: Christian (2020), *The Alignment Problem: Machine Learning and Human Values*

# 6 The police and Big Tech are profiling me!

Classification is everywhere, it's also very useful. Just get over it.

Online Advertising, Justice, Job Applications, Creditworthiness, Getting Insurance (Weapons of Math Destruction), Civic Life, /sideciteOneil2017 ; The Default Male, Invisible Women effects snow clearing schedules and drug discovery

```
cd myproject
docker run tensorflow
#profit!
```

`tex.stackexchange.org` for help.

# 7 The useful chaos of spaghetti code

Interacting Layers of Statistical Understanding Useful Chaos

These layers are totally transparent, but you can't understand them because they're complicated, yo

You can't understand ML layers, but they are useful nonetheless.

Talk about this book and Kasparov's tournament with computers and people. [16]

[16]: Mansharamani (2020), *Think for yourself: Restoring common sense in an age of experts and artificial intelligence*

# Self-stabilizing concept drift 8

Many models need to constantly be retrained

Does progress slow down because we keep reusing the work of the past to generate our work?

# 9 My horse drives itself, thanks!

Can Models bring Incremental or Revolutionary Change?

Respond directly to Jon Krohn's TED talk about monkeys being dumber than us... what about construction equipmenth that's stronger than us, or racism/eugenics people that are dumber than us [17]

"The inhabitant of London could order by telephone, sipping his morning tea in bed, the various products of the whole earth – he could at the same time and by the same means adventure his wealth in the natural resources and new enterprise of any quarter of the world – he could secure forthwith, if he wished, cheap and comfortable means of transit to any country or climate without passport or other formality." - John Maynard Keynes [18]

Who is affected the most?

What should individuals do?

What should governments do?

What should businesses do?

[17]: (2022), *Jon Krohn*

[18]: Keynes et al. (2012), *The Collected Writings of John Maynard Keynes (Volume 5)*

# Skynet: did you try unplugging it? 10

# Appendix

# A

# ????

Let's say we want to build an ensemble model to analyze poetry, put a haiku into craiyon's online shit, then we categorize the resulting photo. [2]

[2]: Andreu et al. (2021), *Humans won't be able to control a superintelligent AI, according to a study*

# Bibliography

Here are the references in citation order.

[1] John von Neumann and Ray Kurzweil. *The Computer and the Brain (The Silliman Memorial Lectures Series)*. New Haven, CT, USA: Yale University Press, Aug. 2012 (cited on page ii).

[2] Abraham Andreu and Qayyah Moynihan. 'Humans won't be able to control a superintelligent AI, according to a study'. In: *Business Insider* (Sept. 24, 2021). (Visited on 09/24/2021) (cited on pages iii, 7, 16).

[3] Péter Érdi. *Ranking: The Unwritten Rules of the Social Game We All Play*. Oxford, England, UK: Oxford University Press, Oct. 2019 (cited on page iii).

[4] Nick Bostrom. *Superintelligence: Paths, Dangers, Strategies*. 1st. USA: Oxford University Press, Inc., 2014 (cited on page iii).

[5] Brian Christian. *The Alignment Problem: Machine Learning and Human Values*. New York, NY, USA: W. W. Norton & Company, Oct. 2020 (cited on pages iii, 9).

[6] Reid Blackman. *Ethical Machines: Your Concise Guide to Totally Unbiased, Transparent, and Respectful AI*. Harvard Business Review Press, July 2022 (cited on pages iii, 9).

[7] Elaine Rich, Kevin Knight, and Shivashankar B. Nair. *Artificial Intelligence*. Tata McGraw-Hill, 2009 (cited on page 1).

[8] MacAskill2022. 'The Case for Longtermism'. In: *The New York Times* (Aug. 5, 2022). (Visited on 08/05/2021) (cited on page 6).

[9] Cade Metz. 'The Long Road to Driverless Trucks'. In: *N.Y. Times* (Sept. 2022) (cited on page 6).

[10] Cade Metz. 'Stuck on the Streets of San Francisco in a Driverless Car'. In: *N.Y. Times* (Sept. 2022) (cited on page 6).

[11] Caglar Aytekin. 'Neural Networks are Decision Trees'. In: (2022). DOI: 10.48550/ARXIV.2210.05189 (cited on page 6).

[12] Jeff Hawkins. *A thousand brains: A new theory of intelligence*. Basic Books, 2022 (cited on pages 7, 9).

[13] David McCullough. *Truman*. Riverside, NJ, USA: Simon & Schuster, June 1992 (cited on page 9).

[14] Cathy O'Neil. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York, NY, USA: Crown, Sept. 2017 (cited on page 9).

[15] Caroline Criado Perez. *Invisible Women: Data Bias in a World Designed for Men*. New York, NY, USA: Abrams Press, Mar. 2019 (cited on page 9).

[16] Vikram Mansharamani. *Think for yourself: Restoring common sense in an age of experts and artificial intelligence*. Harvard Business Review Press, 2020 (cited on page 11).

[17] *Jon Krohn*. [Online; accessed 18. Oct. 2022]. Oct. 2022. URL: https://www.jonkrohn.com/posts/2022/10/7/tedx-talk-how-neuroscience-inspires-ai-breakthroughs-that-will-change-the-world (cited on page 13).

[18] John Maynard Keynes, Elizabeth Johnson, and Donald Moggridge. *The Collected Writings of John Maynard Keynes (Volume 5)*. Cambridge, England, UK: Cambridge University Press, Dec. 2012 (cited on page 13).

# Notation

The next list describes several symbols that will be later used within the body of the document.

| | |
|---|---|
| $c$ | Speed of light in a vacuum inertial frame |
| $h$ | Planck constant |

# Alphabetical Index