

Full Self-Driving Skynet and Other AI Myths

The realities of decision making with deep machine learning models

Brad Flaughner

October 31, 2022

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 Flaughner

Contents

Preface	iii
Contents	iv
UNDERSTANDING MODERN AI	1
1 Is it an Algorithm or Machine Learning?	2
1.1 Algorithms: Codified Human Understanding	2
1.2 Machine Learning: Data-Derived Insights	2
2 How Data is Represented	3
3 Machine Learning Methods	4
4 Classifiers	5
5 Transformers	6
THE CHAOTIC PREDICTION MACHINE	7
6 Transparent, but impossible to understand	8
6.1 Interacting Layers of Statistical Understanding	8
6.2 Useful Chaos	8
7 Ethics and Errors	9
8 Revolutionary for some, mildly helpful for most	10
8.1 Can Models bring Incremental or Revolutionary Change?	10
8.2 Who is affected the most?	10
8.3 What should individuals do?	10
8.4 What should governments do?	10
8.5 What should businesses do?	10
APPENDIX	11
A Poetry Test	12
Bibliography	13
Notation	14
Alphabetical Index	15

List of Figures

1.1 The Mona Lisa	2
-----------------------------	---

List of Tables

List of Listings

UNDERSTANDING MODERN AI

Is it an Algorithm or Machine Learning?

1

1.1 Algorithms: Codified Human Understanding

AI is a shitty term

We tried a lot of things, teaching computers explicit grammar and explicit rules

IMO, this was not AI, this was codified human understanding.

In code, that understanding might look like this....

TODO talk about this book [7]

bradflaughter.com.¹

1.1 Algorithms: Codified Human Understanding 2

1.2 Machine Learning: Data-Derived Insights 2

1.2 Machine Learning: Data-Derived Insights

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]

[7]: Douthat (2022), *Can We Resist the Age of the Algorithm?*

1: Snarky sidenote!



Figure 1.1: The Mona Lisa.
https://commons.wikimedia.org/wiki/File:Mona_Lisa,_by_Leonardo_da_Vinci,_from_C2RMF_retouched.jpg

How Data is Represented

2

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

```
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*

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 [12]

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

TODO talk about these books [13] [14] [6] [5]

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

tex.stackexchange.org for help.

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

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

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.

Transformers

5

GPT-3, BERT and Bloom

[Link some cool shit here, Draw Owl!](#)

THE CHAOTIC PREDICTION MACHINE

Transparent, but impossible to understand

6

6.1 Interacting Layers of Statistical Understanding

6.1 Interacting Layers of Statistical Understanding 8

6.2 Useful Chaos

6.2 Useful Chaos 8

Revolutionary for some, mildly helpful for most

8

8.1 Can Models bring Incremental or Revolutionary Change?

Respond directly to Jon Krohn's TED talk about monkeys being dumber than us... what about construction equipment that's stronger than us, or racism/eugenics people that are dumber than us [15]

"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 [16]

8.2 Who is affected the most?

8.3 What should individuals do?

8.4 What should governments do?

8.5 What should businesses do?

8.1 Can Models bring Incremental or Revolutionary Change?	10
8.2 Who is affected the most? . .	10
8.3 What should individuals do?	10
8.4 What should governments do?	10
8.5 What should businesses do?	10

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

APPENDIX



Poetry Test

Let's say we want to build an ensemble model to analyze poetry, put a haiku into crayon'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, 3, 12).
- [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, 4).
- [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, 4).
- [7] Ross Douthat. ‘Can We Resist the Age of the Algorithm?’ In: *The New York Times* (July 30, 2022). (Visited on 07/30/2022) (cited on page 2).
- [8] MacAskill2022. ‘The Case for Longtermism’. In: *The New York Times* (Aug. 5, 2022). (Visited on 08/05/2021) (cited on page 2).
- [9] Cade Metz. ‘The Long Road to Driverless Trucks’. In: *N.Y. Times* (Sept. 2022) (cited on page 2).
- [10] Cade Metz. ‘Stuck on the Streets of San Francisco in a Driverless Car’. In: *N.Y. Times* (Sept. 2022) (cited on page 2).
- [11] Caglar Aytakin. ‘Neural Networks are Decision Trees’. In: (2022). doi: [10.48550/ARXIV.2210.05189](https://doi.org/10.48550/ARXIV.2210.05189) (cited on page 2).
- [12] David McCullough. *Truman*. Riverside, NJ, USA: Simon & Schuster, June 1992 (cited on page 4).
- [13] 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 4).
- [14] Caroline Criado Perez. *Invisible Women: Data Bias in a World Designed for Men*. New York, NY, USA: Abrams Press, Mar. 2019 (cited on page 4).
- [15] 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 10).
- [16] 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 10).

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

preface, iii