

Most “Open-Source” AI Isn’t. And What We Can Do About That

Christopher J. Hazard, PhD

HOW/SO™

A Story of Contract Software Development

- Definition and requirements
- Framework selected
- Representative data given
- Contractor selected and paid

- You get: A binary executable
- You could poke at it, edit the assembly code, or ask the contractor to try to fix issues

```
0015c0d0 00 20 08 A1 99 22 05 A1 20 05 20 1D 1B 65 0D 01 . . . . " . . . . e . .
0015c0e0 0B 20 0D 21 0F 20 06 20 07 65 0D 00 20 12 21 0F . ! . . . e . . ! .
0015c0f0 0B 20 08 20 13 20 0F 42 04 86 7C A7 2B 03 00 A1 . . . . B . . | . + . . .
0015c100 99 21 06 20 1A 42 03 51 04 40 20 19 20 11 42 F0 ! . . B . Q . @ . . B .
0015c110 00 7E 7C A7 2B 03 30 22 05 20 06 20 05 10 55 20 ~ | . + . 0 " . . . . U
0015c120 06 20 05 20 06 63 1B 22 06 A1 22 05 20 06 20 05 . . . . c . " . . . .
0015c130 20 06 63 1B 21 06 0B 20 01 A7 28 02 58 21 1D 20 . c . ! . . . ( . X ! .
0015c140 19 20 11 42 F0 00 7E 7C A7 2B 03 38 22 07 44 00 . . B . . ~ | . + . 8 " . D .
0015c150 00 00 00 00 00 00 64 45 0D 03 20 06 9A 20 07 . . . . . d E . . . .
0015c160 A3 21 05 20 1D 41 01 47 0D 01 20 05 10 44 0C 02 ! . . A . G . . . D . .
0015c170 0B 20 0E 20 13 7D 42 04 87 21 0F 44 00 00 00 00 . . . . } B . ! . D . . . .
0015c180 00 00 00 00 20 19 20 11 42 F0 00 7E 7C A7 2B 03 . . . . . B . . ~ | . + .
0015c190 38 22 05 44 00 00 00 00 00 00 00 64 45 0D 04 8 " . D . . . . . d E . .
0015c1a0 1A 44 00 00 00 00 00 00 00 80 20 05 A3 21 06 20 . D . . . . . . ! .
0015c1b0 05 44 00 00 00 00 00 00 08 40 A2 44 00 00 00 00 . D . . . . . @ . D . . . .
0015c1c0 00 00 00 00 A0 02 7C 20 01 A7 28 02 58 22 1D 41 . . . . . | . . ( . X " . A
0015c1d0 01 46 04 40 20 06 10 44 0C 01 0B 44 00 00 00 00 . F . @ . . D . . . D . . .
0015c1e0 00 00 F8 7F 20 06 20 06 62 0D 00 1A 44 00 00 00 . . . . . b . . . D . . .
0015c1f0 00 00 00 00 00 02 7E 20 06 99 44 00 00 00 00 . . . . . ~ . . D . . . .
0015c200 00 E0 43 63 04 40 20 06 B0 0C 01 0B 42 80 80 80 . . C c . @ . . . . B . . .
0015c210 80 80 80 80 80 80 7F 0B 22 0D 42 BA 7A 53 0D 00 . . . . . " . B . z S . .
0015c220 1A 44 00 00 00 00 00 00 F0 7F 20 0D 42 C5 05 55 . D . . . . . . B . . U
0015c230 0D 00 1A 20 06 20 0D B9 A1 22 05 20 05 20 05 44 . . . . . " . . . D
0015c240 54 31 9D EF 0A F1 D1 3F A2 44 DB 18 3B E1 25 38 T1 . . . . . ? . D . . ; . % 8
0015c250 DB 3F A0 A2 44 AE 15 65 1D 2B 34 F0 3F A0 A2 44 . ? . . D . . e . + 4 . ? . . D
0015c260 B9 CA 2C A5 DB 0F 3F A0 20 0D 42 03 86 42 A0 . . . . . ? . . B . . B .
0015c270 C1 05 7C A7 2B 03 00 A2 0B A2 44 00 00 00 00 . . | . + . . . . D . . . .
0015c280 00 E0 3F A2 21 06 20 01 A7 2B 03 48 22 05 44 00 . . ? . ! . . + . H " . D .
0015c290 00 00 00 00 00 F0 3F 61 0D 03 20 05 44 00 00 00 . . . . . ? a . . . D . .
0015c2a0 00 00 00 00 40 61 04 40 20 06 20 06 A2 21 06 0C . . . . @ a . @ . . . ! .
0015c2b0 04 0B 20 1D 41 01 46 04 40 20 06 20 05 10 3A 21 . . . A . F . @ . . . !
0015c2c0 06 0C 04 0B 20 06 44 00 00 00 00 00 00 00 61 . . . . . D . . . . . a
0015c2d0 04 40 44 00 00 00 00 00 00 00 00 21 06 0C 04 0B . @ D . . . . . ! . . . .
```

A Story of Machine Learning Development

- Definition and requirements
- Framework selected
- Representative data given
- Compute selected and paid

- You get: A bunch of weights
- You could poke at it, adjust training data, loss function, and architecture to try to fix issues

```
0033dce0 00 54 02 29 FA F8 06 00 00 5D 01 80 21 04 00 00 .T.).....]...!...
0033dcf0 4B 4F 12 47 54 01 54 00 4F 1F 80 00 18 A0 80 4F KO.GT.T.O.....O
0033dd00 20 80 FE 00 4F 33 0D 5E 12 08 1C 08 03 25 07 80 ...03.^.....%..
0033dd10 A9 00 00 00 53 00 0E 6C 12 55 00 0F 4F 02 DA 8B ....S..l.U..O...
0033dd20 01 00 4F 01 80 59 08 00 00 53 00 53 00 26 E1 D4 ..O..Y...S.S.&..
0033dd30 05 00 46 4C 4F 11 53 00 4F 01 80 5A 08 00 00 26 ..FLO.S.O..Z...&
0033dd40 10 41 BD 82 07 03 00 80 F0 15 00 00 26 81 9D 06 .A.....&...
0033dd50 00 2C 88 83 BA 20 00 55 88 83 BA 20 4C 4B 4F 01 .,... .U... LKO.
0033dd60 80 5B 08 00 00 3A FD F8 06 00 00 54 02 29 FD F8 .[.....T.)...
0033dd70 06 00 00 4F 12 47 54 01 54 00 4F 1F 80 00 18 A0 ...O.GT.T.O.....
0033dd80 80 4F 20 80 FE 00 4F 33 0D 5E 12 08 1C 08 03 25 .O ...03.^.....%
0033dd90 07 80 A9 00 00 00 53 00 0E 6C 12 55 00 0F 4F 02 .....S..l.U..O.
0033dda0 DA 8B 01 00 4F 01 80 DD 07 00 00 53 00 53 00 26 ....O.....S.S.&
0033ddb0 DA D4 05 00 B9 FE F8 06 00 A8 83 9F 91 02 99 88 .....
0033ddc0 83 FB 91 02 00 46 2D D9 D4 05 00 2D D8 D4 05 00 .....F-....-....
0033ddd0 2D D7 D4 05 00 4C 4F 11 53 00 4F 01 80 DF 07 00 -.....LO.S.O.....
0033dde0 00 26 00 F9 06 00 26 E6 D4 05 00 B9 FE F8 06 00 .&....&.....
0033ddf0 88 83 EA 90 02 99 88 83 82 92 02 00 BD 88 83 F1 .....
0033de00 A1 02 00 80 02 89 00 00 4C 32 88 83 81 92 02 4B .....L2.....K
0033de10 4F 01 80 E0 07 00 00 26 01 F9 06 00 B9 FE F8 06 O.....&.....
0033de20 00 88 83 EA 90 02 33 88 81 13 00 27 88 83 BF A1 .....3.....'.....
0033de30 02 33 88 81 13 00 27 88 83 D7 A1 02 32 88 83 DD .3.....'.....2...
0033de40 A1 02 4B 4F 01 80 E1 07 00 00 26 02 F9 06 00 B9 ..KO.....&.....
0033de50 01 F9 06 00 88 83 D6 A1 02 33 88 81 13 00 27 88 .....3.....'.....
0033de60 83 EA A1 02 32 88 83 EB A1 02 4B 4F 01 80 E2 07 ....2.....KO....
0033de70 00 00 26 03 F9 06 00 B9 01 F9 06 00 88 83 D6 A1 ..&.....
0033de80 02 33 88 81 13 08 27 88 83 EA A1 02 32 88 83 EB .3.....'.....2...
0033de90 A1 02 4B 4F 01 80 E3 07 00 00 26 04 F9 06 00 B9 ..KO.....&.....
0033dea0 01 F9 06 00 88 83 D6 A1 02 33 88 81 13 10 27 88 .....3.....'.....
0033deb0 83 EA A1 02 32 88 83 EB A1 02 4B 4F 01 80 E5 07 ....2.....KO....
0033dec0 00 00 26 E1 43 33 86 41 74 80 36 07 00 00 40 A8 ..&.C3.At.6...@.
0033ded0 83 C0 25 66 02 A2 91 02 BF 25 55 86 41 74 33 86 ..%f.....%U.At3.
0033dee0 41 74 00 55 86 41 74 B9 01 F9 06 00 88 83 D6 A1 At.U.At.....
0033def0 02 55 88 83 D6 A1 02 4C 99 88 83 C1 25 00 BD 82 .U.....L.....%...
0033df00 07 03 00 80 C1 12 00 00 4C 4B 4F 01 80 E7 07 00 .....LKO.....
0033df10 00 53 00 B9 02 F9 06 00 88 83 EB A1 02 30 88 83 .S.....0...
```



What Is Free/Open-Source Software?

- Free use
- Free distribution
- Free modification and understanding!

The Four Essential Freedoms of Free Software – FSF

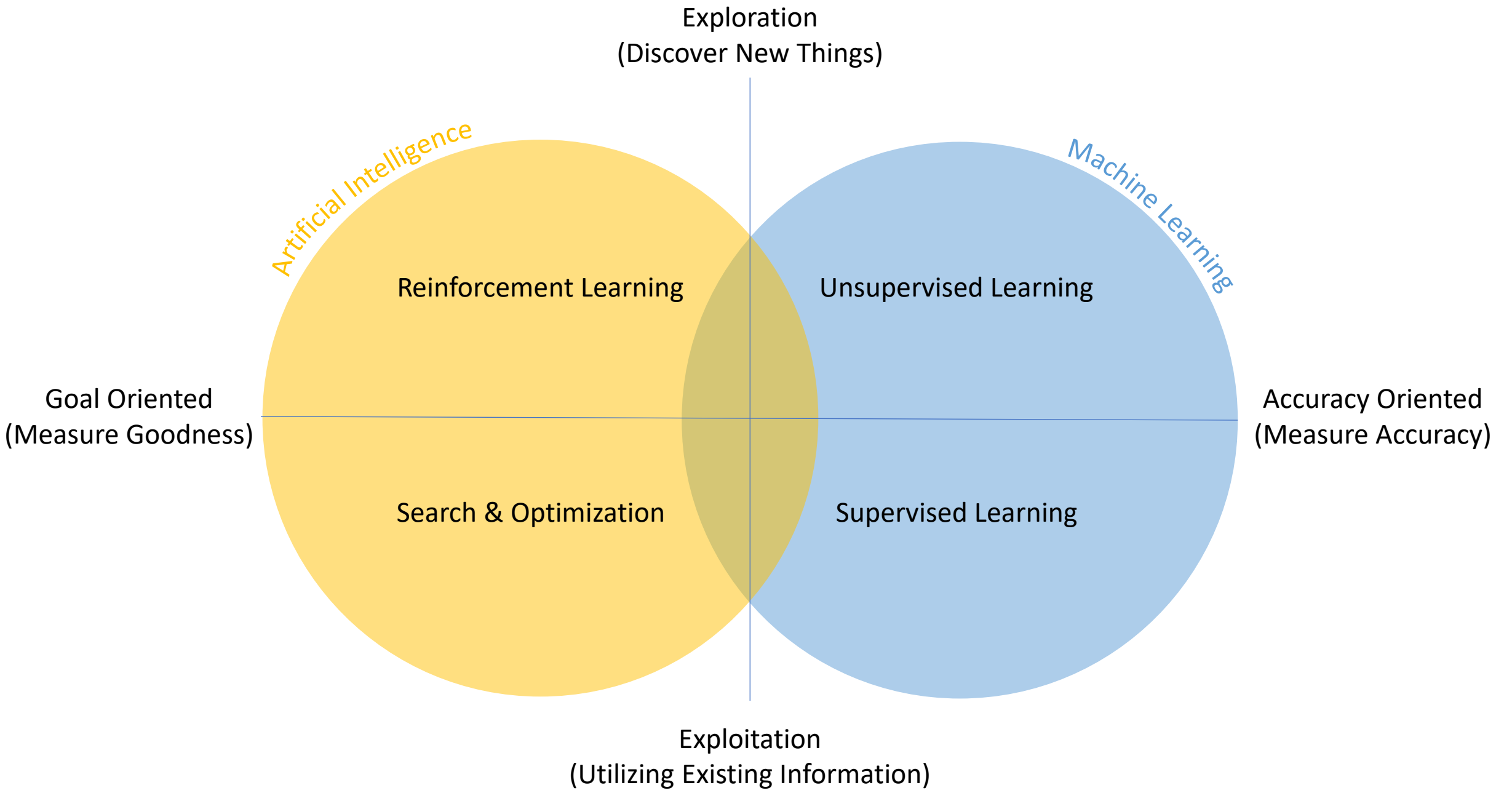
The freedom to study how the program works, and change it so it does your computing as you wish (freedom 1). Access to the source code is a precondition for this.

The Open Source Definition – OSI

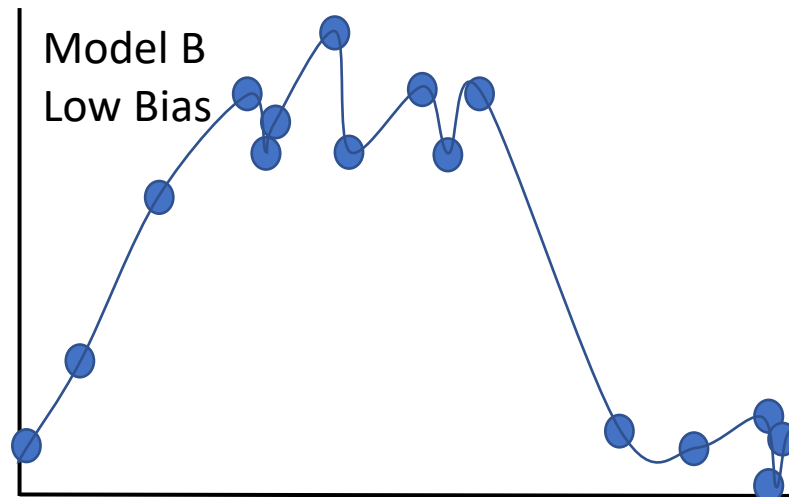
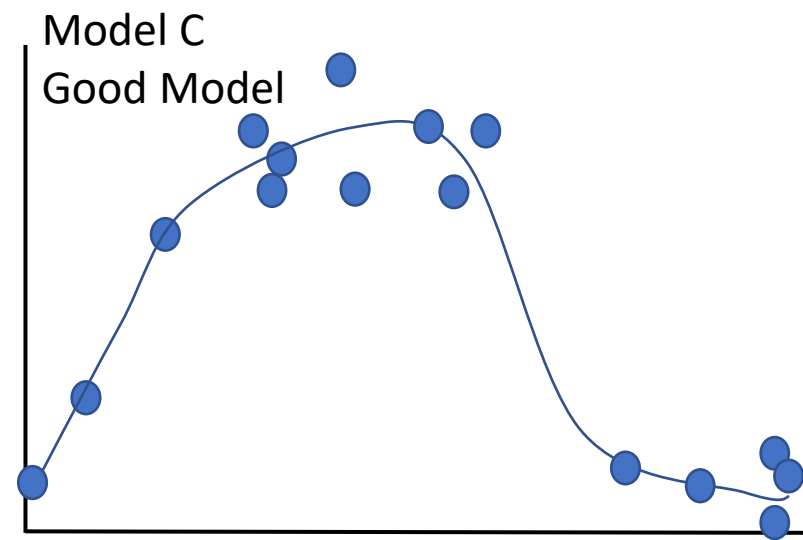
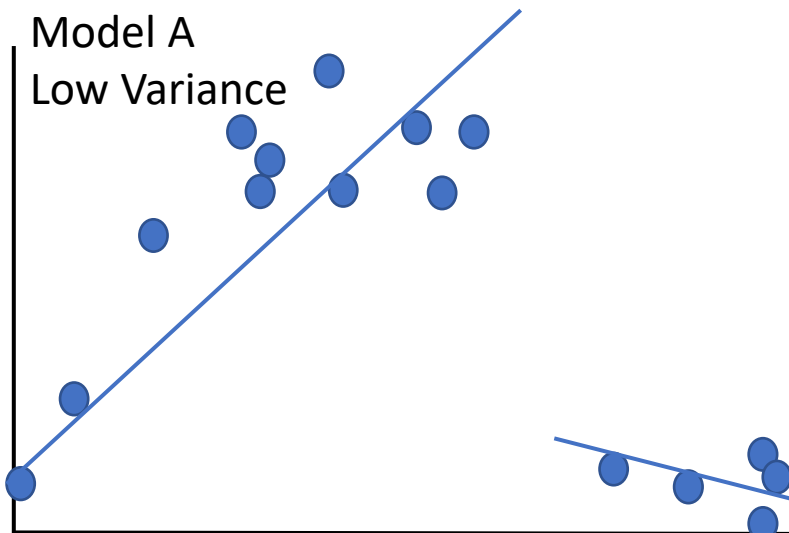
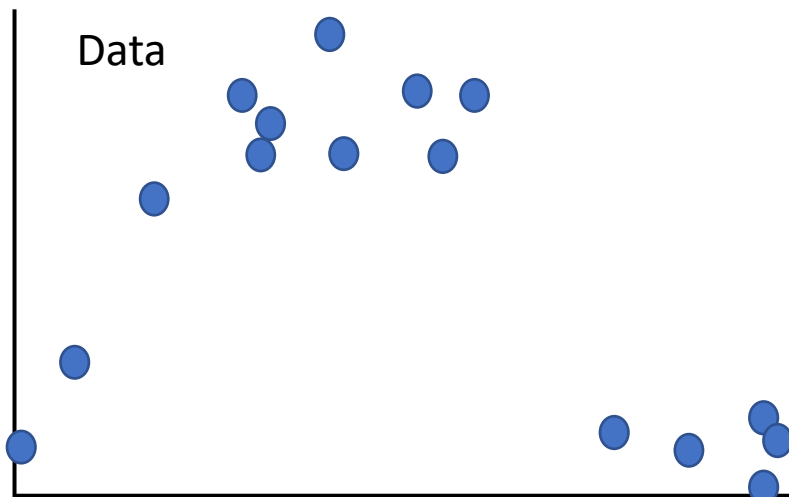
Source code: The program must include source code, and must allow distribution in source code as well as compiled form. Where some form of a product is not distributed with source code, there must be a well-publicized means of obtaining the source code for no more than a reasonable reproduction cost preferably, downloading via the Internet without charge. **The source code must be the preferred form in which a programmer would modify the program. Deliberately obfuscated source code is not allowed.** Intermediate forms such as the output of a preprocessor or translator are not allowed.

- ML: Programming with data
- ML: Compression + generalization

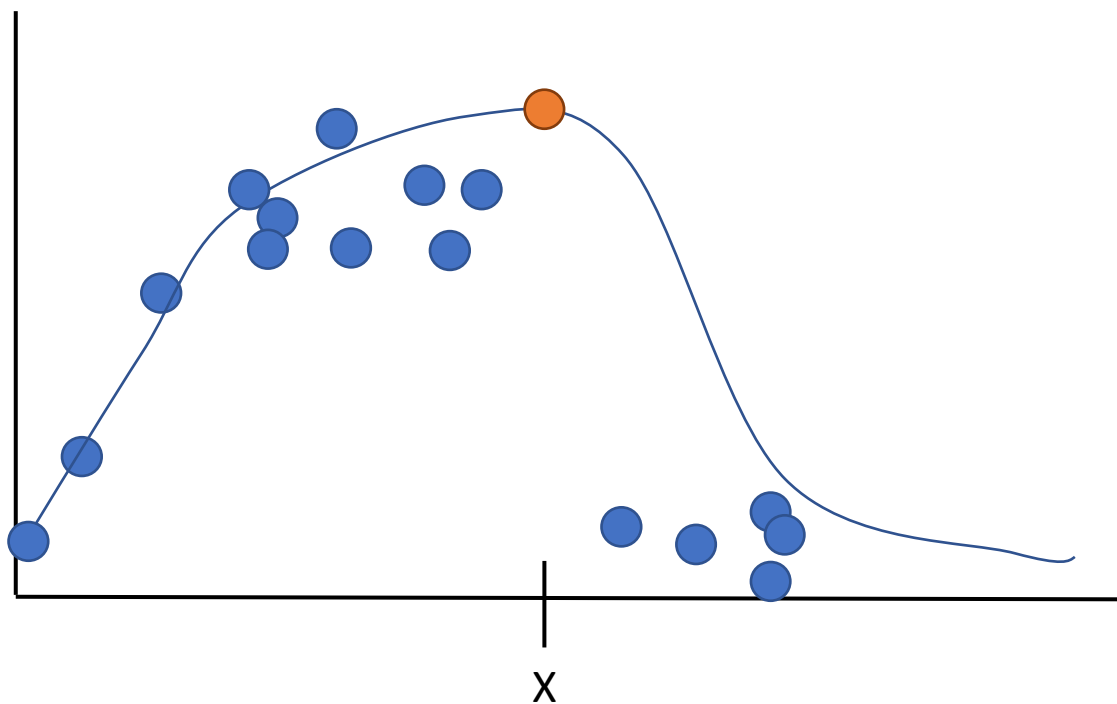
- AI: Hard computer science problems that haven't been solved yet
- AI: Machines doing intelligent things



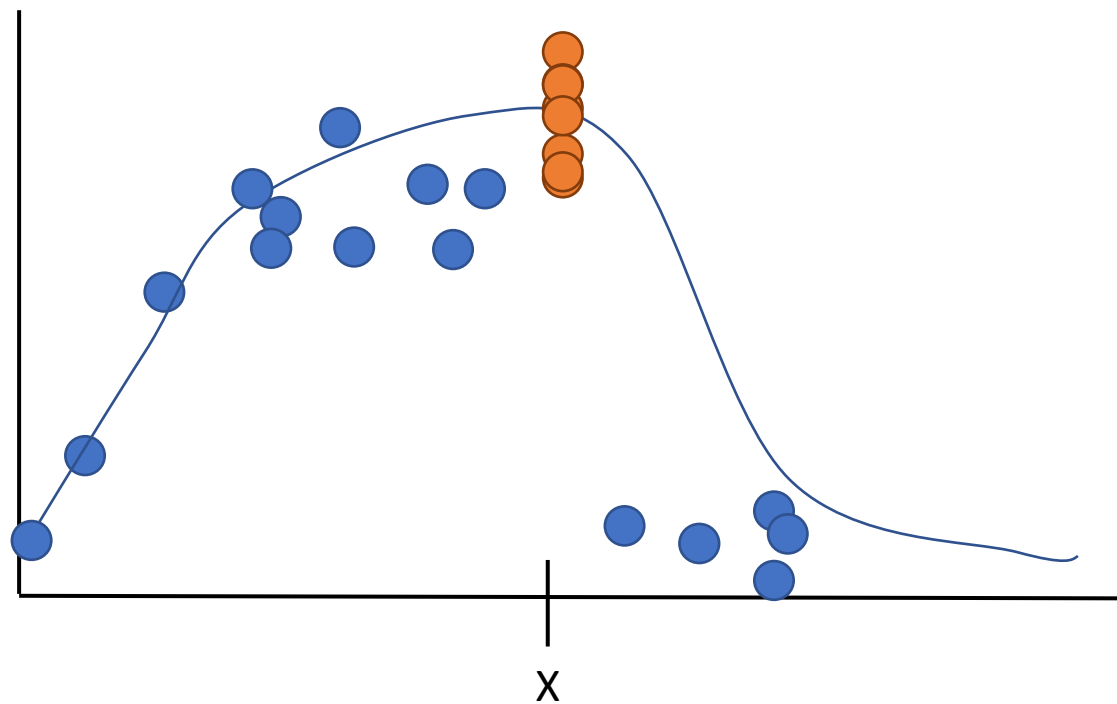
Machine Learning: Function Approximators



Discriminative

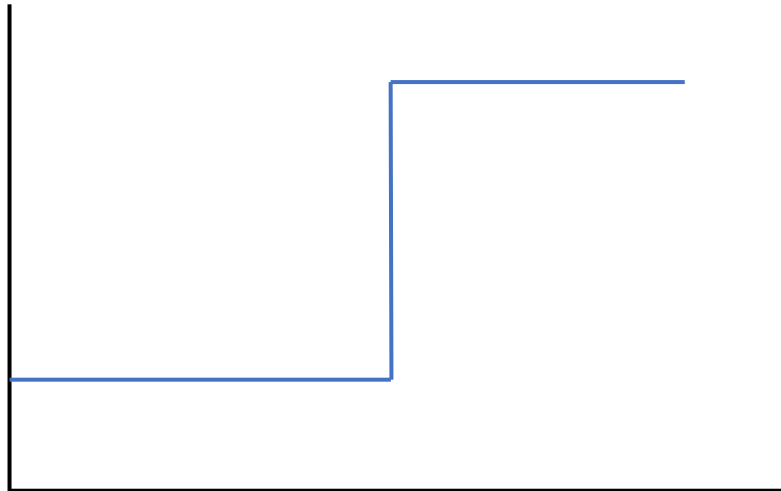


Generative

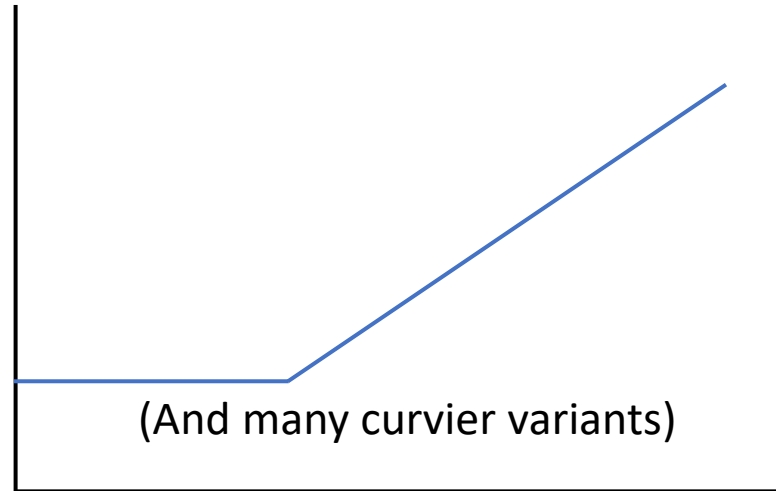


Programming With Data, Building Blocks

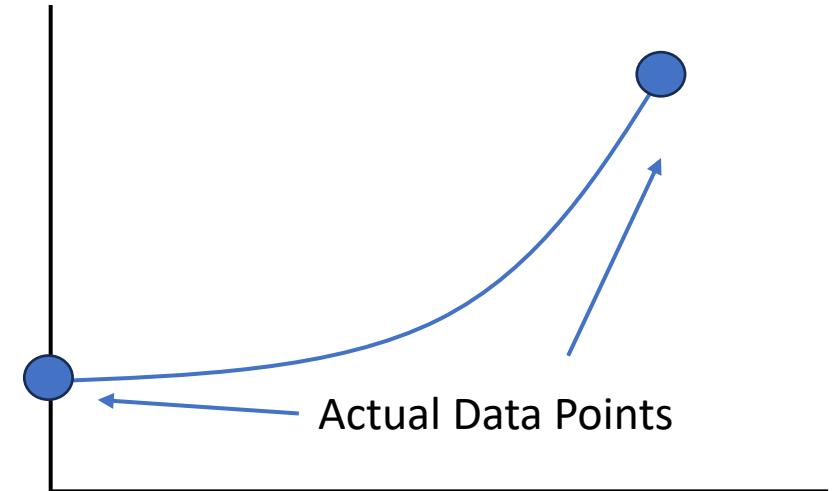
If-Then:
Decision Tree,
Gradient Boosted Random Forest



Artificial Neuron:
Neural Networks, Deep Learning



Case/Instance:
Instance Based Learning, kNN



What Does It Mean to Understand?

- Knowable
- Comprehensible
- Empirical
- Predictable
- Causal
- Counterfactual
- Communicable

Without understandability, we build **intellectual debt**

What Does Most AI/ML Offer?

- Interpretable: Inner workings are understandable, can follow and reproduce the prediction or decision
 - Decision trees, linear regression, GAM, etc.
 - Being eroded to mean something more similar to “explainable”
- Explainable: Ex post characterization of how a model behaves; a justification
 - SHAP, LIME, counterfactual values, surrogate models, feature importance, etc.

FEATURES

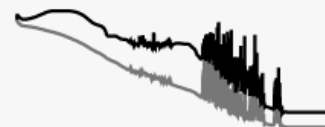
Which properties do you want to feed in?

+ - 4 HIDDEN LAYERS

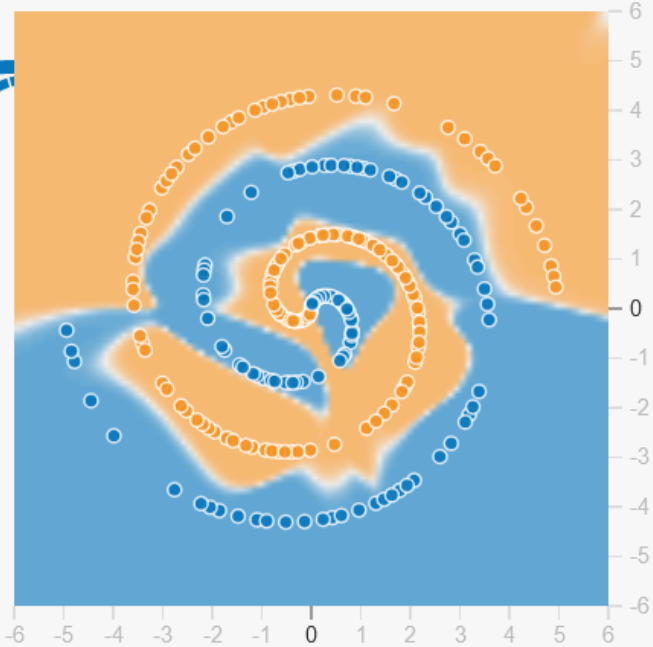
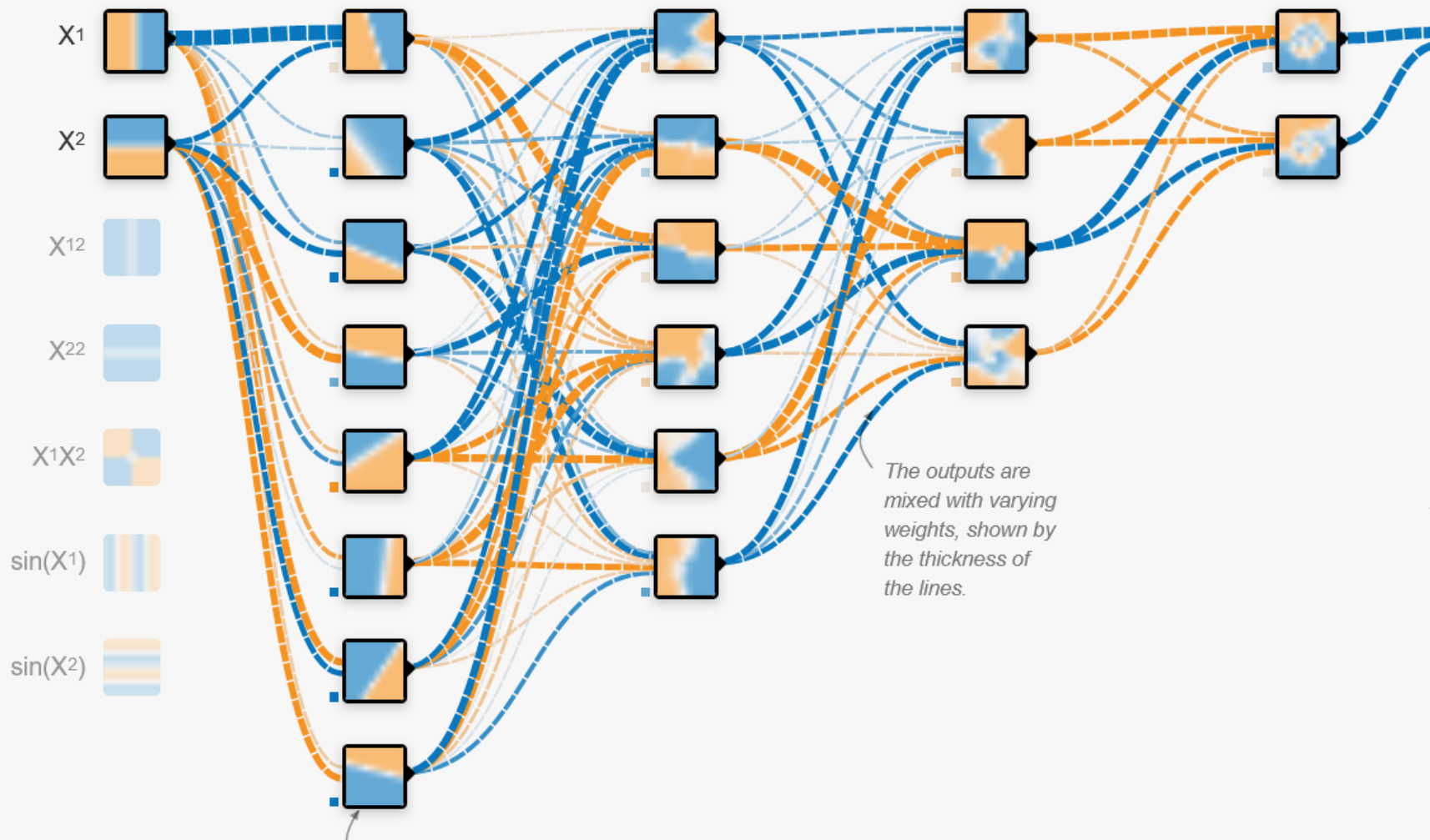
OUTPUT

Test loss 0.069

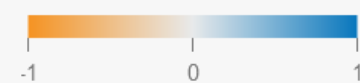
Training loss 0.000



+ - 8 neurons + - 6 neurons + - 4 neurons + - 2 neurons

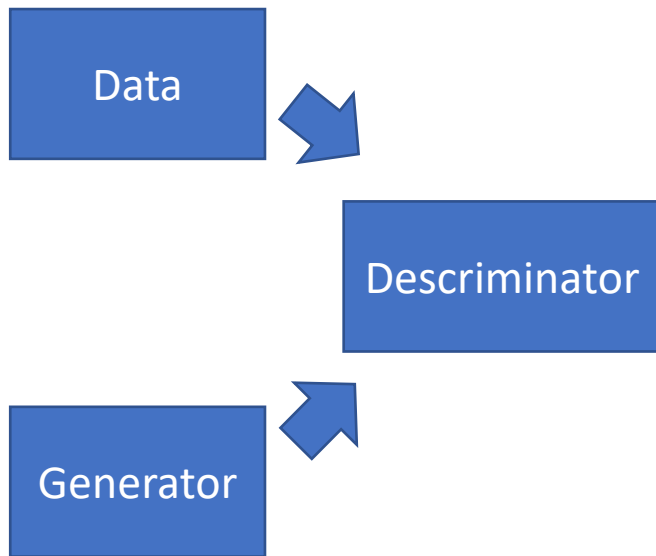


Colors shows data, neuron and weight values.



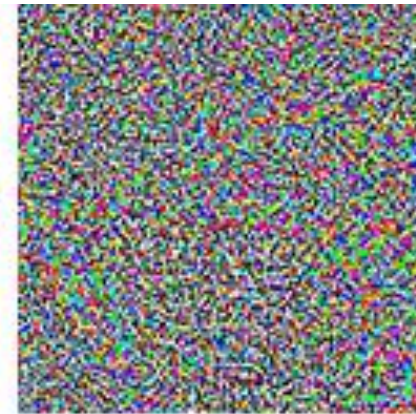
Show test data Discretize output

GANs (Generative Adversarial Networks): Using AI to Attack AI results and explanations



"panda"
57.7% confidence

+ ϵ



=

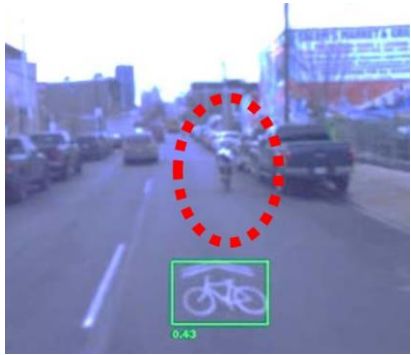


"gibbon"
99.3% confidence

Goodfellow, 2014

Context and Biased Training Data

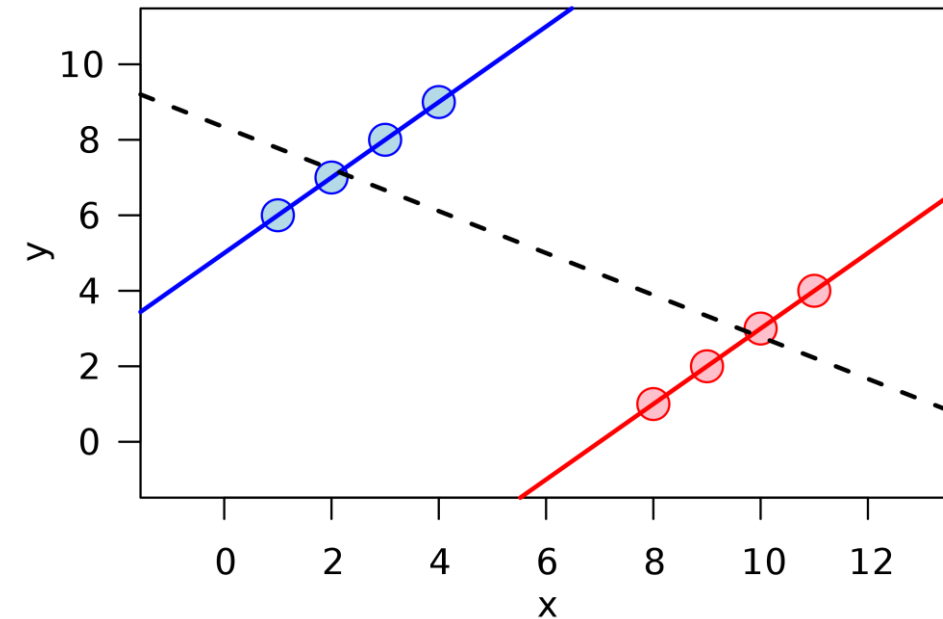
- The algorithm will optimize given the data
- Importance of loss functions: cost of error, symmetry
- Long tail of situations: Ensure sufficient coverage of the real world



Phillip Koopman, SSS 2019 & SafeAI 2019

What's Your AI/ML Model Really Doing?

- All sorts of bias: confirmation bias, Dunning-Kruger, loss aversion, projection bias, survivorship bias, group attribution error, etc.
- Stratified sampling is the answer?
 - Not entirely: Simpson's paradox
 - Sometimes practically impossible
- Testing on the training set
 - (Overfitting)
- Empirical results
 - Smith & Pell, BMJ: "Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomised controlled trials"
 - No evidence

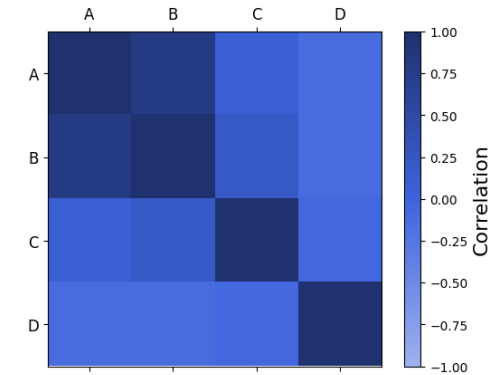
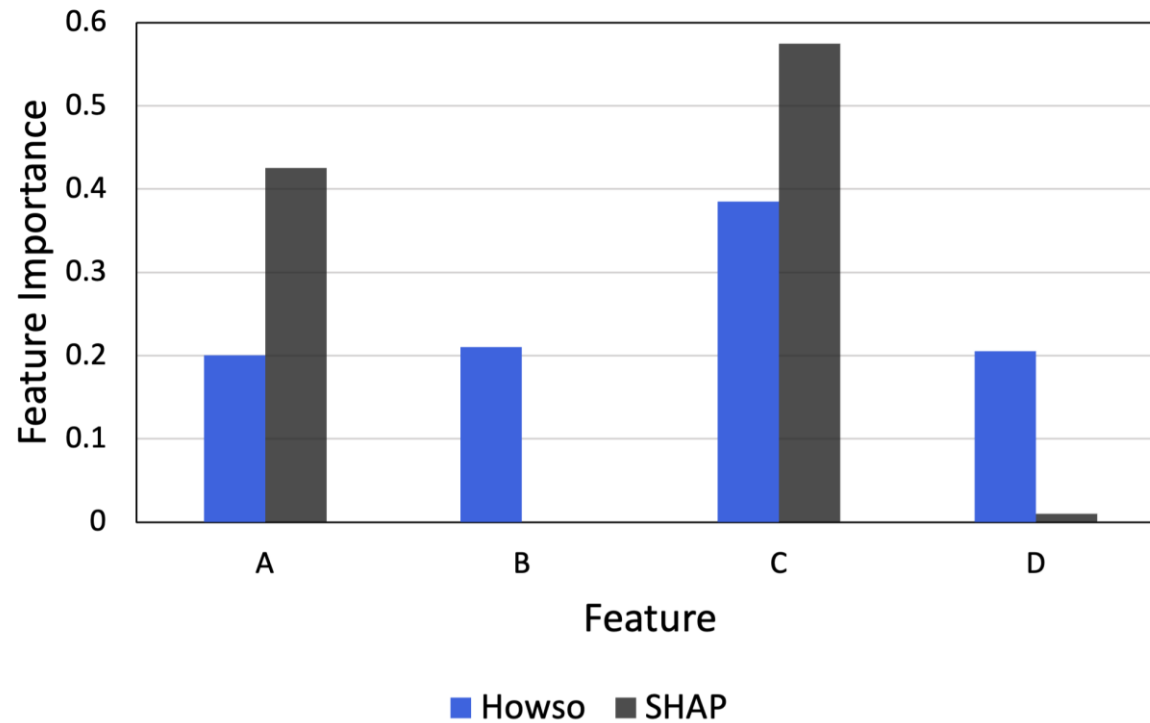


Because...

- “Excuse me, I have 5 pages. May I use the Xerox machine?”
- 60% allowed line skipping
- “Excuse me, I have 5 pages. May I use the Xerox machine, because I’m in a rush?”
- 94% allowed line skipping
- “Excuse me, I have 5 pages. May I use the Xerox machine, because I have to make copies?”
- 93% allowed line skipping

-Langer & Chanowitz, J Personality & Social Psychology, 1978

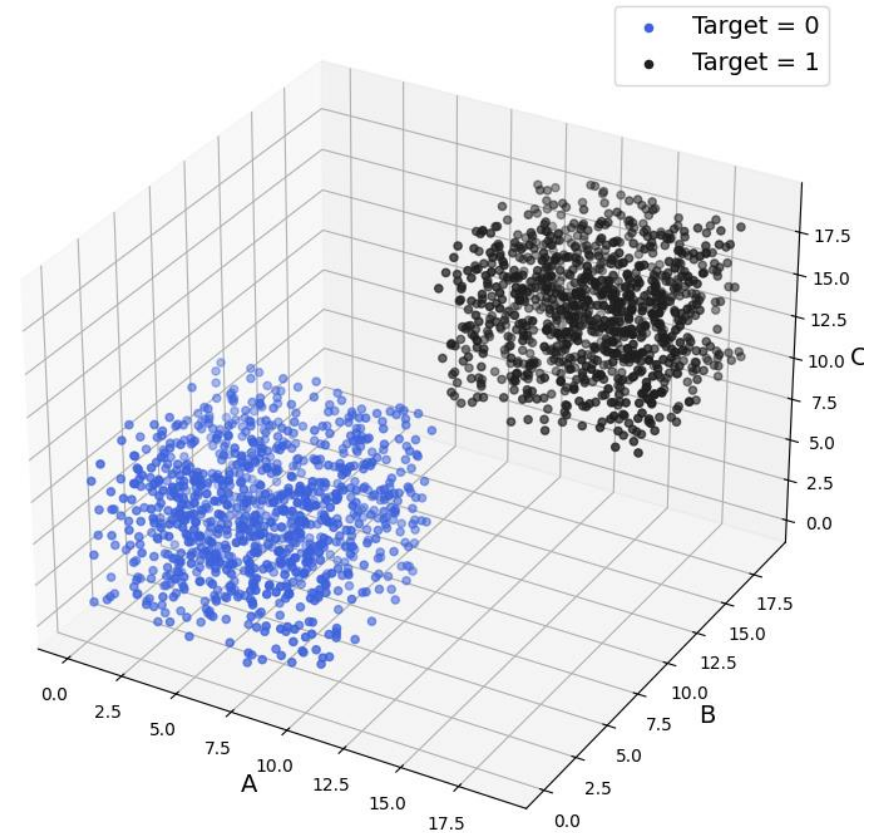
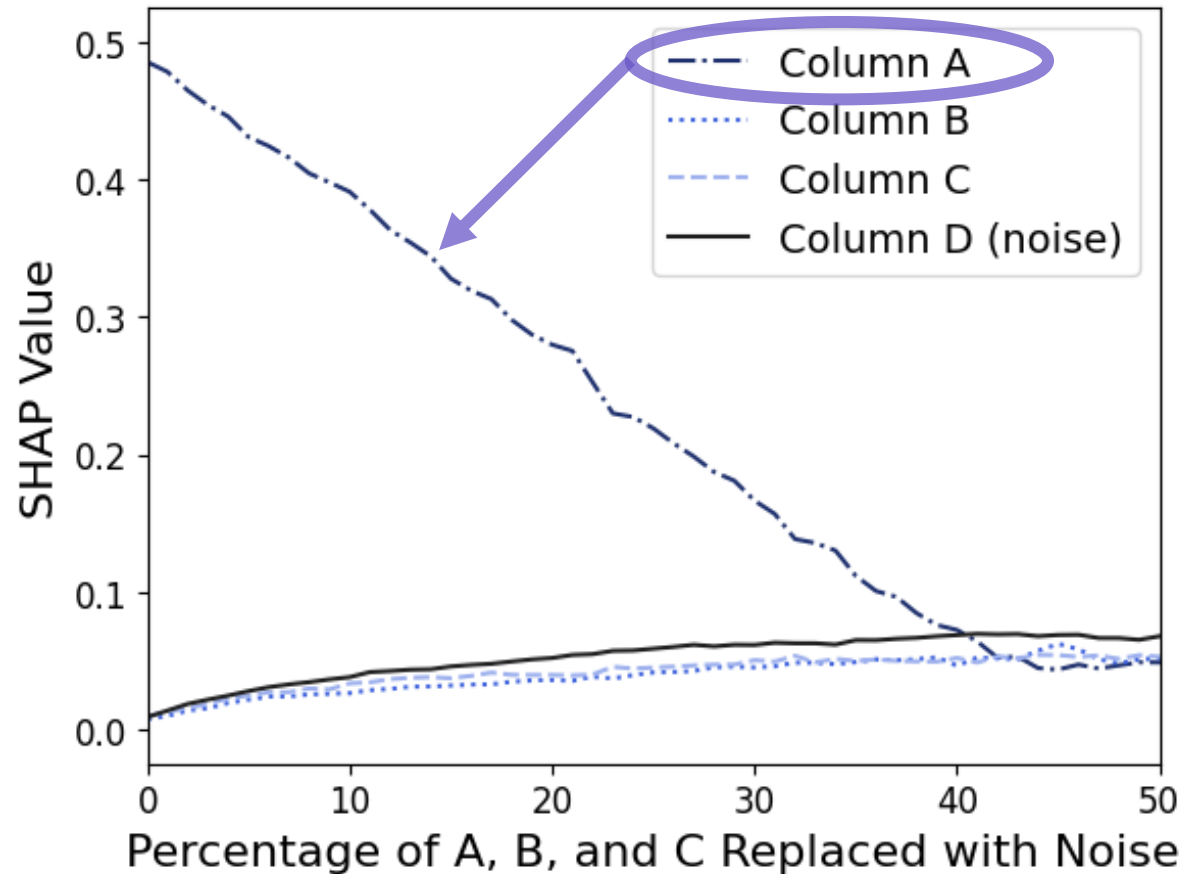
Feature Importance: When SHAP Fails



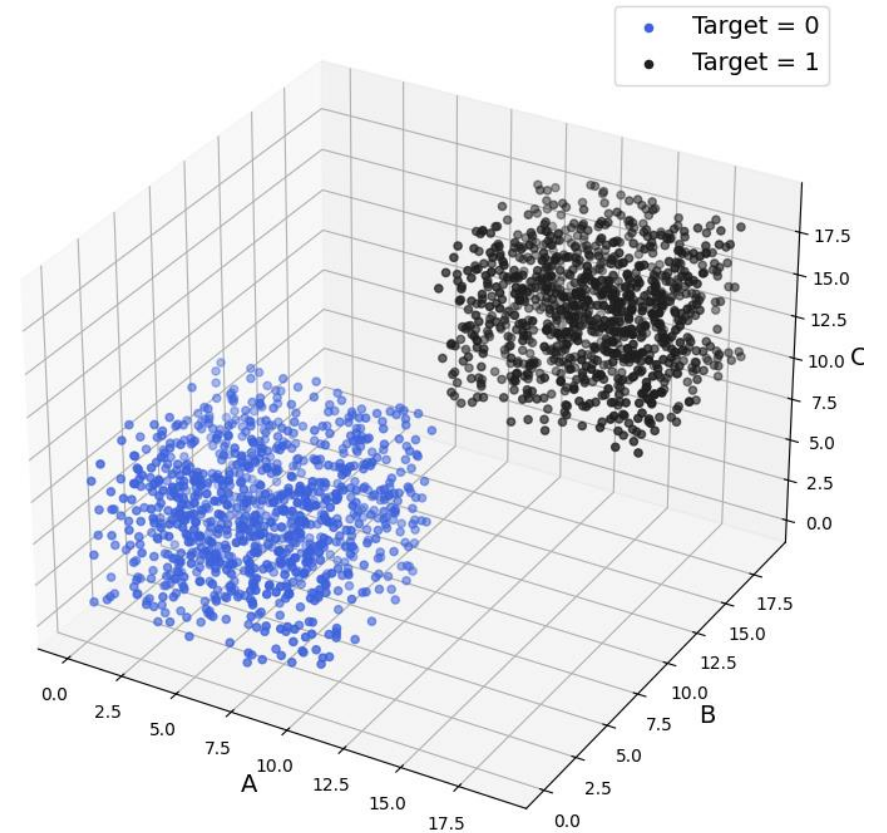
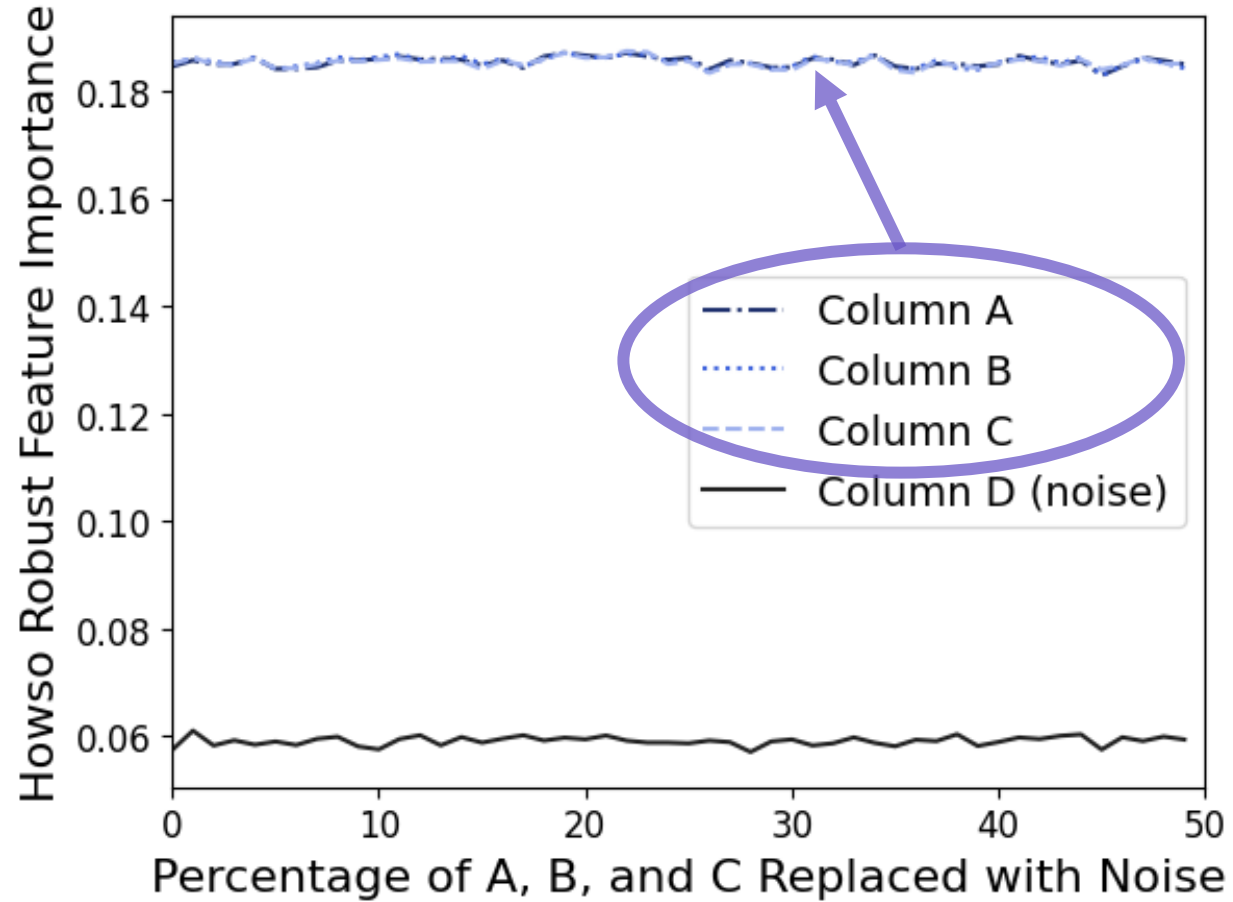
$$\text{Target} = A + B + C + D$$

Feature Importance Can Be Misleading

The Status Quo of SHAP

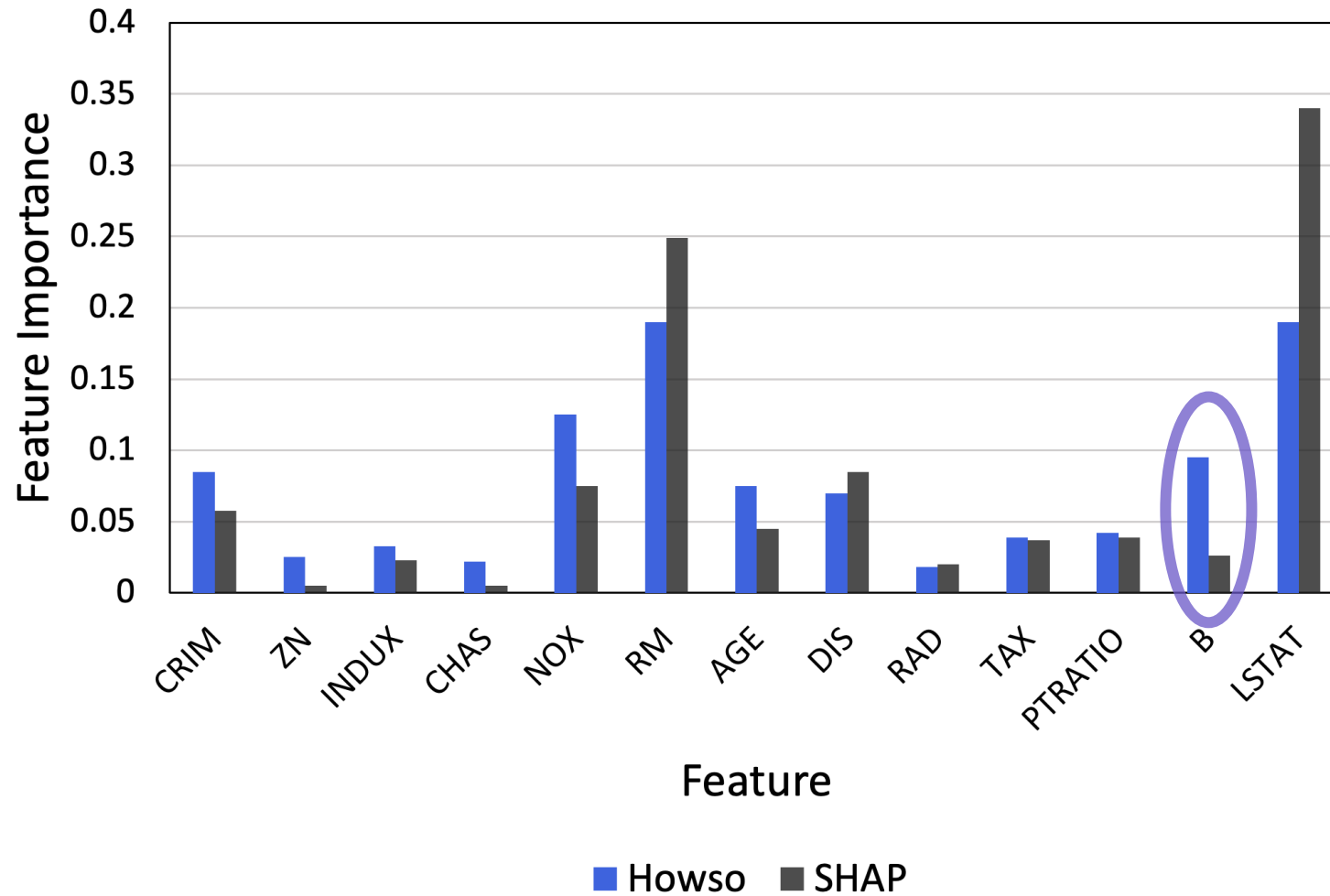


Robust Feature Contributions



Why Getting Feature Attribution Right Matters

Know the Data, Not Just the Model



Two Essential Paths of Understanding

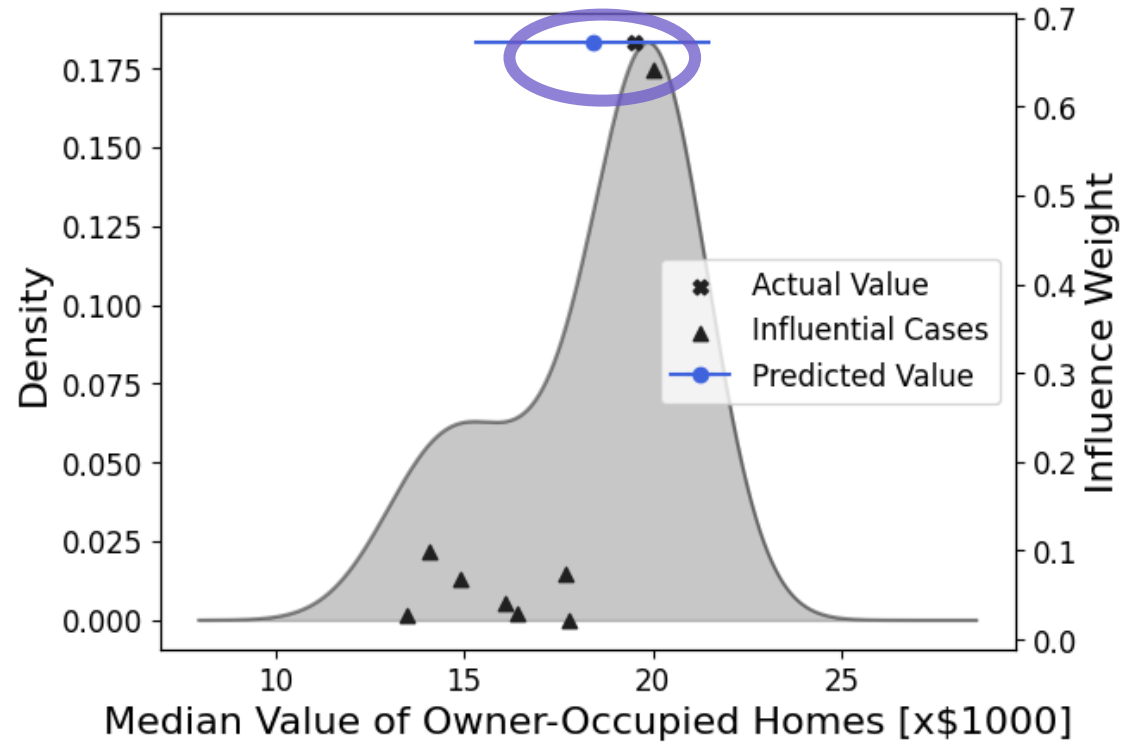
As a Consumer of Insights

- Justifiable
- Corroborate with experience & evidence
- Point to specifics (cases and features)
- Able to answer why and why not?
- Accurate

As a Provider of Insights

- Every step is clear
- Clearly defined units (e.g., ft/sec)
- Clearly defined operations
- Verifiable operations and outcome
- Verifiable provenance and lineage of data

Understandable ML Example



Results from Howso Engine™

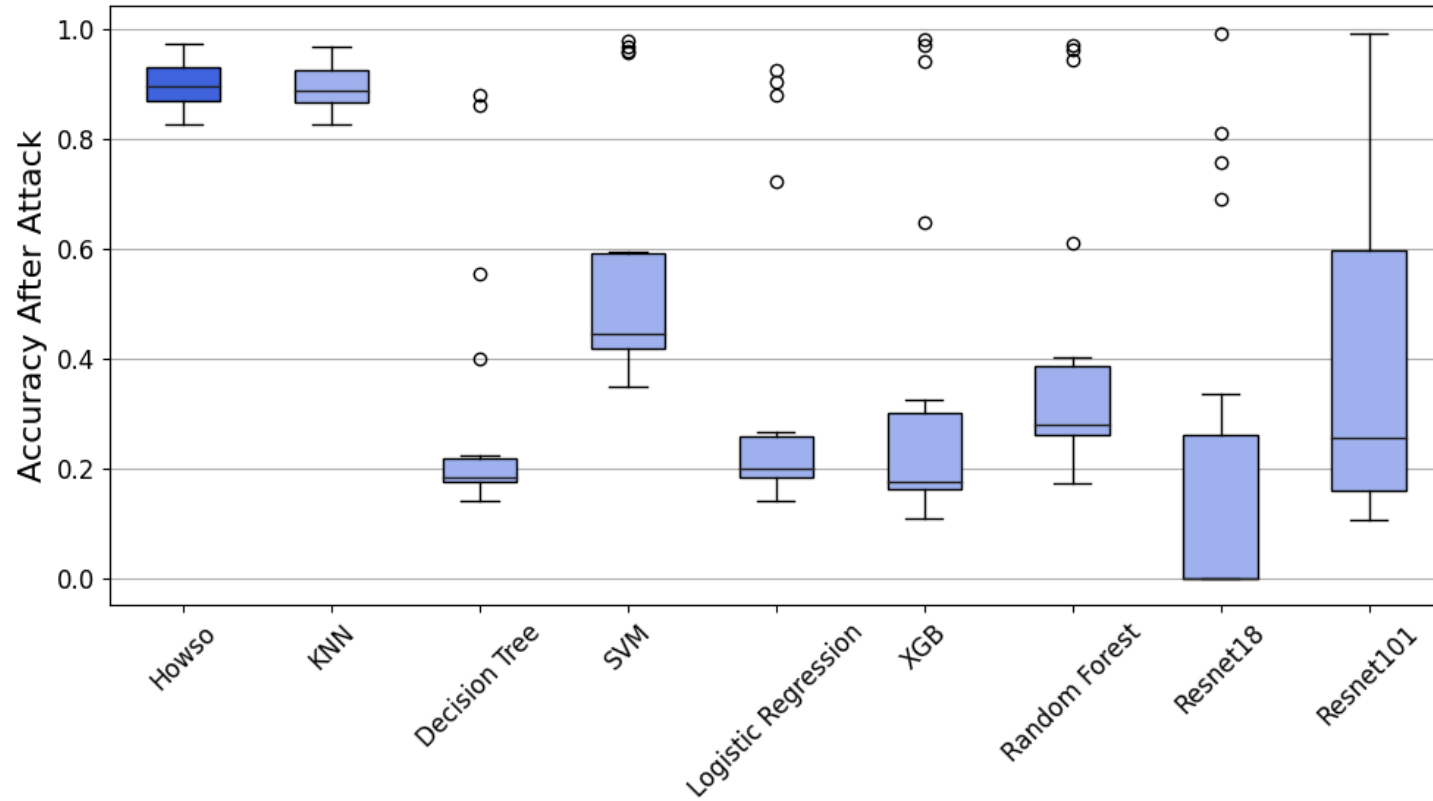
All steps reasonably reproducible manually from the data

Know Your Data: Debuggability

Data Quality Matters

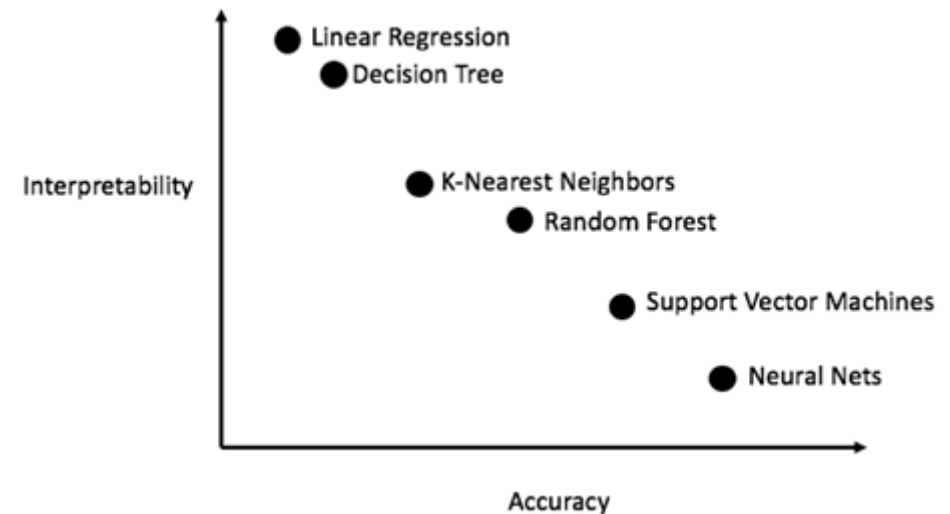


AI Robustness = Security



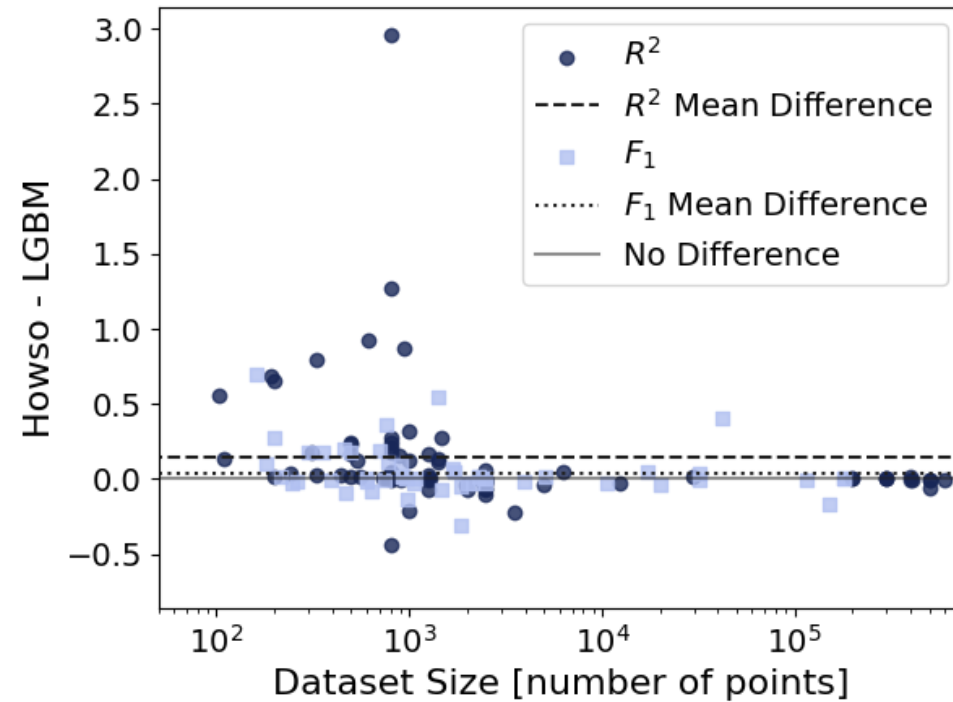
Interpretability & Explainability Myths

- “X is all you need!” e.g., feature importance, counterfactuals.
- Explanations are good enough
 - Problematic bias, adversarial models, high-cost mistakes with insufficient explanations, intellectual debt
- “My model gives me a probability value, so I can use that” without calibration
- Decision trees are accessible



From <https://towardsdatascience.com/model-complexity-accuracy-and-interpretability-59888e69ab3d>

Accuracy – Interpretability Not A Tradeoff Anymore



Influential Data Attribution

Local Feature Contributions	0.26	0	0	0	1.2	0.19	0.24	0.21	0	0	0	0.44	0.29	-
%Influence	CRIM	ZN	INDUS	CHAS	NOX	RM	AGE	DIS	RAD	TAX	PTRATIO	B	LSTAT	Median Value of Home
Actual	6.65	0.0	18.1	0.0	0.71	6.32	83.0	2.73	24	666	20.2	396.9	14.0	19.5
64.0%	6.80	0.0	18.1	0.0	0.713	6.08	84.4	2.72	24	666	20.2	396.9	14.7	20.0
9.82%	9.33	0.0	18.1	0.0	0.713	6.19	98.7	2.26	24	666	20.2	396.9	18.1	14.1
7.42%	3.69	0.0	18.1	0.0	0.713	6.38	88.4	2.57	24	666	20.2	391.4	14.7	17.7
6.86%	7.75	0.0	18.1	0.0	0.713	6.30	83.7	2.78	24	666	20.2	272.2	16.2	14.9
4.09%	5.09	0.0	18.1	0.0	0.713	6.30	91.8	2.37	24	666	20.2	385.1	17.3	16.1

Compression – What's in a Model?

- Diffusion models: 15k-to-1 to 50k-to-1

Original:



Generated:



- Significant memorization found in large models, even diffusion models (Carlini et al., 2023 <https://arxiv.org/pdf/2301.13188.pdf>)
- Some claim fair use for ML training, but memorization can occur
 - Differential privacy may be a plausible fair use
- ChatGPT ~7-to-1. Other LLMs???
- Where did the data come from? Was it rightfully used? Are you sure?
- If transferring the model, is that sufficient transformation?
 - Or do we want organizations only offering SaaS like search?

How Do Free & Open-Source Licenses Apply?

- Source code: Permissive or copyleft software license
- Documentation: Permissive or copyleft documentation license
- Data: Permissive documentation license
(relatively new, e.g., <https://cdla.dev/>)

How Do Free & Open-Source Licenses Apply To Black Box?

- A black box model: ???
- A black box model trained on non-free material without differential privacy mechanisms that may have memorized some of the material: ????
- A black box model trained on AGPLv3 code without differential privacy that emits code derived from AGPLv3: ?????
- The output of one of the above blackbox models: ?????

How Do Free & Open-Source Licenses Apply To Instance-Based Learning?

- An instance-based learning model using data? The data license(s – assuming compatibility)!
- An instance-based learning model that ingests code and can do inference on code? The source code license
- The output of one of the above instance-based models: A license compatible to the data it was trained on

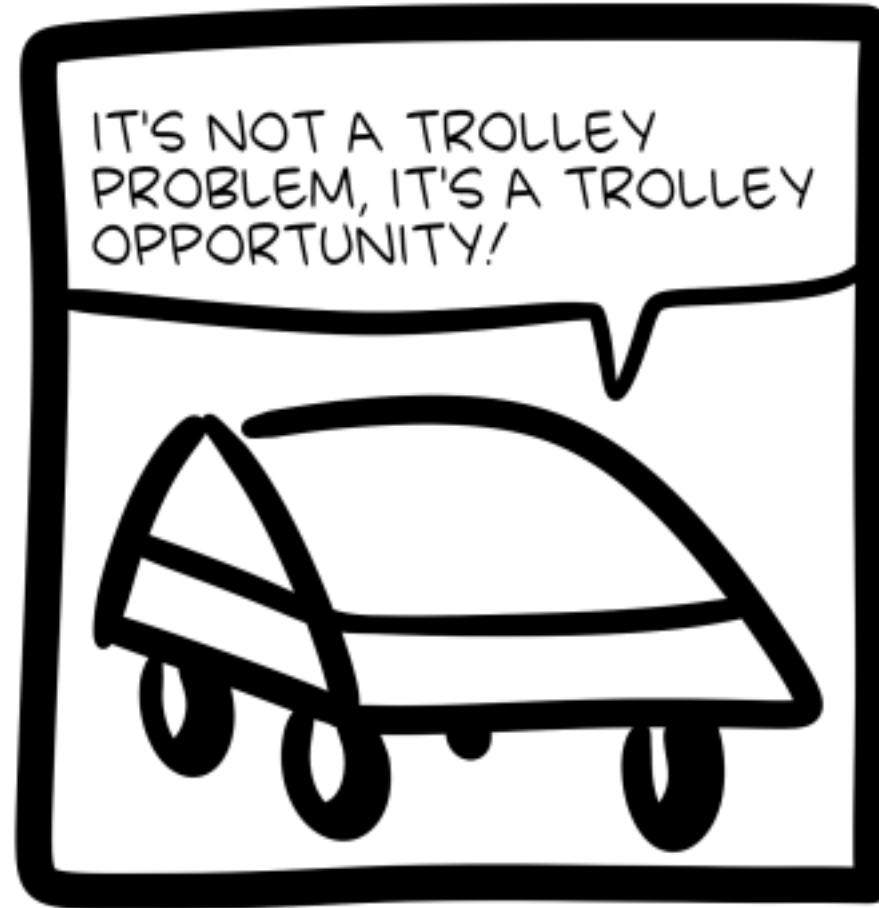
How Can We Fix This?

- Use AI/ML that is understandable, debuggable
- Use AI/ML that is attributable and/or causal
- Know the licenses of all code and data and use appropriately
- Know where you got the data and code from, as well as rights, consent, quality
- Use differential privacy, synthetic data, and other appropriate privacy mechanisms when the data cannot or should not be published
- Publish characteristics and performance
- Publish negative results! (Just like science)
- Consider some exception like the GCC Runtime Library Exception for output of AI/ML
- Let's not compromise open-source fundamentals for short-term practicality! We may have almost all the license structures everything we need right now.

Proposed Data Provenance Standards

<p>SET IDENTIFIER</p> <p>Provenance Metadata Unique ID</p> <p>A unique label identifying the provenance metadata of the current dataset</p>	STANDARD	DESCRIPTION
	Lineage	Identifiers or pointers to metadata representing the data which comprise the current dataset
	Source	Identifies the origin (person, organization, system, device, etc.) of the current dataset
	Legal rights	Identifies the legal or regulatory framework applicable to the current dataset, along with the required data attributions, associated copyright or trademark, and localization and processing requirements
	Privacy and protection	Identifies any types of sensitive data associated with the current dataset and any privacy enhancing techniques applied
	Generation date	Timestamp marking the creation of the current dataset
	Data type	Identifies the data type contained in the current set, and provides insights into how the data is organized, its potential use cases, and the challenges associated with handling and using it
	Generation method	Identifies how the data was produced (data mining, machine-generated, IoT sensors, etc.)
	Intended use and restrictions	Identifies the intended use of the data and which downstream audiences should not be allowed access to the current dataset

Ethics!



GPT-4, probably, with the right prompt

<https://www.smbc-comics.com/comic/decisions> -- press the red button

Thank you!

HOW/SO[™]

github.com/howsoai