BISHOP FOX

Weaponizing Machine Learning

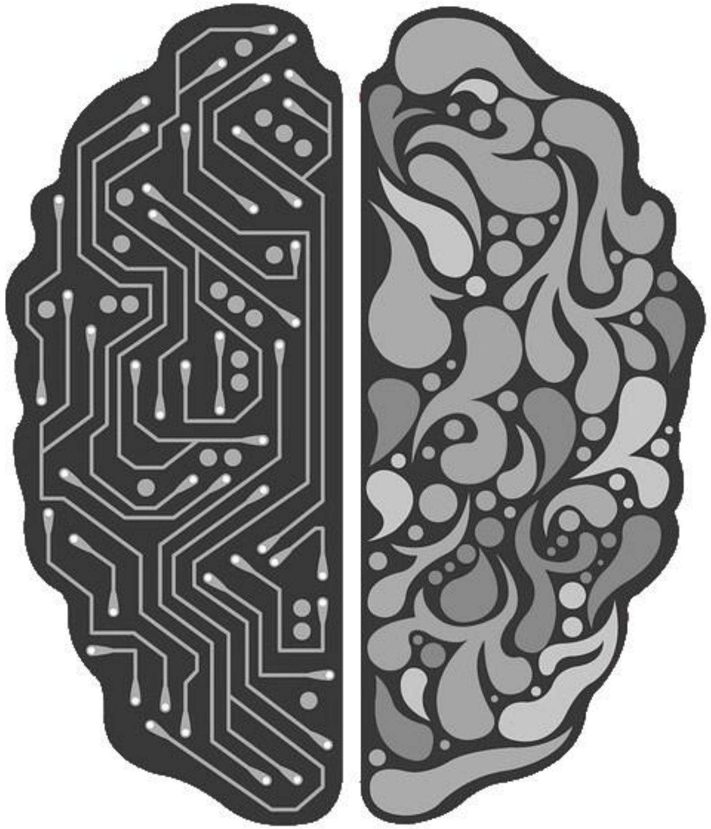
HUMANITY WAS OVERRATED ANYWAY

DEF CON

July 30, 2017

DeepHack

ISN'T THAT KIND OF DANGEROUS?

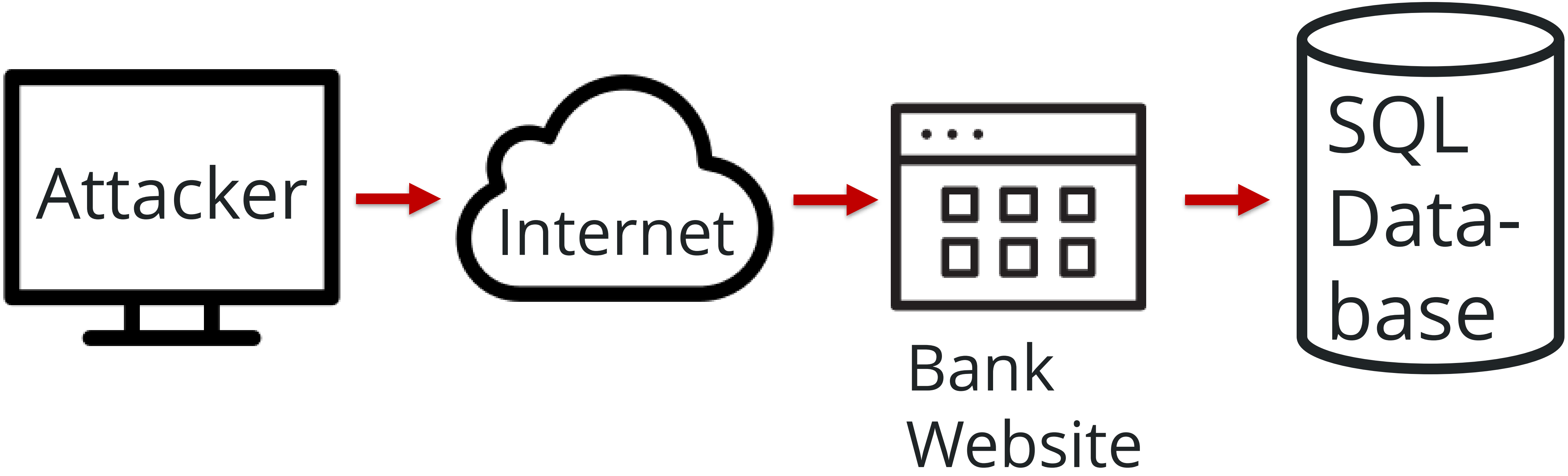


Artificial Intelligence

Hacking

The Scenario

I'D LIKE TO OPEN AN ACCOUNT

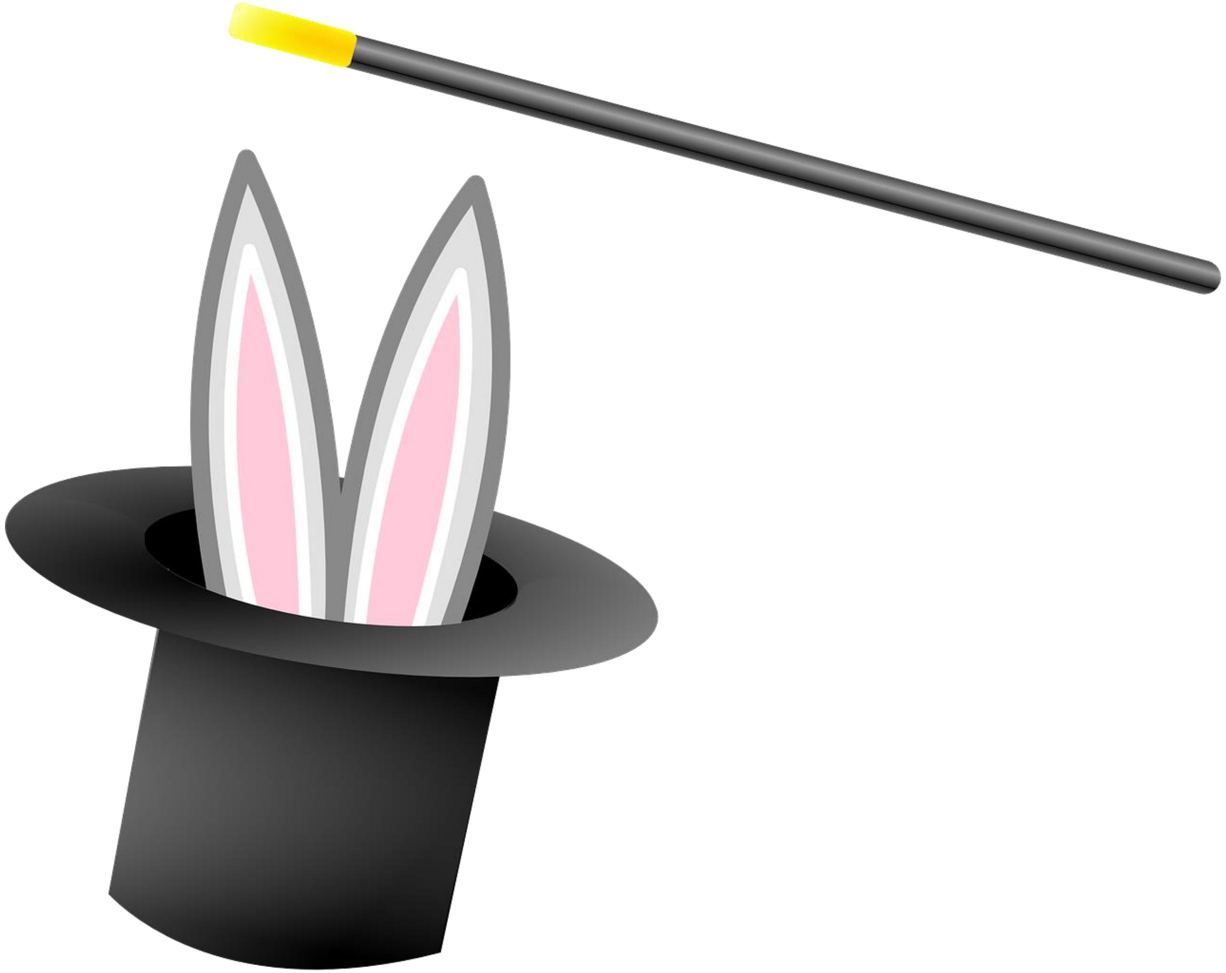


DEMONSTRATION

CROSSES FINGERS

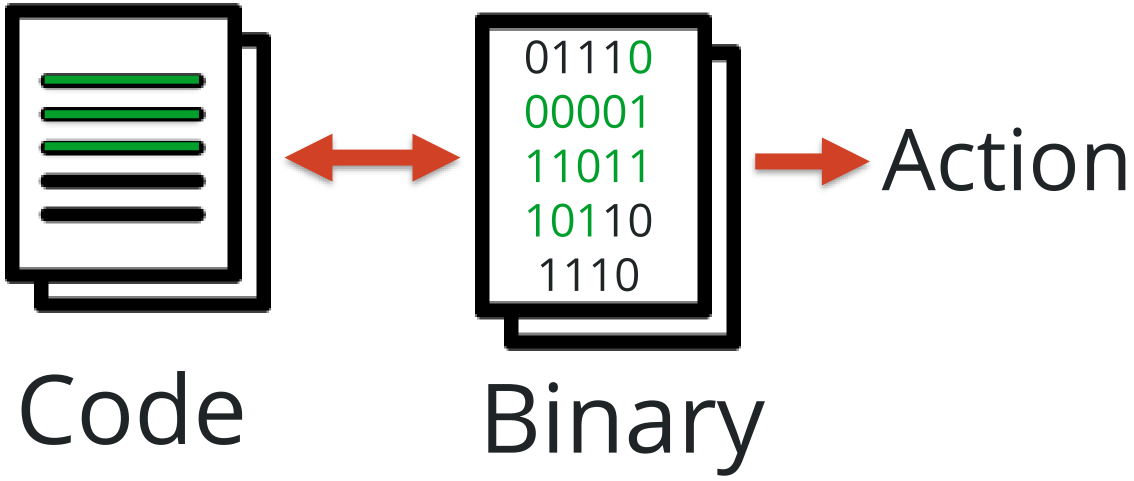
No Tricks Up Our Sleeve

FOOL ME ONCE



The Model

HE'S SO HOT RIGHT NOW



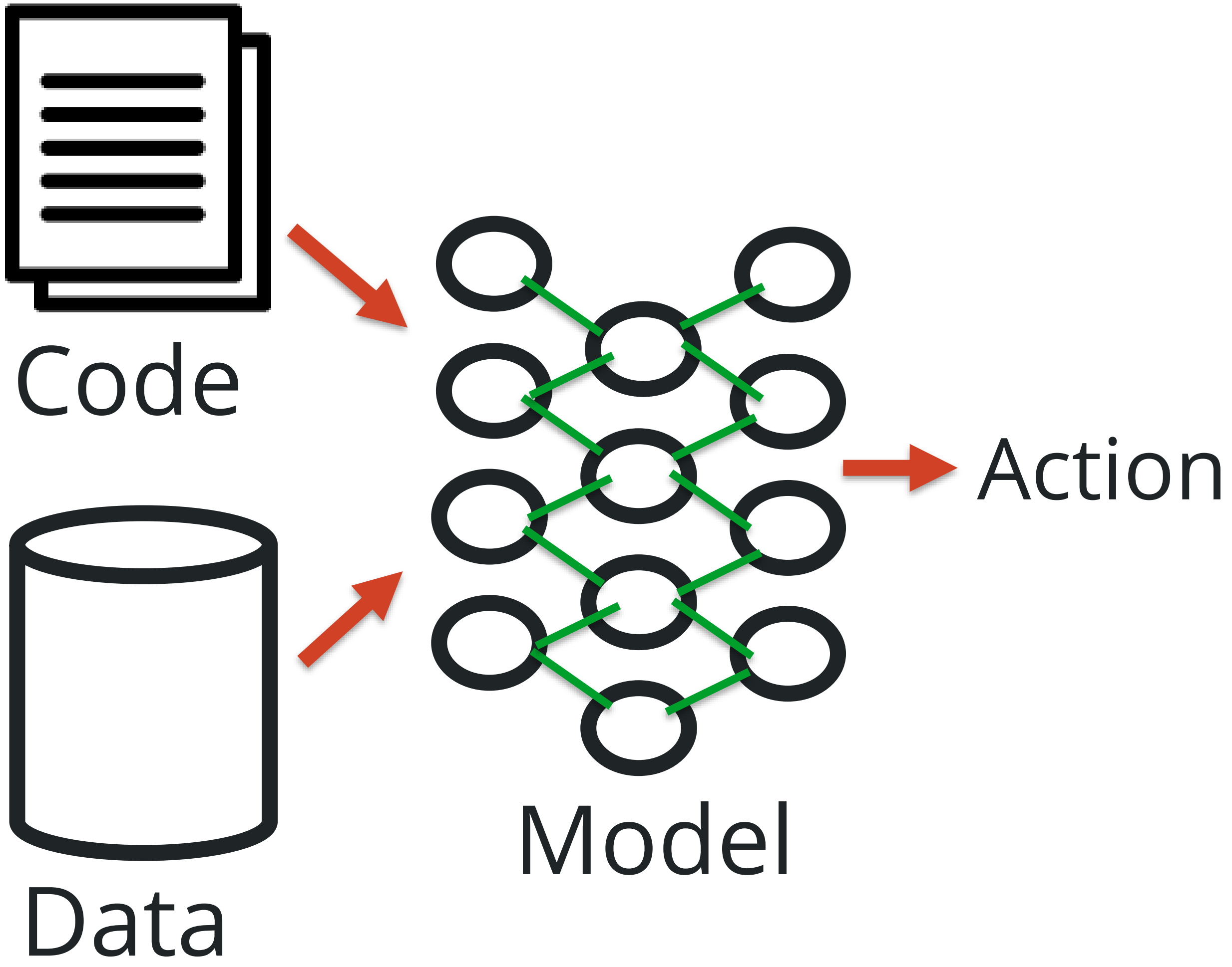
01110
00001
11011
10110
1110

Code

Binary

Action

Regular Program



Code

Data

Model

Action

AI Program

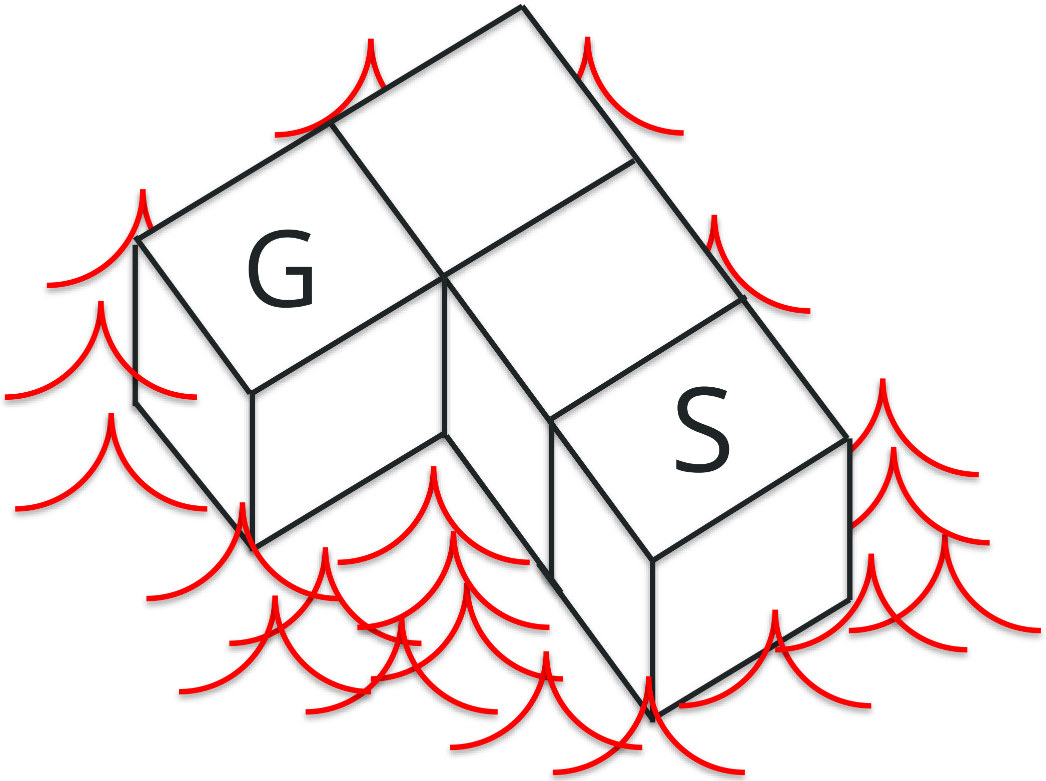
AI Programming

NOBODY LIKED LISP ANYWAY

1. Object Oriented
2. Functional
3. Machine Learning

Machine Learning 101

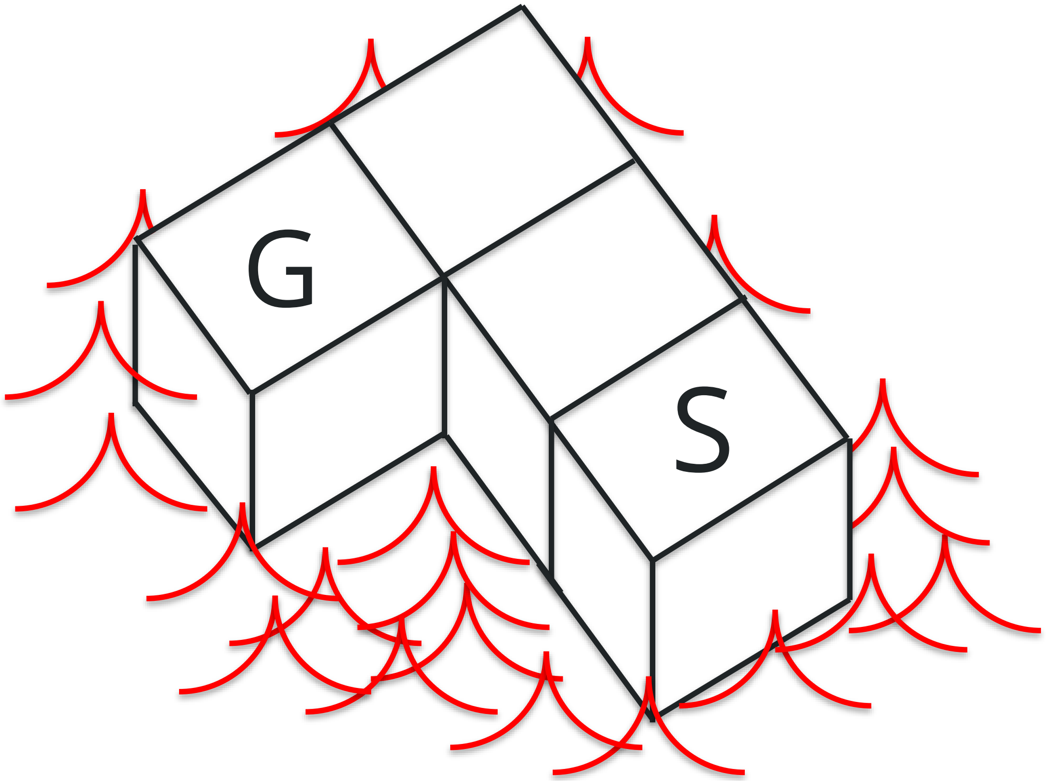
THAT'S SOME HOT LAVA



Maze Solving Robot

Machine Learning 101

A WILD TABLE APPEARS

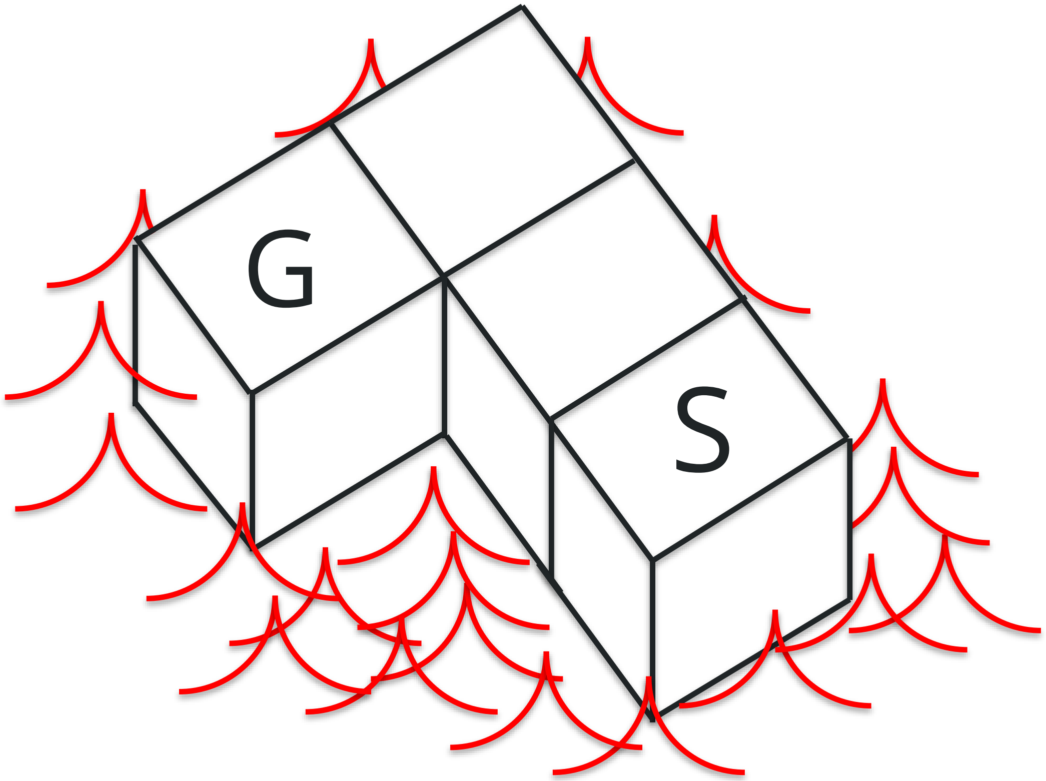


Maze Solving Robot

State	Action	Reward
0,0	Up Right Down Left	
0,1	Up Right Down Left	
0,2	Up Right Down Left	

Machine Learning 101

FELL INTO THE LAVA

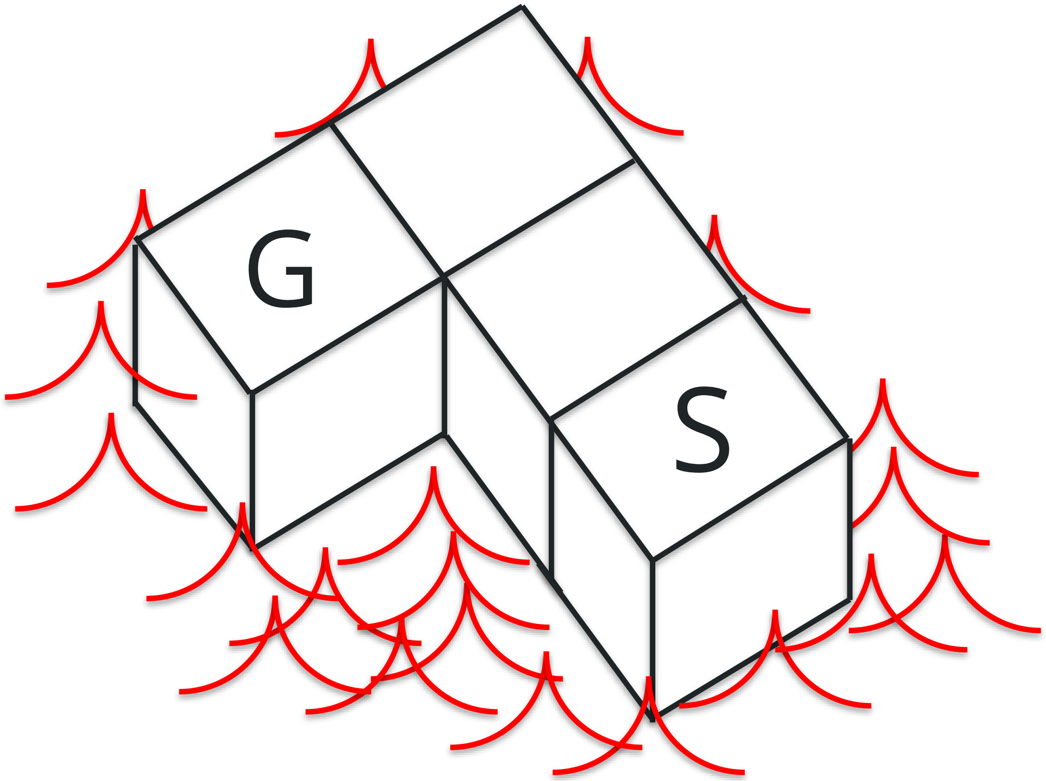


Maze Solving Robot

State	Action	Reward
0,0	Up Right Down Left	-50
0,1	Up Right Down Left	
0,2	Up Right Down Left	

Machine Learning 101

GOT FURTHER

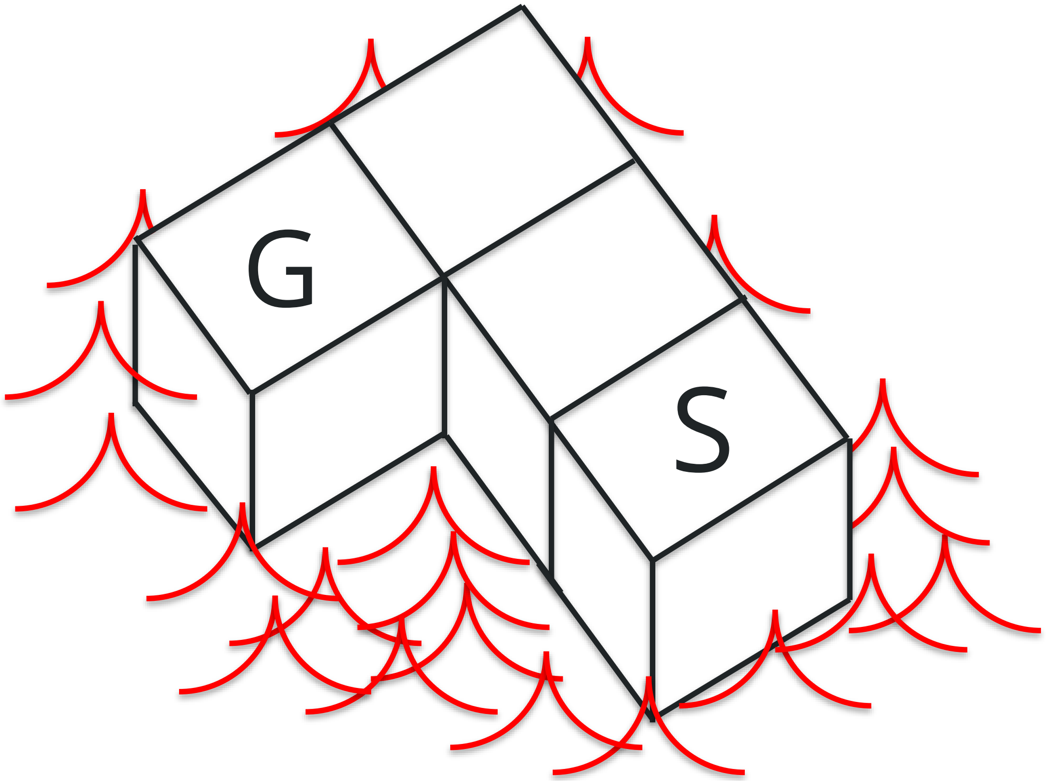


Maze Solving Robot

State	Action	Reward
0,0	Up Right Down Left	-1 -50
0,1	Up Right Down Left	
0,2	Up Right Down Left	

Machine Learning 101

FELL INTO THE DAMN LAVA AGAIN

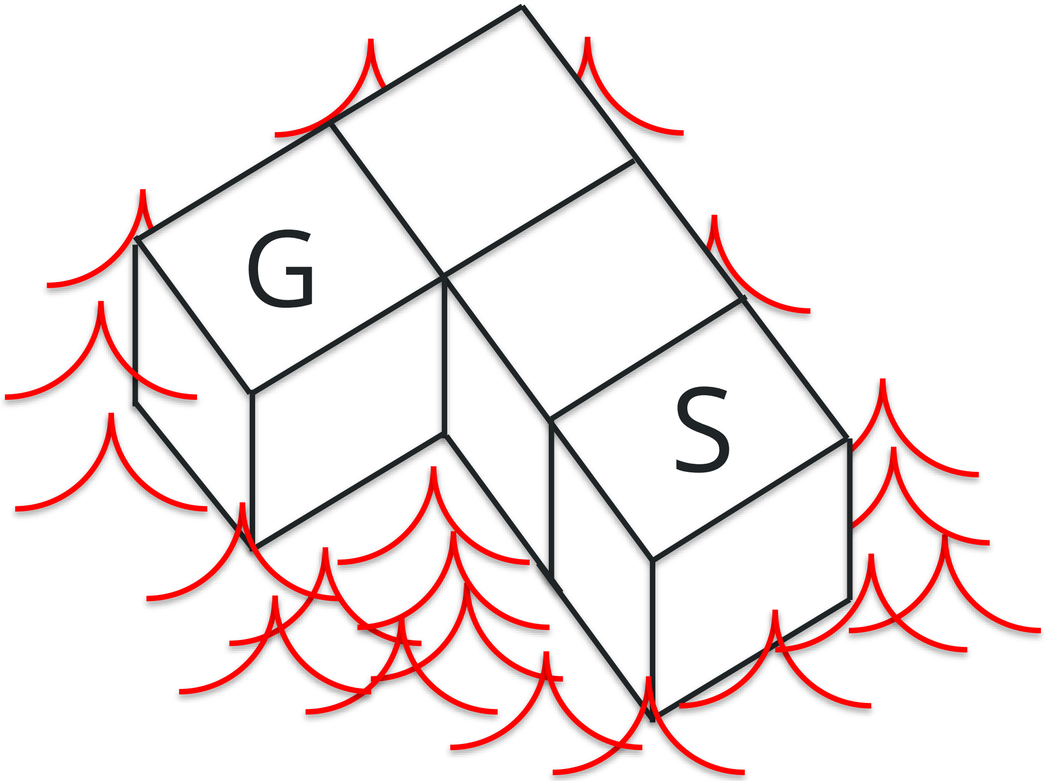


Maze Solving Robot

State	Action	Reward
0,0	Up Right Down Left	-1 -50
0,1	Up Right Down Left	-50
0,2	Up Right Down Left	

Machine Learning 101

FINISHED!

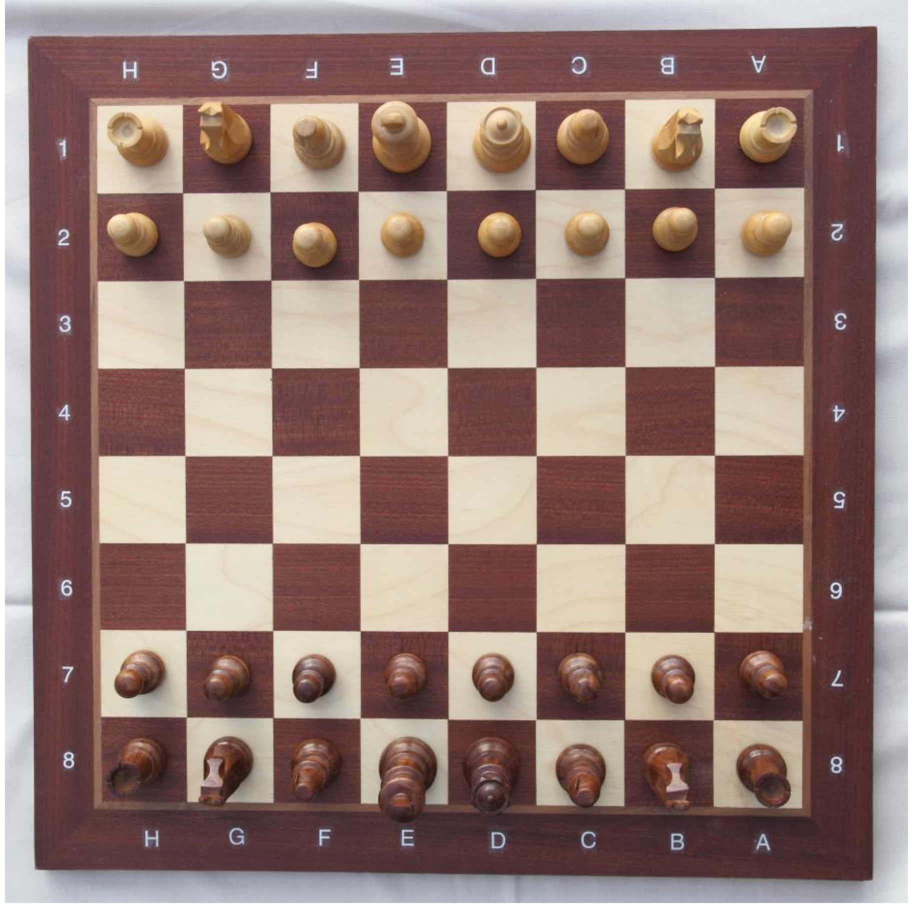


Maze Solving Robot

State	Action	Reward
0,0	Up Right Down Left	-1 -50 -50 -50
0,1	Up Right Down Left	-1 -50 -1 -50
0,2	Up Right Down Left	-50 -50 -1 +50

Machine Learning 101

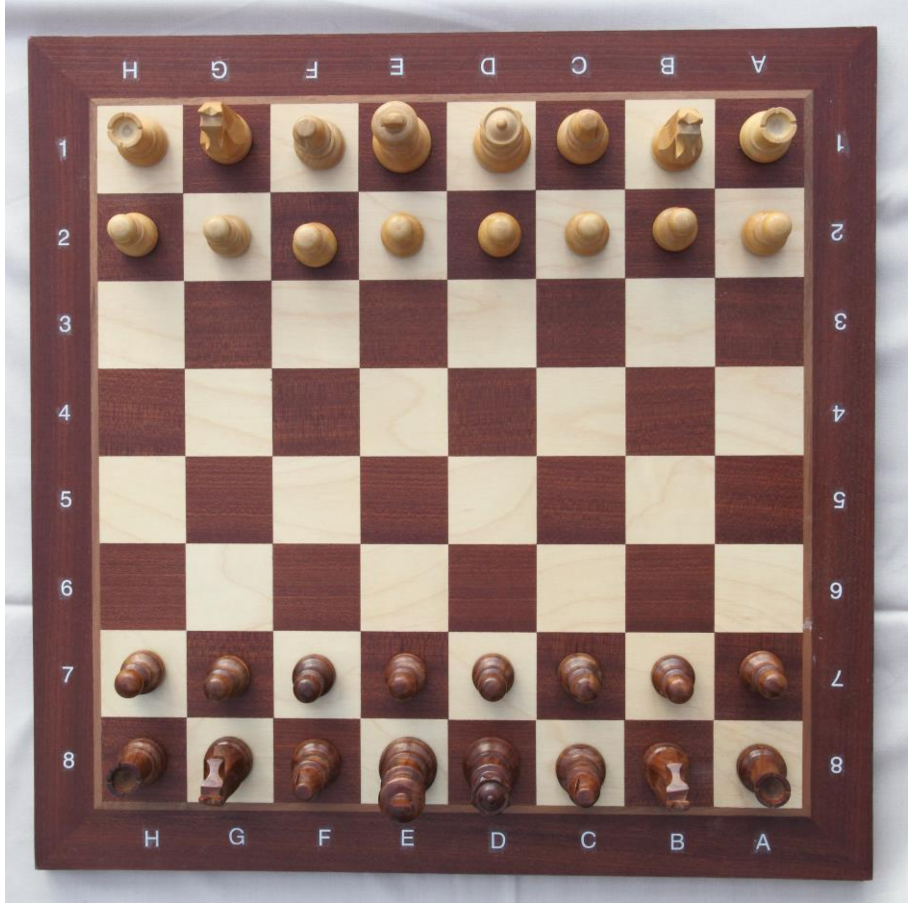
THIS LOOKS HARD



Chess Playing Robot

Machine Learning 101

THIS LOOKS HARD



Chess Playing Robot

State	Action	Reward
???	??? ??? ??? ???	-?? -?? -?? -??

~10^{47} States in Chess

Can’t store it all

Function Approximation

IS NOT MAGIC

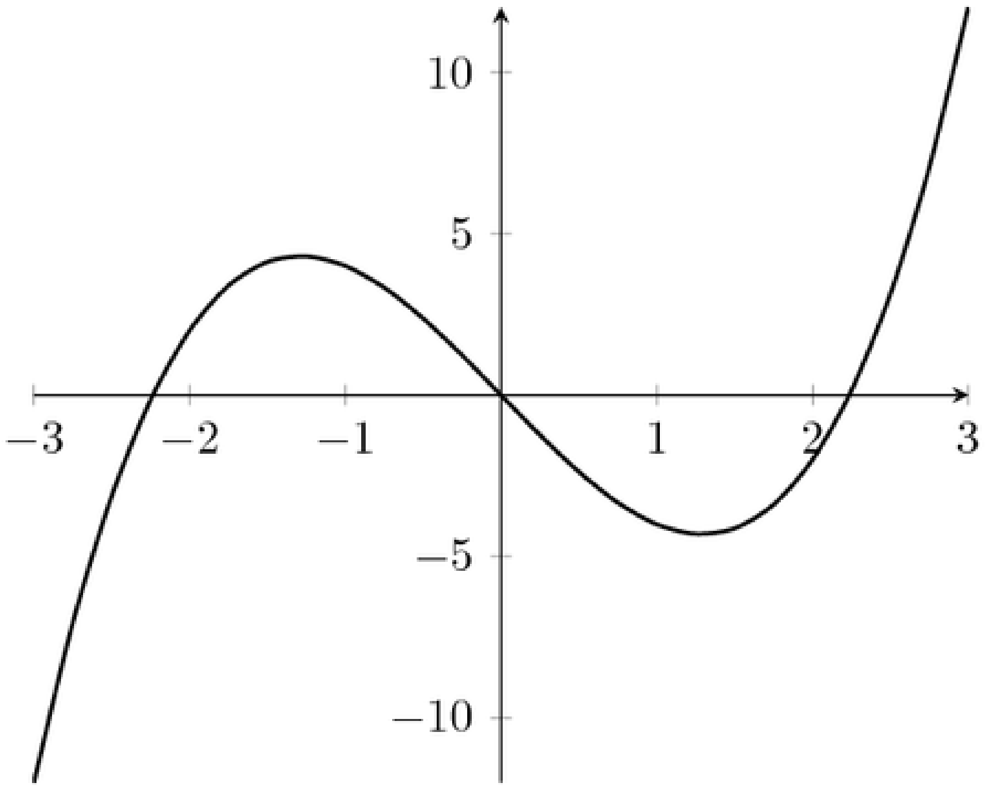
State	Action	Reward
0,0	Up	-1
	Right	-50
	Down	-50
	Left	-50
0,1	Up	-1
	Right	-50
	Down	-1
	Left	-50
0,2	Up	-50
	Right	-50
	Down	-1
	Left	+50

Function Approximation

IS NOT MAGIC

State	Action	Reward
0,0	Up Right Down Left	-1 -50 -50 -50
0,1	Up Right Down Left	-1 -50 -1 -50
0,2	Up Right Down Left	-50 -50 -1 +50

Is ***just*** one of these:



(A function)

Math

NOT EVEN ONCE

STOCHASTIC GRADIENT DESCENT

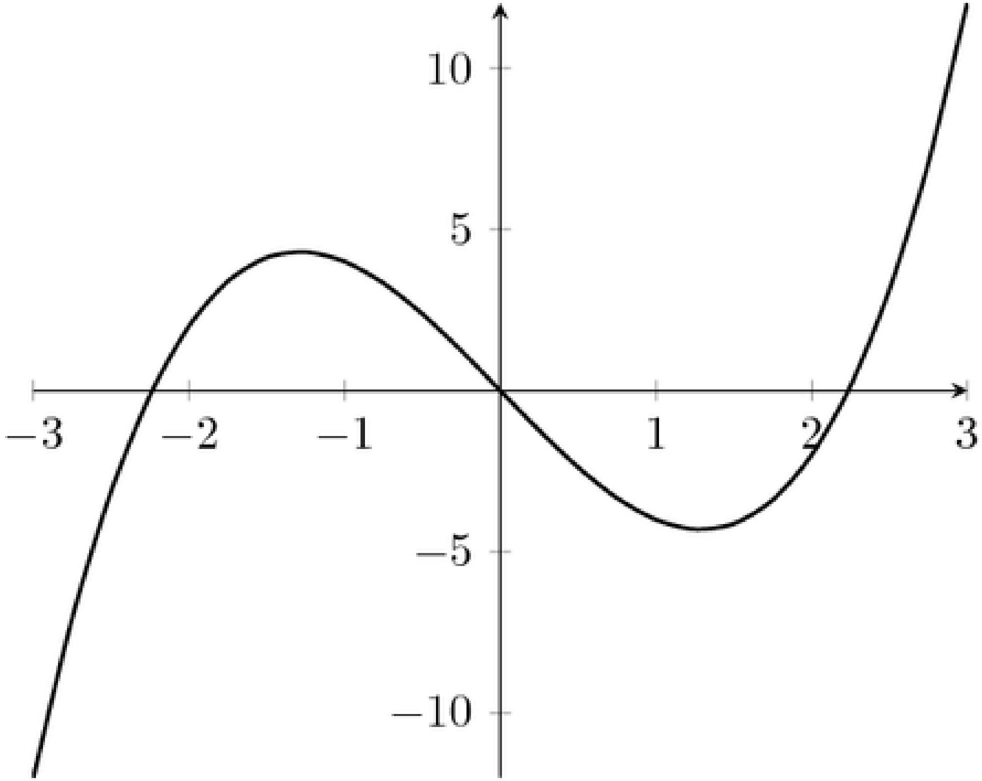
BILINEAR INTERPOLATION

LINEAR REGRESSION

CONVOLUTIONAL TRANSPOSITION

Function Approximation

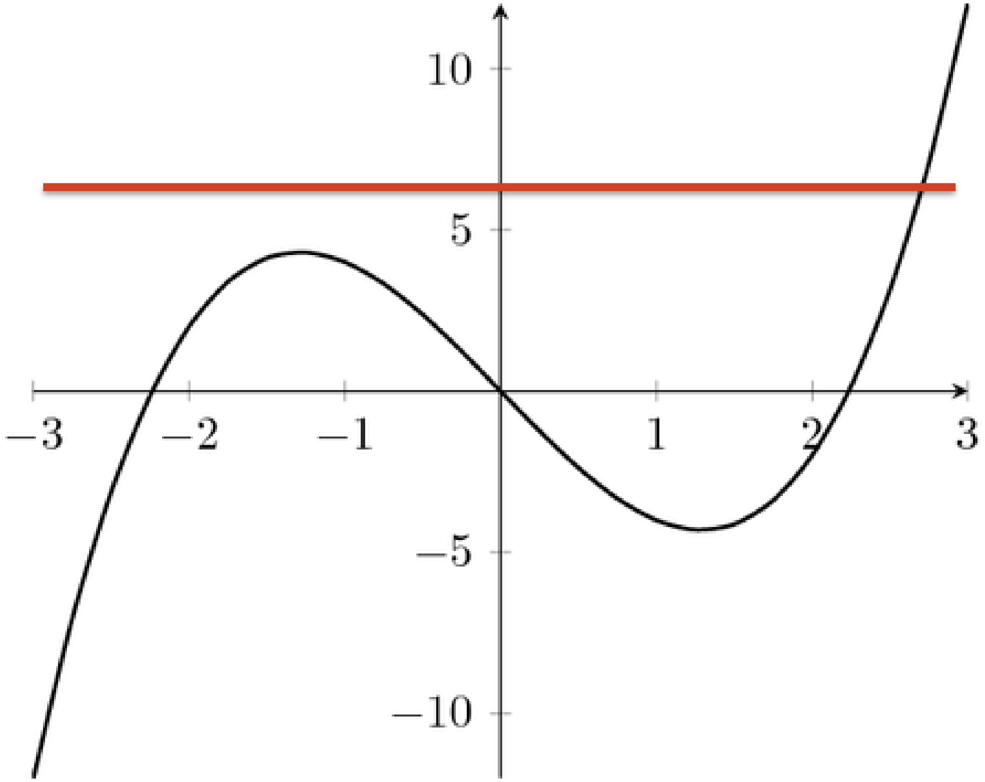
IS NOT MAGIC



(A function)

Function Approximation

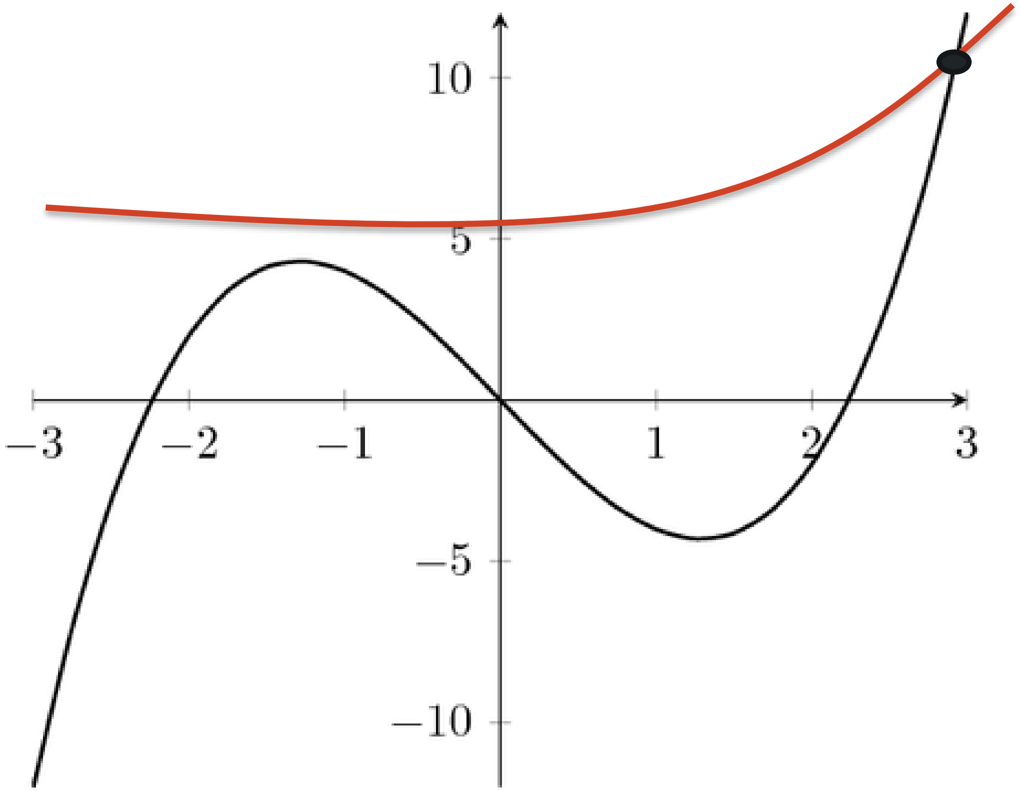
IS NOT MAGIC



(A function)

Function Approximation

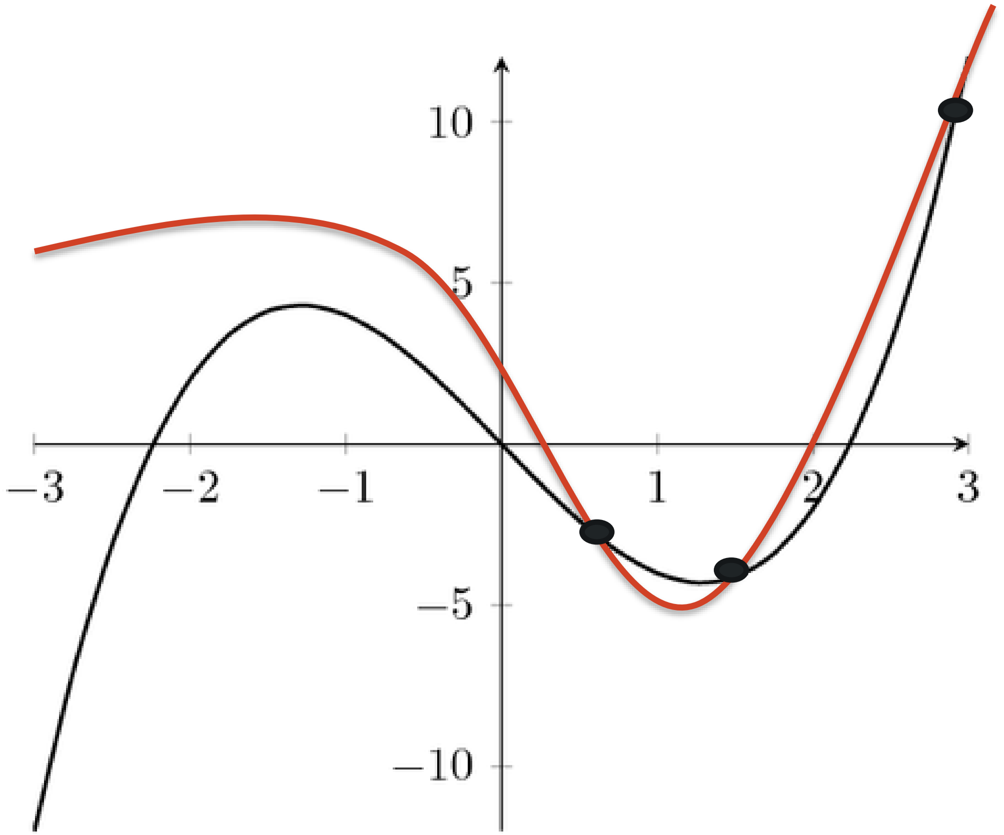
IS NOT MAGIC



(A function)

Function Approximation

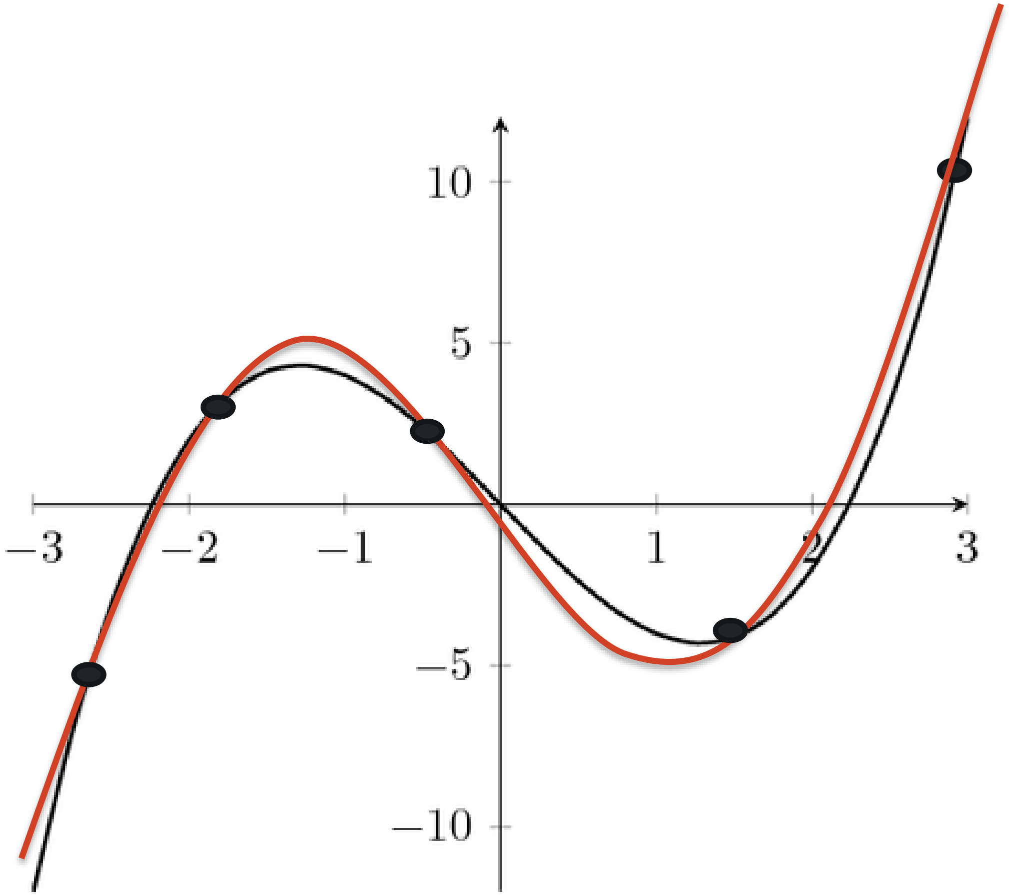
IS NOT MAGIC



(A function)

Function Approximation

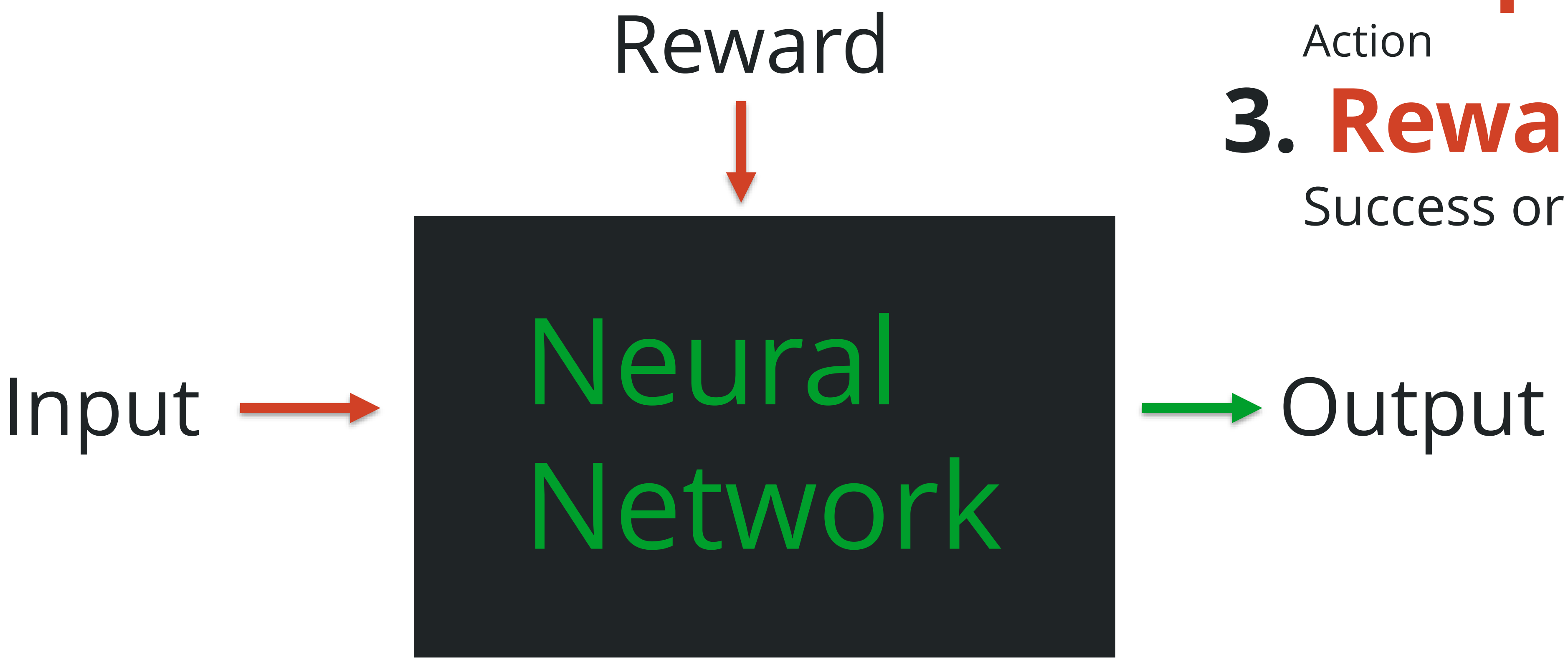
IS NOT MAGIC



(A function)

The Neural Network

REINFORCEMENT LEARNING

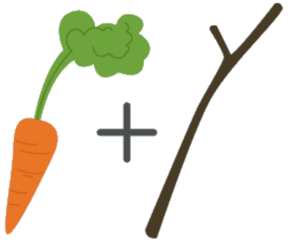


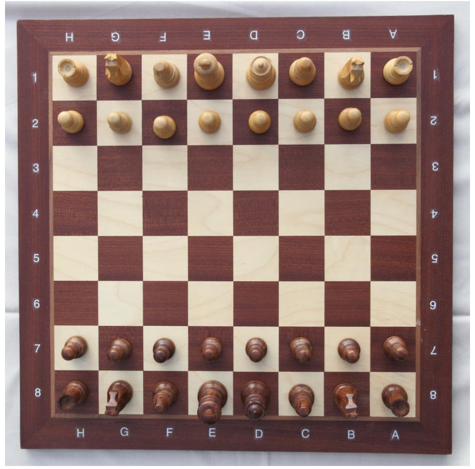
1. **Input**
   Environment
2. **Output**
   Action
3. **Reward**
   Success or Failure

Example: Chess

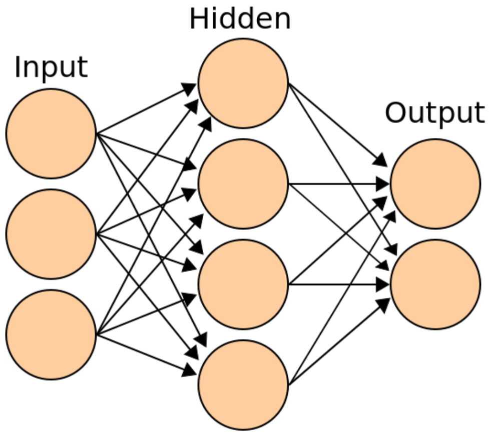
IS NOT MAGIC

Reward: Material gained / lost





Input: Current Piece Positions



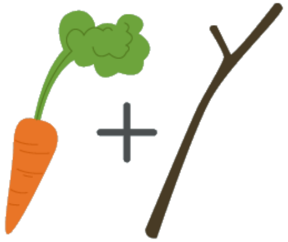
Neural Network

Output: Move one piece

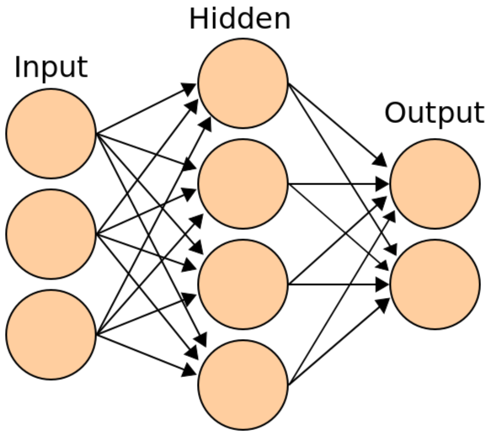
Example: DeepHack

TEXT GENERATION

Reward: HTTP Status (200/500)



SELECT * FRO



Neural Network

M

Input: Query String

Output: Next Character

Autocomplete Game

DO TRY THIS AT HOME

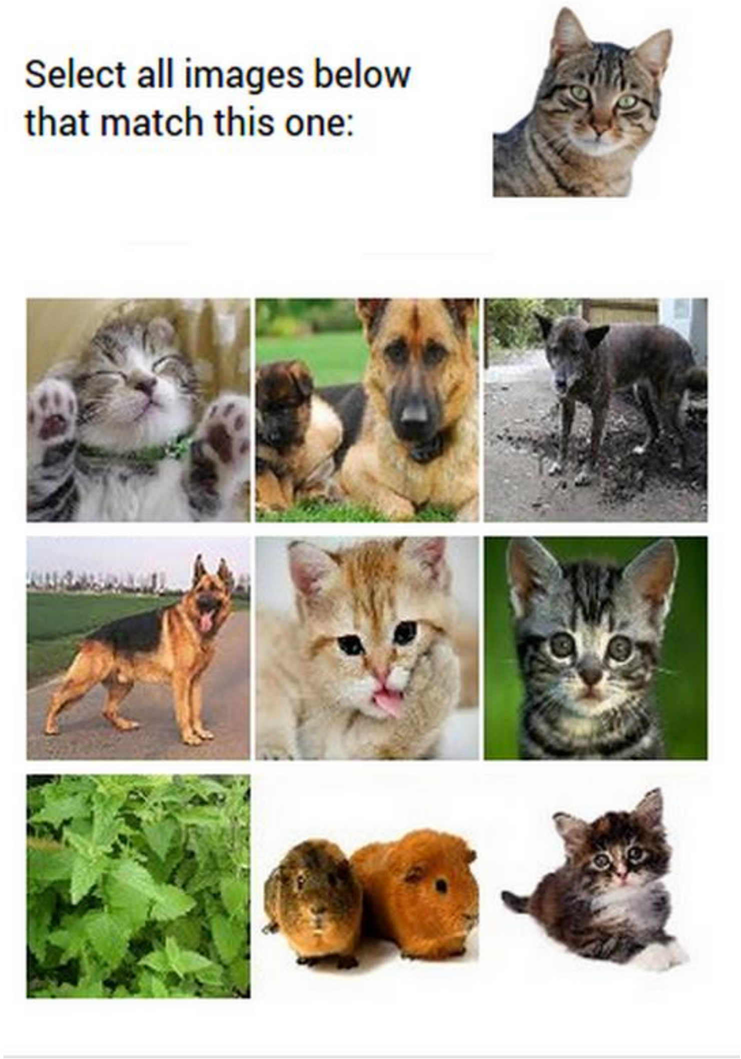
A lot of foo□

d? t? s?

Training

WE'RE GONNA NEED A MONTAGE

- Harvest good labeled data
- Bootstrap your model with experiences
- … Or get your users to do it for you



Select all images below that match this one:

So What?

I DON'T GET IT

DEMONSTRATION

CROSSES FINGERS

Lessons Learned

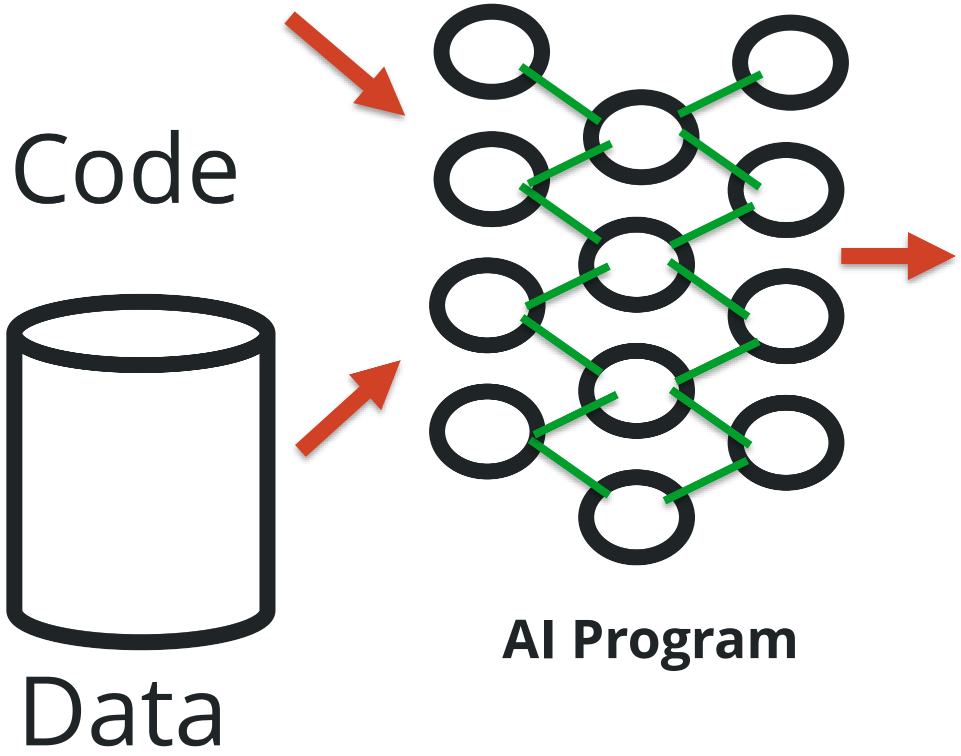
MISTAKES WERE MADE

- **Quality training data is important**
  - And hard to get
  - Garbage in, Garbage out
- **Be careful about what you reward**
  - You will get more of it
- **Get a GPU**
  - Or better yet, a lot of them

Other Considerations

BY THE NUMBERS

- **Inherent Proprietary-ness**
- **Unreliable factor**
  - Stochastic
  - Undebuggable
    - Trained too much?
    - Not enough?
- **Power Imbalance**
  - Computational / Data



Future Work

TODO

- **Instrumented Webapp Fuzzer**
  - Available data?
- **Password Bruteforcing**
  - Context aware
- **Service Identification**
  - What's behind an open port?
- **Bad at**:
  - Finding new *classes* of vulnerabilities

Questions!

www.bishopfox.com

github.com/bishopfox

careers@bishopfox.com

Attributions (Images in Slides)

Artificial intelligence brain image

Hacker image

Magic hat image

Chessboard image

Function plot graph image

Artificial neural network image

Chess pieces image

Carrot+Stick<Love image