

Neural Nets as a Model of “Knowing How”

Kenny Easwaran
UC Irvine, Logic and Philosophy of Science

Stanley and Williamson, “Knowing How” (2001)

- “Hannah knows how to ride a bike”
- “Hannah knows that Jesse rides a bike”

Stanley and Williamson, “Knowing How” (2001)

- “Hannah knows who to call.”
- “Hannah knows what to order.”
- “Hannah knows when to arrive.”
- “Hannah knows where to go.”
- “Hannah knows why to stay.”
- “Hannah knows how to swim.”
- “Hannah knows which class to take.”
- “Hannah knows whether to knock.”

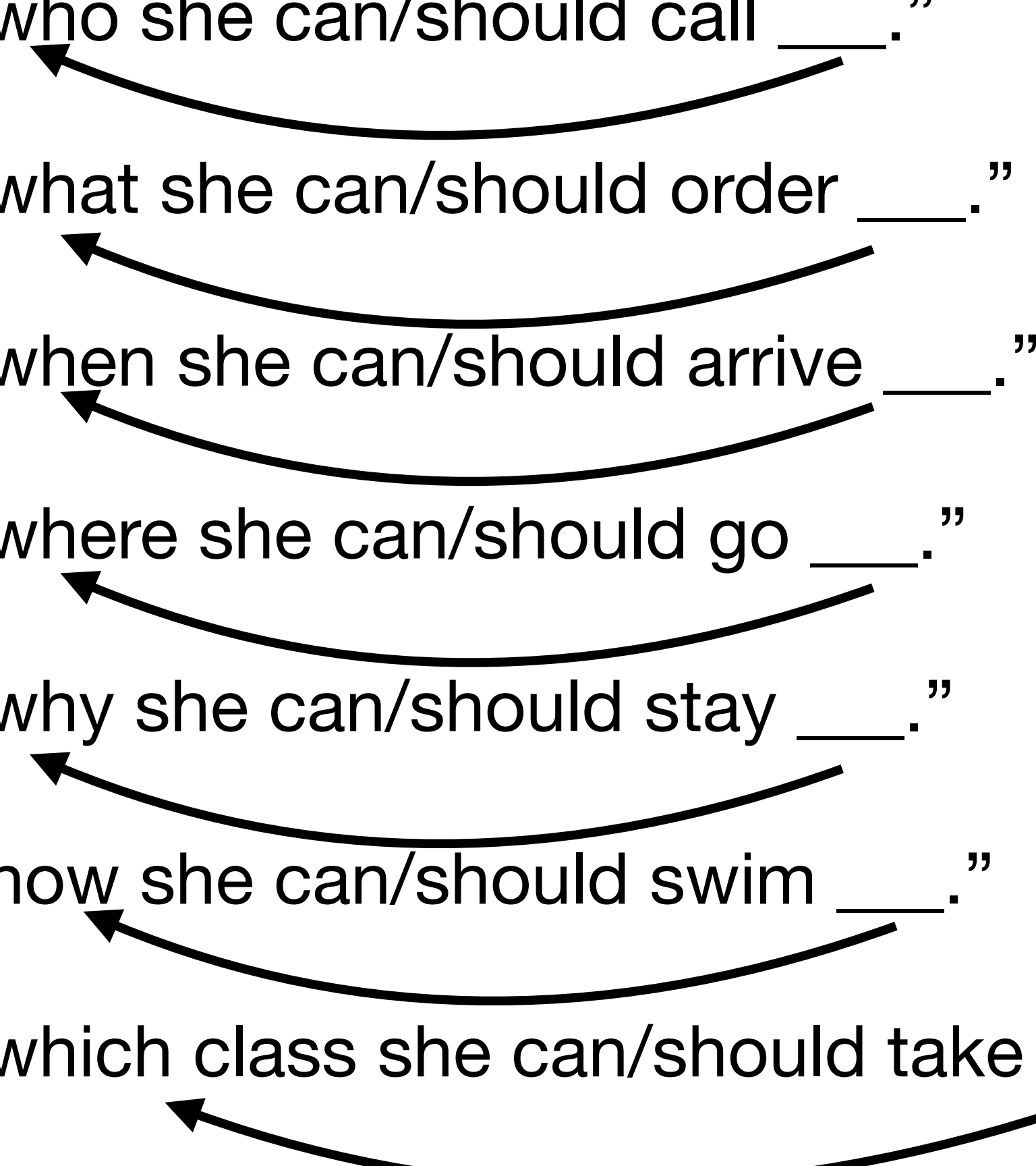
Stanley and Williamson, “Knowing How” (2001)

- “Hannah knows who PRO to call.”
- “Hannah knows what PRO to order.”
- “Hannah knows when PRO to arrive.”
- “Hannah knows where PRO to go.”
- “Hannah knows why PRO to stay.”
- “Hannah knows how PRO to swim.”
- “Hannah knows which class PRO to take.”
- “Hannah knows whether PRO to knock.”

Stanley and Williamson, “Knowing How” (2001)

- “Hannah knows who she can/should call.”
- “Hannah knows what she can/should order.”
- “Hannah knows when she can/should arrive.”
- “Hannah knows where she can/should go.”
- “Hannah knows why she can/should stay.”
- “Hannah knows how she can/should swim.”
- “Hannah knows which class she can/should take.”
- “Hannah knows whether she can/should knock.”

Stanley and Williamson, “Knowing How” (2001)

- “Hannah knows who she can/should call ____.”
 - “Hannah knows what she can/should order ____.”
 - “Hannah knows when she can/should arrive ____.”
 - “Hannah knows where she can/should go ____.”
 - “Hannah knows why she can/should stay ____.”
 - “Hannah knows how she can/should swim ____.”
 - “Hannah knows which class she can/should take ____.”
 - “Hannah knows whether she can/should knock.”
- 

Stanley and Williamson, “Knowing How” (2001)

- “Hannah knows that she can/should call (X).”
- “Hannah knows that she can/should order (X).”
- “Hannah knows that she can/should arrive (at X time).”
- “Hannah knows that she can/should go (to X place).”
- “Hannah knows that she can/should stay (for X reason).”
- “Hannah knows that she can/should swim (in X way).”
- “Hannah knows that she can/should take (X class).”
- “Hannah knows that she can/should(n’t) knock.”

Stanley and Williamson, “Knowing How” (2001)

- “Know that p ” is the more general case
- “Know + wh- + PRO + inf.” is the special case
 - True answer to the question, “wh- can/should I V”
 - Under a “practical mode of presentation”
 - “Hannah knows that X is how she can/should ride a bike.”

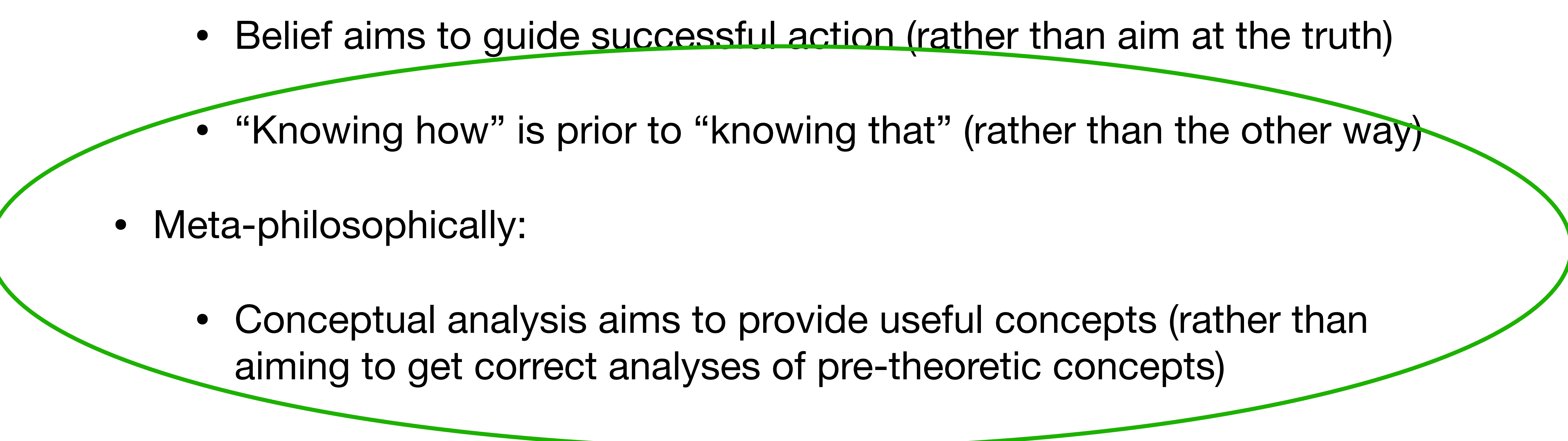
“Intellectualism”

- Ryle; Stanley and Williamson - identifying “know how” as a type of “know that” is “intellectualism”
 - But Stanley is an “anti-intellectualist” about pragmatic encroachment
- Stalnaker, “Intellectualism and the Objects of Knowledge” (2012)
 - Stanley and Williamson’s propositions must not be put in words
 - Is the “mode of presentation” part of the proposition or not?

“Intellectualism”

- Hyper-intellectualism - “knowing that” is primary, propositions can be put in words, what Ryle’s regress rules out
- Thin intellectualism - Stanley and Williamson, “knowing that” is primary
- Pragmatism - ability is primary

“Pragmatism”

- In epistemology:
 - Knowledge depends on practical interests (rather than just evidence)
 - Belief aims to guide successful action (rather than aim at the truth)
 - “Knowing how” is prior to “knowing that” (rather than the other way)
 - Meta-philosophically:
 - Conceptual analysis aims to provide useful concepts (rather than aiming to get correct analyses of pre-theoretic concepts)
- 

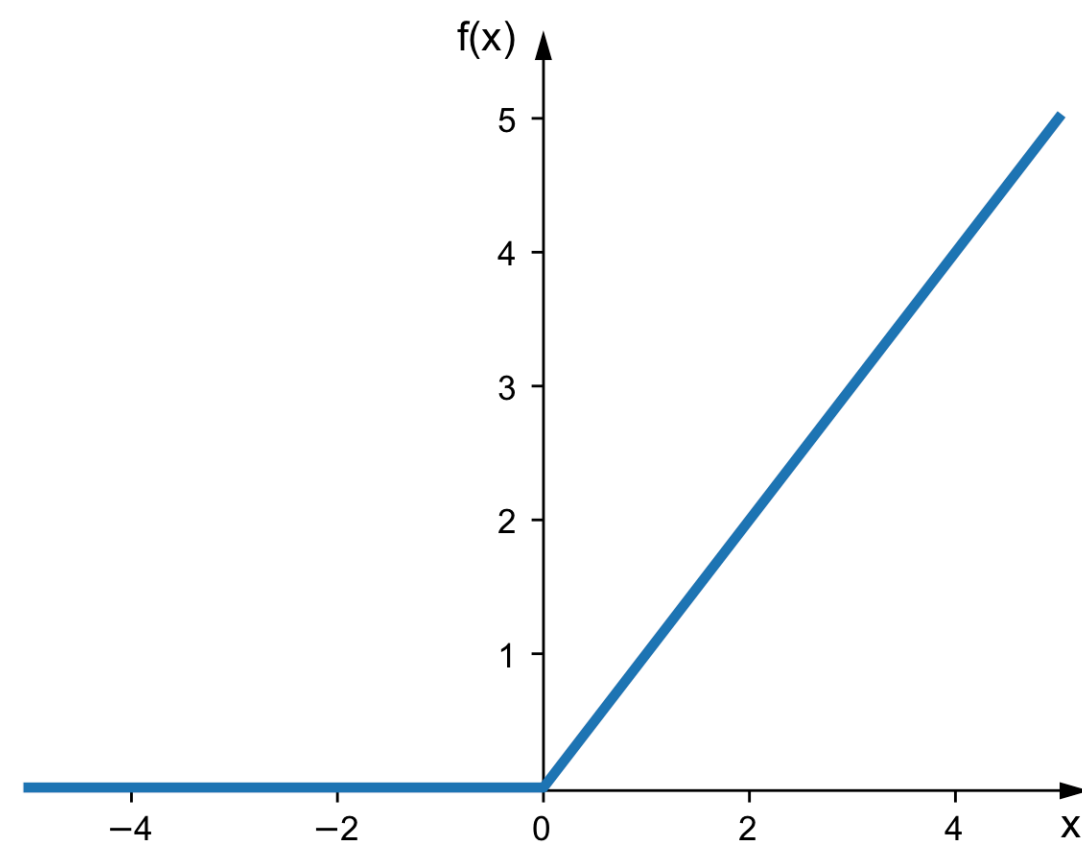
The use of “know”

- Gilbert Ryle: Explaining intelligence (both theoretical and practical)
- Edward Craig: Pooling information
- Joshua Habgood-Coote: (“know how”)
 - Pooling capacities
 - Mutual reliance

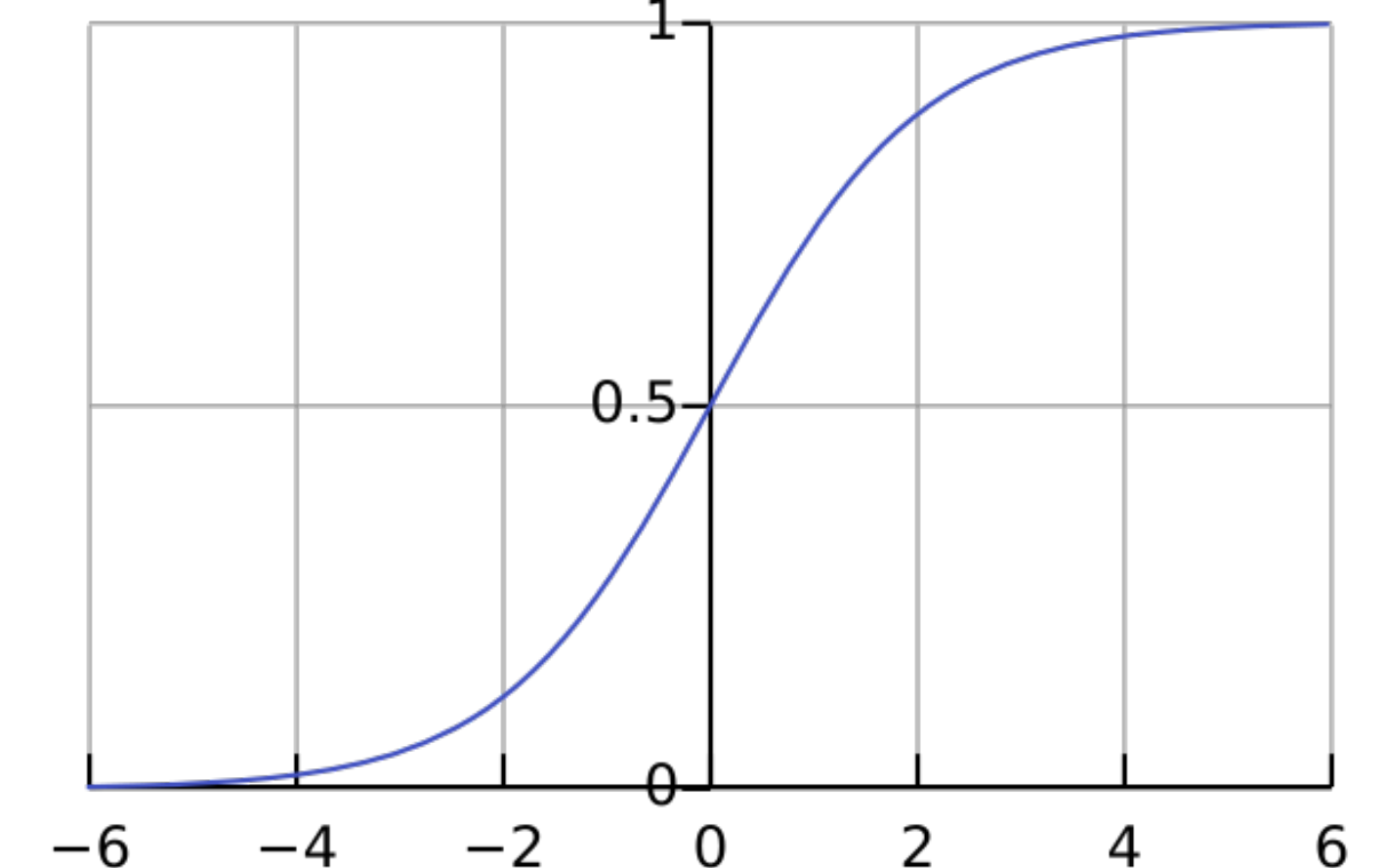
The “Five A’s”

- Animals
- Artificial Intelligences
- Associations (i.e., groups)
- Aliens
- Angels

Neural nets



ReLU



Sigmoid

Learning rate: 0.03
Activation: Sigmoid
Regularization: None
Regula: 0

DATA

Which dataset do you want to use?



Ratio of training to test data: 50%

Noise: 0

Batch size: 10

REGENERATE

FEATURES

Which properties do you want to feed in?

X_1
 X_2
 X_1^2
 X_2^2
 $X_1 X_2$
 $\sin(X_1)$
 $\sin(X_2)$

+ - 4 HIDDEN LAYERS

+ -

6 neurons

+ -

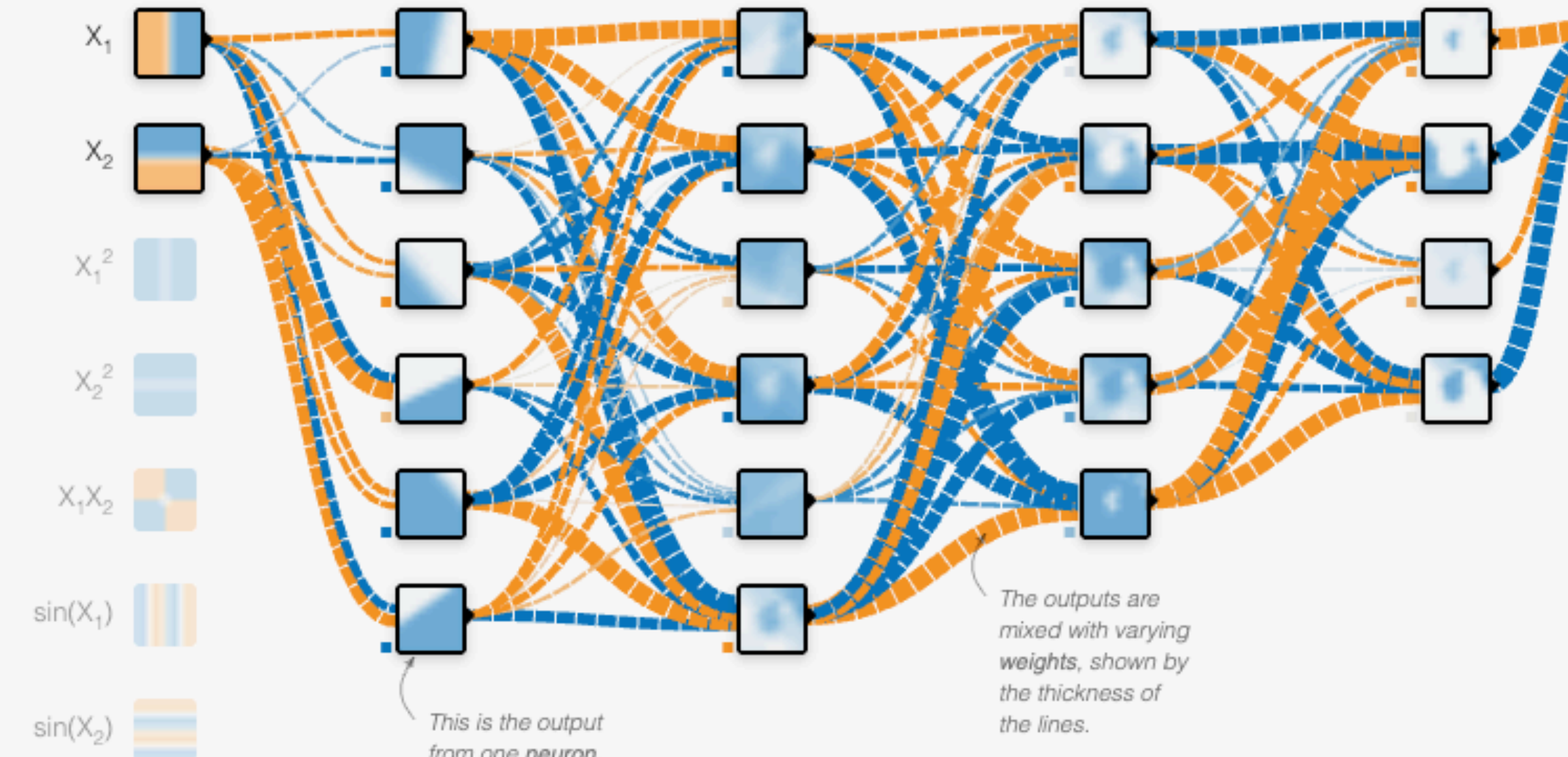
6 neurons

+ -

5 neurons

+ -

4 neurons

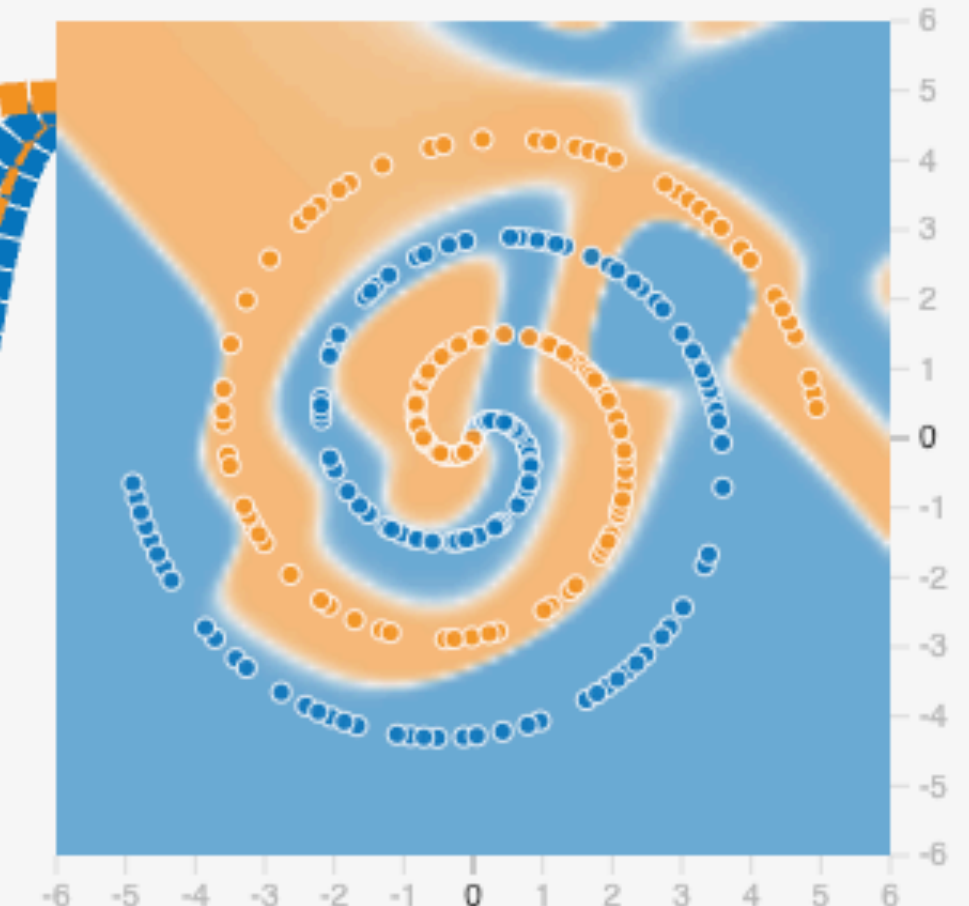


The outputs are mixed with varying weights, shown by the thickness of the lines.

OUTPUT

Test loss 0.095

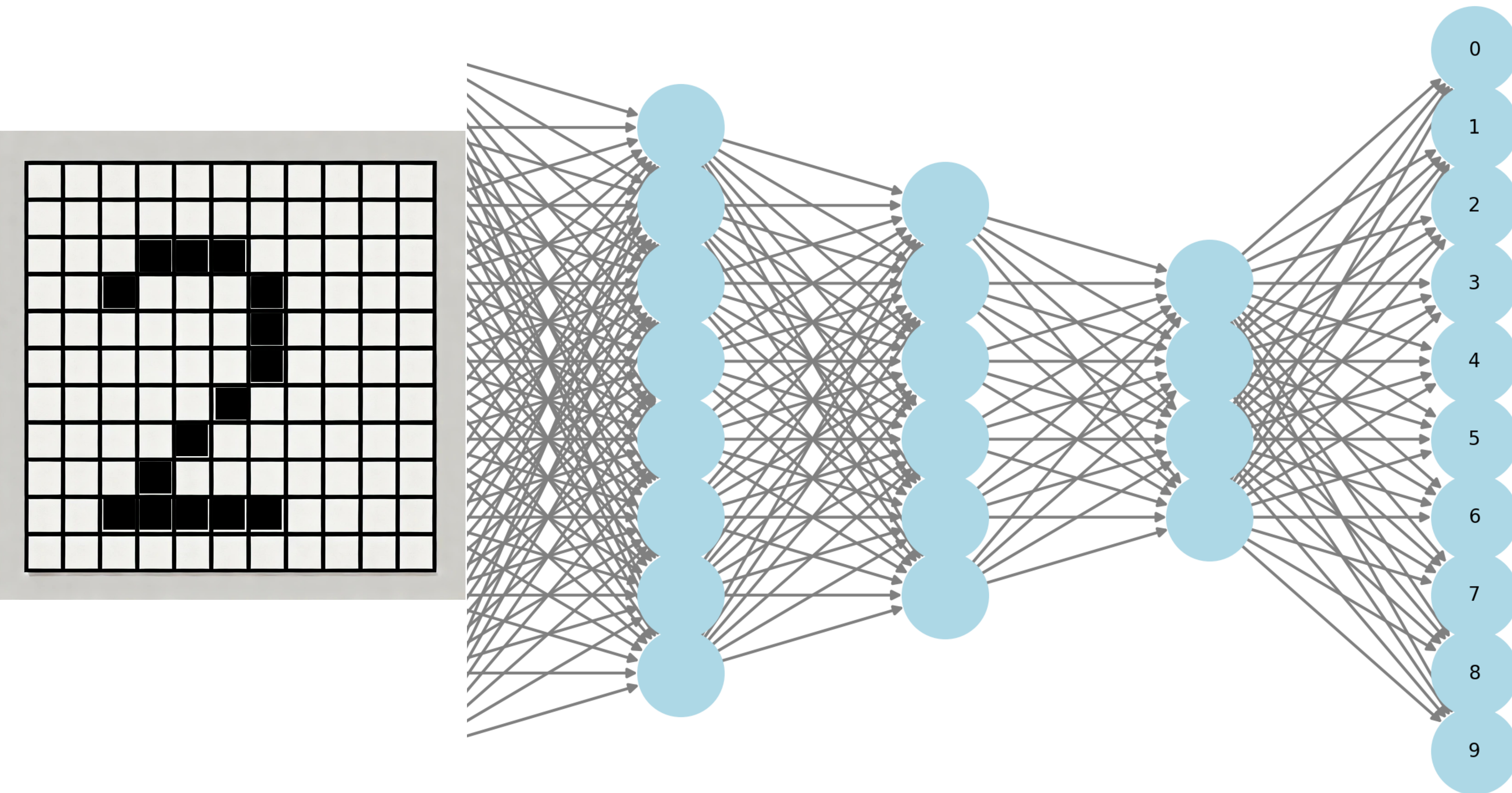
Training loss 0.046



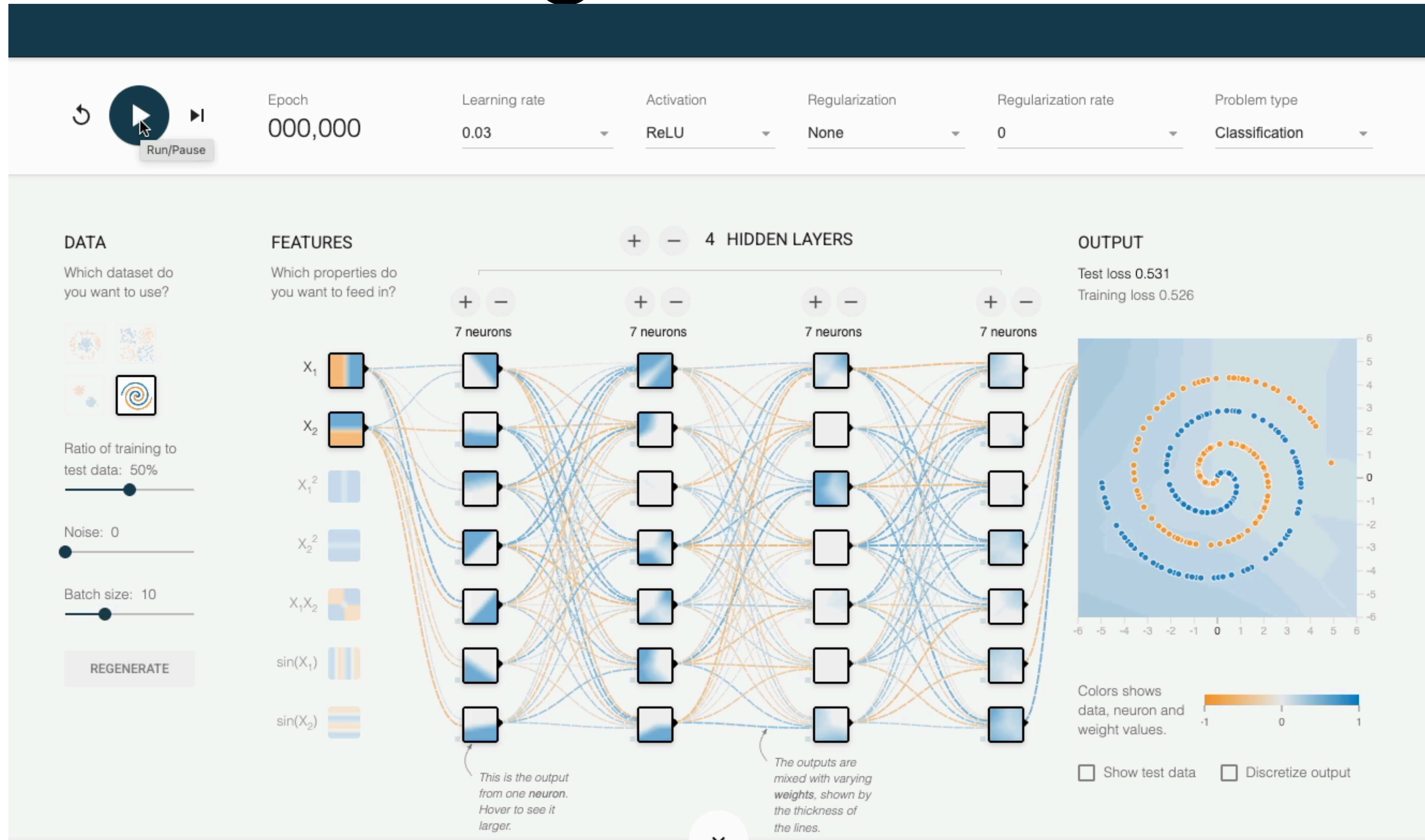
Colors shows data, neuron and weight values.

☐ Show test data ☐ Discretize output

Digit Recognition



Training a Neural Net



Knowing wh- to V

- Complex information-processing ability
- Typically learned through practice or evolved
- Only has to work in normal contexts
- Can work with opaque neurons/intuitions, not just concepts

Examples



“My phone knows who to unlock for.”

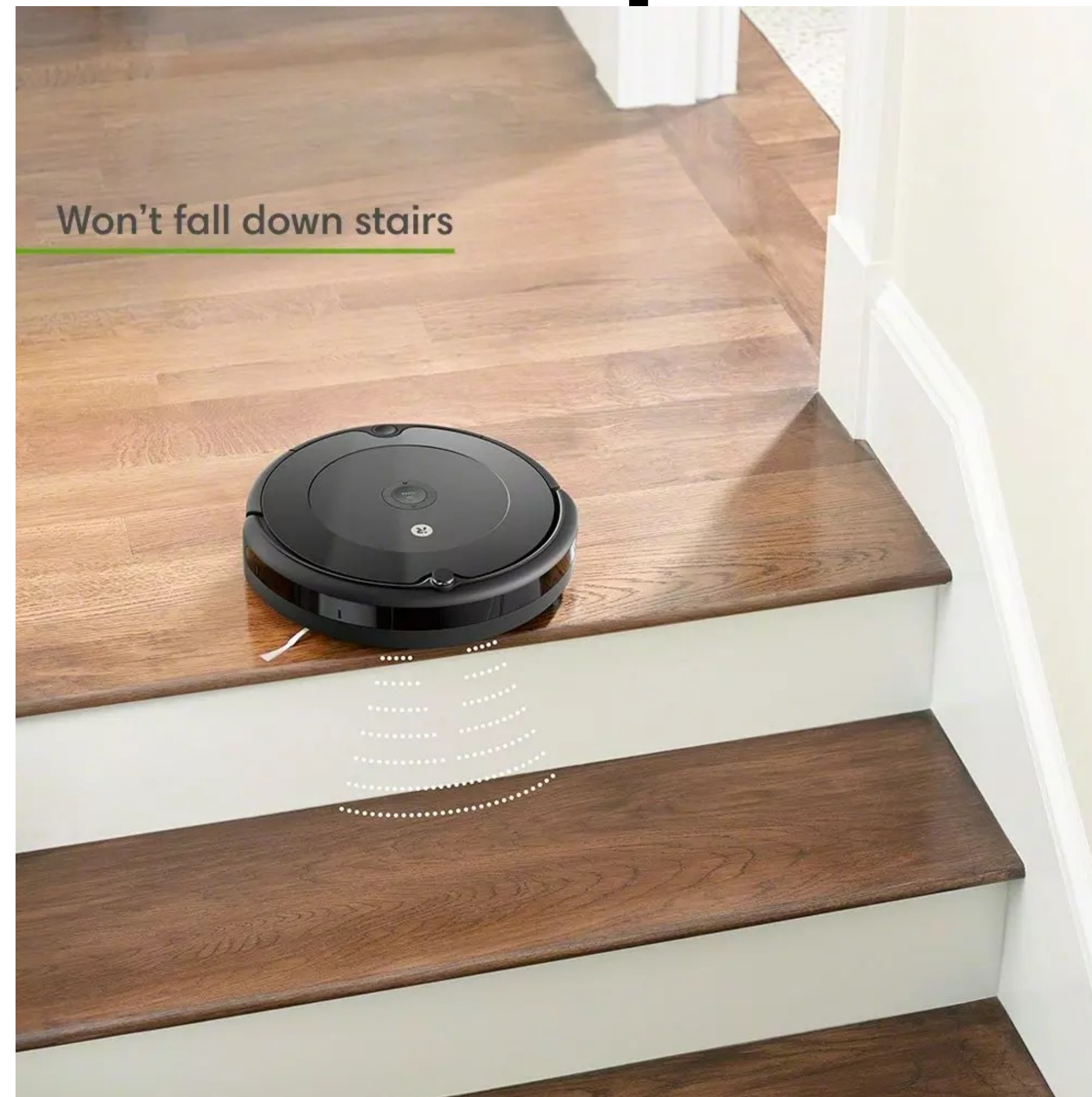
Examples



“The post office scanner knows where to send the package.”

“The post office scanner knows which digits to send where.”

Examples



“The Roomba knows how to get around the house.”

“The Roomba doesn't know how to go down stairs.”

“The Roomba knows not to go down the stairs.”

Examples

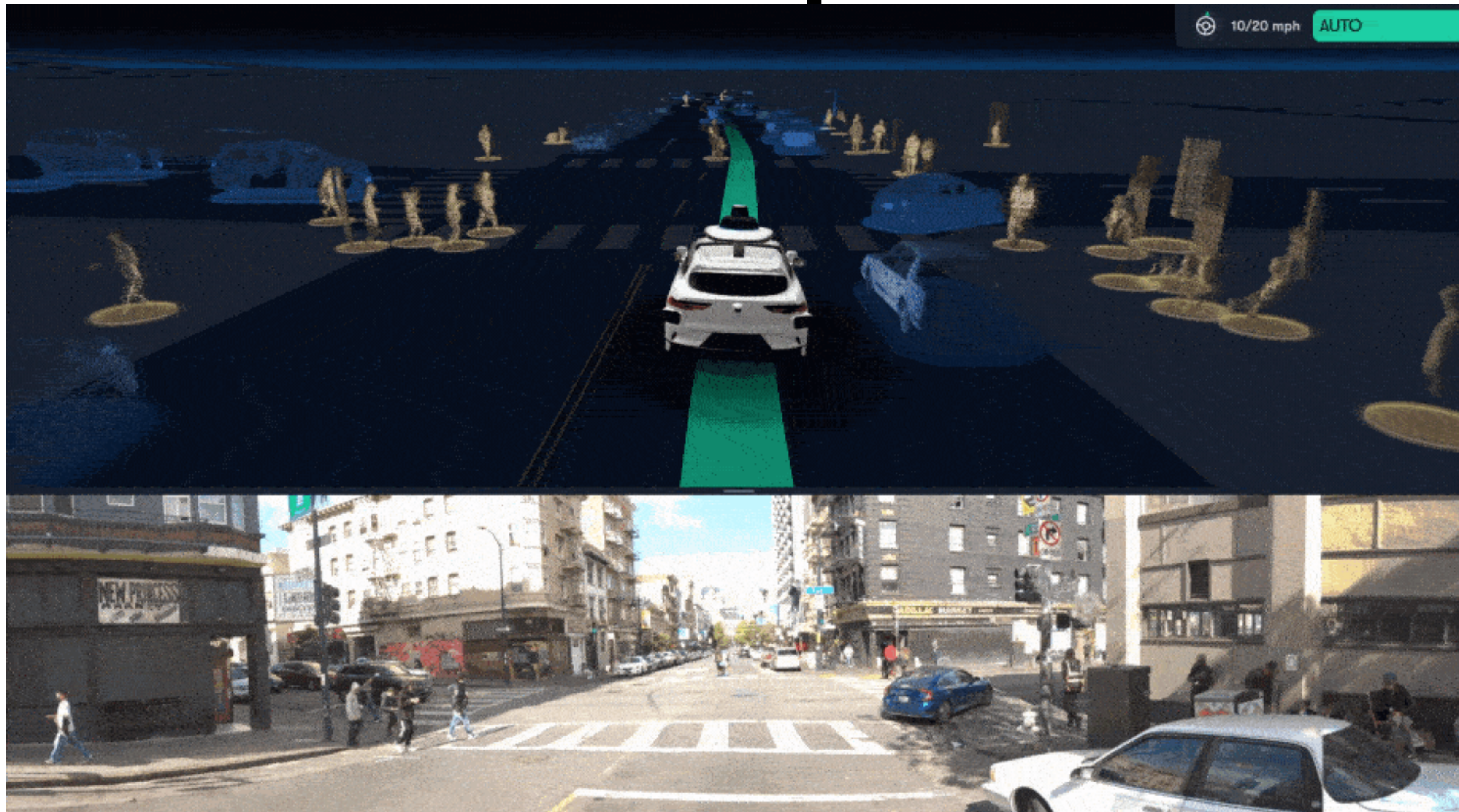


“The delivery robot knows how to navigate campus”

“The delivery robot knows who to open for”

“The delivery robot knows where to go”

Examples



“The Waymo...

knows what to stop for
knows when/why to slow down
knows where to pull over

knows how to get around
knows which lane to take
knows whether to go”

Safety vs Sensitivity

- Nozick (1981) - To know p , belief must be “sensitive”: in the nearest worlds where p is false, one must not believe p
- Sosa (1999) - To know p , belief must be “safe”: in all nearby worlds where one believes p , p must be true
- Distinctions:
 - Safety without sensitivity: there are no nearby worlds at all where p is false, but the nearest world where p is false is one in which the subject still believes p
 - Sensitivity without safety: The nearest worlds where p is false are ones in which the subject doesn't believe p , but there are some nearby worlds where p is believed and false

Examples



“The bird knows what to feed.”
“The bird knows how to feed its young.”

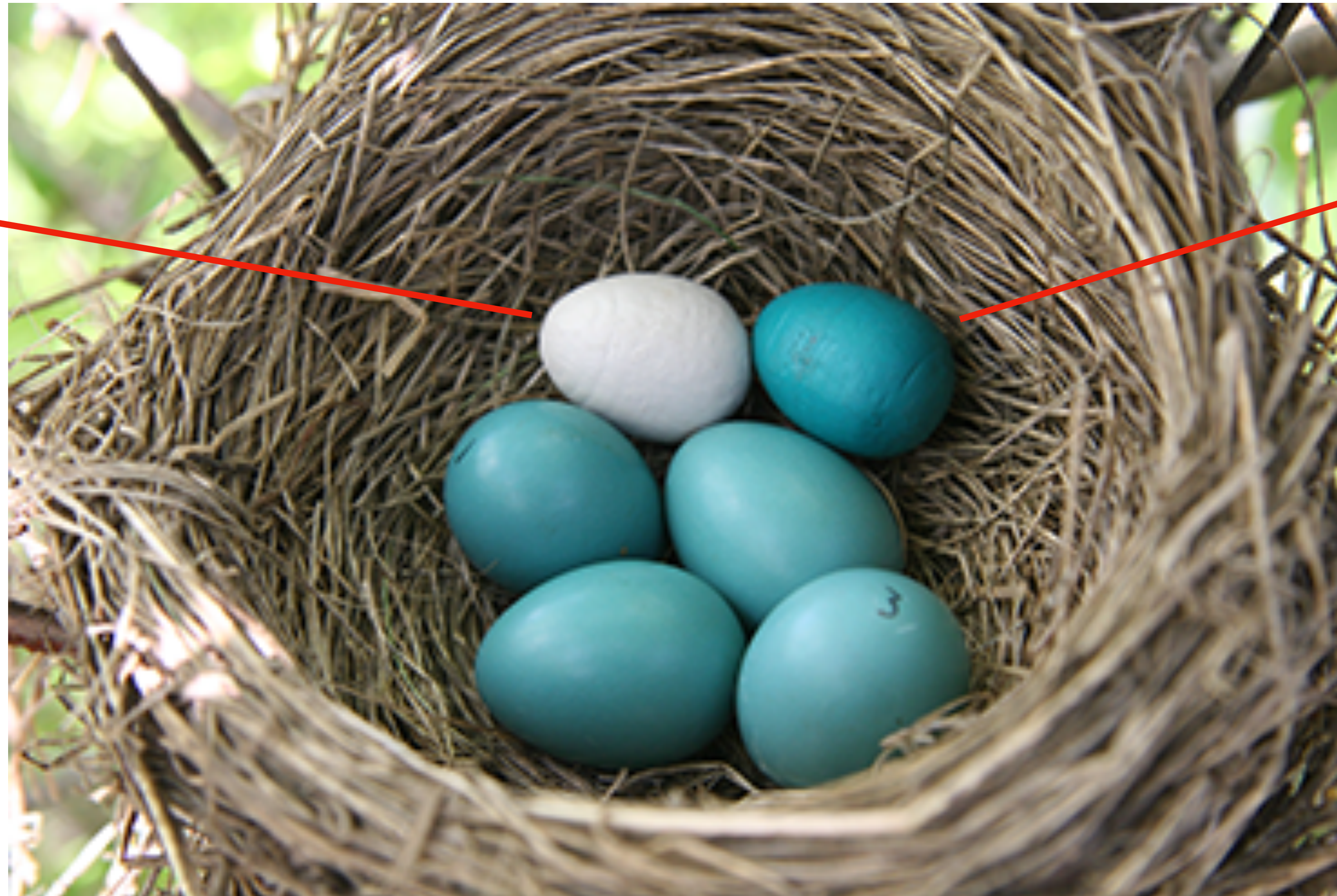
Imperfect Proxies



~~"The bird knows what to feed."~~
~~"The bird knows how to feed its young."~~

Imperfect Proxies

Wrong color,
rejected



Wrong size,
accepted

Does the robin know which eggs to reject?

Imperfect Proxies

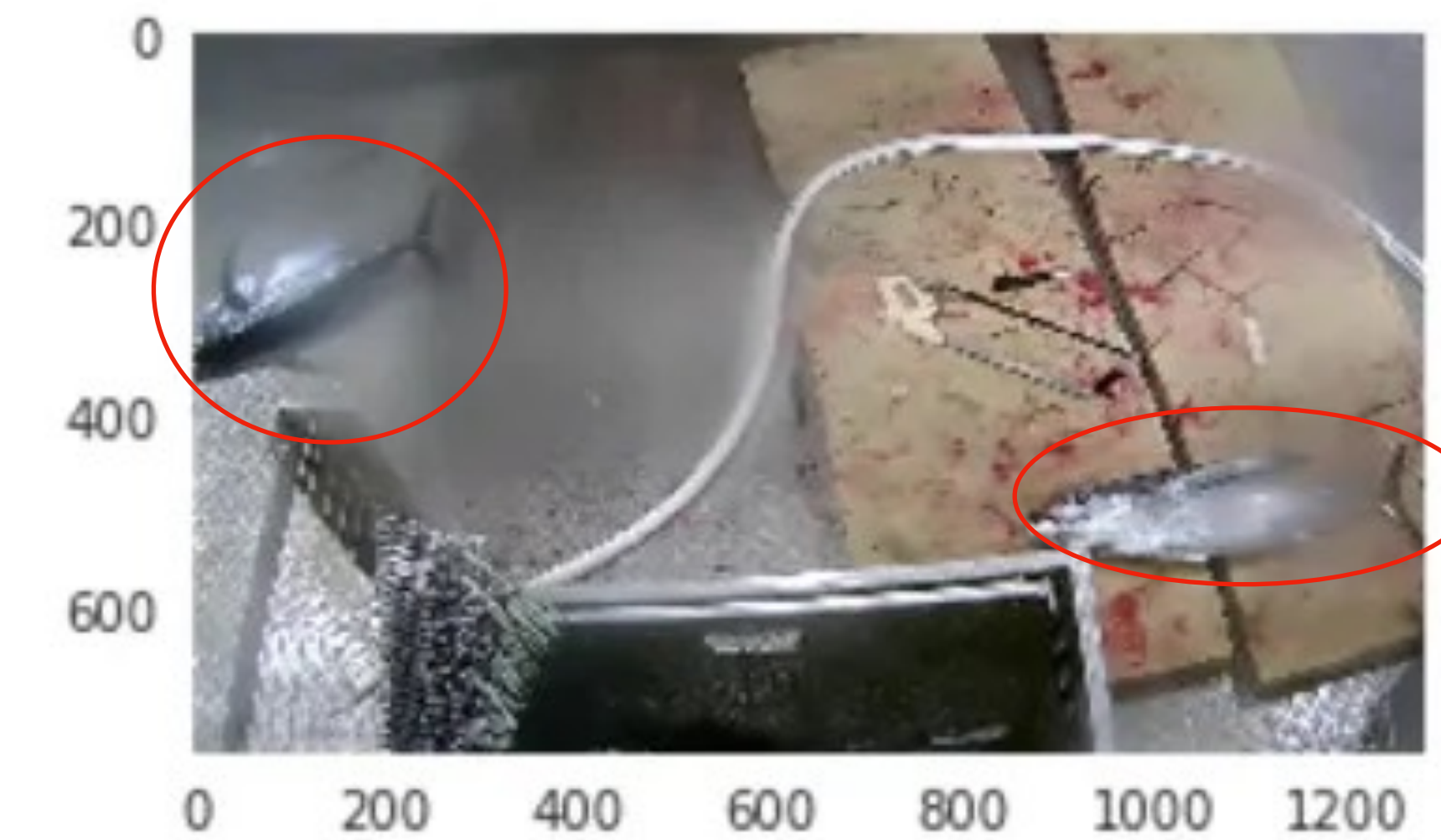
A.I. Deep Fake Phone Scams

How cyber criminals are using
loved ones' familiar voices to
trick people

*INCOMING
CALL*

Does the grandparent know who to trust?

Imperfect Proxies



Imperfect Proxies



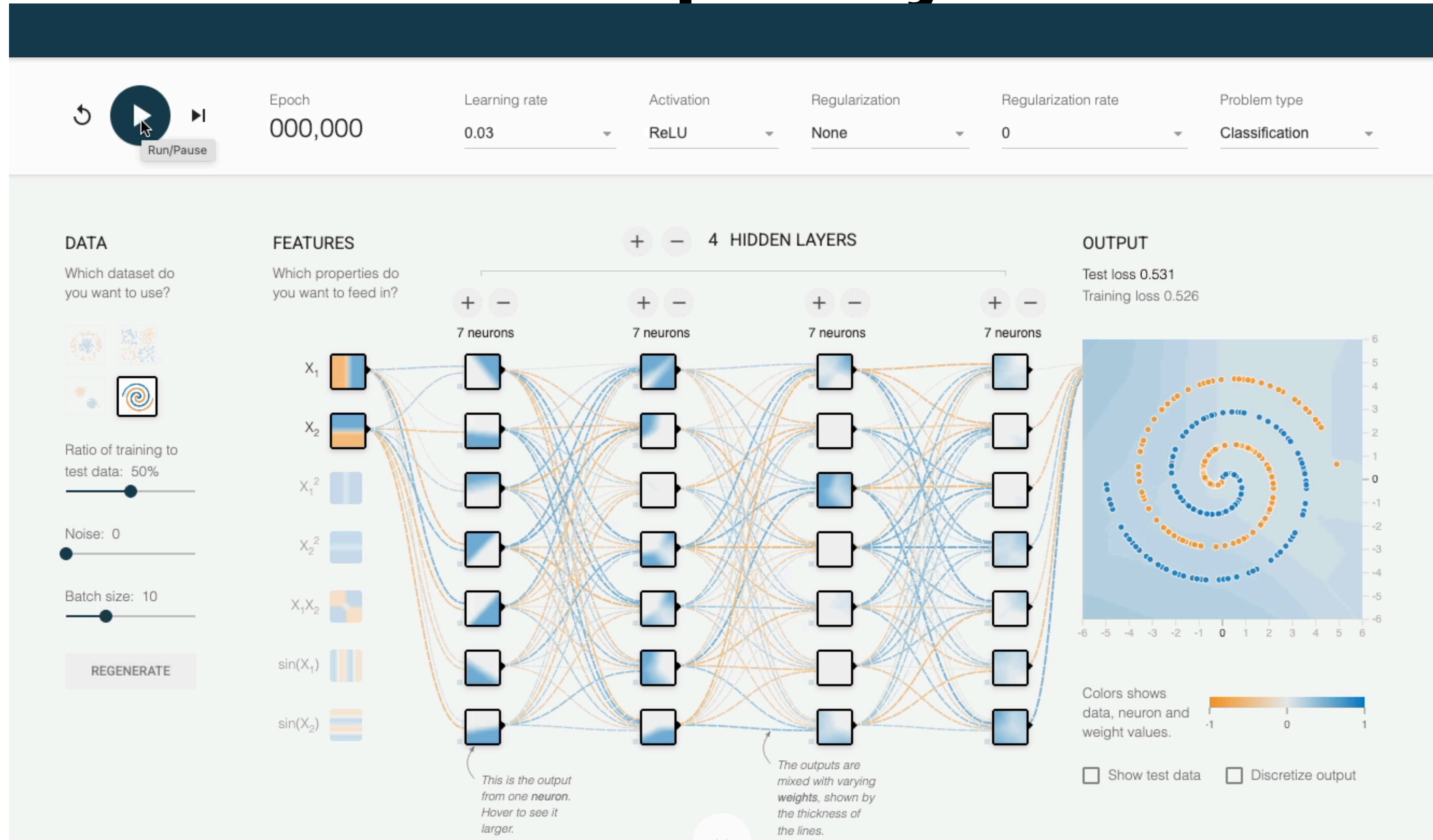
Who is the doctor?

Trick question - stock photo actors!

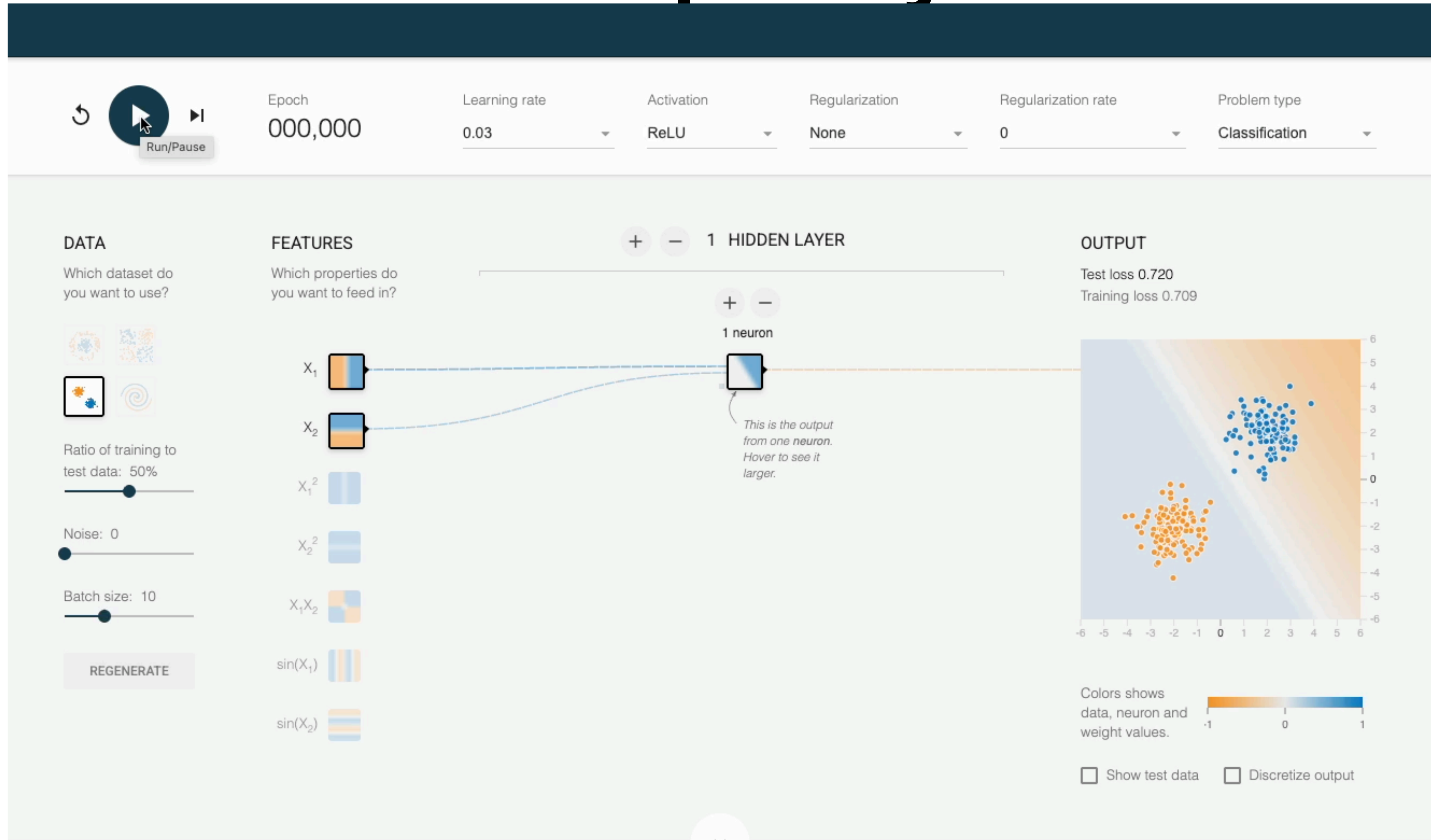
Simplicity

- I know how to keep my flip flops on while I walk
 - (I needed to learn, because they fell off easily)
- Do I “know how” to keep gloves on?

Simplicity

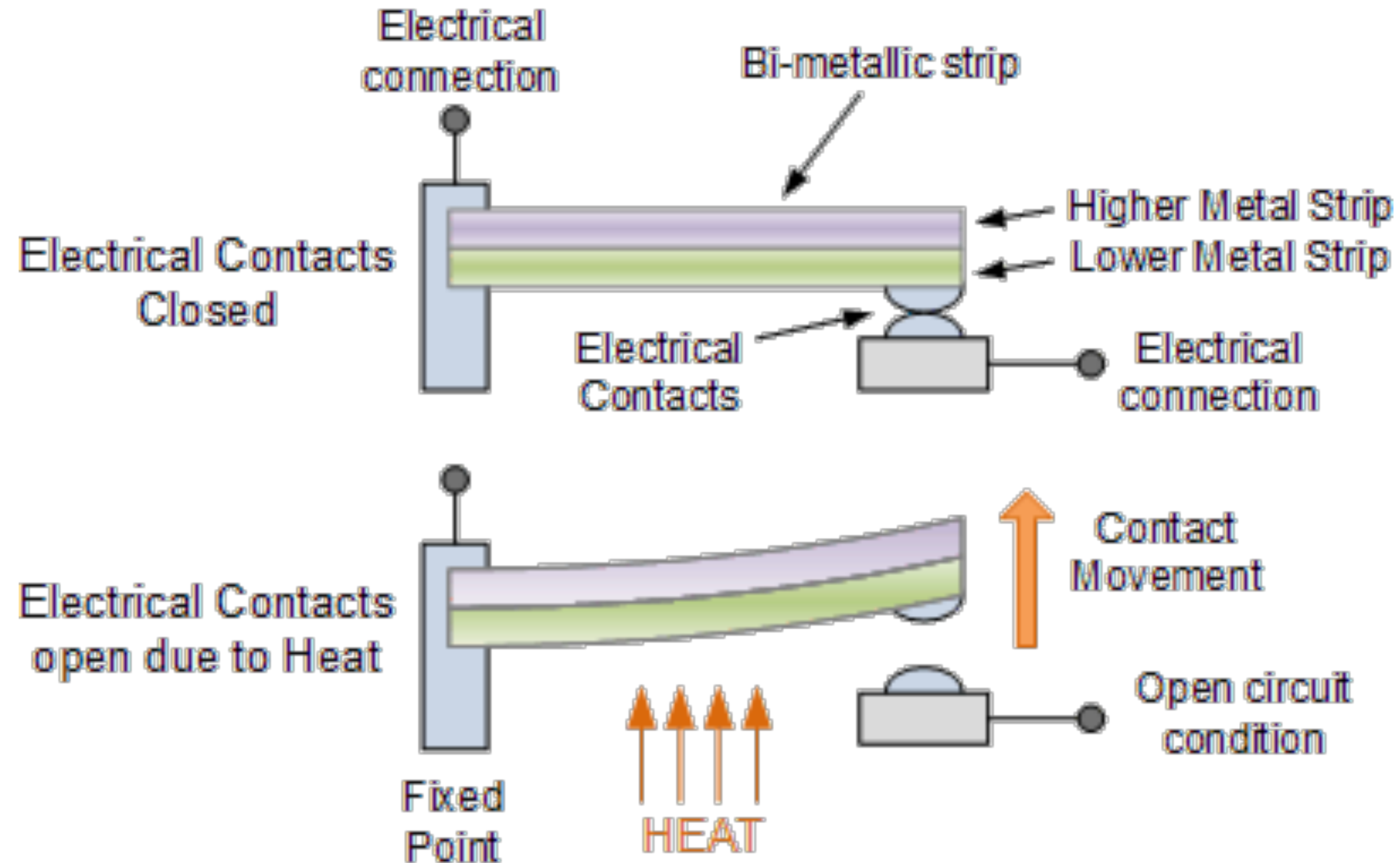


Simplicity



Simplicity

Thermostat



Knowing wh- to V

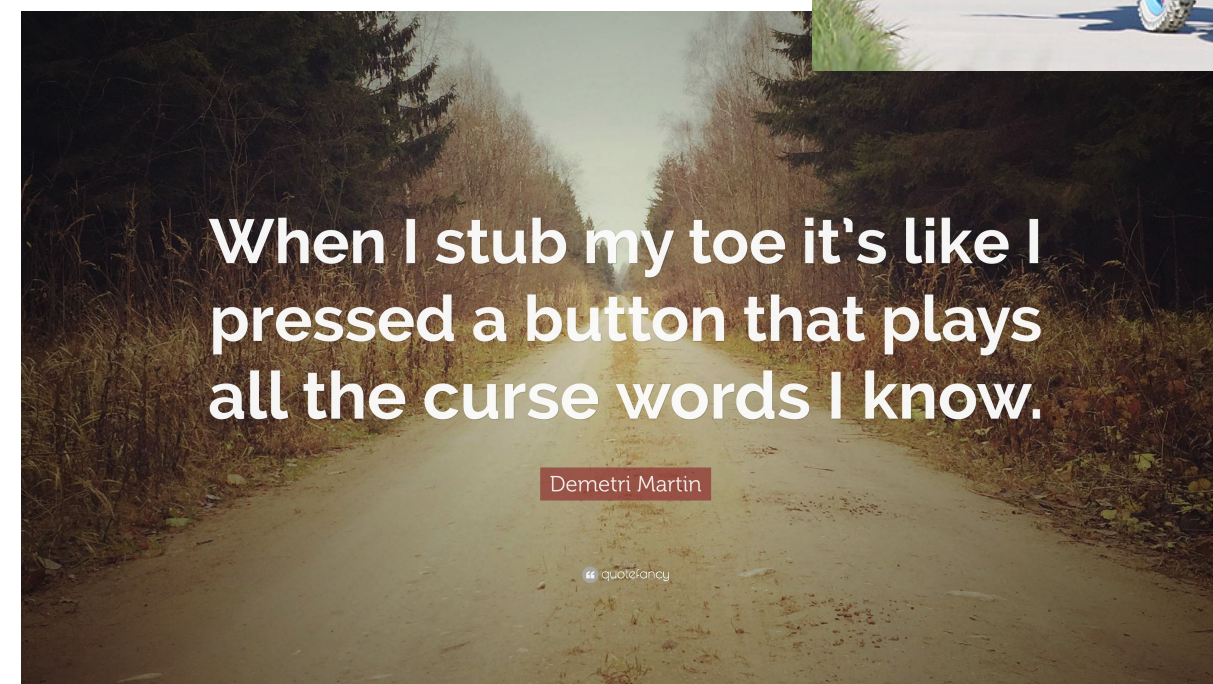
- Complex information-processing ability
- Typically learned through practice or evolved
- Only has to work in normal contexts
- Can work with opaque neurons/intuitions, not just concepts

Objections

- Stanley and Williamson, “Digesting food is not the sort of action that one knows how to do.”
 - I say: “Cow stomachs know how to digest grass but human stomachs don’t.”
- Pavese, “Alice cannot intentionally have ended up at the Colosseum without true beliefs about how to get there.”
 - If she walked “on autopilot” while daydreaming

System 1 and System 2

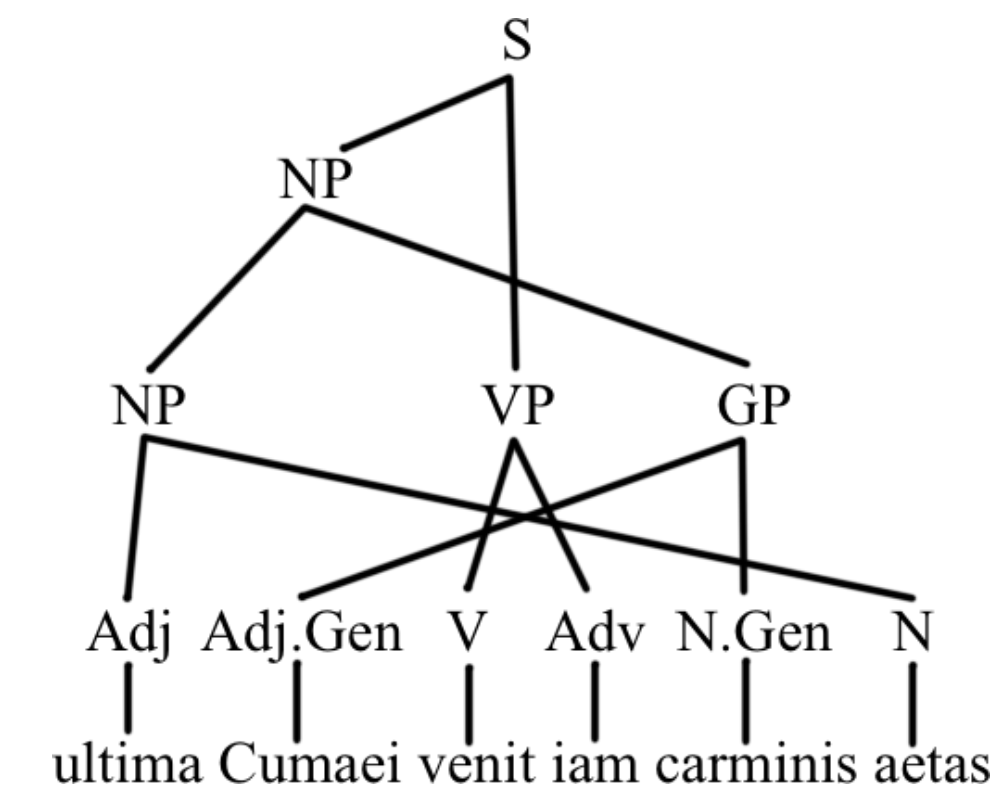
- System 1 - Fast



- System 2 - Slow



$$\begin{array}{r} 21 \\ \cancel{4} \cancel{1} \\ 573 \\ \times 46 \\ \hline \textcircled{1} 3438 \\ + 22920 \\ \hline 26358 \end{array}$$



“The Yips”



YIPS:

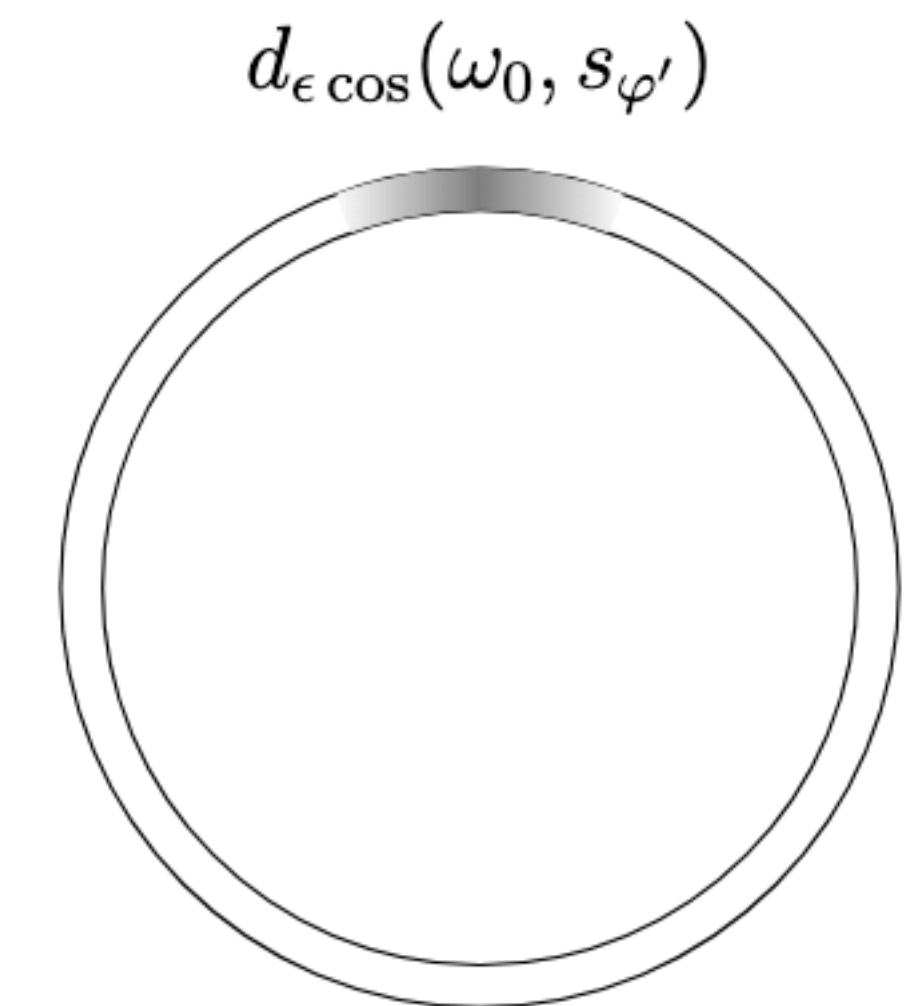
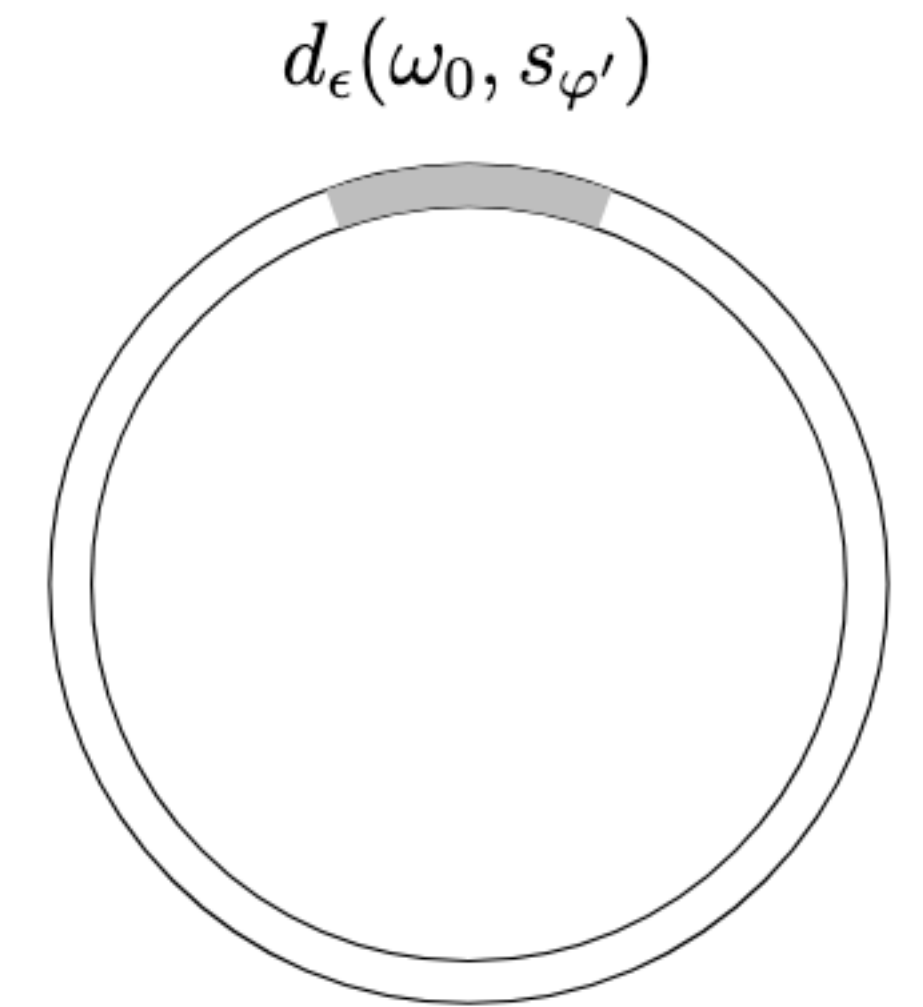
TYPES, CAUSES, SYMPTOMS,
RISK FACTORS, DIAGNOSIS AND
TREATMENT.

Orthopedics



Easwaran and Nielsen, 2025

- “Updating by Maximizing Expected Accuracy in the Infinite, Non-Partitional Case”
- The agent has a credence distribution over propositions
- The agent will look at an unmarked clock - no one-to-one correspondence between signals and worlds
- The optimal update uses the likelihood of the signal in each world
- If flat (like the top) this is equivalent to conditioning on a proposition, but otherwise not
- Propositional knowing is a special case of knowing how to make your credences most accurate



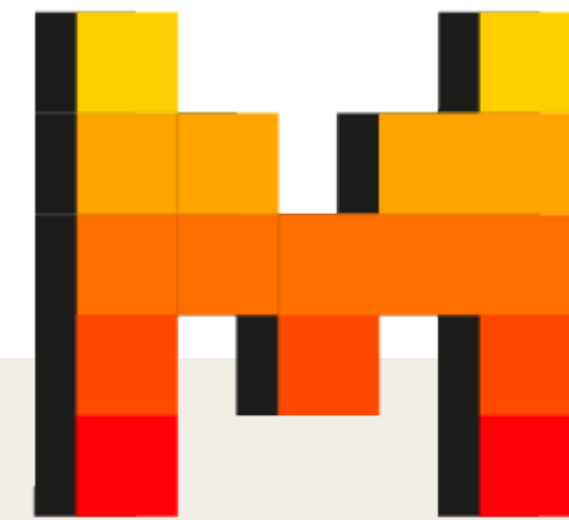
“Knowing that”

- “Knowing that” is just a special cluster of abilities (not Hetherington’s account, which is one ability)
- Eric Schwitzgebel, Tamar Gendler, both give examples where people have mixtures of these abilities
- The general cluster often goes together, because of our System 2 (symbol processing system)
- Once you have a symbol processing system, “knowing that” can be achieved very easily through testimony, rather than practice

Can a Neural Net “Know That”?



ChatGPT



MISTRAL
AI_



 Claude

BY ANTHROPIC



deepseek

Gemini

Conclusion

- “Knowing wh- to V” - having the ability to process information usefully in a way that allows one to V appropriately in normal contexts
 - Animals, neural nets, groups, people can do this fine
- “Knowing that p ” - having concepts that can be manipulated syntactically, connected to appropriate abilities, in a way that flexibly allows appropriate actions
 - Humans can, well-organized groups can, maybe animals?
 - Maybe Large Language Models?

Crosswords

ChatGPT 5 ▾



Aug1225.pdf
PDF

Can you process the grid from this crossword into a spreadsheet or some other data type you can interact with?

...

Notes and next steps:

- The PDF text didn't include the exact block layout, so the grid is a clean 15×15 template. If you want, I can help infer the black-square pattern (and fill lengths) from the clue numbering/structure—or you can provide a photo/screenshot of the printed grid and I'll map it precisely.

Crosswords

There is an image of the grid in the pdf file - can you see it?

Yes — the PDF does contain an embedded image, though it looks quite small (about **254 × 41 pixels**) and is on the first page. That's most likely the crossword grid (compressed into a low-resolution bitmap).

If you'd like, I can try extracting that image, cleaning it up, and then either:

- Run OCR/pattern recognition to reconstruct the actual **15×15 grid with black squares** (so we can integrate it with the spreadsheet), or
- Just give you the extracted image directly, so you can confirm whether it's the grid.

Do you want me to attempt the reconstruction into a usable grid, or first just show you

Crosswords

let's start by showing me the image

Here's the extracted image from the PDF — likely the crossword grid:

The New York Times

Do you want me to try reconstructing this into a clean, interactive **15×15 grid with block positions?** `</>`

