We Are AI: Taking control of technology

We are AI #2 : Learning from Data

Cover-alt:

An embodied AI/robot grows increasingly large on the screen. At the bottom is a pile of ‘data’, which the AI has its palm upon, and is drawing its strength from.

Terms of use:

Cite as:
Julia Stoyanovich and Falaah Arif Khan. “Learning from Data”. We are AI Comics, Vol 2 (2021) https://dataresponsibly.github.io/we-are-ai/comics/vol2_en.pdf

Page 1

When talking about Artificial Intelligence (AI), we usually make a distinction between 'classical AI' systems - those that operate according to predetermined rules - such as Roomba, the smart vacuum,

A round, metallic robotic floor-cleaner, spins around

and those that 'learn' rules from data - such as chess-playing AI

Four different chess pieces are placed on screen - from left to right: a pawn, the king, the queen, and a horse.

or a self-driving car.

Picture a woman sitting in a self-driving car – she's in the driver's seat, but has kicked back with her arms behind her head and her legs on the dashboard. The car is in self-driving mode, with the steering wheel lit up, and a robot assistant appearing on the LED screen on the dash board.

This distinction is not strict.

In fact, most AI systems out there in the world today combine hand-crafted rules with some form of learning.

Let's see examples of such rules, and see how machines learn them from data!

Page 2

Suppose that you are tasked with designing a smart lighting system- that autonomously turns the lights in your house on and off.

An AI-genie appears, and conjured up a light bulb, that it is mystically controlling by turning on/off

How should your AI decide when to take these actions?

We can start designing such a system based on our own everyday 'intelligence'.

Let's call this Rule 1. It simply says – turn on the lights when it's dark outside, Picture a simple caricatured house with a door and one window. It is dark outside – the moon is out, and the sky is purple, while the interior of the house is bright – the window shows yellow light filling the room

otherwise keep the lights off.

The house with the door is now dark – the window shows a dark purple hue settling over the house, while outside the sun is out, and the sky is a bright blue

This algorithm is very simple, it's just one step –

It takes the outside conditions (whether it's dark outside) as an input, and, using Rule 1, predicts the appropriate output - whether to turn on the light.

A simple symbolic depiction of rule 1: The sun and moon on the left, and arrow in the middle, and a lit (yellow) bulb and/or a dim (grey) bulb on the right

This rule is simple, but is it any good?

Does it appropriately capture how you would manually operate the lights in your house?

Let's run an experiment to find out!

Page 3

We'll collect data about whether it's dark outside and whether the lights are on in the house.

For each observation, we will check whether the prediction made by our algorithm in fact matches what we observed.

The screen is split into four panels: a, b, c and d

Panel a: It's late in the evening, a woman is lounging on her couch, channel-surfing casually. It's dark outside, but the room is well lit.

Panel b: Two women sit at a dinner table: plates of delectable delights sit in front of them, as they raise their glasses in celebration. It is dark outside, but the table is well-lit with an overhead hanging lamp-shade.

Panel c: A woman sits at her table and reads at her with a morning cup of coffee in one hand, and an iPad in the other. A lamp is in view behind her, but it is off. There's a window on the other side behind her that is open, and the pleasant morning sun lights up the room.

Panel d: A woman is deep in sleep in her bed, snuggled into her duvet. The room is dark and the lights are off, and so we see her in the reflected hues of purple and blue.

For observations a and b, it's dark out but we're watching TV and eating dinner, and so we've left the lights on.

And this is indeed what Rule 1 has predicted!

This means that a and b both support the hypothesis encoded by Rule 1.

Observation c also supports the hypothesis - it's nice and bright out, and so we've kept the lights off.

Let's look at observation d now:

it's dark out and we're turning in for the night, and have turned off the lights.

But, as per Rule 1, the algorithm has predicted the lights to be 'on' instead. Rule 1 fails here!

Page 4.

Let's look at what we've learned about Rule 1 from our experiment -

Our rule was able to predict the output correctly 3 times out of 4.

Dark Out

Light On

Prediction

There's a table summarizing the observations of our experiment:

There are 3 columns for each observation, namely "Dark out" [moon/sun], "Light on"[yellow bulb/gray bulb] and "prediction" [yellow bulb/gray bulb]. The table reads as follows: A: Dark out, light on, light on, B: dark out, light on, light on, C:bright out (or) not dark out, light off, light off, D: dark out, light off, light on.

That's the rule's predictive accuracy - 75%.

Our AI-genie is flexing at the camera, but its bicep is only 4 of the way "filled" with color. The numbers "75%" are written across the flexed bicep.

Let's fix some more terminology here -

It's customary to refer to the input of the rule - whether it's dark out, in our example - as the input feature, or just feature.

If it's dark out, we represent it by a moon, whereas if it's bright, we draw the sun.

The output - the state of our lights (on/off) - is usually called the outcome or the label.

The different possible outcomes - turning the lights off or leaving them on, in our example - are called classes'.

If the light is on, the bulb appears yellow, whereas, if the light is off, the bulb appears gray.

The rule we are working with is called a 'classifier' - it assigns a class label to an observation.

The classifier we designed using Rule 1 can make one of two choices - turn the lights either on or off.

We call such classifiers 'binary'.

Page 5

Back to observation d, which rule 1 misclassified:

Recall panel d: A woman is deep in sleep in her bed, snuggled into her duvet. The room is dark and the lights are off, and so we see her in the reflected hues of purple and blue.

The rule would take in the input – "dark outside", and return the output "light on", thereby incorrectly switching on the light and waking the woman up.

What is the cost of this mistake?

The sleeping woman is abruptly woken up by blinding lights– zombie-like eyes and disheveled hair mirror the exasperated expression on her face and her arms thrown in the air.

As you probably guessed, it’s that the smart light woke you up in the middle of the night!

What can we do about this?

If we compare observations a, b and d, we see that they have the same value of the input feature (it’s dark out), but different outputs.

Dark Out

Light On

Prediction

This makes us think that we need some additional features to distinguish between these situations.

Now, a decision that you, as the designer of this system, have to make is:

What are some other inputs that may be useful?

A woman is deep in thought with her arms crossed and her fist resting on her chin – above her a thought bubble appears with the images of an alarm clock, a cup and saucer of tea, and a thermometer.

Based on your experience, what features do you think will be helpful in predicting the outcome?

The outside temperature? No!

The price of tea in China? No!

Well, what about - whether it’s bedtime?

Let’s get to work to refine our rule.

Page 6

Here’s our refined rule - we’ll call this Rule 2 “turn on the lights if it’s dark outside, and it’s not yet bedtime. Otherwise keep the lights off.”

This time the caricatured rule contains two inputs: symbols of either the sun or moon – representing whether it’s light or dark outside, and an alarm clock – representing whether it is bedtime yet or not. There is an arrow in the middle, and the outputs on the right are the same – a yellow bulb (light on) and a gray bulb (light off)

Let’s re-run our experiment, collect some more observations, and evaluate our new rule.

There are 4 different snapshots on the screen: e, f, g and h

Panel e: A woman sits with her with a morning cup of coffee in one hand, and an iPad in the other. A lamp is in view behind her, but it is off. There’s a window on the other side behind her that is open, and the pleasant morning sun lights up the room.

Panel f: A woman lies on her couch, with her feet resting on the arm-rest. She is on the phone and is gesturing animatedly as she talks. Her window reveals a dark sky, while the room she is in is well-lit by an overhead light.

Panel g: A woman sits at her desk in front of a laptop. She gazes intently into her laptop, while holding a cup of tea/coffee in one hand. Her table lamp is next to her, but is turned off, as the window behind her shows a bright blue sky that is lighting up the room inside.

Panel h: A woman is curled up asleep on her couch – she is lying on her stomach with her legs resting against the armrest, and seems to have dozed off in front of a laptop that is open in front of her. Her table lamp is on and is shining light across the room. The AI-genie has appeared and is hovering between the light and the sleeping woman, magically turning the lamp off.

Voila! It seems to be working perfectly!

We don’t need the lights on unless it’s dark outside and we’re awake doing something!

And look, this time it's turning off the lights for us if we fall asleep on the couch (past our bedtime) - so caring!

Page 7

Rule 2 has shown perfect accuracy - our algorithm is able to predict the outcome correctly 100% of the time!

Dark out

Bed time

Light on

Prediction

We have a new table summarizing the observations of this experiment:

There are 4 columns for each observation, namely "Dark out" [moon/sun], "Bed time", "Light on"[yellow bulb/gray bulb] and "prediction" [yellow bulb/gray bulb]. The table reads as follows: E: bright out, not bed time, light off, light off, F: dark out, not bed time, light on, light on, G: bright out, not bed time, light off, light off, H: dark out, bed time, dark out, light off.

Our AI-genie is once again flexing at the camera, this time its bicep is completely "filled" with color. The numbers "100%" are written across the flexed bicep.

We've used our intuition and experience to devise a suitable rule, collected data to validate its performance, and made the necessary adjustments until we reached a performance that we're happy with!

In fact, we are so happy with our smart light — with how convenient it is, and how it's helping us conserve energy — that we want to offer it to others to use.

We're introducing our AI-genie to other humans! The AI-genie shakes the arm of a woman. Our protagonist embraces both the AI-genie and the woman, making the introduction.

Let's see if it works for our friend Ann to use at her office.

But before we go ahead and rewire Ann's entire office building, let's collect some usage patterns to see whether our rule would be effective.

You know the drill: let’s run an experiment!

Page 8

Recall rule 2 - “turn on the lights if it’s dark outside, and it’s not yet bedtime. Otherwise, keep the lights off.”

The screen is divided into 4 panels: I, J, K and L

Panel I: A diverse group of people sit on a table, looking at a presentation on a laptop. The table has other technological gadgets, as well as a couple of cups of tea/coffee. They are working in an open, modern workspace, with a large window behind them that supplements the lights inside the office.

Panel J: Zoomed in view of an office building from the outside – it is dark and so we see the two floors of offices in a purple-pink hue. The offices are empty, with several empty chairs and tables on each floor.

Panel K: Inside view of the office – it seems to be the middle of the day since the sky is bright and the office is lit with natural lights. The office is empty and so we only see large, long office meeting tables and a slew of empty chairs, but no people!

Panel L: A woman sits on her task frantically typing on her laptop. The table lamp at her desk is on, as are the other lights in the office. She looks tired, and seems to be working overtime as the window behind her reveals a dark night sky.

At the office, we keep the lights on through the day when people are working.

But Rule 2 would incorrectly turn them off, since it’s bright outside.

Most evenings, when everyone’s gone home, the lights are turned off.

But Rule 2 would incorrectly turn them on, since it’s dark out!

On weekends, the office is empty and so the lights are off throughout the day!

In this case Rule 2 works correctly - it keeps the lights off, since it is bright outside.

Every so often, someone stays late at the office, finishing up deliverables in time for a critical deadline.

We need the lights on in this scenario, but Rule 2 is wrong again: it would turn the lights off, since it's late (past our usual bedtime).

Page 9

Oh, no! That means we got only 1 correct prediction out of 4!

If we assume that lights are on about half of the time, we would have done better had we just flipped a coin to decide whether to turn on the lights!

Why is the accuracy of Rule 2 so low?

This is because the dataset on which we tested the classifier represents a different situation than the dataset on which we trained it.

We 'trained' the classifier – decided on the appropriate rules – based on our consumption patterns at home,

Over-the-shoulder snapshot of the woman at home – she sits with a cup of coffee, her legs crossed on top of the desk in front of her. The bright sky lights the room, and so the table lamp is off.

but Ann 'tested' it at her office.

Snapshot of Ann working on her laptop at the office, with her table lamp on, despite it being late in the day, and dark outside.

Whether it's bed time or if it's dark outside doesn't matter much anymore!

What matters is whether the office is open!

Page 10

What's the cost of a mistake here?

The cost of a false positive - unnecessarily turning on the light when it should be off - is that we waste energy and roll up a hefty bill for the company.

Ann holds a massively long paper of the electricity bill for the company, and exclaims upon looking at the estimate. Her hand is on her head in astonishment, and she looks distressed.

The cost of a false negative - turning off the light when it's needed — is that we keep interrupting people in the middle of their work, and their productivity suffers.

Snapshot of an ongoing meeting at an office – the lights have suddenly been turned off and so we can only see the purple-blue hues of the people in the room. They all have expressions of exclamation at the sudden outage, and/or outright disdain at the interruption in their discussion.

What should we do? Go back to the drawing board again?

Think about what additional features to use, collect that data,

and run an experiment to check whether our rule in fact works?

But this is tedious!

The protagonist is lost in thought – she is holding her face with one hand, and looking upwards deep in thought. We see small robot heads in the reflection in her irises as she contemplates the use of AI.

And most importantly - it will be very hard for us to continue deriving these rules that are getting more and more complex,

depending on the locations where the smart light is used and their different requirements!

The answer is - use a machine learning algorithm to identify statistical patterns in our data - to 'learn' the rules automatically!

Page 11

But just as we designed experiments to check whether the rules we came up with were any good, so, too, should we design experiments to test whether the rules that are automatically learned work when deployed.

This is because what rule is learned is fully dependent on historical data - if that data is representative of future use of the smart light system, then the rule will work well. But if the actual use case is different - then the rule will make lots of mistakes,

Picture the AI-genie holding an orange, but thinking of it as an apple (as seen in its thought bubble) since it has only ever seen apples.

such as turning on the light in empty office buildings,

The AI-genie has conjured itself into the empty office. The entire office is devoid of people, but the AI-genie is magically turning on all of the bulbs in the office.

or waking up your toddler in the middle of his afternoon nap.

The AI-genie has conjured itself into the house and is magically controlling all the lights – it looks sheepishly at the camera, while the blinding light that it has conjured makes the sleeping toddler sit up and bawl uncontrollably.

Whether the rules are written down by humans, or learned from historical data by machines, we should be sure to use the scientific method - the formulation and testing of a falsifiable hypothesis.

As philosopher of science Karl Popper famously said:

“A theory or idea shouldn’t be scientific unless it could, in principle, be proven false.”

A caricatured Karl Popper flips a coin in the air.

Page 12

The scientific method starts with an observation.

Our protagonist is dressed up like Sherlock Holmes – she holds up a magnifying glass, and looks straight at the camera through it – this distorts her eye and makes it look massive to the viewer through the glass.

It is followed by the formulation of a falsifiable hypothesis - one that can be proven false.

The protagonist sits contemplatively – both her hands are crossed on the mouth and her expression is severe.

Next, an experiment is designed to check whether the hypothesis is falsified.

The protagonist holds two beakers on top of a large pot – she frantically pours their contents into the concoction below, with an expression of anticipation.

And, if so, we refine our hypothesis, design a new experiment, rinse and repeat.

The protagonist is having a “Eureka” moment – one of her hands is on her lips, while the other points upwards in the air.

Page 13

The crucial question: “Does the classifier work?” is formalized as “Are the classifier’s predictions more accurate than a random guess”?

This is the lowest bar for accuracy – we don’t want to end up building fancy AI to be flipping coins.

The AI-genie is conjuring up a bulb and controlling its wattage. One snapshot behind it shows a person working in the dark and in need of light, whereas other snapshots show a sleeping woman and an empty office, both of which require the lights to be turned off.

When we have ground-truth information (observations of whether the light indeed should or should not be on) - we can check.

And if the hypothesis is falsified, we don’t use it.

We design a new data generation procedure and we retrain the model to learn a new rule.

Importantly, even with all this additional sophistication, it is still unlikely that we’ll see perfect accuracy.

One of the reasons for this is that there is uncertainty in the world.

Snapshot of a busy office – people work on different workstations, in a large open space. There are several question marks around the space, pointing to the uncertainty about whether they need the light, whether they prefer to work in a dim light and/or whether they’ll be leaving soon or working late.

And another is that rules are sometimes broken: people may come to work over the weekend, or they stay at work longer to meet a deadline.

And so the best we can hope for, in real-life situations, is that the classifier will work most of the time.

But once in a while, someone will need to get up and turn on or turn off the light.

The protagonist walks up to the light switch, and is about to flip it. She turns to the camera/viewer and winks knowingly.