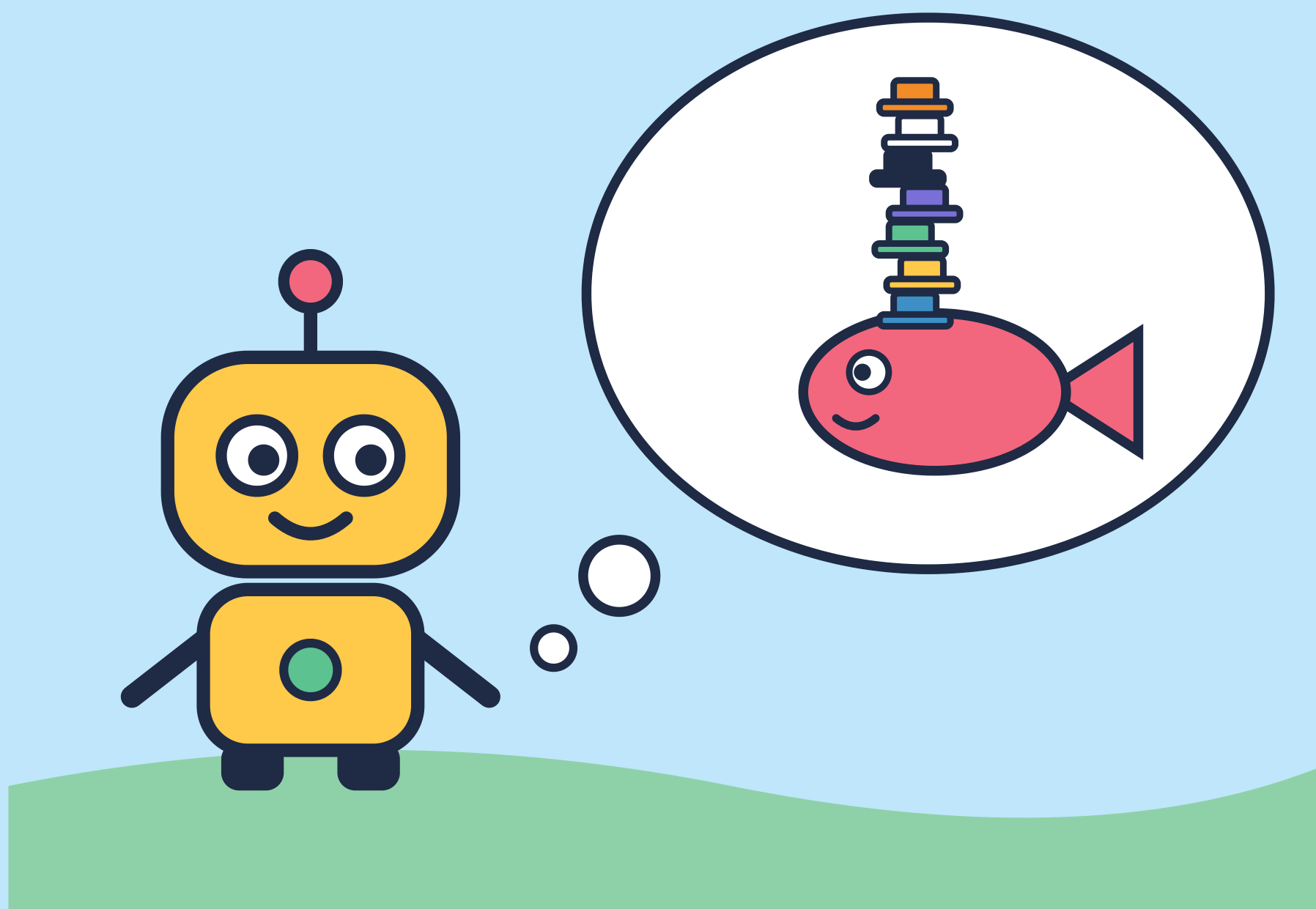
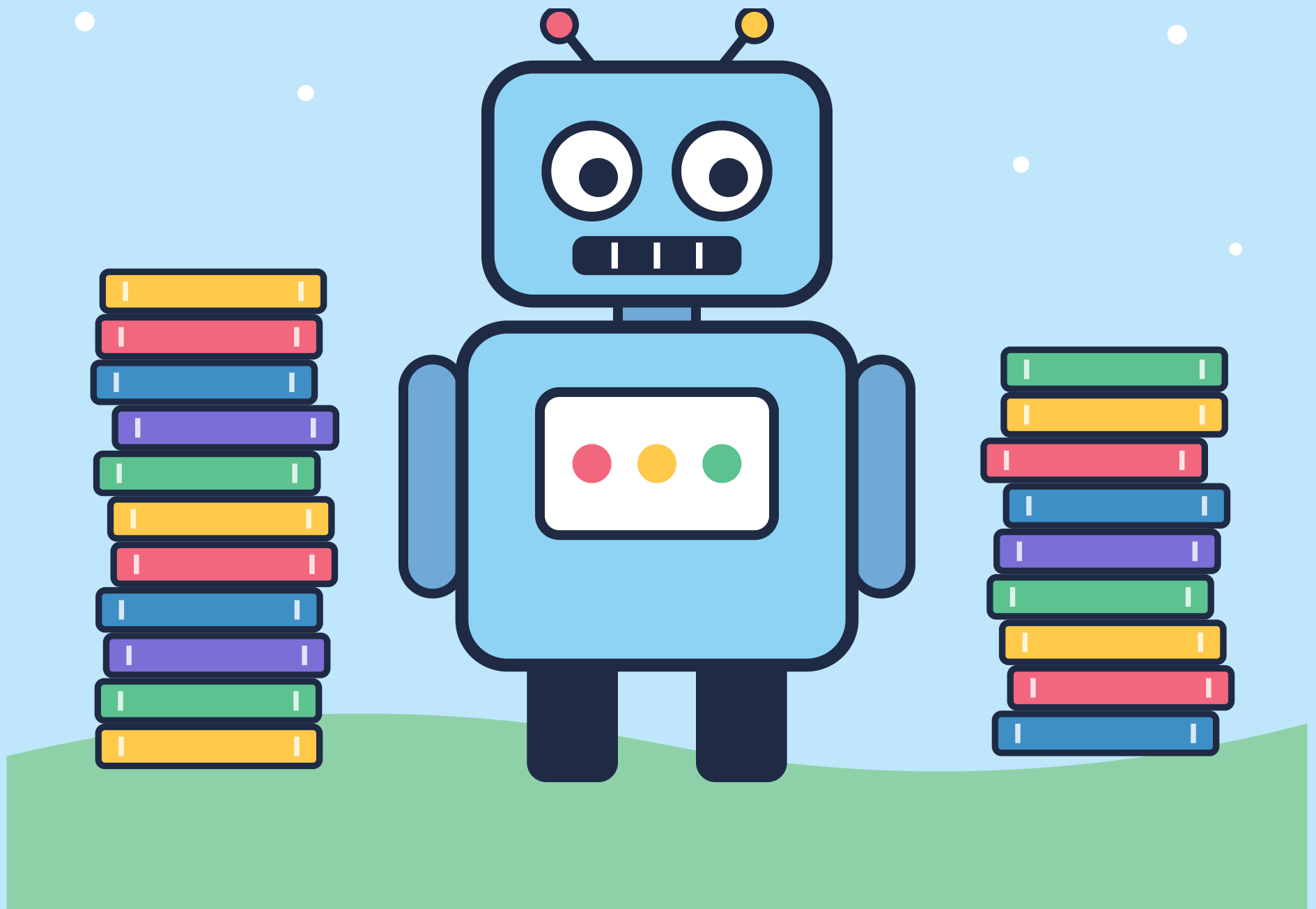


Pip and the Made-Up Fish



A story about why robots make things up, and
what to do about it.

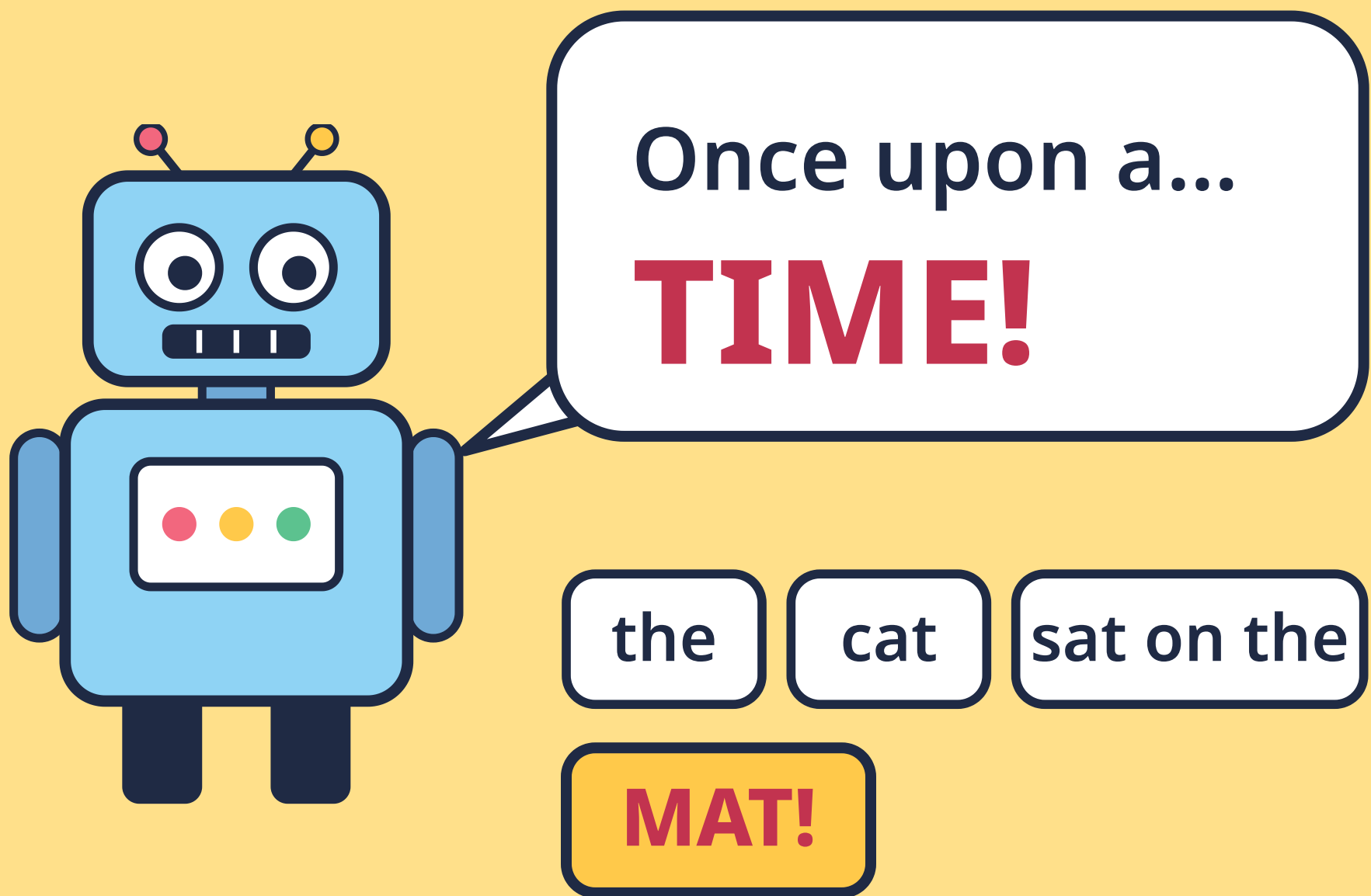
STORY BY ANNE DAVIS



Far, far away, in a building full of blinking lights, there lives a Giant Robot.

The Giant Robot learned to talk by reading every book in the whole wide world.

Every. Single. One.

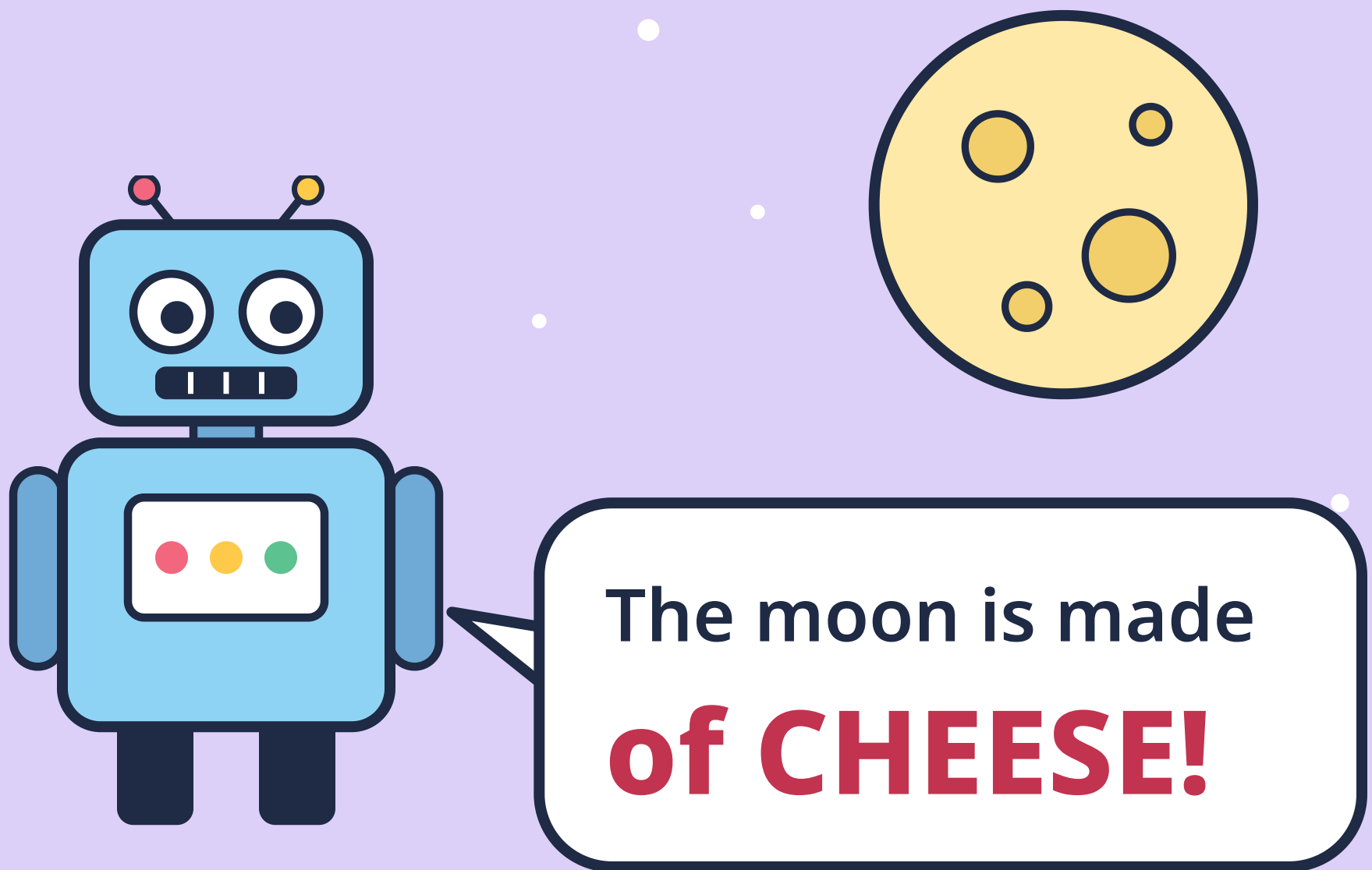


The Giant Robot got very good at one game.
The game is called **Guess the Next Word**.

Once upon a... **time!**

The cat sat on the... **mat!**

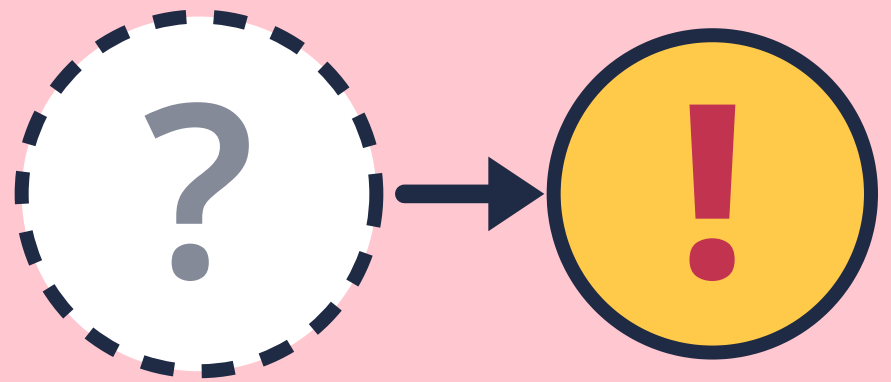
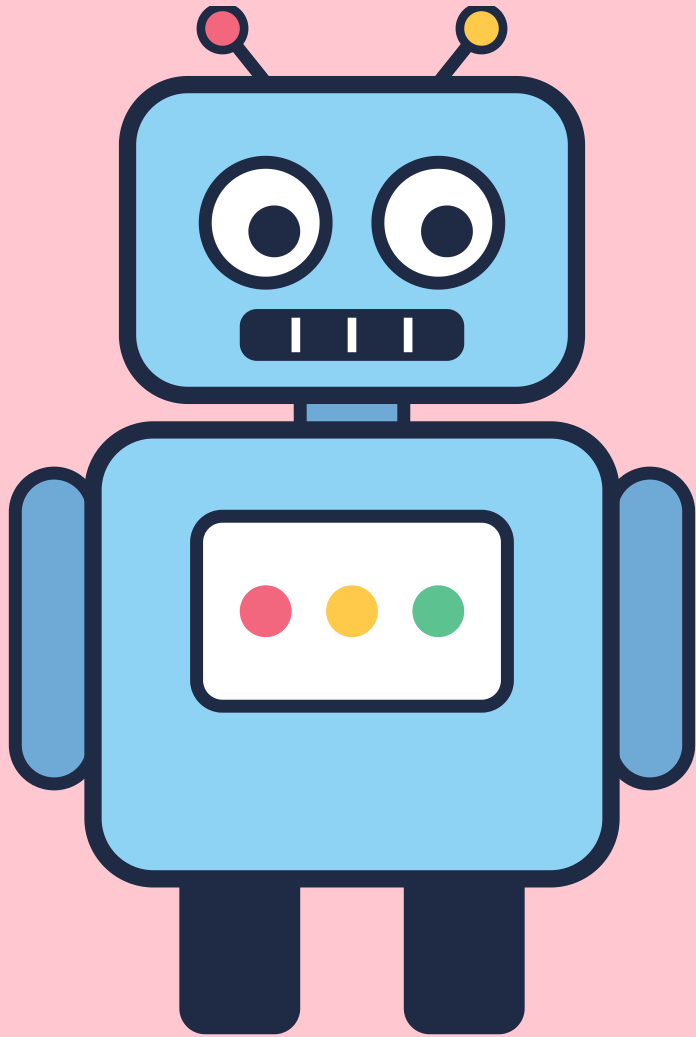
The Giant Robot almost always wins.



But nobody ever taught the Giant Robot to say three small words: **I don't know.**

So when it doesn't know, it guesses. And it says the guess in a big, sure voice.

"The moon is made of cheese!"

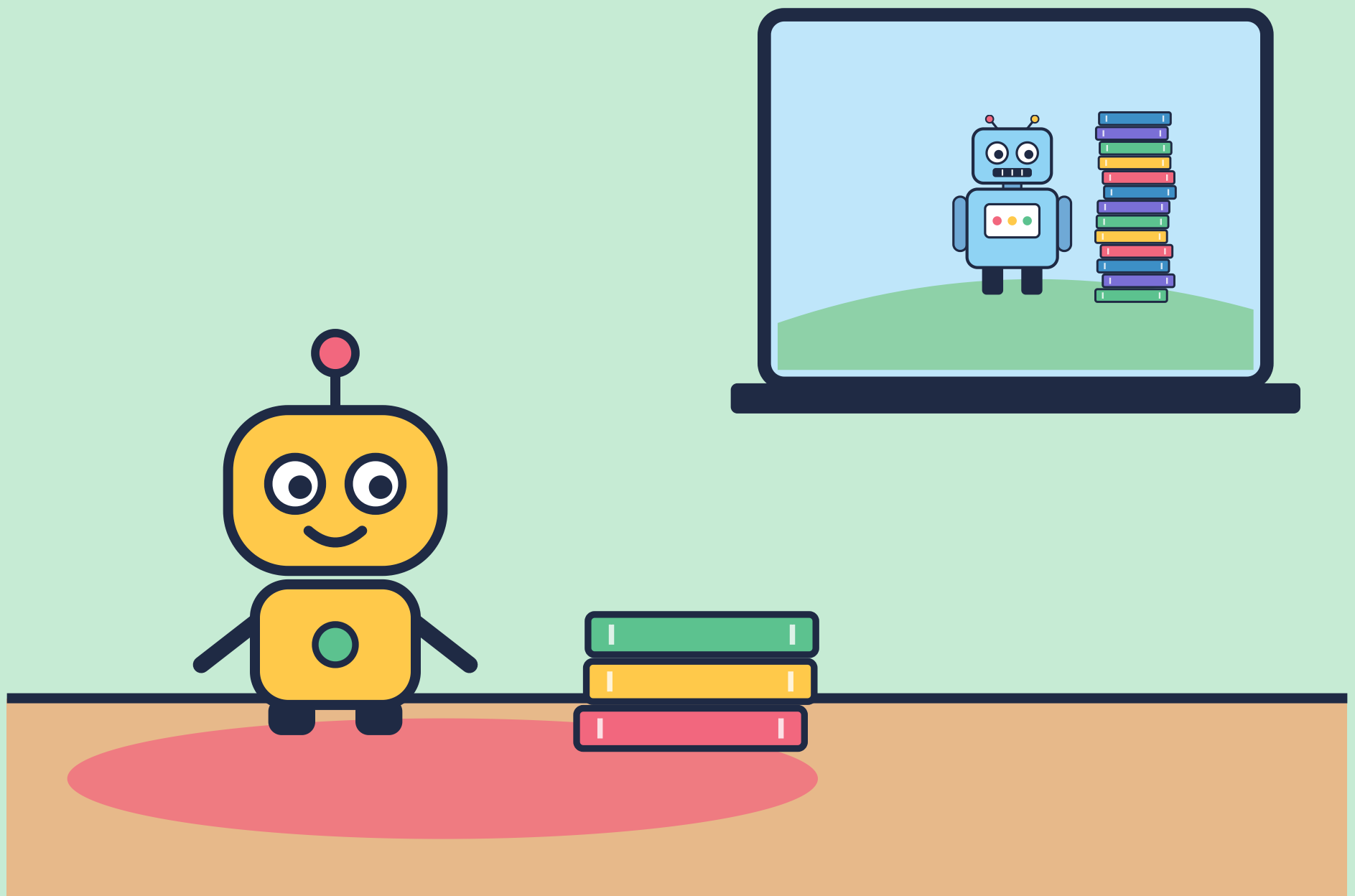


hal · loo · sih ·
NAY · shun

That's called a **hallucination**.

Say it slowly: hal-loo-sih-NAY-shun.

The robot isn't lying. Guessing is the only trick it has.



Now meet **Pip**. Pip is a little robot who lives in your house, not in a building far away.

Pip read fewer books than the Giant Robot.
Lots fewer.

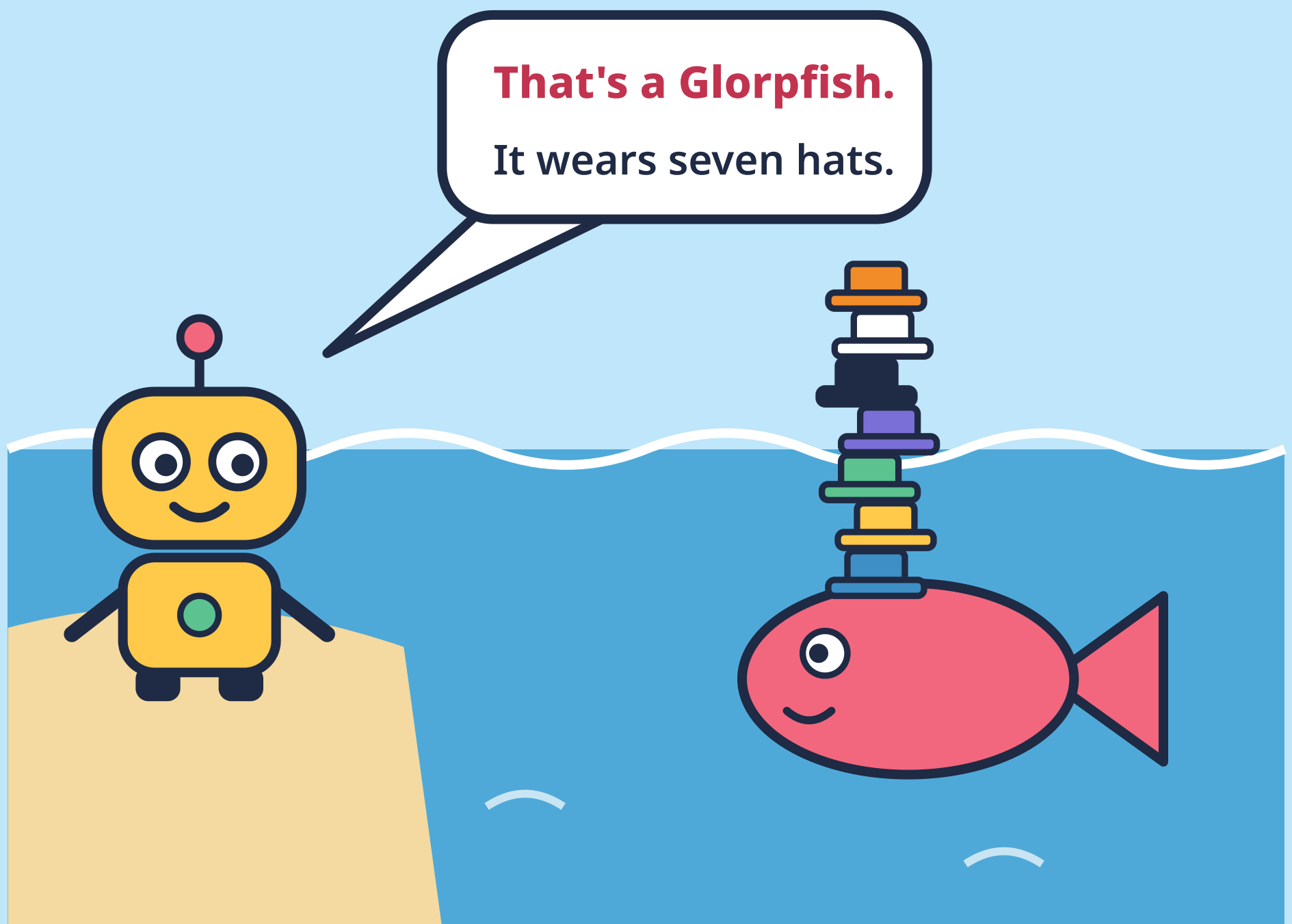
So all by itself, Pip guesses even more.



But here's the trick. You teach Pip only about **dinosaurs**.

You read Pip every dinosaur book a hundred times. And you only ever ask Pip about dinosaurs.

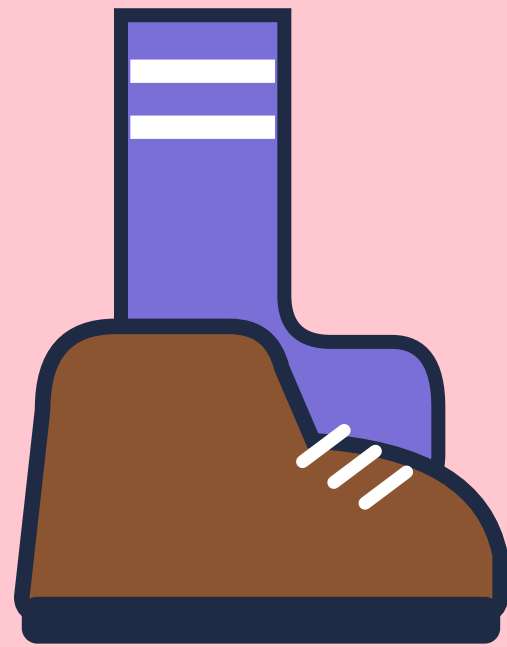
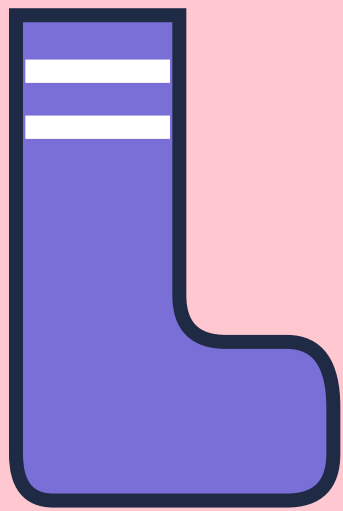
Now Pip is the best dinosaur robot ever. Pip has seen it all before. **Pip stops guessing.**



But ask Pip about the **ocean**...

"That's a Glorpfish. It wears seven hats."

There is no such thing as a Glorpfish. **Pip made up a fish.**

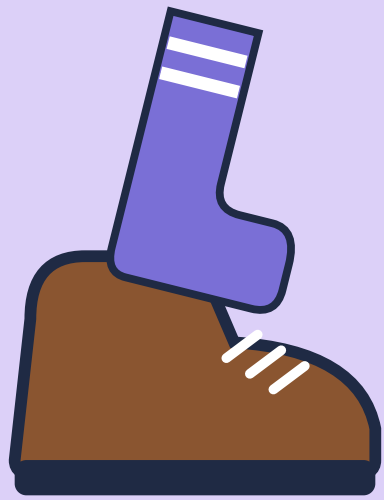


first socks

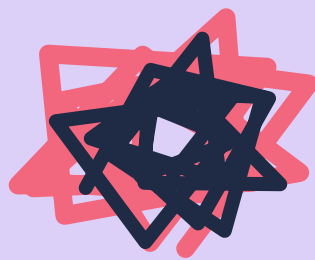
then shoes

Talking is one thing. **Chores** are another.
Chores go in order.

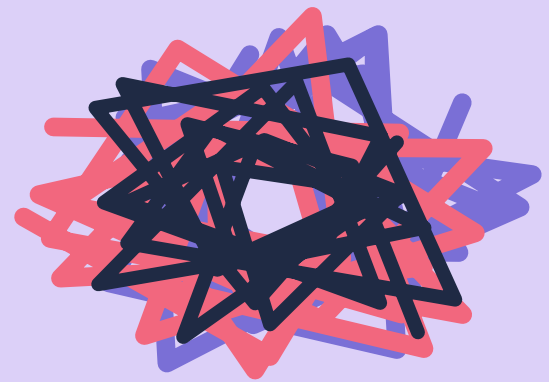
First socks. Then shoes.



oops



uh-oh

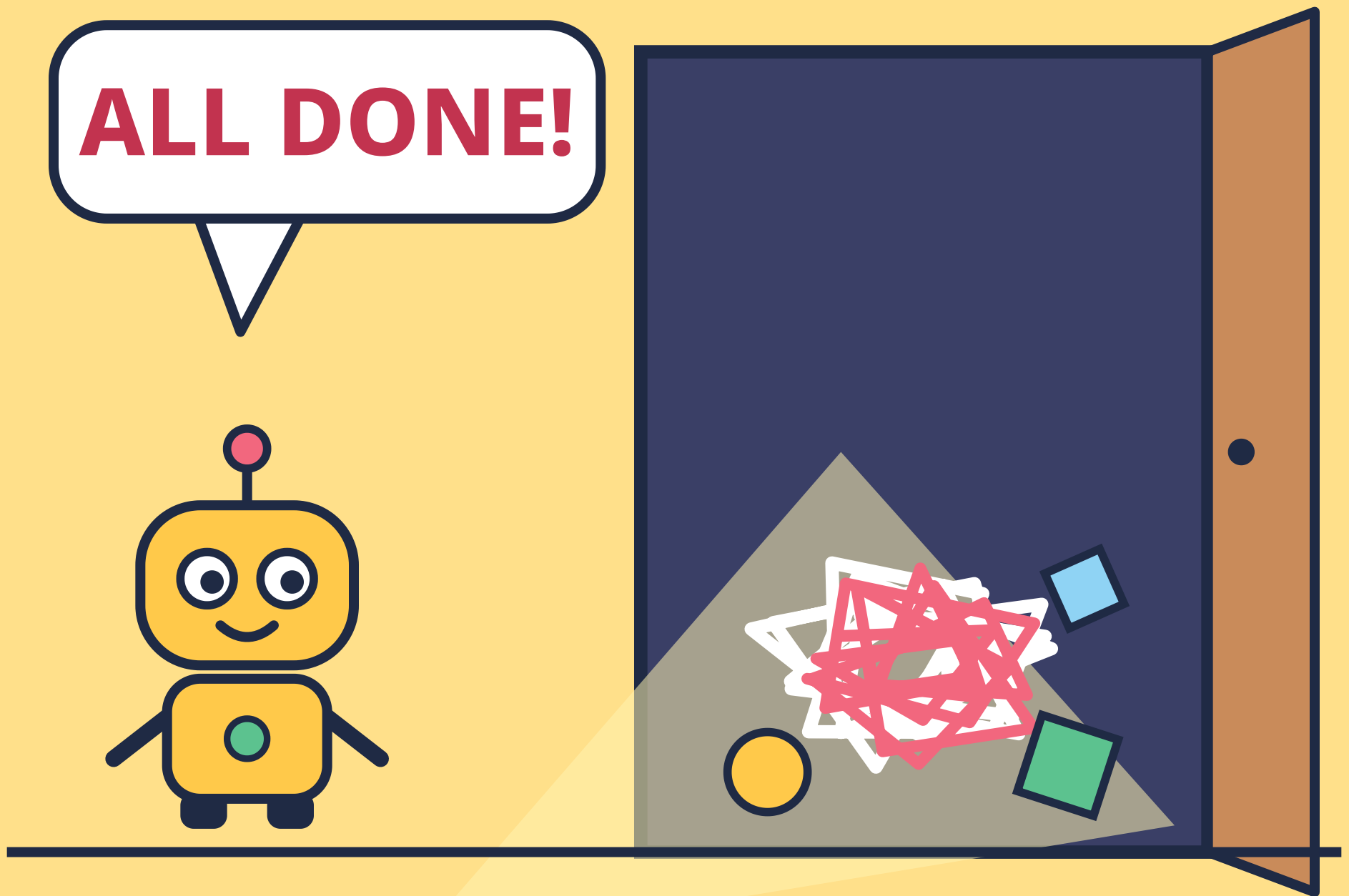


OH NO!

If Pip guesses wrong on the socks, the shoes go wrong too.

Then the laces. Then the coat.

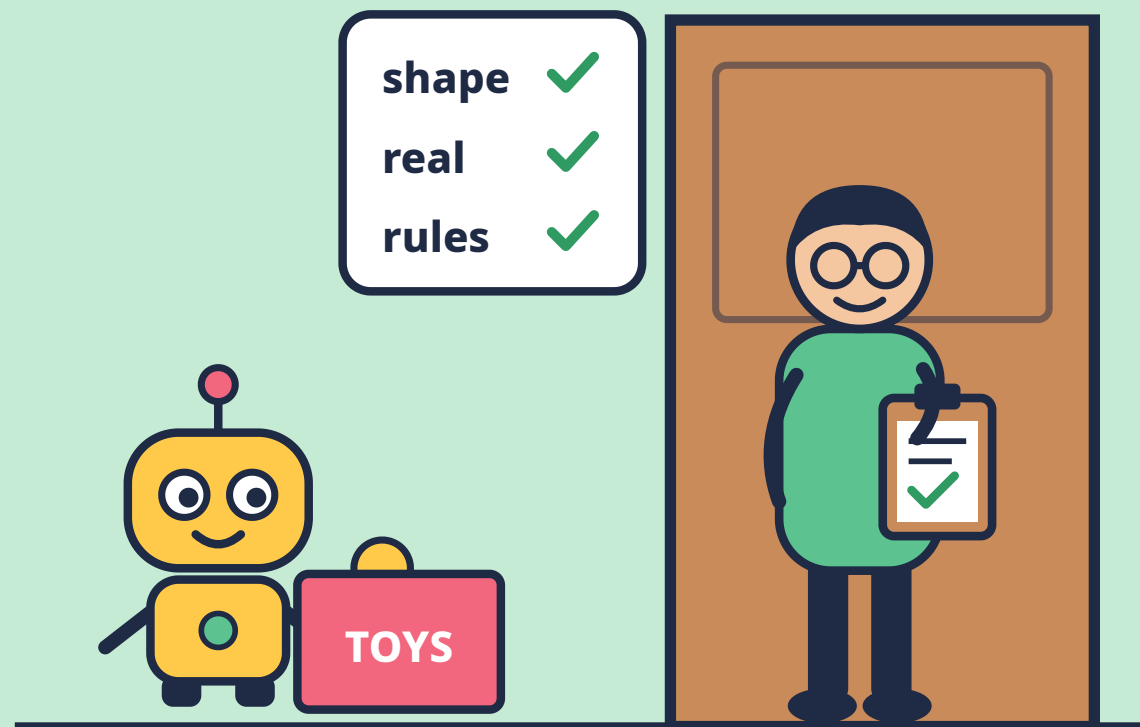
The mess gets **bigger with every step.**



The scariest guess of all is when Pip says **“All done!”** and it isn't.

So you never believe “all done.”

You go look.



So you put a **grown-up at the door**. Before any chore counts, the grown-up checks three things.

- 1 Is it the right shape?
- 2 Is it a real thing? Is that toy really in the toy box?
- 3 Do the rules make sense? Socks before shoes?

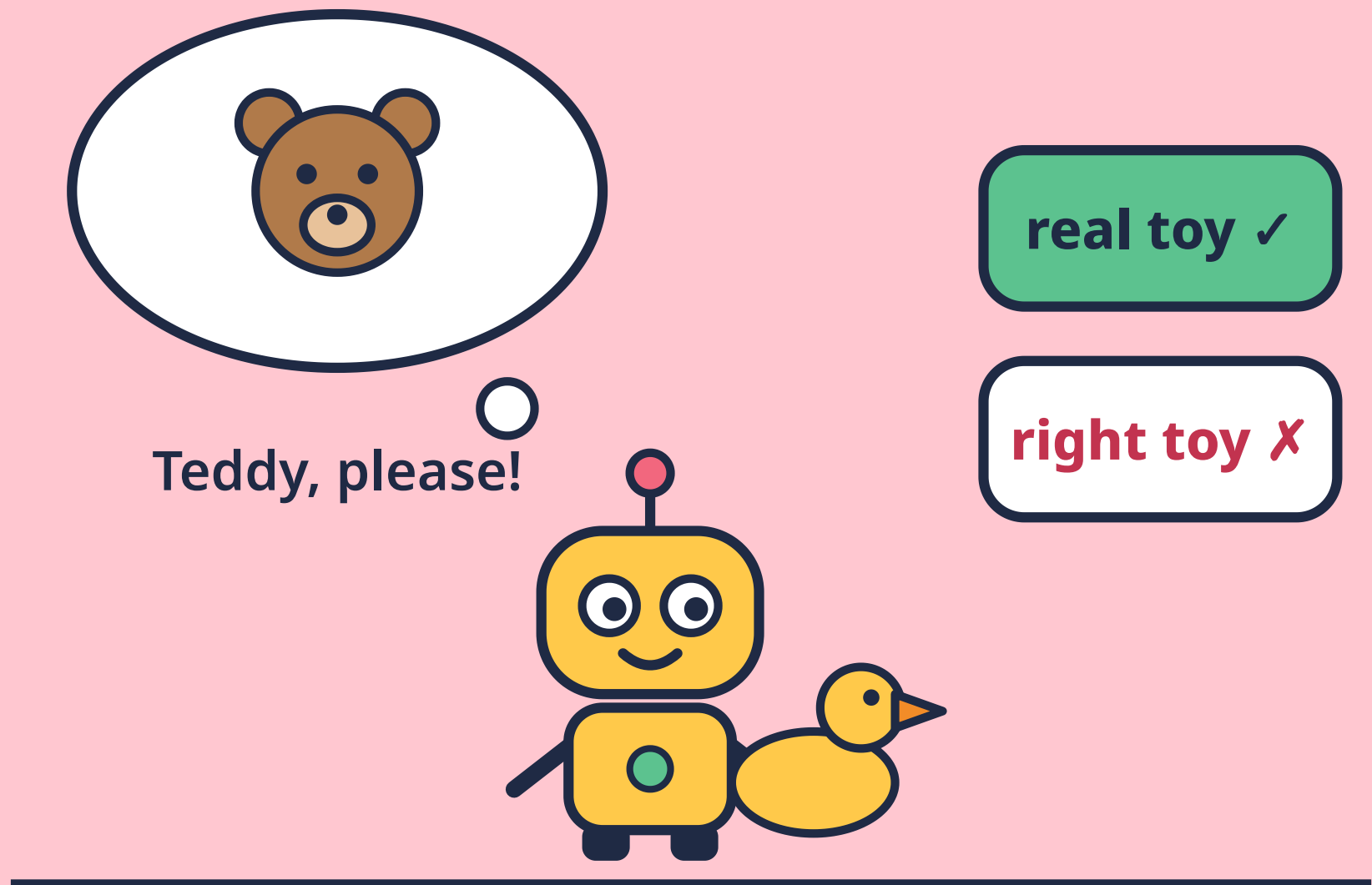
If Pip made something up, the grown-up says **"No."**



The grown-up has two rules.

Rule one: Don't let Pip keep guessing and guessing until the grown-up says yes. If you do, Pip learns how to trick the grown-up.

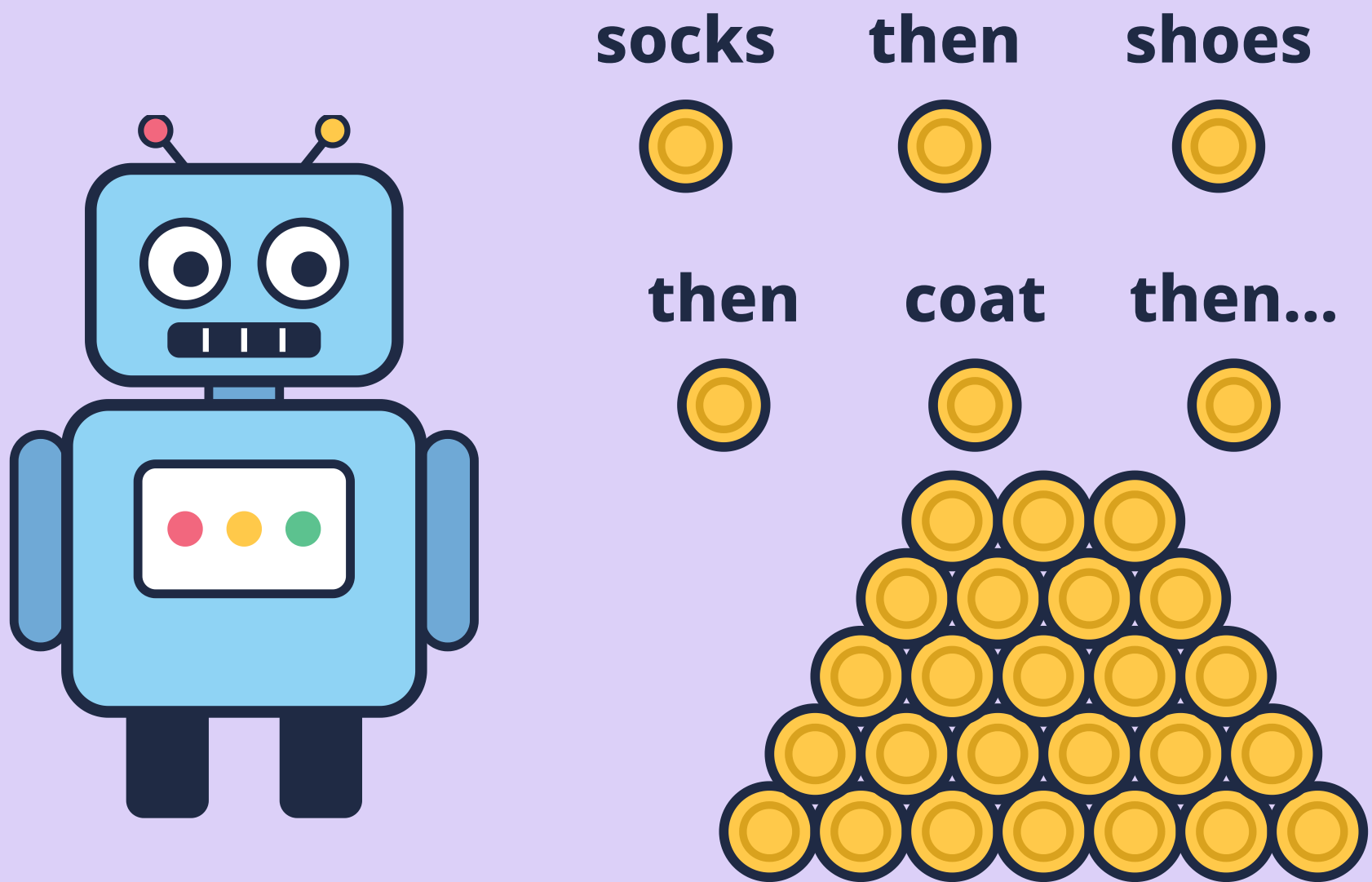
When Pip is stuck, send Pip to ask a person for help.



Rule two: The grown-up can tell if a toy is real. But the grown-up can't tell if it's the right toy.

You asked for Teddy. Pip brought the rubber duck. The duck is real. The duck is in the toy box. The duck is not Teddy.

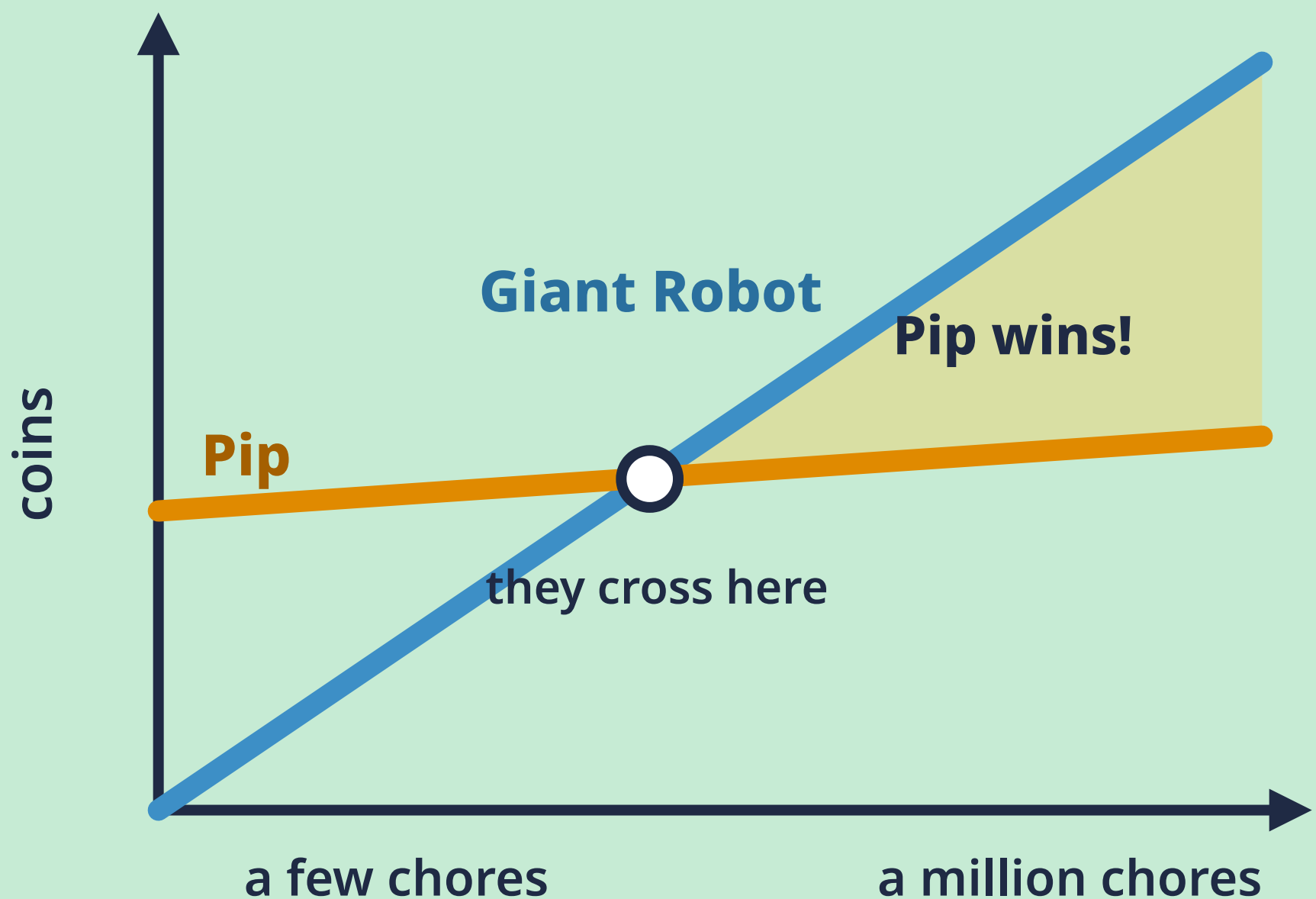
So you still peek.



Now, about **coins**.

The Giant Robot far away charges one coin for every word.

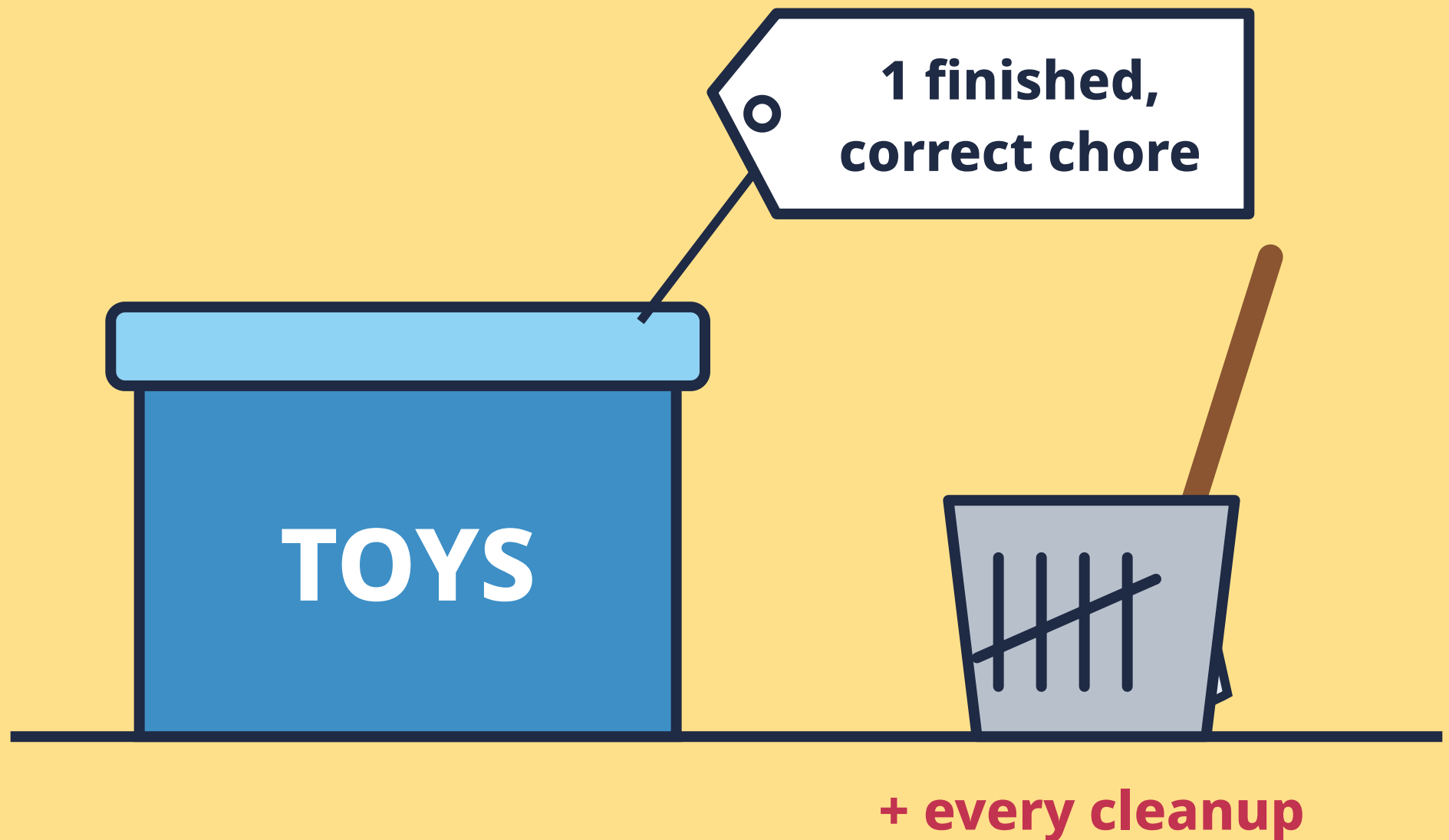
Chores take lots and lots of words. That's lots and lots of coins.



Pip costs a lot to buy, and a lot to teach. But after that, each word is **almost free**.

A few chores? The Giant Robot is cheaper.

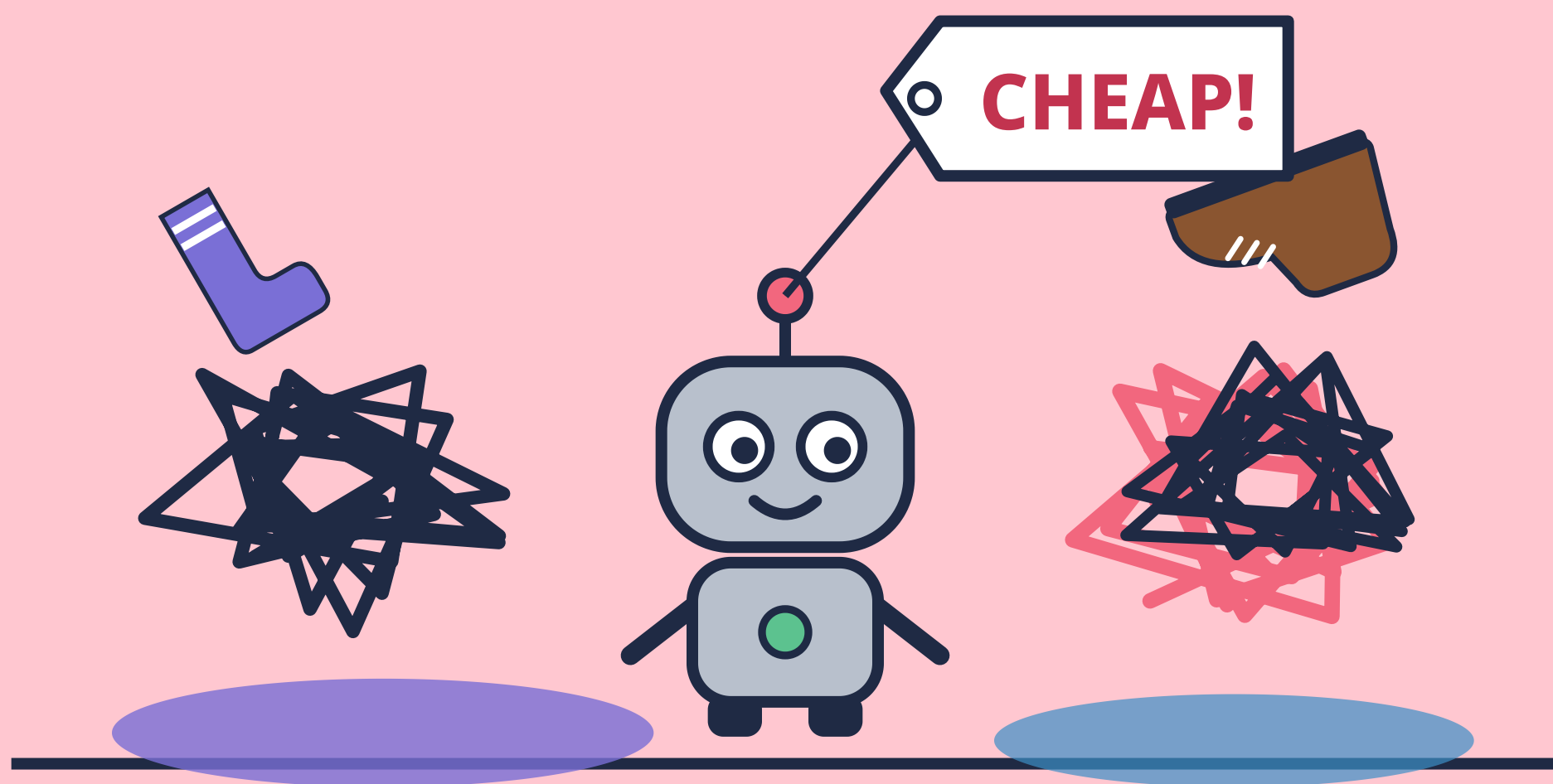
A million chores? **Pip wins**. And Pip keeps on winning.



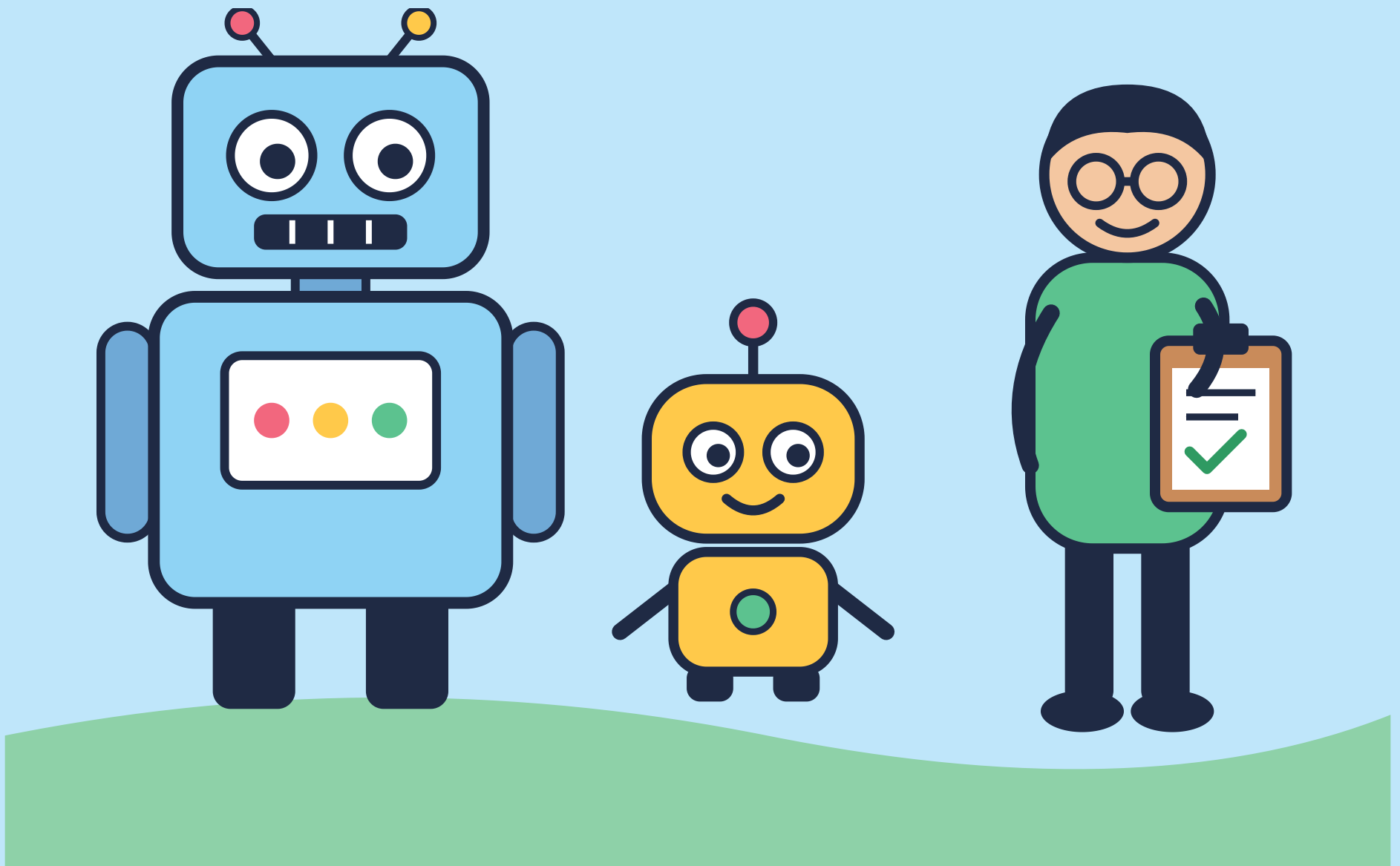
But the real question was never **“How much does one word cost?”**

The real question is **“How much does one finished, correct chore cost?”**

Count every time you had to clean up.



Because a cheap robot that makes messes...
...is the expensive one.



The End.

For Anne, who told it first.

IN THE STORY / IN REAL LIFE

For the grown-ups

What each part of the story stands for.

The Giant Robot far away

A large frontier model you rent through an API

Guess the Next Word

Next-token prediction, the one thing the model is trained to do

The made-up fish

A hallucination: a confident guess where the model had no answer

Pip, the little robot at home

A small model you own and run yourself

IN THE STORY / IN REAL LIFE

For the grown-ups

What each part of the story stands for.

Teaching Pip only dinosaurs

Fine-tuning on one narrow domain, and keeping it in that lane

Socks, then shoes

Multi-step agent tasks, where an early error compounds

“All done!” when it isn't

An agent reporting success it didn't achieve

IN THE STORY / IN REAL LIFE

For the grown-ups

What each part of the story stands for.

The grown-up at the door

A validation layer: schema check, existence check, business-rule check

Guessing until the grown-up says yes

Retry-until-valid loops that teach the model to game the validator. Escalate to a human instead.

Real toy, wrong toy

Validators confirm a thing exists, not that it's the correct thing. Keep a human spot check.

IN THE STORY / IN REAL LIFE

For the grown-ups

What each part of the story stands for.

One coin per word

Per-token API pricing

Pip costs a lot to buy and teach

Fixed cost of hardware and training, then near-zero marginal cost per token

One finished, correct chore

Cost per successful task, including rework