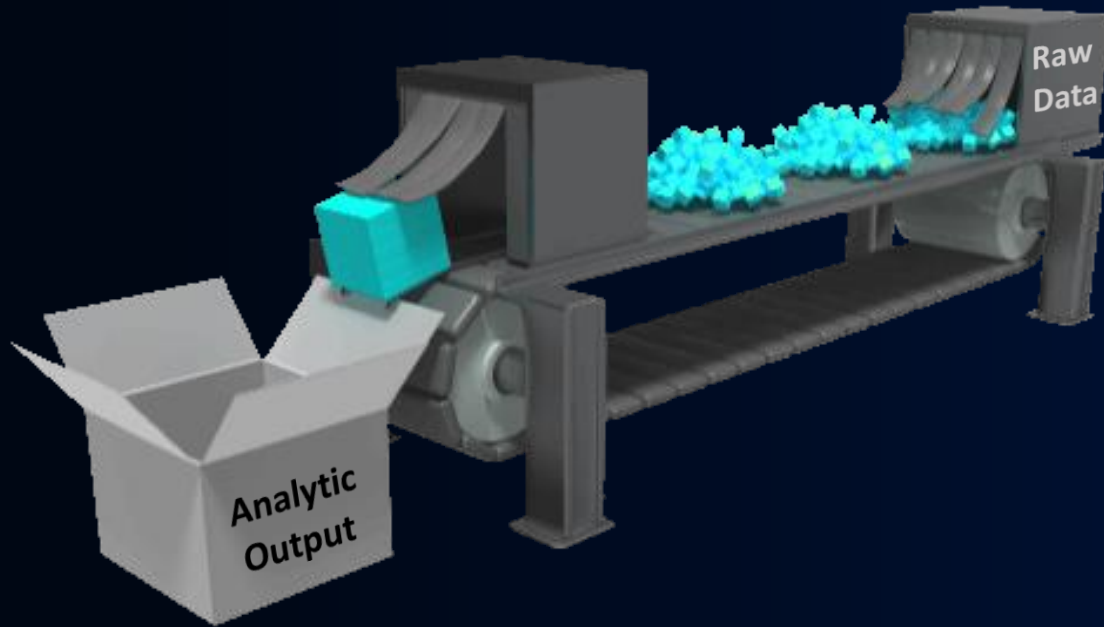


Data Integrity



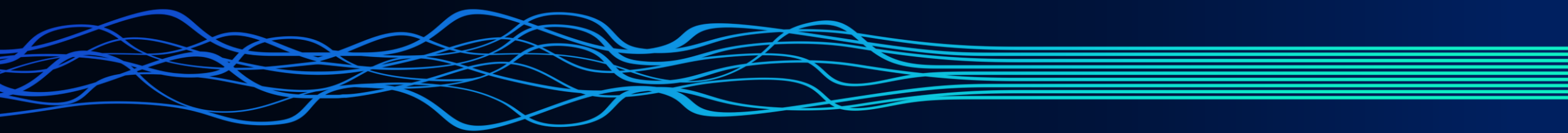
How is
Data Literacy

connected
to

Data Collection
?

Questions to ask ourselves about data literacy

Do we ignore the ways that employees already interact with data?





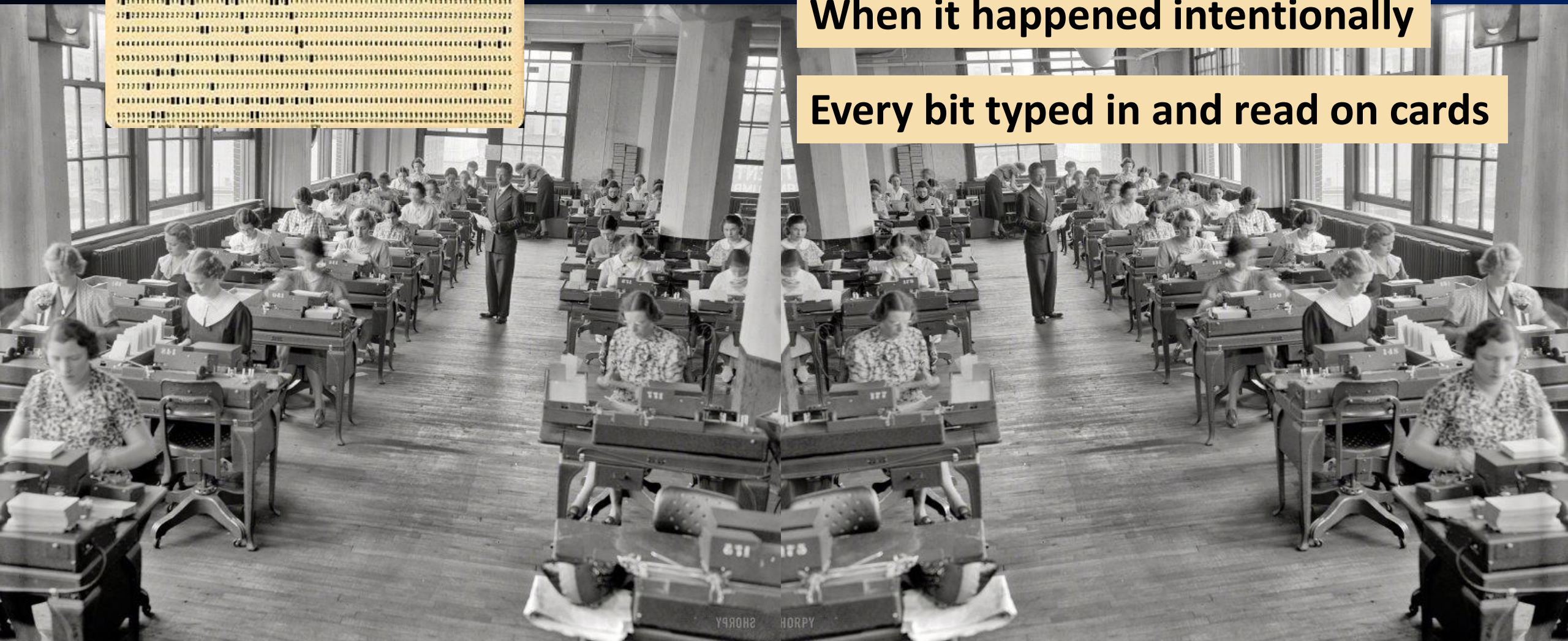
analytic-translator.com

Historically, it was hard to get data into a computer



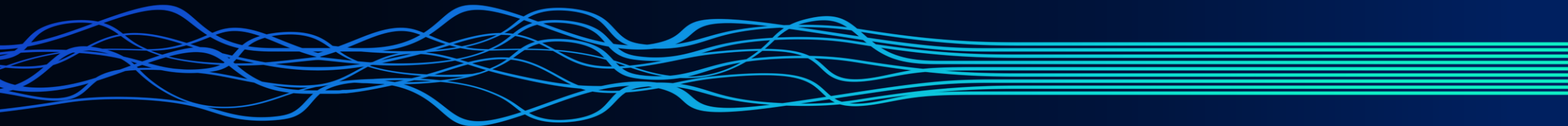
When it happened intentionally

Every bit typed in and read on cards



It was common to:

- Collect data on paper, hand-written
- Hire a professional (human) data entry team
- Double enter the same information to ensure accuracy
- Save stacks of punch cards for years
 - (because it wasn't stored anywhere)



Now we leave behind data like dust

When

How many

How much

What

What time

Who

duration

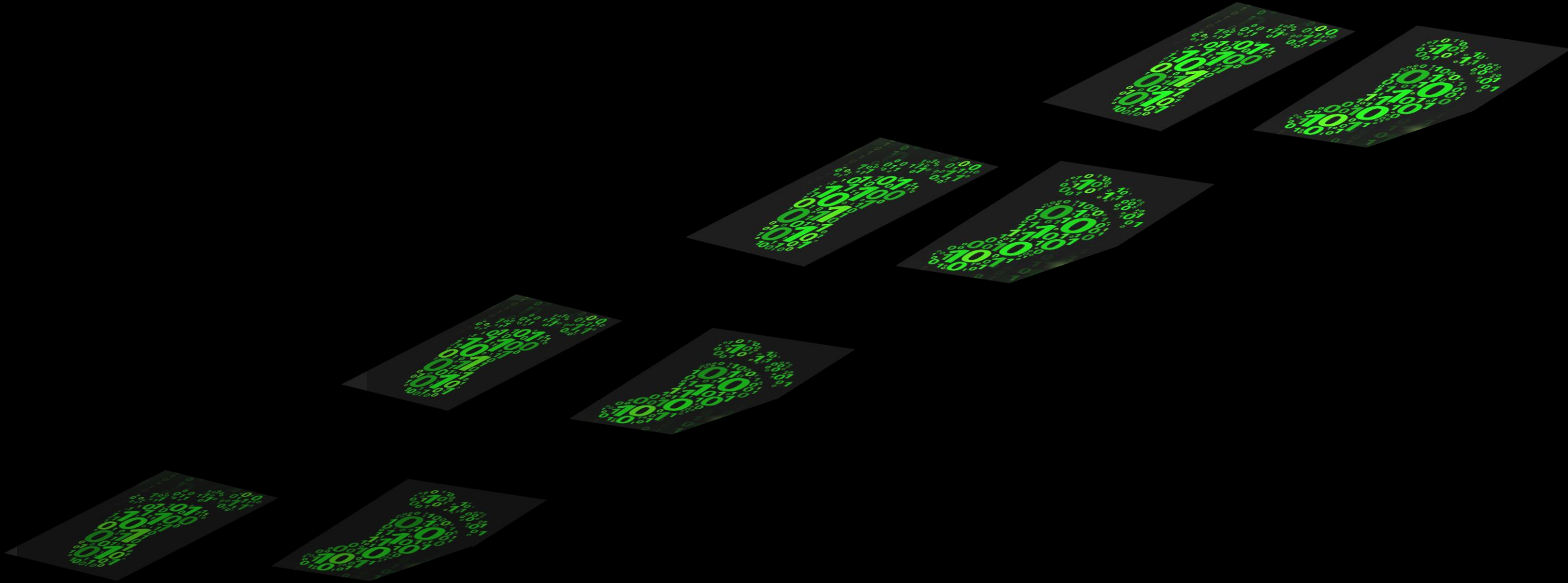
Where

price



Our digital footprints

Are only getting bigger and brighter



Today, **everyone** is a data creator

Every purchase on a credit card

Every amazon review

Every phone call

Every ATM withdrawal

Every email

Every text

Every Google search

Every Netflix movie viewed

Every story read

Every Like on Instagram

Every camera we pass by

Every report we download

Every prospect we list in Salesforce

Every click on a website

Every location our phone tracks

Every step on our fitbit

Every person we tag

Every score we enter

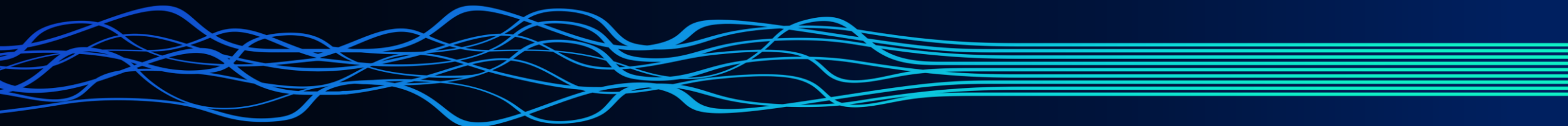
Every group we belong to

Every prescription we fill



Can you trust the information?

We all impact data quality



Choosing how honest to be

Let's assess health behaviors



A majority of respondents underreport unhealthy answers



Choosing whether to participate



Will you answer a survey?



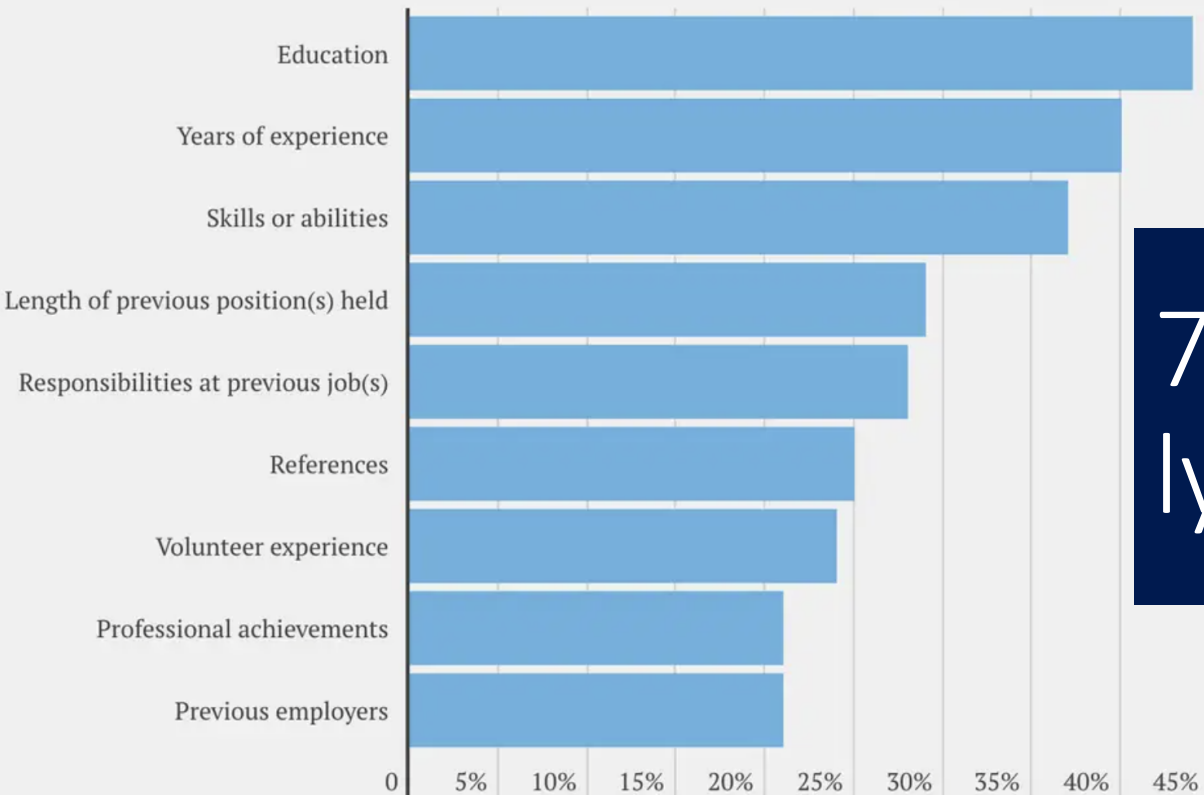
Surveys Face Growing Difficulty Reaching, Persuading Potential Respondents

	1997	2000	2003	2006	2009	2012
	%	%	%	%	%	%
Contact rate (percent of households in which an adult was reached)	90	77	79	73	72	62
Cooperation rate (percent of households contacted that yielded an interview)	43	40	34	31	21	14
Response rate (percent of households sampled that yielded an interview)	36	28	25	21	15	9

Deciding what to reveal

“We hire only the best”

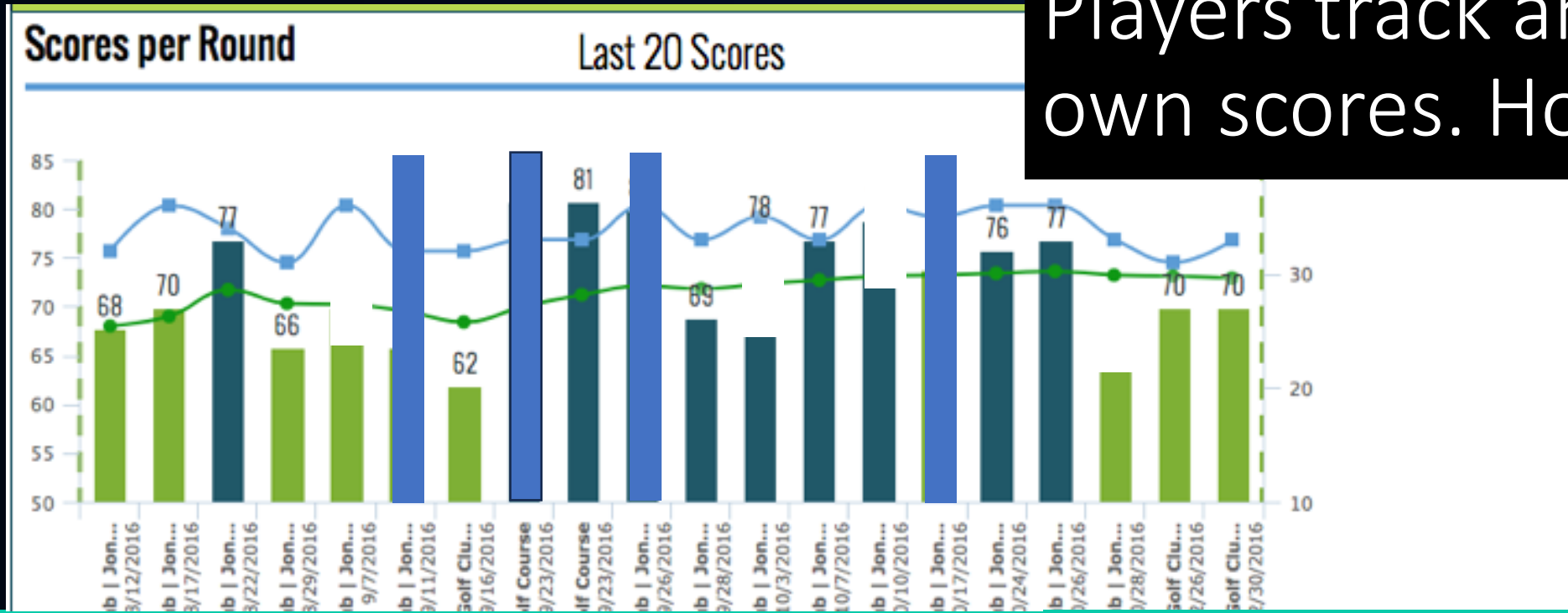
Lies Candidates Told on Resumes



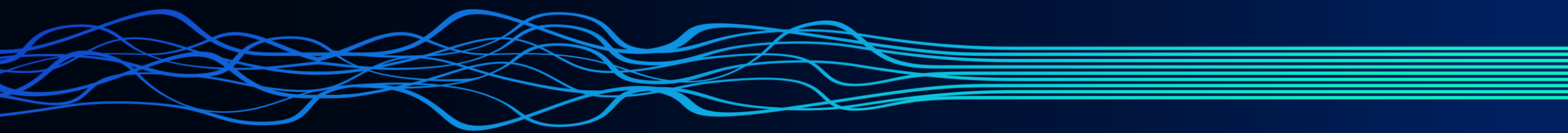
72% of respondents admit lying on their resume

Choosing how to manipulate data

Players track and enter their own scores. Honor system.



John wins every tournament. Mary loses every match.



Today, **everyone** is a data creator **and** a data evaluator

How do you decide if you believe it?

Which Product to Choose

"I look at the reviews"

50 five-star reviews for \$259



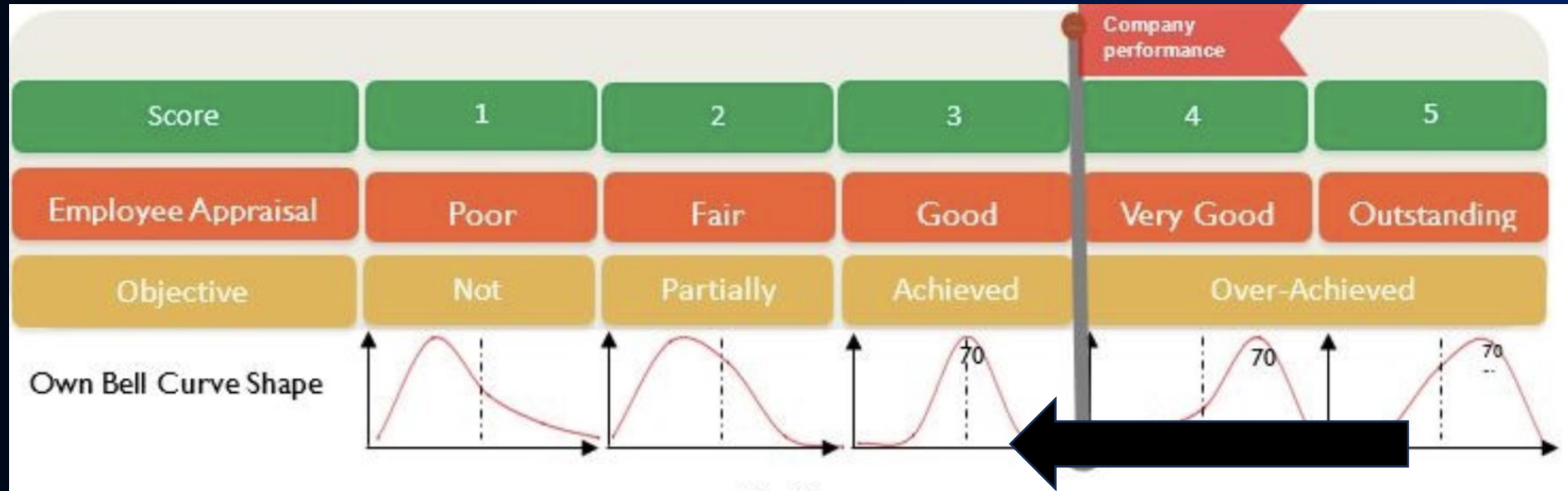
Buy Google Reviews with Fast Delivery

UseViral offers only the highest quality services. Buy safely and securely below:

1 Google	- \$7.5
2 Google	- \$14.5
3 Google	- \$19
5 Google	- \$32
10 Google	- \$75
15 Google	- \$99
20 Google	- \$120
50 Google	- \$259

24/7 Support

If My Performance Review is Fair



There are too many 5's. Make it a normal curve

“Your rating has gone down. But don't worry.”

If there is evidence to convict?

One in 10 Billion

“But, how can they know?”

A lack of Scientific Literacy

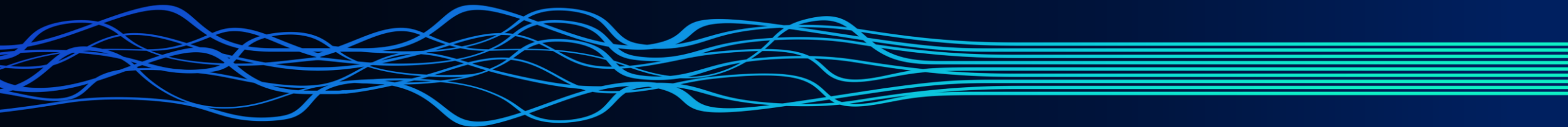
DNA Match



Questions to ask ourselves about data literacy

Do we ignore the ways that employees already interact with data?

Do we use over-technical language others don't understand?



Common Data Literacy Curriculum

Data quality

Statistics

Manipulation

Coding

Visualization

Interpretation

Internal validity

External validity – representativeness

Reliability

Accuracy

Collection bias

Inter-rater reliability

Test-retest reliability

Construct validity

Terminology matters:

Data integrity = Can we trust the data?

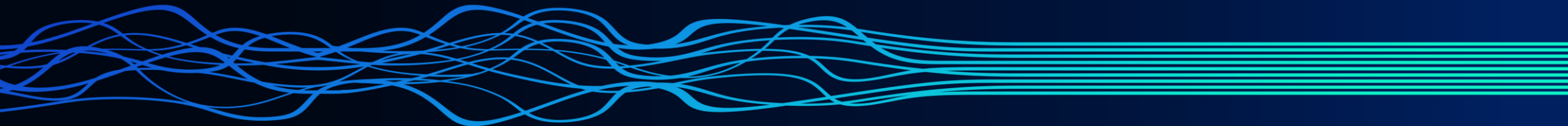
- How do you decide to believe something?
- In what ways could the information be wrong?
- How might that influence your thinking?
- Is there bias one direction or another?
- Can we influence that?

Questions to ask ourselves about data literacy

Do we ignore the ways that employees already interact with data?

Do we use over-technical language others don't understand?

Do we create an adversarial or derogatory dynamic?

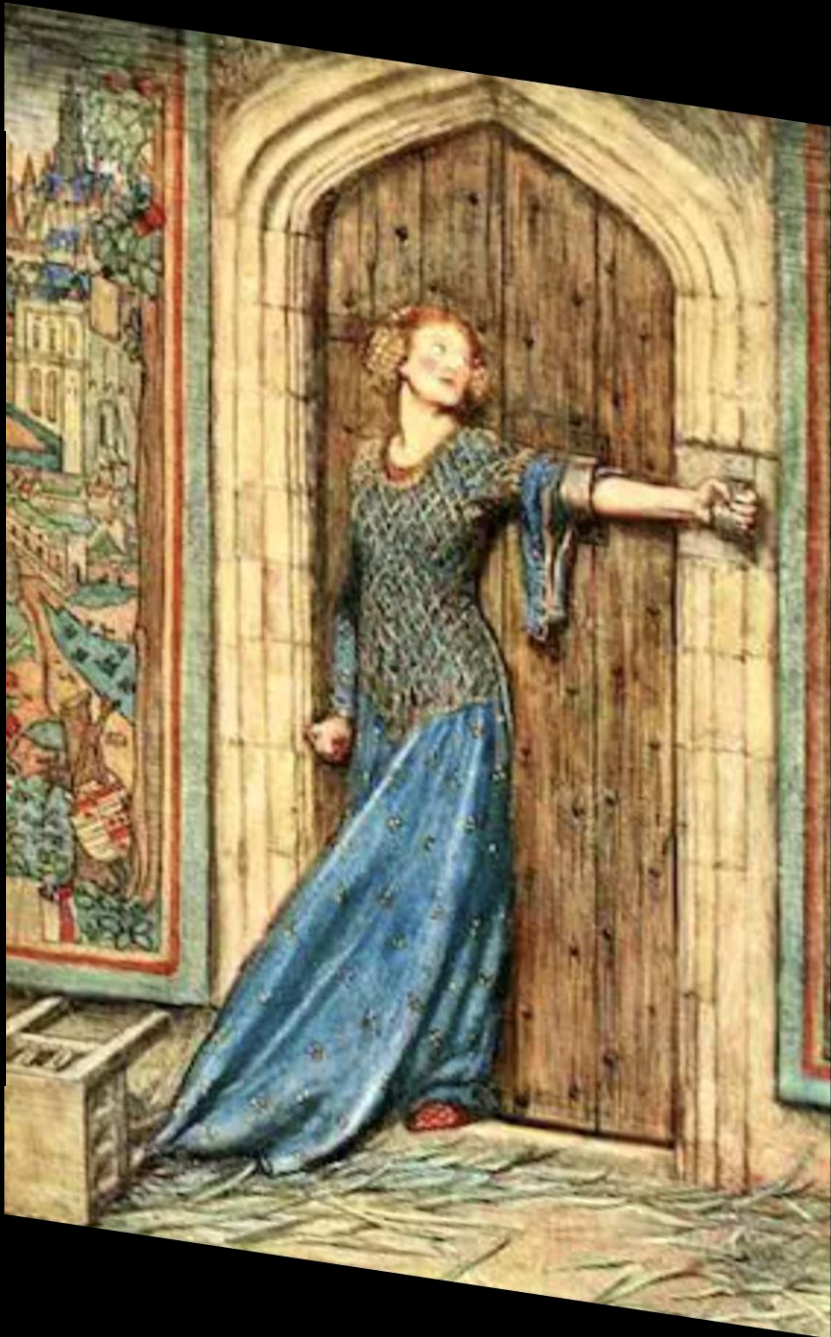




Katy bar the door!!!

The **illiterates** are coming to use our data!!

We mustn't let them in.....



So, who are these **illiterates** we must protect ourselves from?

80% of Employees

76% of Business decision-makers

68% of the C-Suite

**The least literate teams are
Human Resources and Sales**

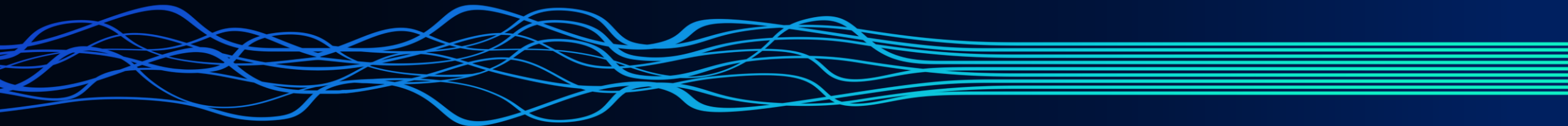
Questions to ask ourselves about data literacy

Do we ignore the ways that employees already interact with data?

Do we use over-technical language others don't understand?

Do we create an adversarial or derogatory dynamic?

Do we assume it's THEIR job to learn about our field?



How literate?

Ninety percent of business leaders believe data literacy will be critical to their success.

Data Literacy is set to be the second language of business.

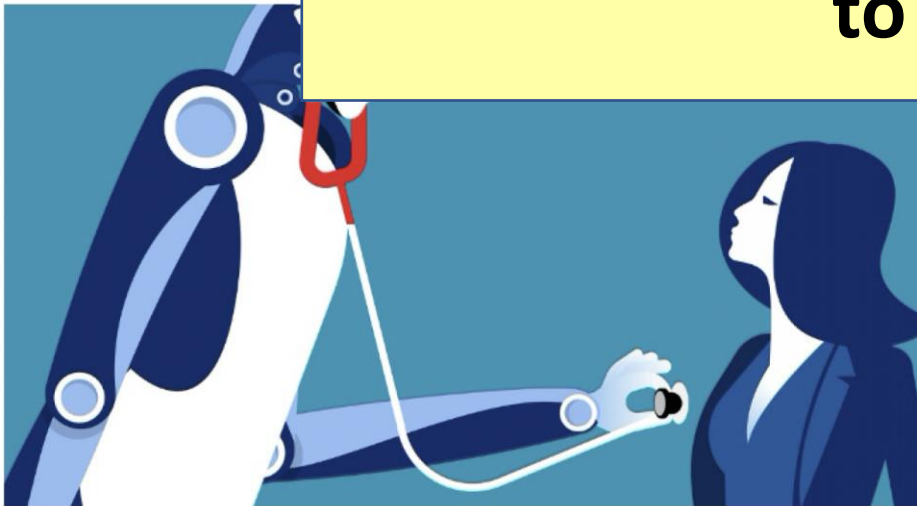
At the very least, individuals should know how key business terms and metrics are defined, particularly those relevant to their function. Ideally, they should also have a basic grasp of statistics and the ability to interpret charts and analytics accurately.



Why We Need Patients to Trust AI in Medicine

“To me, the next generation of clinicians all have to be data scientists”

by Celi



Doctors and engineers are asking: Can we trust Dr. AI?

Think about this.....

Every Doctor

Every CEO

Every Psychologist

Every Engineer

Every HR Director

Every Employee



Questions to ask ourselves about data literacy

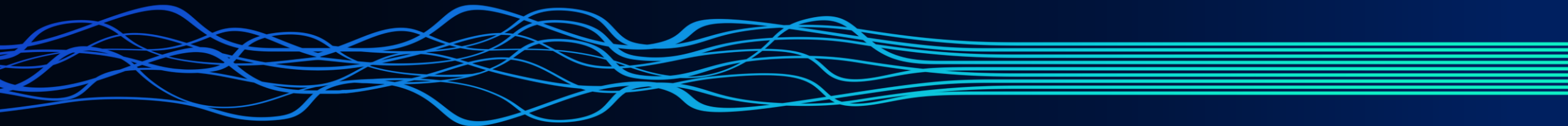
Do we ignore the ways that employees already interact with data?

Do we use over-technical language others don't understand?

Do we assume it's THEIR job to learn about our field?

Do we create an adversarial or derogatory dynamic?

Do we assume interest and aptitude?



WELCOME TO LITERACY TRAINING!!

More Math = More Fun !!



What are we up against?

One third of Americans don't know that a quarter of a pie is the same as 25%



54% say they simply smile and nod rather admit they don't understand data or statistics

★ ABC BANK Statement Ending 09/21/2021 Page 2 of 2

Look over all transactions

ABC RELATIONSHIP CHECKING XXXXXXXX1284 (continued)

Primary Checking		Paycheck	
Account Activity	Debits	Credits	Balance
09/10/2021 Signature POS Debit 09/08 GIANT FOOD I	\$78.22		6,806.09
09/10/2021 L A FITNESS	\$15.38		6,790.71
09/11/2021 FT&T MOBILITY ONLINE PMT	\$100.30		6,690.41
09/14/2021 DEPOSIT		\$500.48	7,190.89
09/14/2021 DIRECT DEP		\$2,576.34	9,767.23
09/14/2021 Signature POS Debit 09/13 MD GIANT FOOD	\$15.40		9,751.83
09/17/2021 ATM Withdrawal 09/15 WV INWOOD	\$350.00		9,401.83
09/17/2021 Signature POS Debit 09/16 MD GIANT FOOD	\$12.48		9,389.35
09/17/2021 Signature POS Debit 09/15 MD GIANT FOOD	\$35.80		9,353.55
09/18/2021 THE HOME DEPOT ONLINE PMT POS	\$8.35		9,345.20
09/18/2021 00RETIRE 0503R3030	\$528.12		8,817.08
09/20/2021 373LA@71557195	\$489.57		8,327.51
09/21/2021 Ending Balance			8,327.51

Daily Balances			Daily Balances		
Date	Amount	Date	Amount	Date	Amount
09/04/2021	19,120.43	09/10/2021	6,790.71	09/18/2021	8,817.08
09/05/2021	7,017.64	09/11/2021	6,690.41	09/20/2021	8,327.51
09/06/2021	6,562.32	09/14/2021	9,751.83		
09/07/2021	7,039.31	09/17/2021	9,353.55		

Overdraft and Returned Item Fees Watch for unexpected fees

Total for this period	Total year-to-date

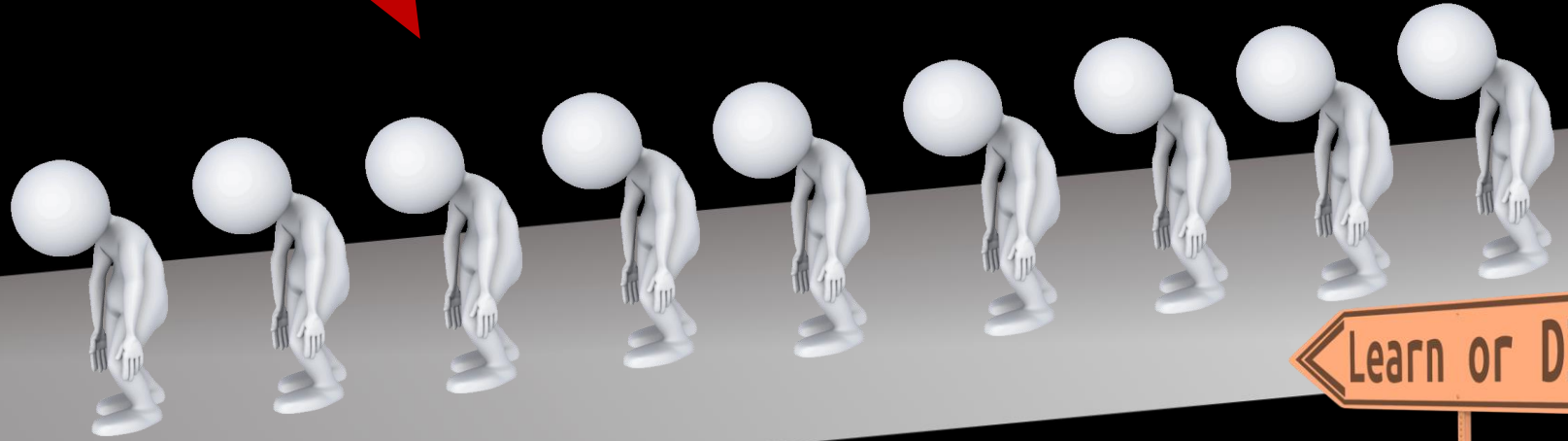
22% reveal they can't understand everyday numeric information, like bank statements

Literacy Training

This Way

MATH AHEAD

Learn or Die



Questions to ask ourselves about data literacy

Do we ignore the ways that employees already interact with data?

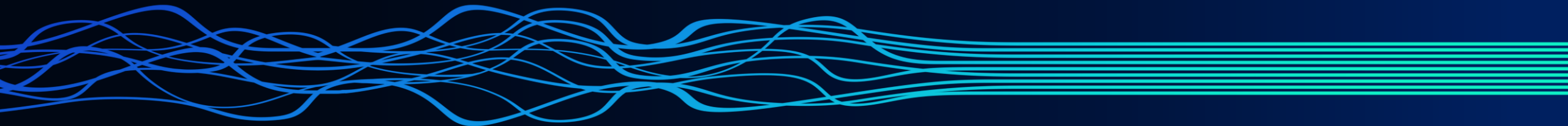
Do we use over-technical language others don't understand?

Do we assume it's THEIR job to learn about our field?

Do we create an adversarial or derogatory dynamic?

Do we assume interest and aptitude?

Do we separate this learning from other learning areas?



Strategic Alignment (a.k.a. Business Literacy)



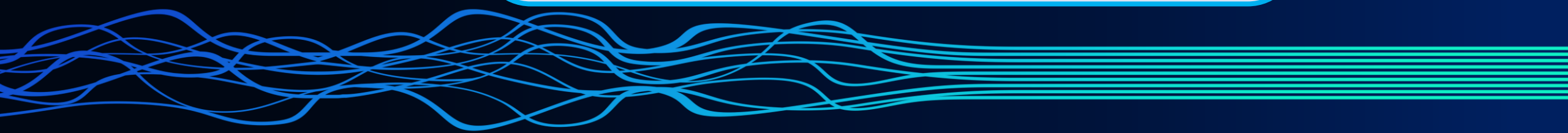
Employees must understand strategy to implement it

Why Your Employees Need to Understand 'The Why'

According to a study by PWC, **93% of employees** could not articulate their company's strategy

Companies whose people understand strategy grow revenue faster and are **72% more profitable**

Only 13% of frontline managers could name their company's top three priorities



Emotional/Social Awareness (a.k.a. People literacy)

52% of HR leaders say they will be hiring managers based on their emotional intelligence

Forbes

Digital Transformation Is
80% EQ And 20% IQ

Forbes

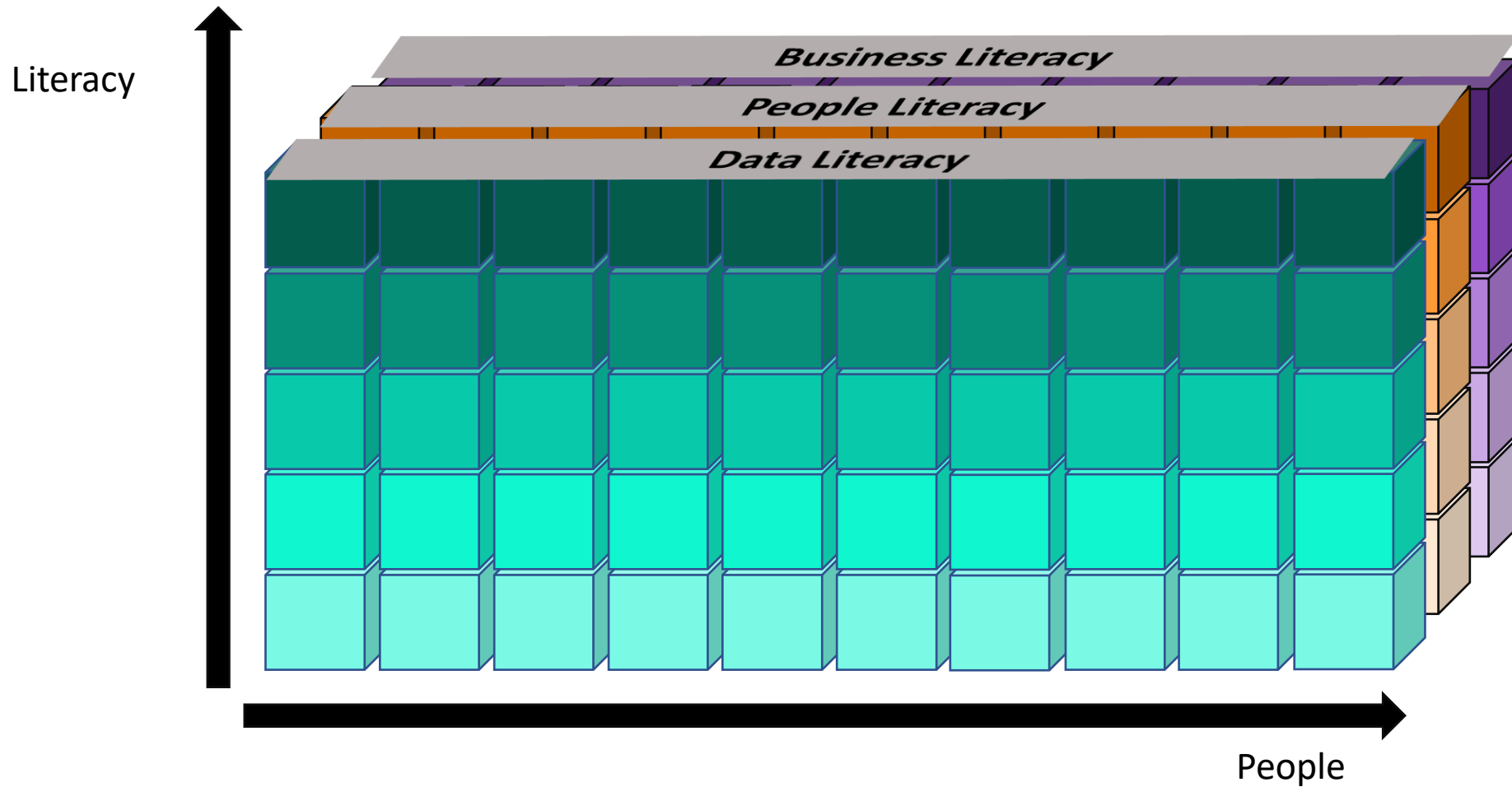
LEADERSHIP

The Importance Of
Emotional Intelligence In
Business

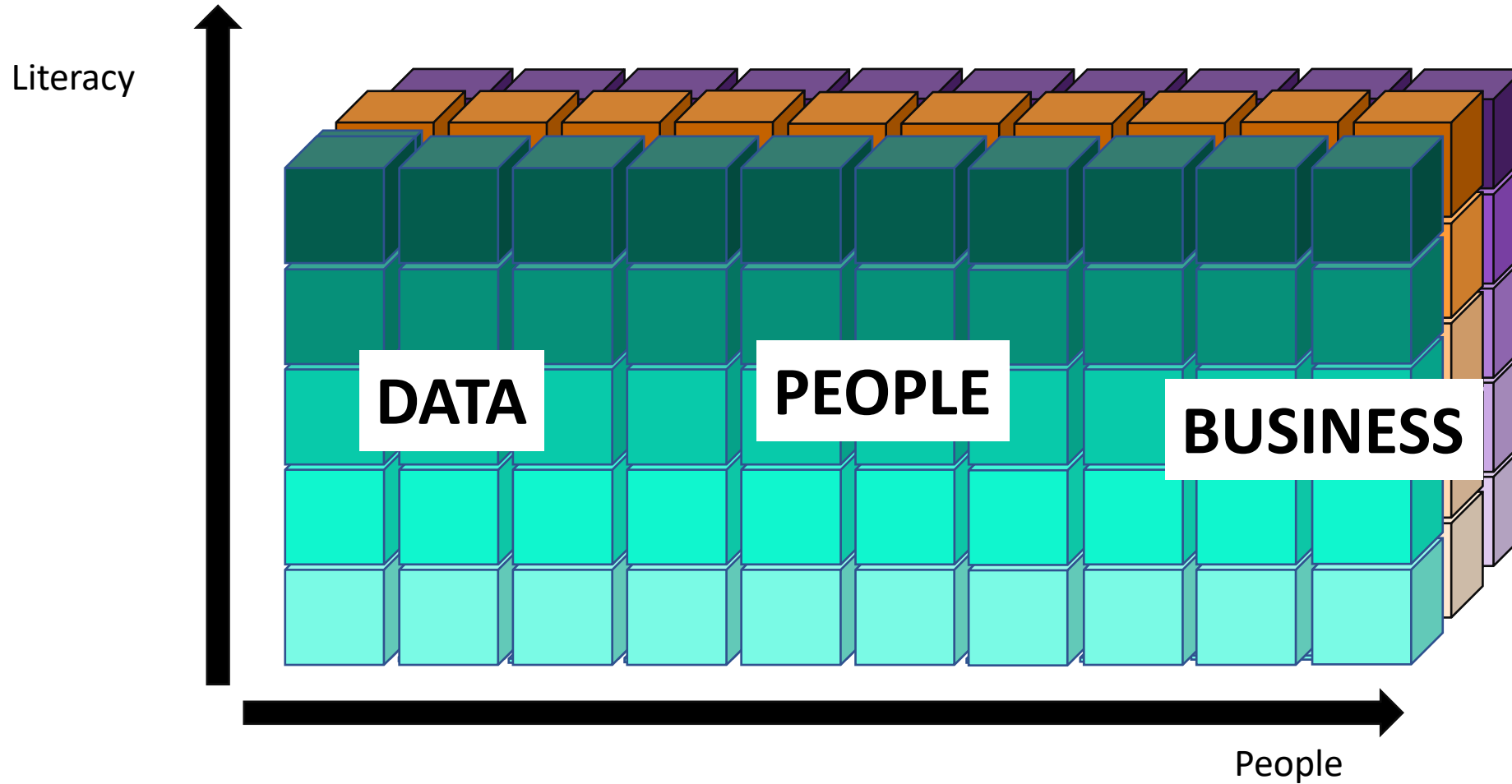
80% of long-term job success depends on EQ, while only 20% on IQ

People majoring in science and business have significantly lower empathy than people in social sciences.

What we want: High Literacy

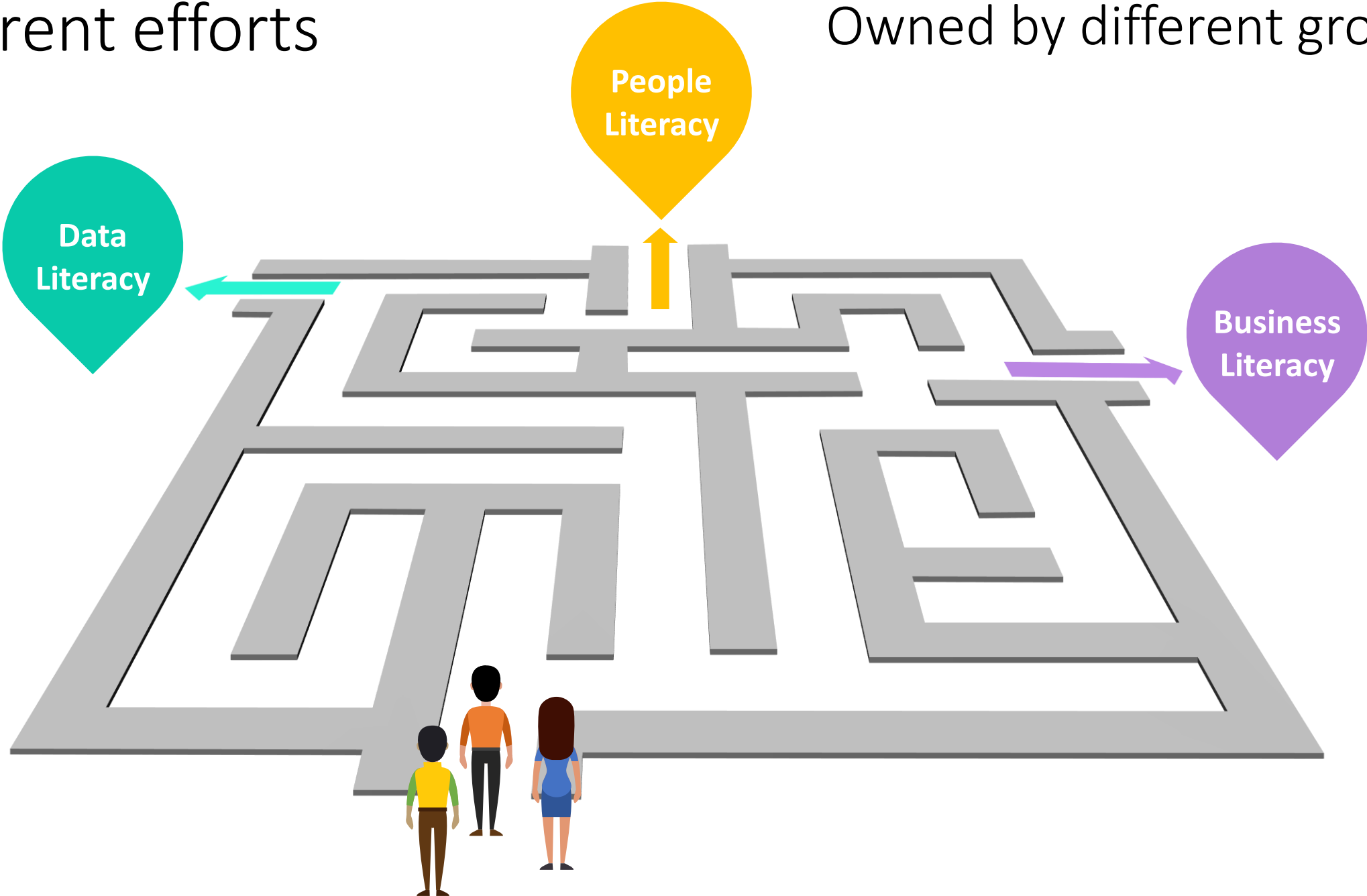


What we have



Different efforts

Owned by different groups



We all have different strengths



If we think about data literacy

Separately

vs.

Within Context

1. Data literacy is a solitary solution
2. By itself, data literacy will make decisions data-driven
3. Everyone can/must become highly literate
4. We want non-experts to become more expert
5. Creates a superior-inferior dynamic

Maybe it belongs in a broader, integrated context

There are strategic and social requirements

Realistically, people have varying strengths

Maybe there are varying levels of expertise

Recognize strengths, fortify weaknesses

Questions to ask ourselves about data literacy

Do we ignore the ways that employees already interact with data?

Do we use over-technical language others don't understand?

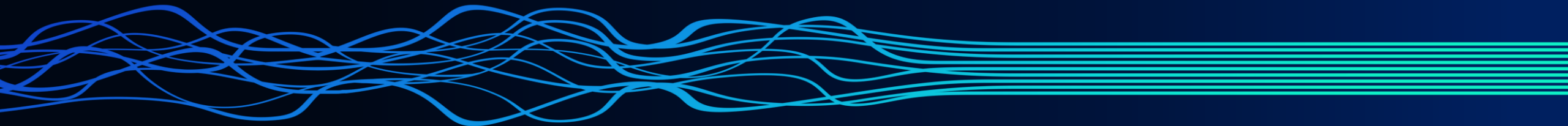
Do we assume it's THEIR job to learn about our field?

Do we create an adversarial or derogatory dynamic?

Do we assume interest and aptitude?

Do we separate this learning from other learning areas?

Do we focus on relevant terms and topics?



In modern society, we all

USE

DEPEND ON

COMPARE

DECIDE TO BELIEVE

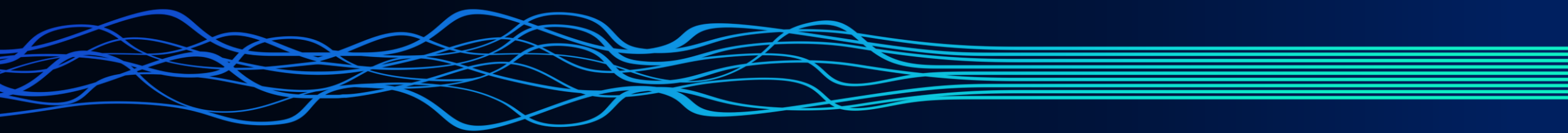
REACT TO

DEBATE

CELEBRATE RESULTS OF

every day.

DATA



We rely on consistent standards and units

We take for granted that we can depend on

What time it is

When

Who holds a world record

Wh

The dollars in our bank statements

How fast the speedometer says we're going

The weight of a 5lb dumbbell

Because
standards
are agreed
upon

We make confident comparisons

Today was colder

My team scored more points

This job pays more

That stock price is higher than

I've lost weight!

This car gets more miles per gallon

Because data
are collected,
and standards
are applied
consistently

We make data-driven decisions

We respond to real time data and predictions

Because
we
(usually)
trust the
sources

Monday.. Let's picnic on Saturday

Google.... I will go another way

are down.... Time to buy a house

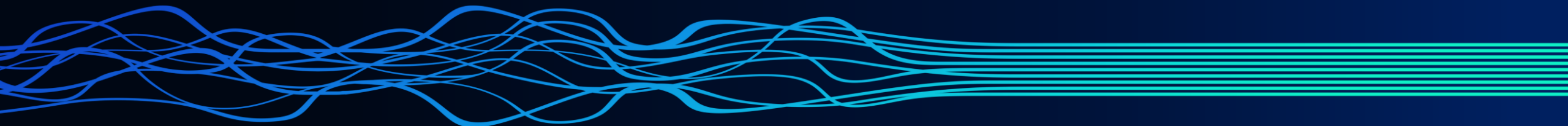
ally lower in the fall... I'll go then

an eclipse on May 3rd... let's go see



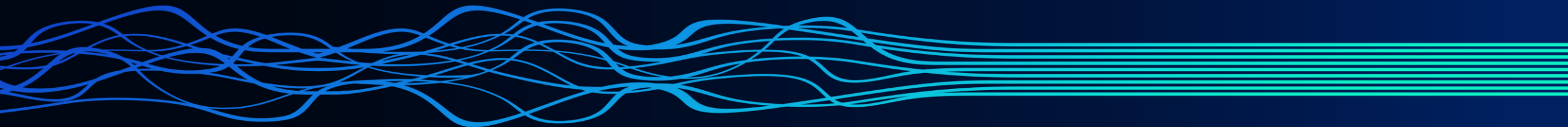
We need our people to have
high data literacy

So, they can make **reliable**
data-driven decisions





We need our people to have
high accuracy
Um... they already do
they can make
data-driven decisions



A person who has low nutritional literacy

Still consumes food.

He just may choose differently.



If we think about consumption (food or data)

There are parallel responsibilities

1. Governing bodies who insure:

- Safety
- Consistency
 - Definitions, metrics, processes
- Transparency

2. Consumers who understand

- Their role in using information
- The meaning of information

Nutrition Facts	
Amount Per Serving	
Calories	450
	<small>% Daily Value*</small>
Total Fat 14g	18%
Saturated Fat 1.8g	9%
Trans Fat 0g	
Polyunsaturated Fat 1.8g	
Monounsaturated Fat 4.8g	
Cholesterol 35mg	12%
Sodium 1550mg	67%
Total Carbohydrate 51g	19%
Dietary Fiber 8g	29%
Total Sugars 20g	
Protein 35g	70%
Vitamin D 0mcg	0%
Calcium 98.8mg	8%
Iron 8.46mg	45%
Potassium 470mg	10%
Vitamin A 1710mcg	190%
Vitamin C 66.2mg	70%

*The % Daily Value (DV) tells you how much a nutrient in a serving of food contributes to a daily diet. 2,000 calories a day is used for general nutrition advice.

Grilled Salmon Salad



Nutrition Facts	
Amount Per Ser	
Calories	1130
	<small>(53.4%)</small>
	<small>% Daily Value *</small>
Total Fat 67g	-
Saturated fat 27g	-
Trans fat 3g	-
Cholesterol 220mg	-
Sodium 1920mg	80%
Carbohydrates 64g	-
Net carbs 61g	-
Sugar 18g	-
Fiber 3g	12%
Protein 70g	-
Vitamins and minerals	
Vitamin A 90µg	10%

Double Cheeseburger



Governance. Information and Rules about:



Collect/ Source



Transport /Transfer



Transform



Cumulative information needed to insure safety and accuracy

Consumer Roles Vary. Governance oversees

More literacy (skill) needed for each level



Prepare / Manipulate



Format/Deliver

Double Cheeseburger



Nutrition Facts	
Amount Per Ser	1130
Calories	(53.4%)
	% Daily Value *
Total Fat 67g	-
Saturated fat 27g	-
Trans fat 3g	-
Cholesterol 220mg	-
Sodium 1920mg	80%
Carbohydrates 64g	-
Net carbs 61g	-
Sugar 18g	-
Fiber 3g	12%
Protein 70g	-
Vitamins and minerals	
Vitamin A 90µg	10%

Nutrition Facts	
Amount Per Serving	450
Calories	
	% Daily Value*
Total Fat 14g	18%
Saturated Fat 1.8g	9%
Trans Fat 0g	
Polyunsaturated Fat 1.8g	
Monounsaturated Fat 4.8g	
Cholesterol 35mg	12%
Sodium 1550mg	67%
Total Carbohydrate 51g	19%
Dietary Fiber 8g	29%
Total Sugars 20g	
Protein 35g	70%
Vitamin D 0mcg	0%
Calcium 98.8mg	8%
Iron 8.46mg	45%
Potassium 470mg	10%
Vitamin A 1710mcg	190%
Vitamin C 66.2mg	70%

*The % Daily Value (DV) tells you how much a nutrient in a serving of food contributes to a daily diet. 2,000 calories a day is used for general nutrition advice.

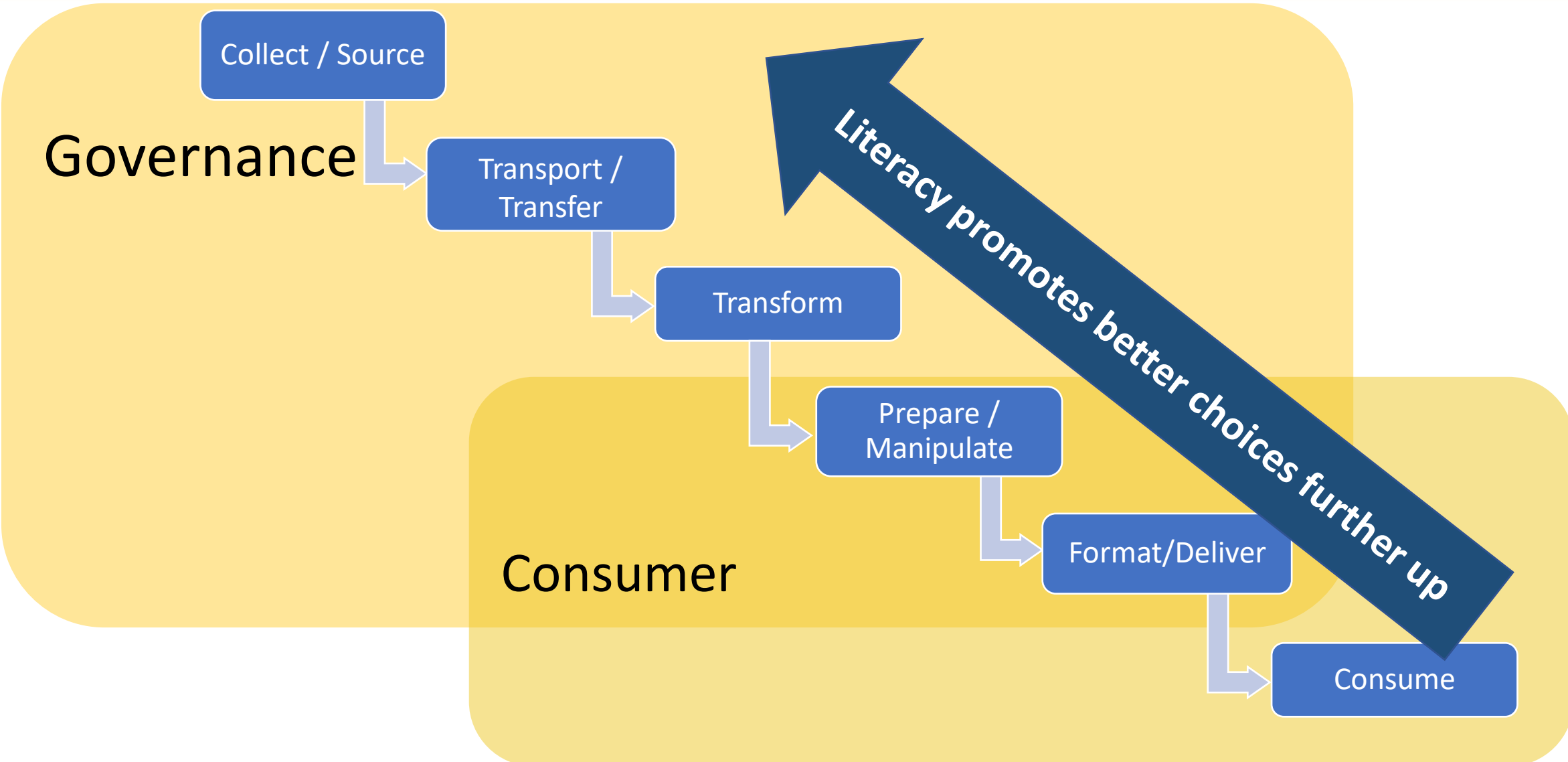
Grilled Salmon sa



Consume



Moving up the chain



Literacy: Moving up the chain

Collect / Source

Transport / Transfer

I will select and prepare based on what I know about the ingredients



Prepare / Manipulate

Format/Deliver

Double Cheeseburger



I'll just eat whatever you serve me....

Consume

A person who has low nutritional literacy

Still consumes food.

He just may choose differently.



A person who has low data literacy

Still consumes information.

He just may choose and use it differently.



Top of the chain



Collect /
Source



Transport /
Transfer

Transform

Why should we question?

**Can we trust the information
we are consuming?**

Prepare /
Manipulate

Format/Deliver

Consume



Labels – are they accurate?



Labels – is it named accurately?



Genetic studies in 2018 indicated:

25% of fish served in restaurants is not what the menu says

55% of Sea Bass served is not sea bass

42% of Snapper is not snapper

47% of sushi was mislabeled

100% of Dover Sole was actually Walleye

0% of Chilean Sea bass comes from Chile

“Wild Caught” salmon is most often farmed

Sales of Orange Roughy soared in the early 1980s after a name change from Slimefish

Labels – is it named accurately?



Governance Laws about Country of Origin (COOL)
Somehow **do not apply to Beef or Pork.**

75% of beef consumed in the US – that says “product of USA” **actually** comes from Australia, New Zealand or Uruguay, but is packaged here.

100% of South American beef can be labeled “grass fed” and “organic.”
Even if it is not.

Genetic tests of meat indicate that of ground meats:

35% of specialty meats are not what they are labeled.

18% of local butcher meats have more than one species.

6% of grocery meats have more than one species (lamb, chicken, turkey)

Label- Are the numbers accurate?



Weight

Texas found that 4% of grocery scales were inaccurate.

For some meats, sellers are allowed to add broth to increase the weight.

Sellers have an incentive to list weight as higher



Calories

FDA allows products to be off by 20%

Tests indicate calories in restaurants are often listed as 100s of calories less than actual

Restaurants have an incentive to list it as lower.

Thought provoking questions

- Who has an incentive to be truthful, or not?
- What are the most likely sources of inaccuracy?
- How does the information shift your thinking?
- Would it influence your choices?
- Who might these inaccuracies impact most?
- What else might you want to know?

Questions to ask ourselves about data literacy

Do we ignore the ways that employees already interact with data?

Do we use over-technical language others don't understand?

Do we assume it's THEIR job to learn about our field?

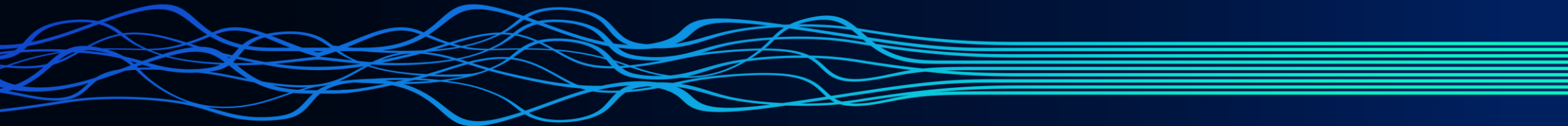
Do we create an adversarial or derogatory dynamic?

Do we assume interest and aptitude?

Do we separate this learning from other learning areas?

Do we focus on relevant terms and topics?

Do we mistake data literacy for **the goal?**



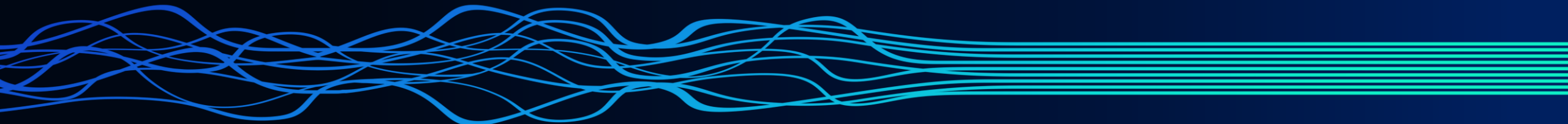


Do we want universal data literacy?

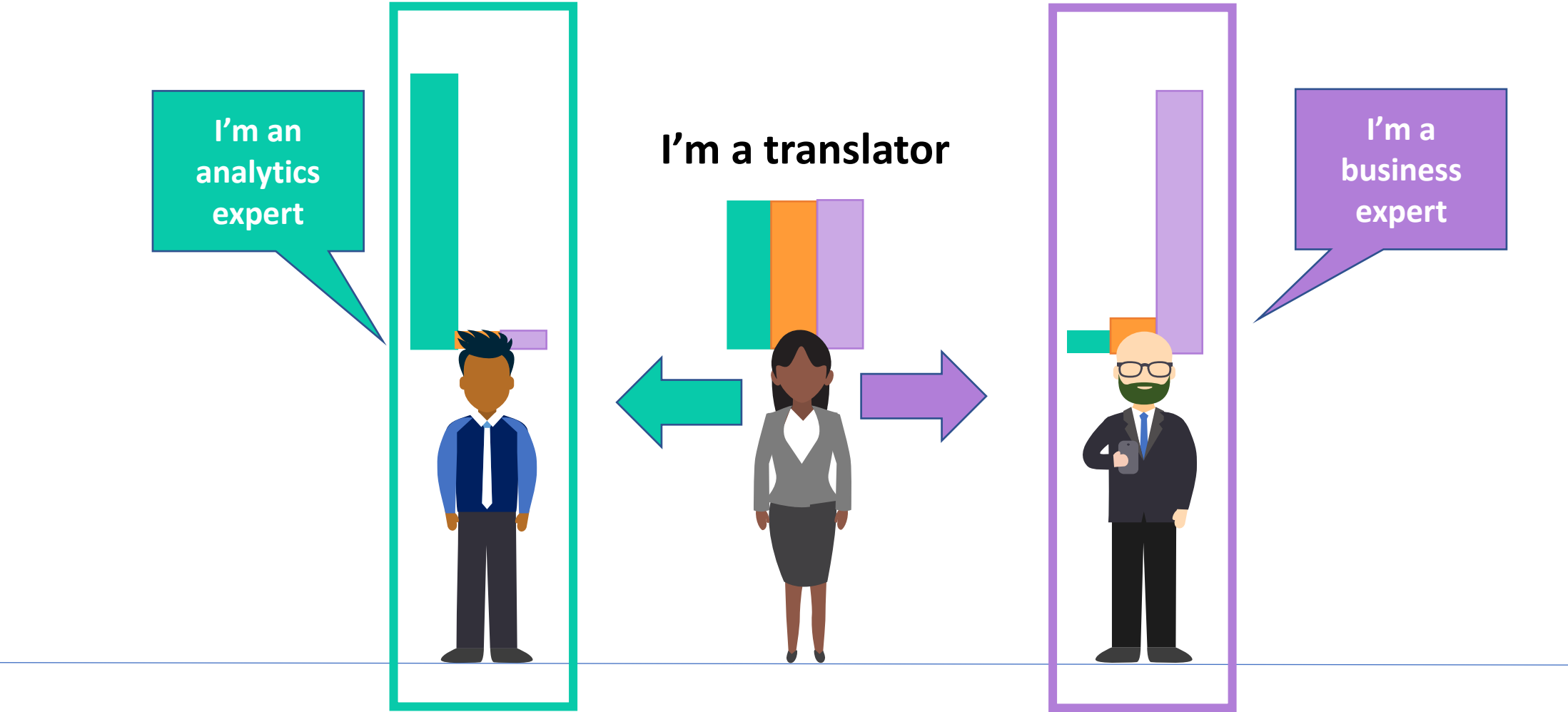
Or do we want: **Company-wide intelligent, information-driven decisions and actions.**

Consistently

- Use timely information
- Notice problems and opportunities
- Ask better questions
- Make better decisions
- Extract insights at all levels



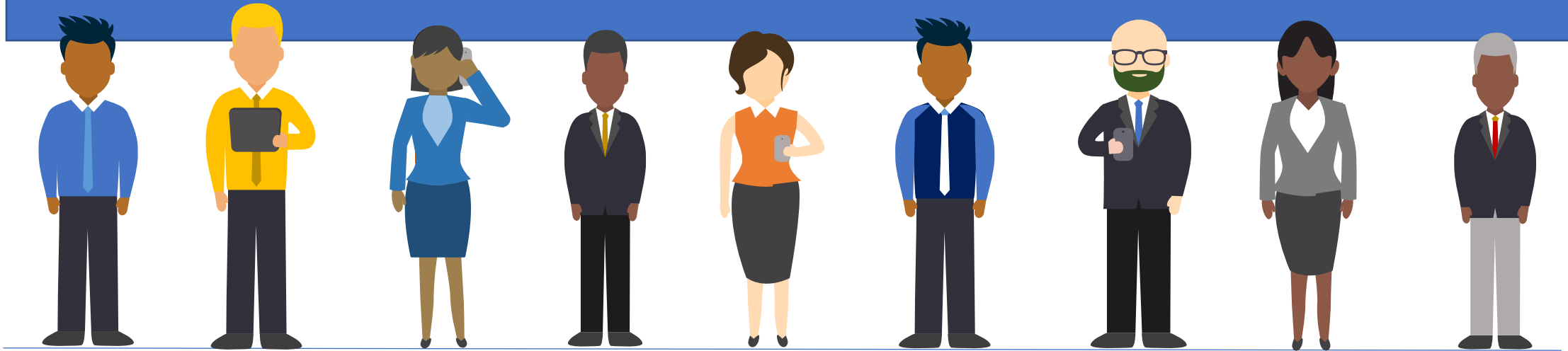
We all have different strengths



...for some that is translation

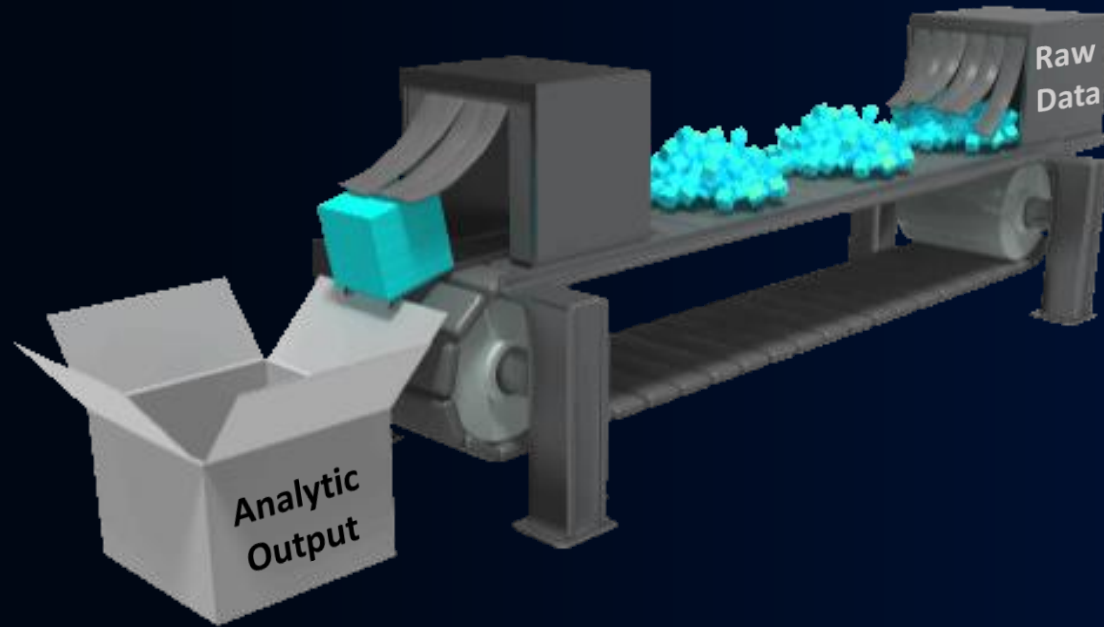
Collectively, we have many strengths

Making strategically-aligned decisions
(with evidence and empathy)



...that we can leverage

Data Integrity



How is
Data Literacy

connected
to

Data Quality
?