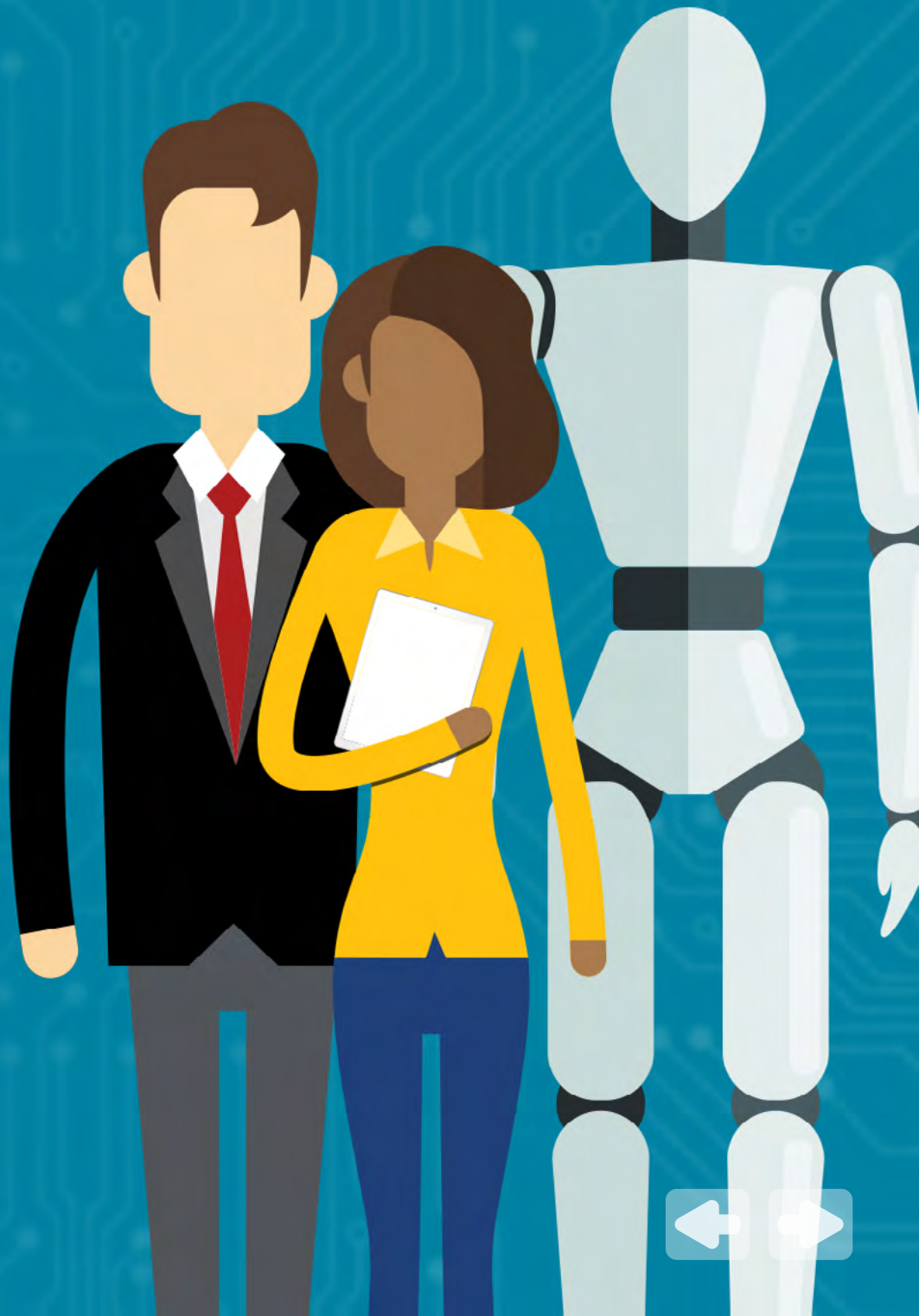
ARTIFICIAL INTELLIGENCE:

WHY HUMANS (STILL) NEED TO APPLY

collibra®



IT'S THE YEAR OF AI

Suddenly AI seems to be everywhere. Alexa, Siri, and other virtual personal assistants help us organize and track information. Facebook and Google automatically tag people in the photos we upload. And our smart homes start the coffee when we get out of bed and turn up the thermostat before we arrive home from work.

Transformative technologies like these are also driving decisions in finance, manufacturing, medicine, marketing, and more.

AI can flag fraudulent credit card charges, predict internal failures along an assembly line, diagnose dangerous lung disease, or anticipate what we're most likely to buy online.

Andrew Ng, one of the foremost thinkers in the field of AI, has called it the “new electricity,” a pervasive technology that will transform how we negotiate the world.

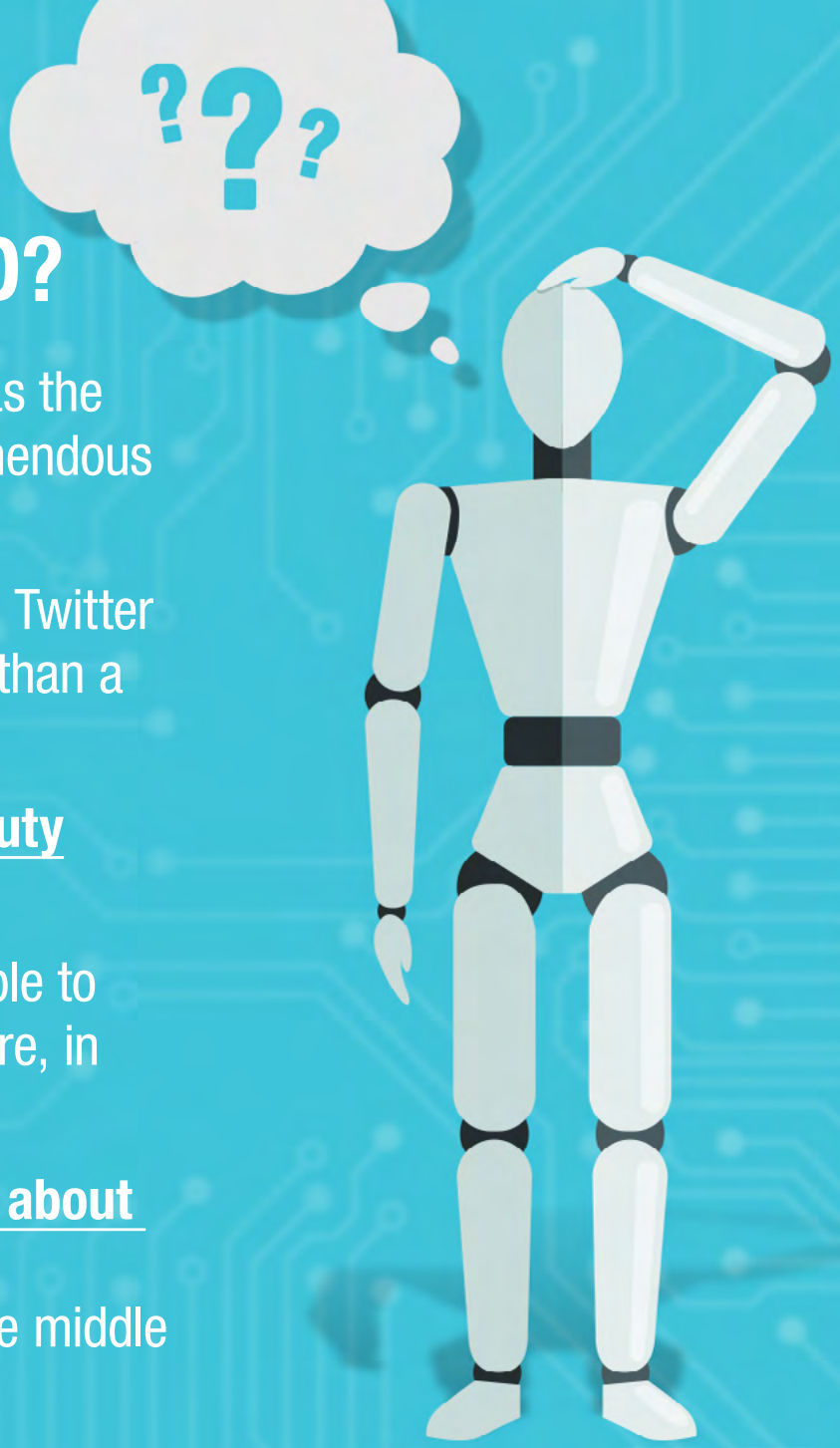
“The most important general-purpose technology of our era is artificial intelligence.”

– Harvard Business Review

AI: WUNDERKIND OR BAD SEED?

AI is pretty smart. And like any precocious child, it has the potential to make tremendous contributions – or tremendous mistakes. Take these examples.

- **Tay**, a chatbot designed to learn by interacting with Twitter users, began spewing inflammatory tweets in less than a day and had to be taken offline.
- Virtually all the "winners" chosen out of 6,000 **beauty contest** participants by an AI program were white.
- By wearing goofy glasses, five researchers were able to convince a **facial biometrics system** that they were, in fact, each other.
- German police officers, responding to a **complaint about a loud party**, discovered that an Amazon Echo had spontaneously decided to start blasting music in the middle of the night.



Predictive analytics has never been a perfect science. But in the world of AI, things happen at lightning speed, **intensifying the results of a bad decision.**

Yes, AI is a wunderkind. But it depends on data to learn how to think. And if the data it learns from is insufficient, inconsistent, poorly defined, or biased, its thinking can be terribly flawed.

“We have to teach our algorithms which are good associations and which are bad the same way we teach our kids.”

– Bloomberg Technology

WAIT, ISN'T AI ALL ABOUT THE ALGORITHM?

IN A WORD, NO.

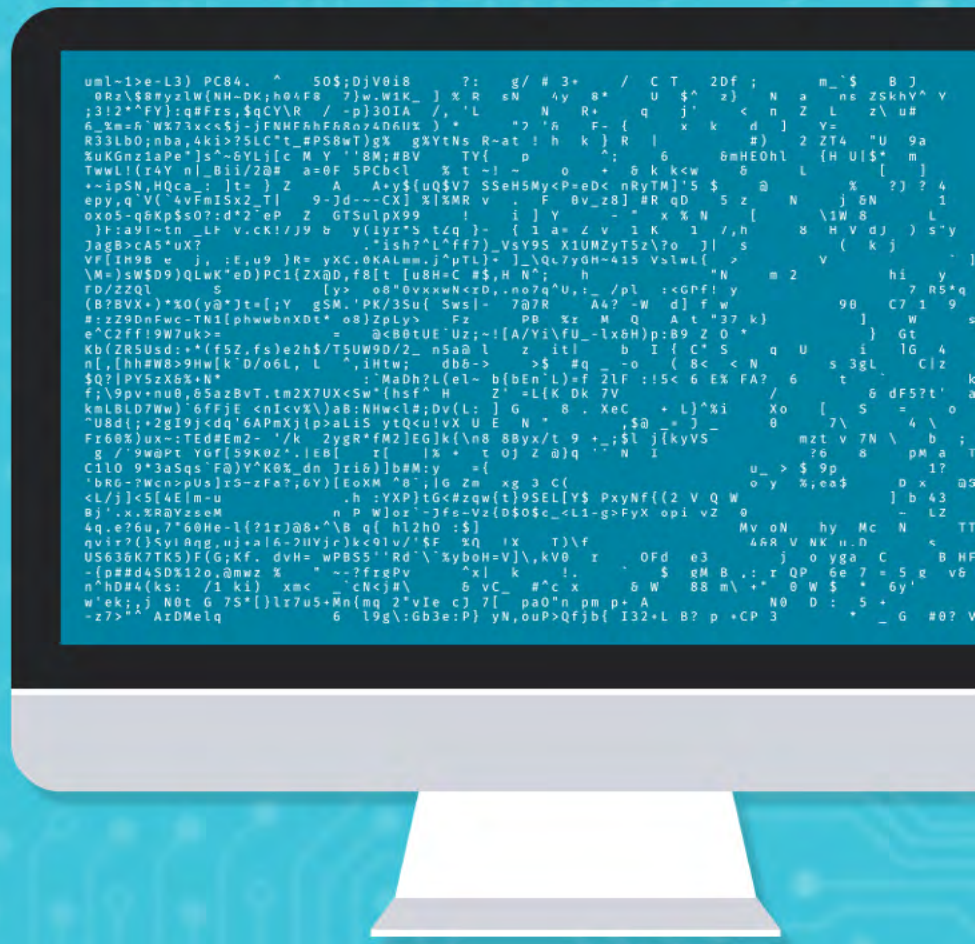
AI's a complicated subject, but we can break it down into three basic buckets:

101

THE DATA

THE MODEL

THE DECISIONS



THE DATA

When we talk about AI, we're often (though not always) talking about machine learning, a subset of AI that uses data to train its algorithms.

THE MODEL

By ingesting large sets of data, these algorithms learn progressively, finding relationships and patterns in the training data and eventually developing a new set of rules – a model.

THE DECISIONS

Once trained, the model can be applied to any data in order to generate an outcome – the decision. (That is, the information we are likely to act on.)

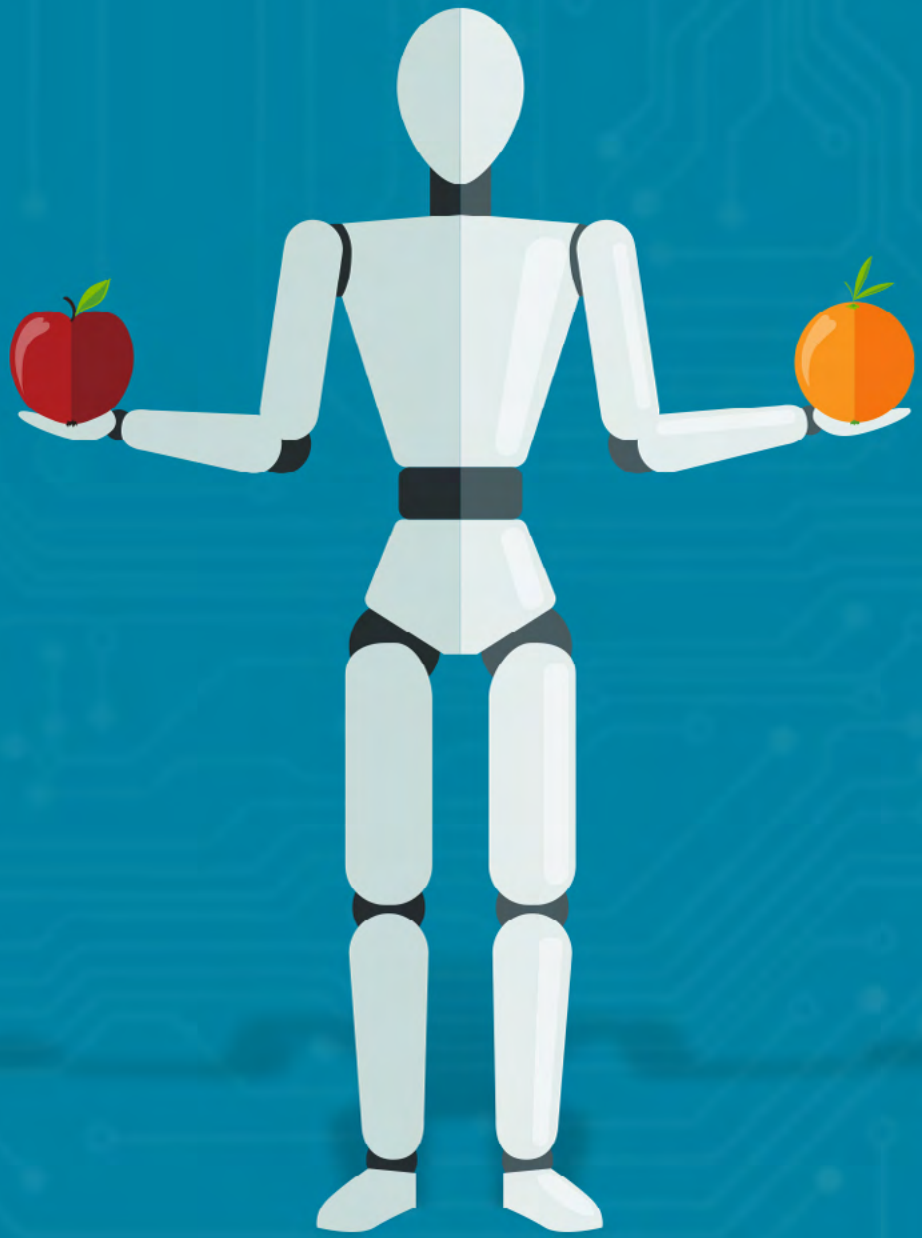
No question that, among these three buckets, algorithms take center stage. They're sexy, mysterious, and, for most of us, synonymous with AI.

But here's the thing. **The algorithm is only as good as the data that fuels it.**

Sure, the better the algorithm, the smarter it can be about finding inferences in the data. But without that data, even the most elegant algorithms will founder – and that can bring innovation to a standstill.

“Simple models and a lot of data trump more elaborate models based on less data.”

– Peter Norvig



As AI emerges as a competitive differentiator, it will rapidly expose your organization's data deficiencies. In particular, it will shine a bright light on poor data quality and faulty data policies and processes.

Simply put, without the ability to find, understand, and trust your data, your AI projects – and your ability to integrate those outputs into your day-to-day business operations – are at risk.

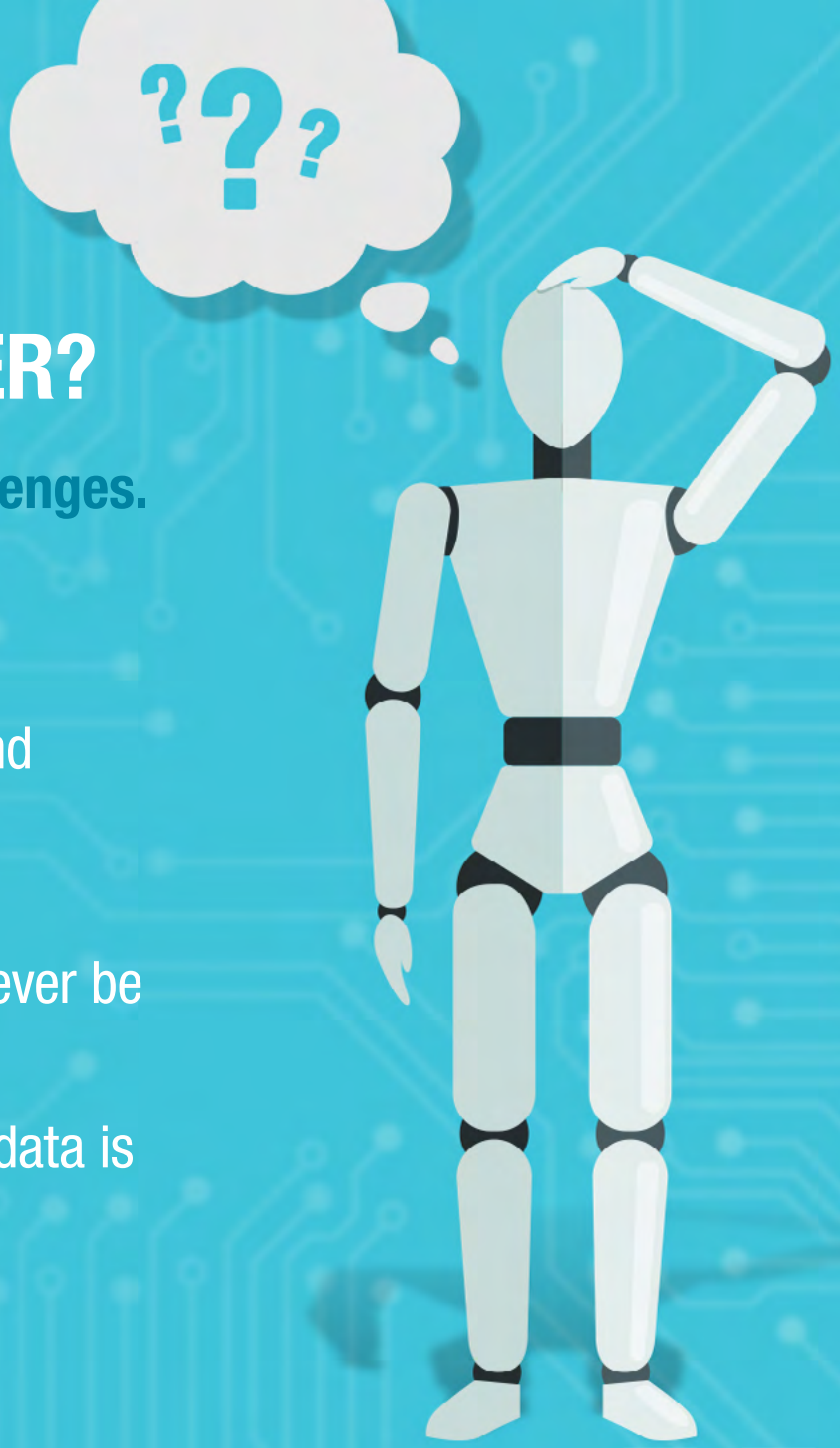
HOW MUCH DOES DATA MATTER?

While its potential is great, AI brings new challenges.

- How can we prevent biased outcomes that can negatively affect individuals and groups?
- When millions of data points can be collected and combined in an instant, how can we protect an individual's privacy?
- As AI gets smarter, can its 'thought processes' ever be explainable?

Clearly, the case for a better understanding of our data is becoming urgent.

Let's take a closer look.



BIAS

When machine learning fails, it's typically because an error has been introduced into the training data, which then gets embedded in the model. That's bad enough. But when the model creates new training data, it replicates – and indeed amplifies – the original bias. And the results can be devastating.

Data bias has the potential to affect the kind of medical treatment we receive, whether we are approved for a loan, or how fairly we're treated by the courts. Remember our earlier examples?

The good news is that data scientists are thinking hard about how to eliminate data bias. But they need help. Finding the right data, understanding what it means, and trusting the integrity of your data sets will be an important first step.

PRIVACY

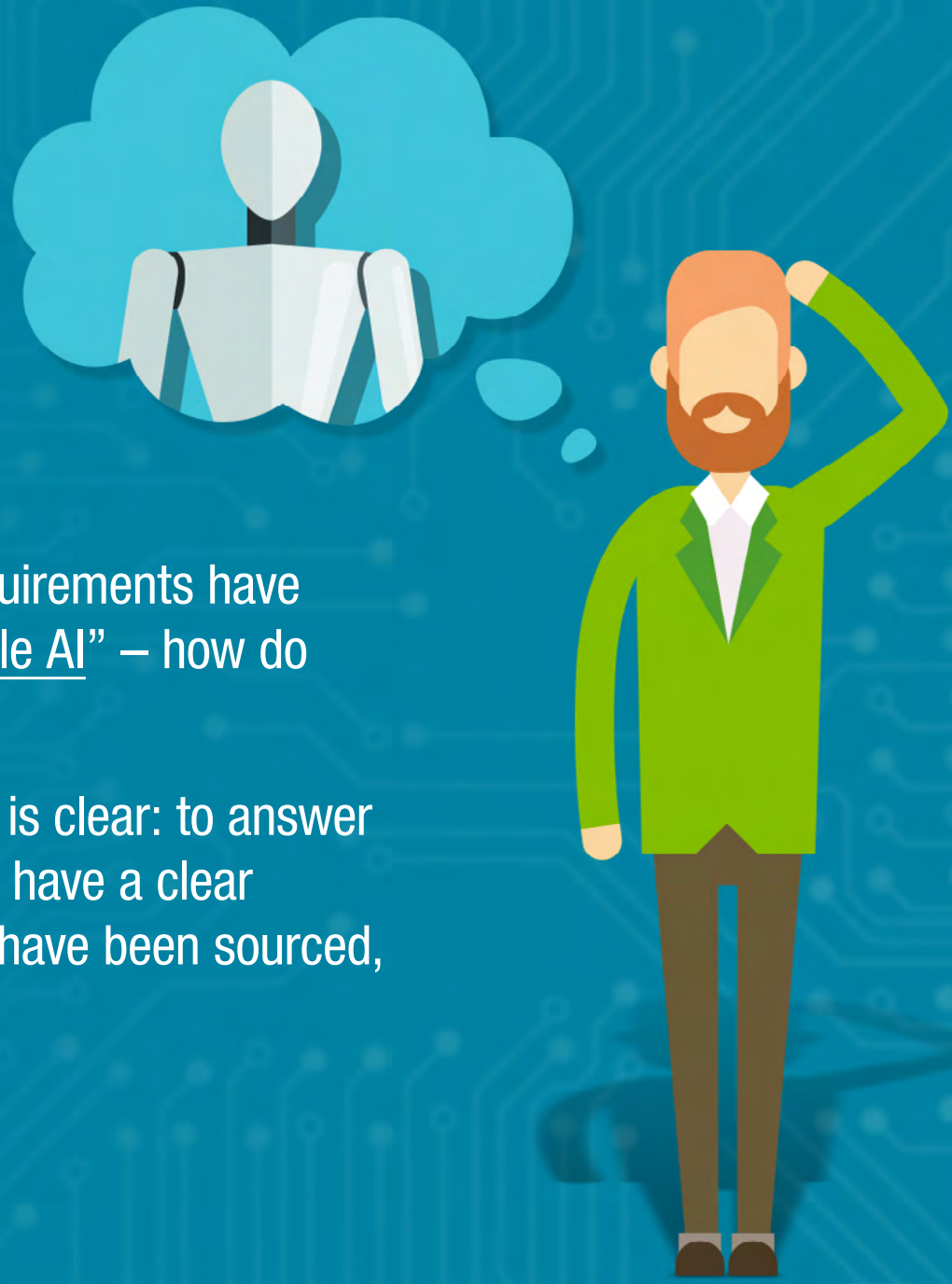
If a web browser can identify who we are by using three presumably anonymous data points, imagine how privacy can be compromised in the age of AI.

Today, endlessly-available data indicators can be parsed in myriad ways to reveal our past behaviors and predict our future actions.

And governments are taking notice.

New regulations like the GDPR are redefining "personal information" to include any information or data indicator specific to a particular person. And it requires that you know the purpose for which you've captured that data.

While the regulation doesn't prevent the application of AI, it may also require businesses to explain – simply and clearly – how machine learning processes reached a particular decision.



EXPLAINABILITY

These ethical, practical, and legal requirements have prompted a new desire for "explainable AI" – how do algorithms arrive at their decisions?

The science is nascent, but one thing is clear: to answer that question, businesses will need to have a clear understanding of how their data sets have been sourced, curated, and trained.

BUILDING TRUST IN AI WITH DATA GOVERNANCE

In its 2017 Data Scientist Report, AI platform provider Crowdflower determined that "the biggest bottleneck in successfully completing AI projects" was access to quality data.

Indeed, according to a new study from MIT, the difference between AI leaders and laggards largely hinges on data: "While most leaders are investing in AI talent and have built robust information infrastructures, other companies lack analytics expertise and easy access to their data."

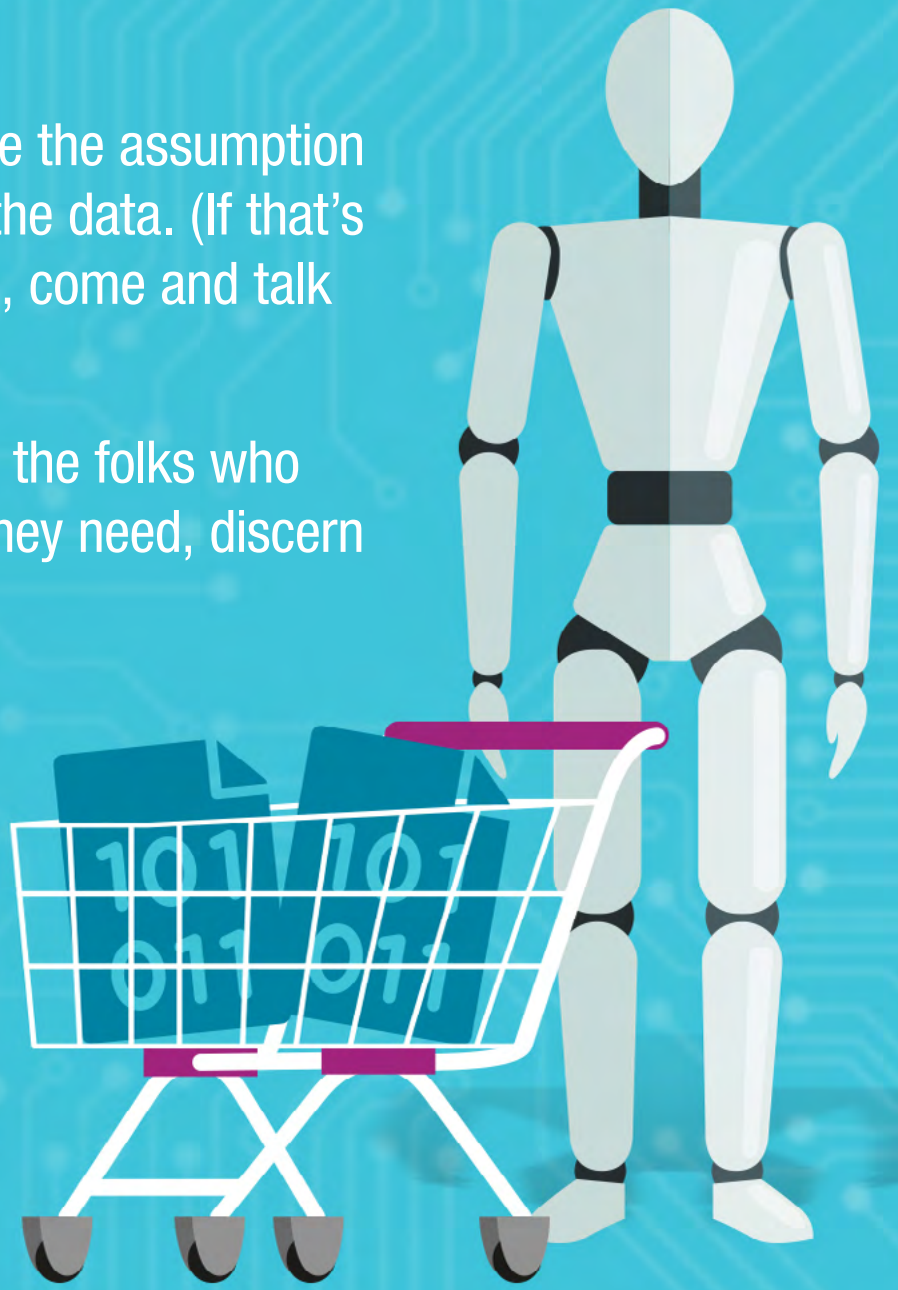
For AI projects to succeed, data scientists, data analysts, software engineers, and, yes, even business users will need to find, understand, and trust the data that their projects depend on.

That's data governance.

But before we go any further, let's first dismantle the assumption that data governance is all about locking down the data. (If that's all your data governance solution can do for you, come and talk to us…immediately.)

Smart data governance does way more. It helps the folks who are building your AI systems discover the data they need, discern how that data has been used in the past and by whom, and determine whether it is fit for purpose and appropriate to use.

Data governance can help your development teams build bias-free, super-smart AI. It's all about access, enablement, and clarity.



ACCESS

Data is everywhere. But sourcing enough high quality data for a viable AI project isn't child's play. Some organizations will have a wealth of data. Others will come up short. Some will restrict how their data can be used; others will have incomplete, poor quality data. Publicly-available data sources may be increasingly ubiquitous, but they cannot always be combined in meaningful ways.

Finding the data you will need for a successful AI project starts with identifying a clear business challenge. Once you've identified what data you need, data governance and its embedded catalog tools can help you determine what data exists and whether it is the right data for your needs.

If you're considering a new AI project, ask yourself these questions:

- Are you able to search for your data using semantic search tools?
- Can you rely on automated data discovery, personalize search results, and uncover hidden insights?
- Do you have the correct permissions to use that data in the way you intend?
- Can you see, at a glance, what you need to know about that data: who owns it, where it comes from, what it means, and how reliable it is?

Powerful data governance and a governed data catalog can provide the right people with the tools they need to discover, access, and assess the best data for their projects.

ENABLEMENT

Data governance is fundamentally about enabling highly skilled people to work more confidently and creatively.

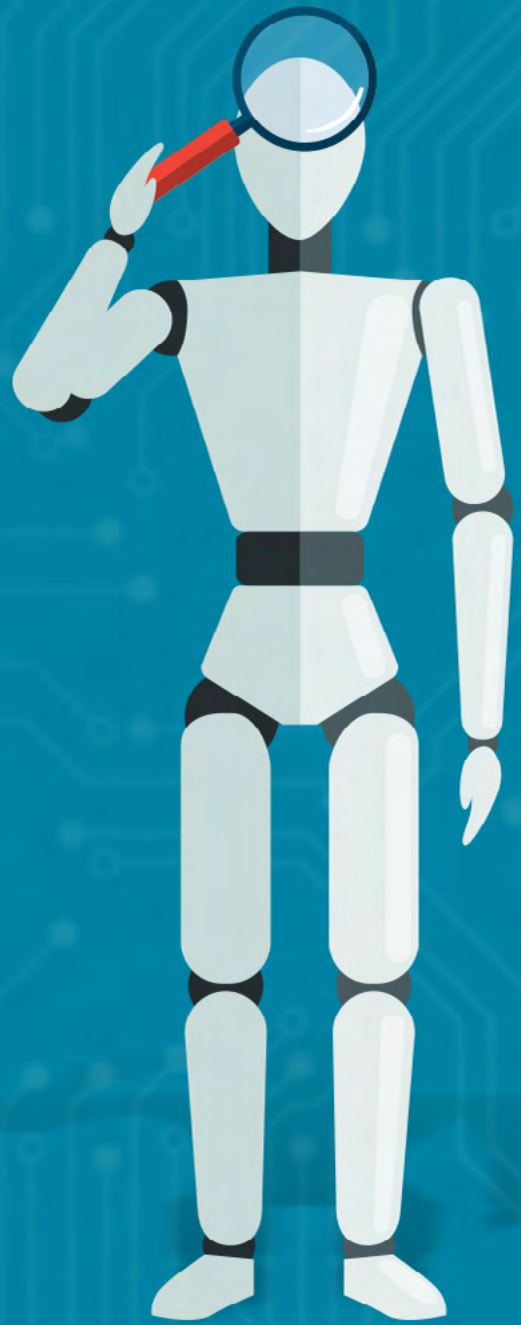
According to a recent Forrester report, that can often require developers to shift their focus from defining the algorithm to discovering, creating, or curating good training data. Data governance can help AI developers be more successful by providing information about what the data means and how it can be used.

New partnerships between development teams and data scientists are quite likely to emerge. Data policies established in a governed environment will provide them with a common language, helping everyone understand when and how the data was collected, trace its lineage, and assess whether the data is likely to produce unbiased and predicted results.

CLARITY

AI models can be opaque about the way they reach decisions – impenetrable “black boxes” that refuse to yield their secrets.

Data governance can’t untangle the web of learning points that results in a model’s decision. But it can provide a level of visibility into key characteristics of the model and the pieces of data that led to its decisions.

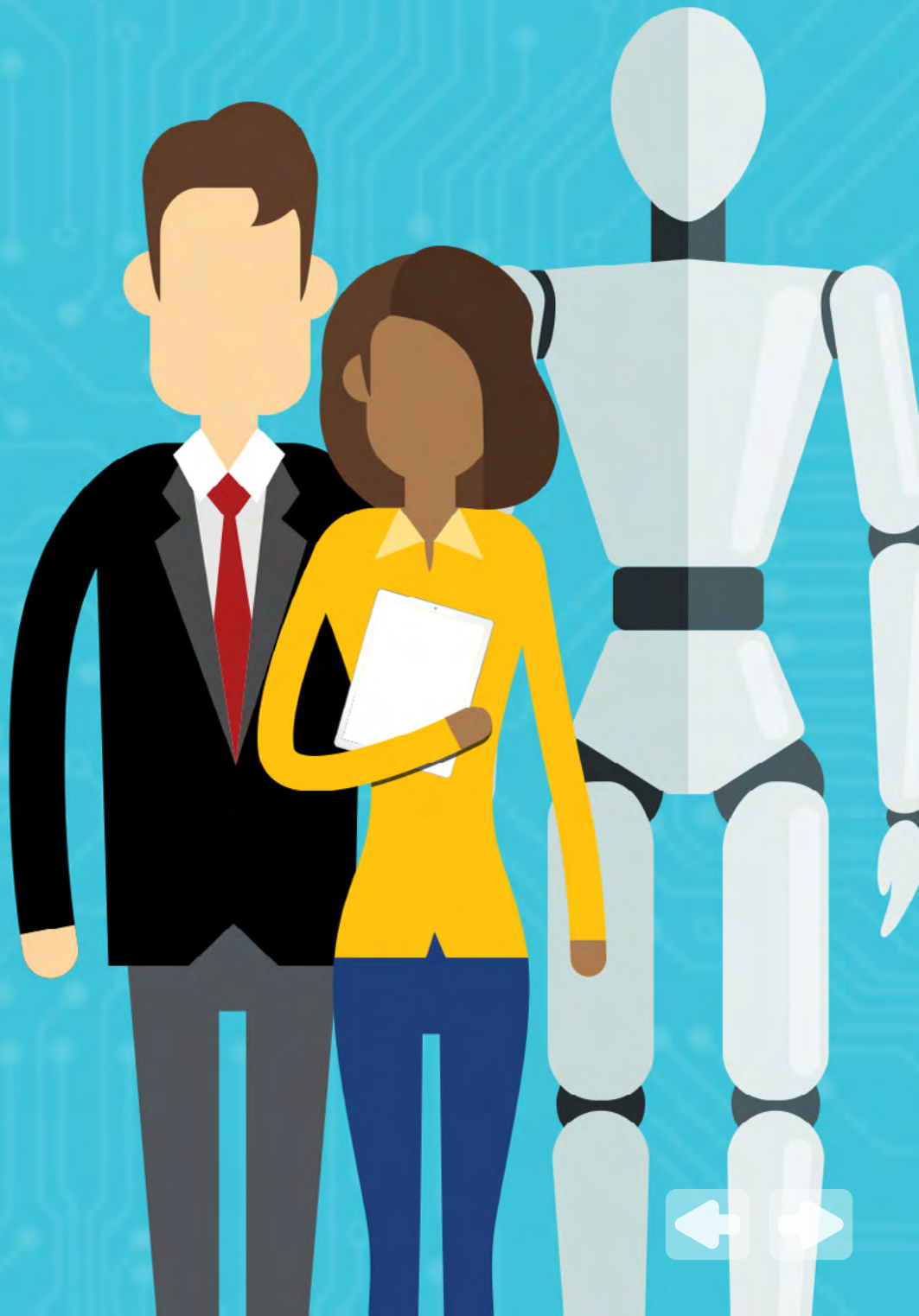


Data governance can provide visibility into the “black box” of AI.

A good data governance platform can also provide critical information about who developed the model, who owns the model now, what data that model uses, and how trustworthy it is.

AI isn't just a new tool. It's a new way of conducting business. Governed AI won't just be a differentiator. Access to the right data and insight into how that data informs the decisions reached by your AI models will soon be foundational business practices.

By governing your data today, you'll be prepared to master the art of AI to solve pressing business problems and drive real business value.



collibra®