

The Seven Most Important Takeaways from This Class

By Prof. Fernández-Loría
(huge thanks to Foster Provost for many lessons and slides)

But first a few things regarding the final exam...

Hands-on part (6% of the 35%)

Questions?

- It will be released on Thursday 14 at 00:01 am.
- Once you start, you have 24 hours to submit your answer or **until December 22nd at 11:59pm**, whichever occurs first.
- Must be done **individually**.
- Ensure stable connection and proper equipment. No retests available.
- Upload two files: your decisions (1s and 0s) and your Dataiku project or Python files (I won't grade your project; it's just a control check).
- The better your decisions, the better your grade.
- TAs and I won't answer content-related questions about the exam.

Written part (29% of the 35%)

Questions?

- It will be released on Thursday 14 at 00:01 am.
- Once you start, you have 3 hours to submit your answer or **until December 22nd at 11:59pm**, whichever occurs first.
- Must be done **individually**.
- Ensure stable connection and proper equipment. No retests available.
- Open-book: notes + book + Internet + ChatGPT
- Questions are multiple choice and fill in the blanks.
- TAs and I won't answer content-related questions about the exam.

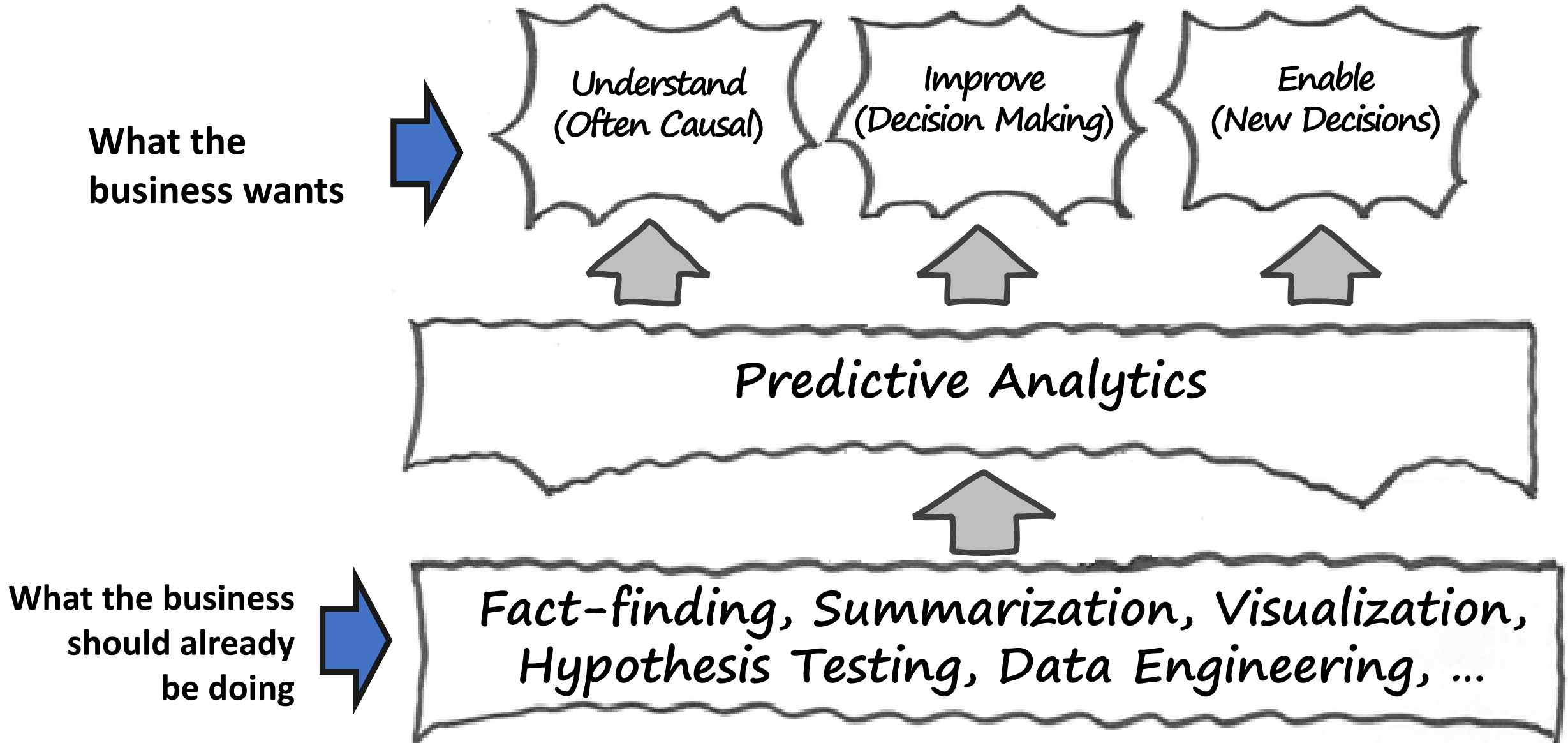
One Last Thing:

- My last office hours are on Dec 14th.
- I will have very limited access to my computer from Dec 15th onwards, so try the TAs first. Of course, you can always copy me.

#1

Craft your solutions with a strong emphasis on the dimensions of value you want to deliver.

Framework #2: Dimensions of Value



Framework #2: Dimensions of Value



Most of this course was devoted to understanding how to use data to maximize profits, minimize costs, etc.

We discussed churn, targeted marketing, demand prediction, investment decisions, customer service, etc.

More on this as we move forward today...

Framework #2: Dimensions of Value

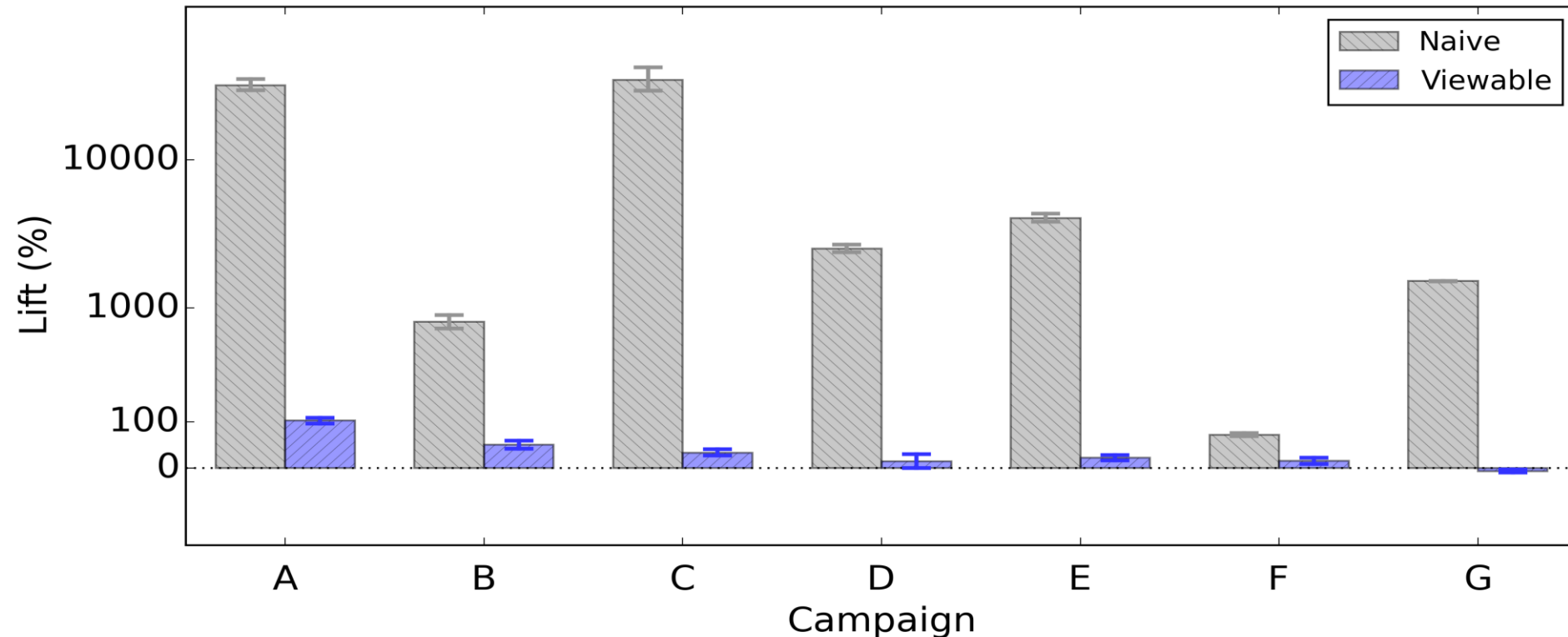


We can also build models that describe which variables correlate with a variable of interest. In these cases, the goal is often to infer causal associations. Causal inference involves predicting what would have happened with & without the cause.

Be mindful about making apples-to-apples comparisons.

Ad visibility as a “natural experiment”

The naive analysis of comparing users who received ads with users who didn't receive ads results in a drastic overestimation compared to an analysis that takes visibility into account.



Framework #2: Dimensions of Value



As we invest in data and analytics for a project, we enable new applications and the opportunity to create other data-driven solutions with substantially less investment.



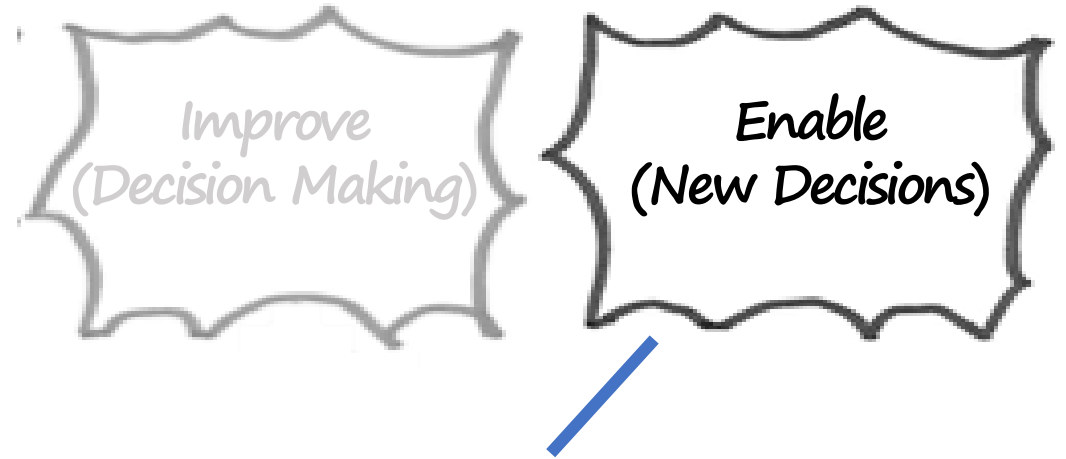
NETFLIX

HBO®



NETFLIX

Framework #2: Dimensions of Value



As we invest in data and analytics for a project, we enable new applications and the opportunity to create other data-driven solutions with substantially less investment.



NETFLIX

HBO®



NETFLIX

Framework #2: Dimensions of Value



As we invest in data and analytics for a project, we enable new applications and the opportunity to create other data-driven solutions with substantially less investment.

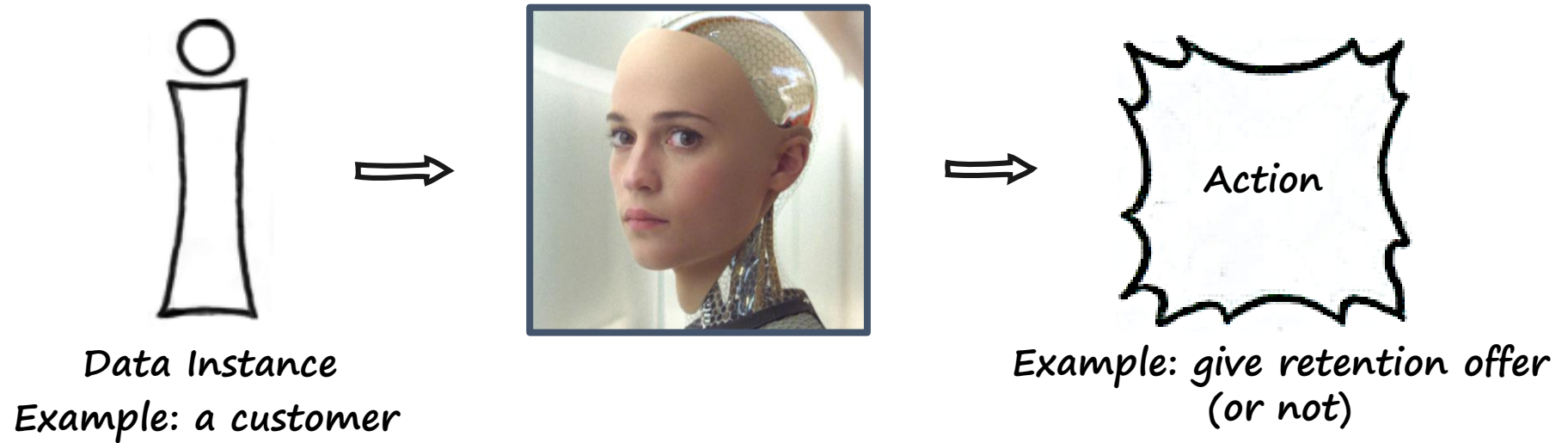


#2

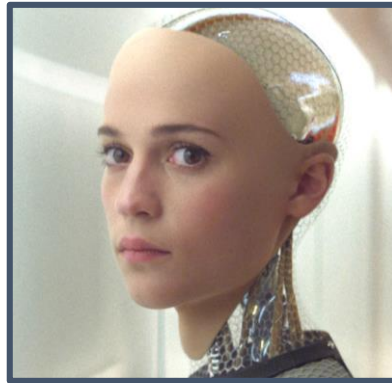
There are two important processes in predictive analytics: creating and using models.

Understanding how they differ and relate is key for formulating predictive analytics solutions.

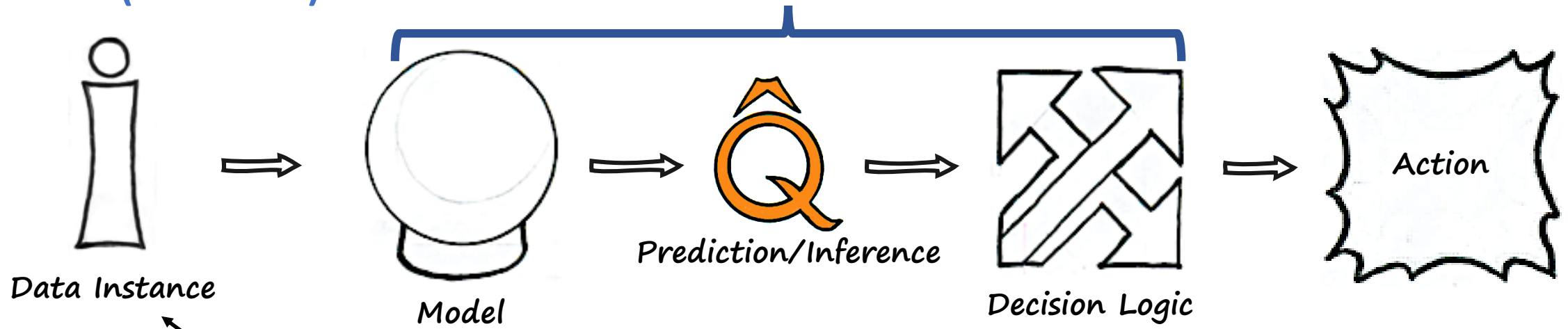
Using an AI system for decision-making



What is really happening inside the AI's head:



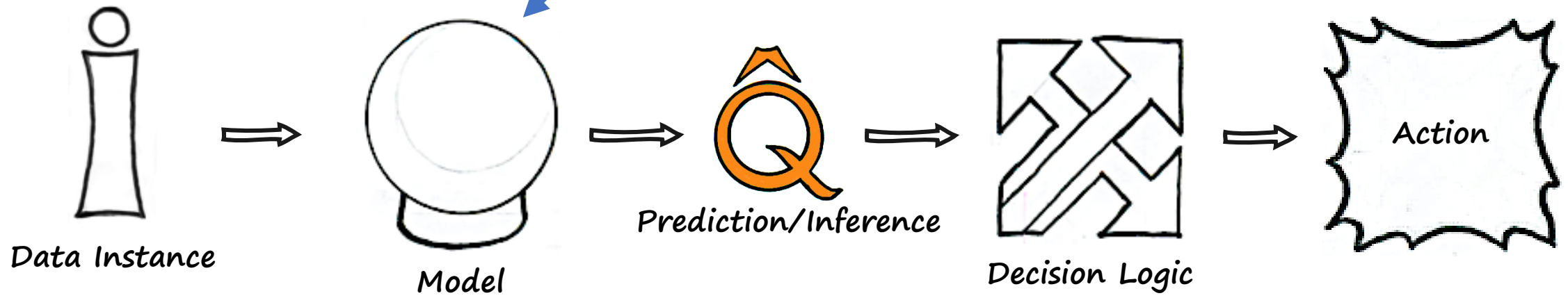
AI in Use (Inference):



Data for inference – some target value is unknown

Where does the knowledge to infer come from?

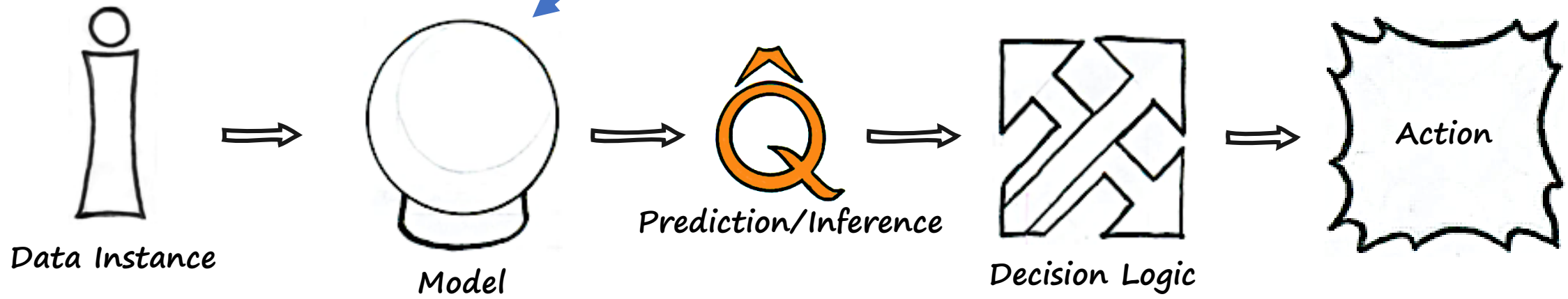
AI in Use (Inference):



Part I

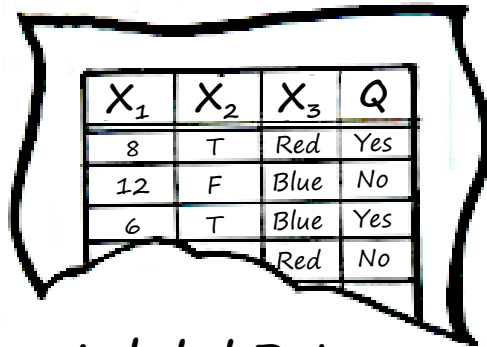


AI in Use (Inference):



Part II

Machine Learning:

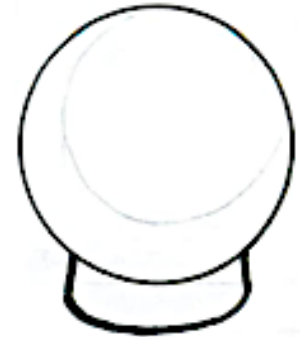


X_1	X_2	X_3	Q
8	T	Red	Yes
12	F	Blue	No
6	T	Blue	Yes
		Red	No

Labeled Data

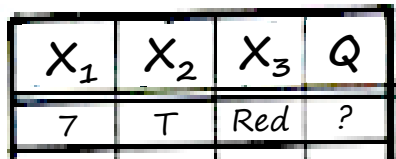


Machine Learning Algorithm



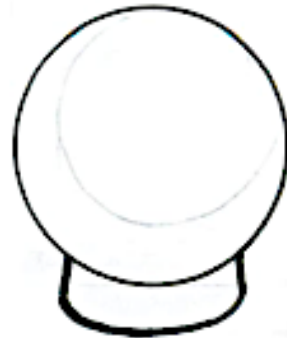
Learned Model

AI in Use (Inference):



X_1	X_2	X_3	Q
7	T	Red	?

Data Instance



Model



Predicted Quantity



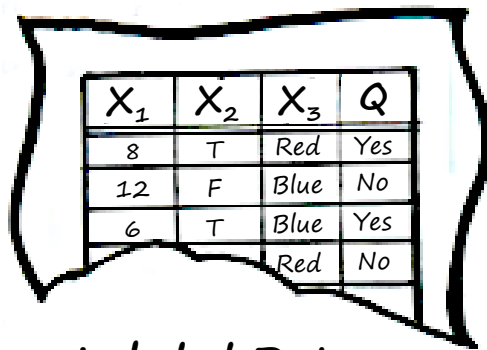
Decision
Logic



Action

Framework #3: AI in Use vs. Machine Learning

Machine Learning:

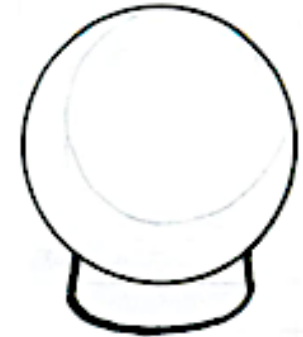


X_1	X_2	X_3	Q
8	T	Red	Yes
12	F	Blue	No
6	T	Blue	Yes
		Red	No

Labeled Data

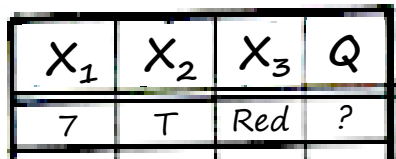


Machine Learning Algorithm



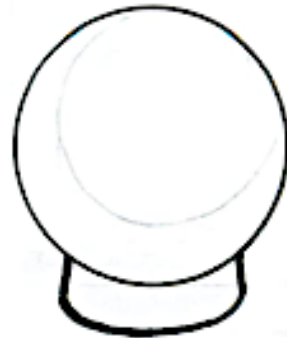
Learned Model

AI in Use (Inference):



X_1	X_2	X_3	Q
7	T	Red	?

Data Instance



Model



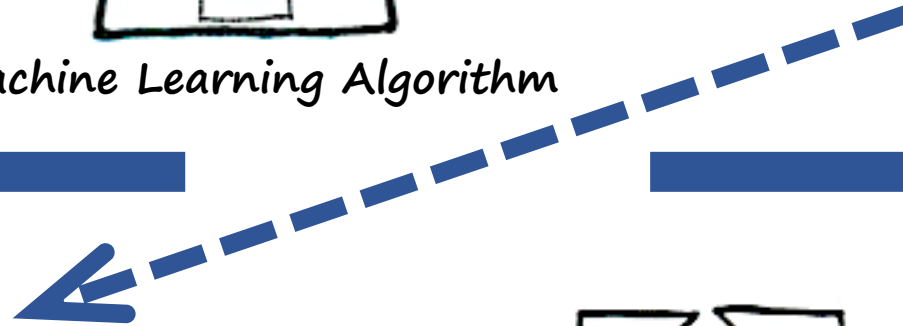
Predicted Quantity



Decision Logic



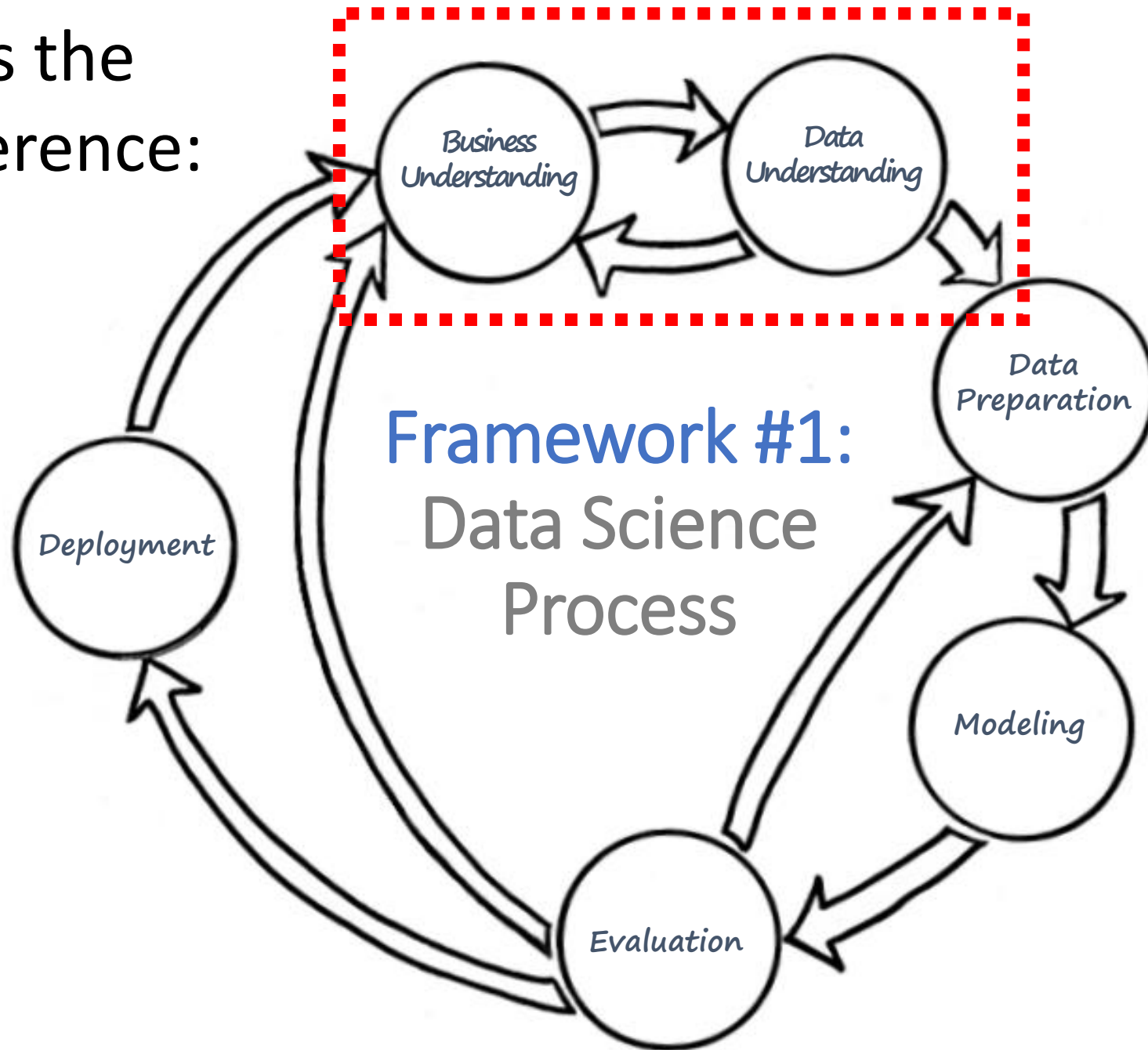
Action



#3

A data analytics undertaking involves a ton of stuff. The data science process can help us organize it all in our minds.

What makes the
biggest difference:



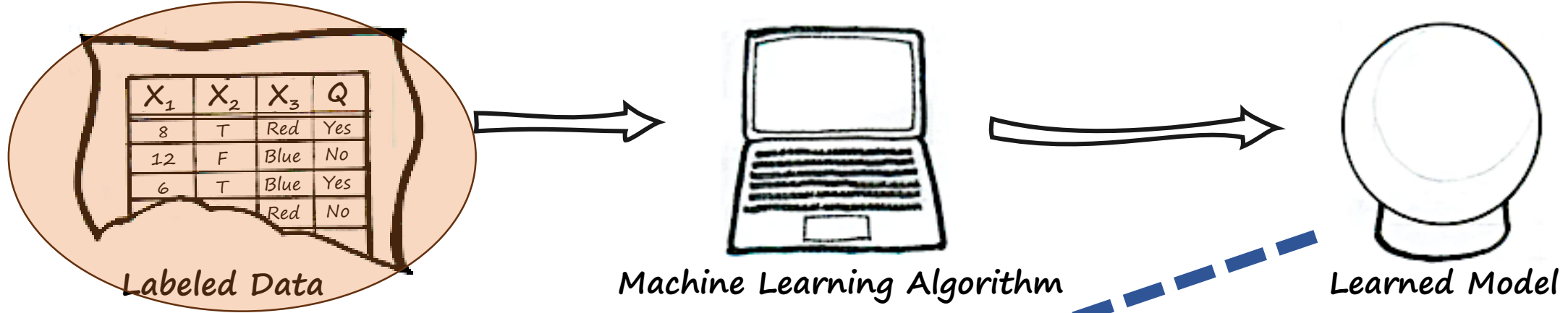
Framework #4: Questions for Problem Formulation

- **Q1:** What decision is the solution intended to support?
- **Q2:** What will be the entity to be represented by a data instance?
- **Q3:** How will the decision be evaluated?
- **Q4:** What quantity would be useful to know or predict to inform the decision (target variable)?
- **Q5:** What other factors are relevant for decision making (costs, benefits, constraints, etc.)?
- **Q6:** (How) Can we get target labels for supervised machine learning?
- **Q7:** What data (features) could be useful to predict the target?

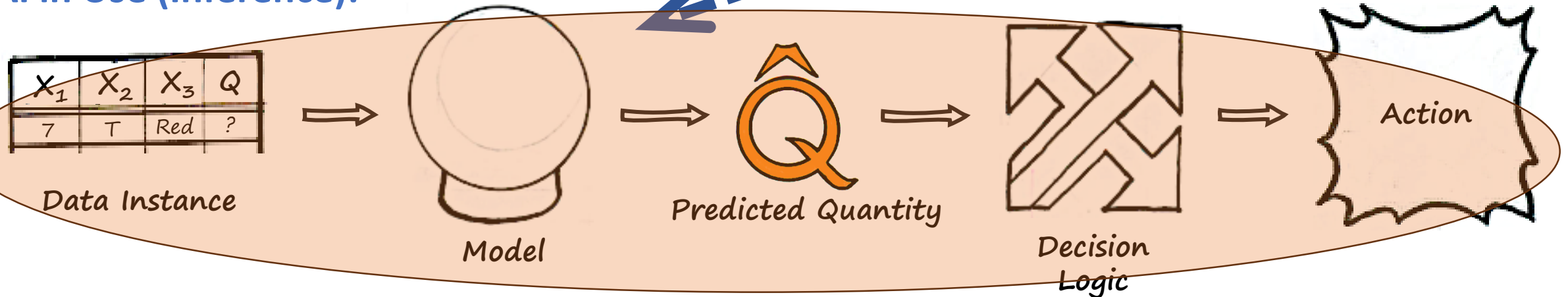
Framework #3: AI in Use vs. Machine Learning

#1 Problem Formulation and #2 Data Engineering.

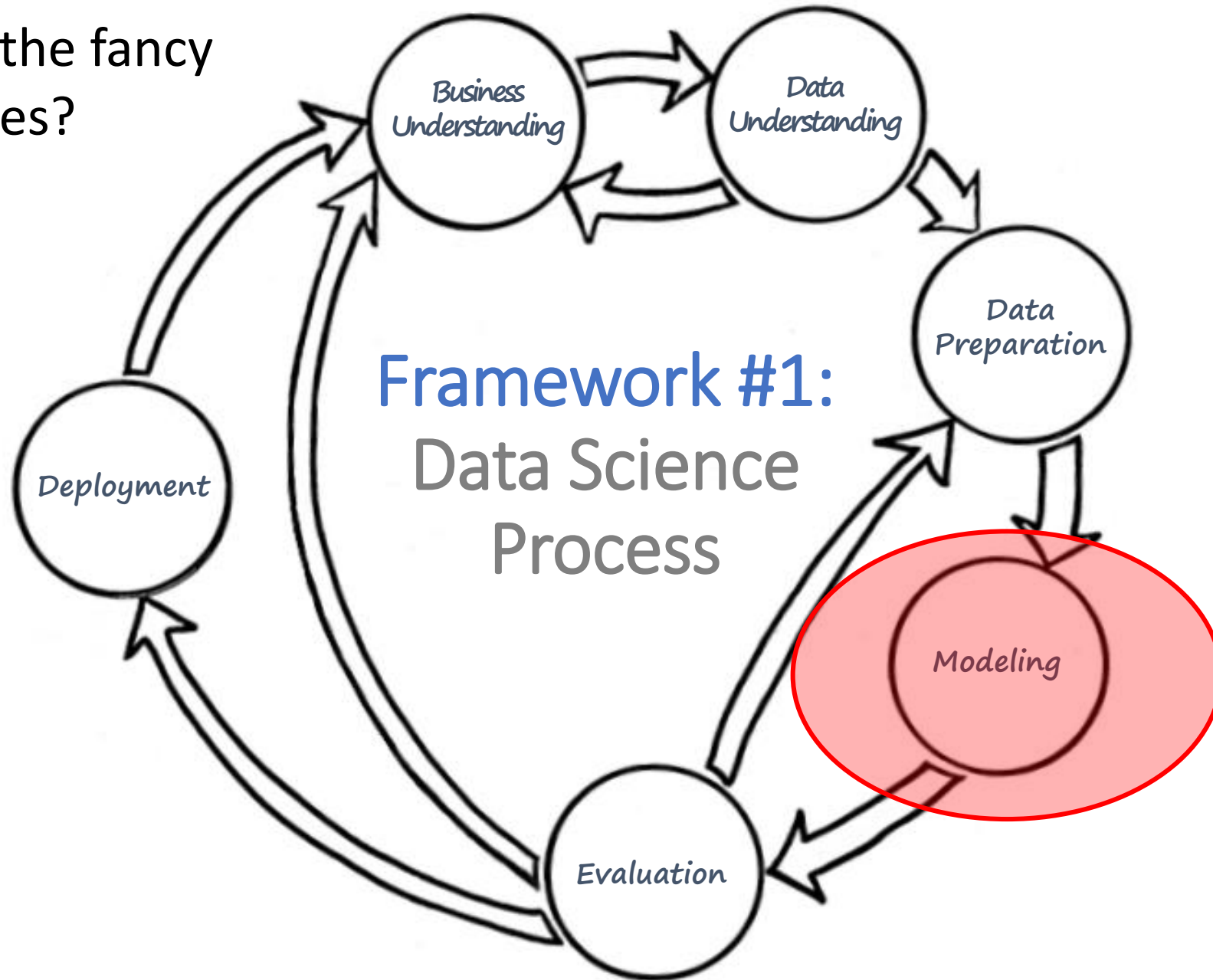
Machine Learning:



AI in Use (Inference):



What about the fancy
ML techniques?



Framework #5: Components of Machine Learning

Machine learning algorithms are programs that **fit** a **model** to **data** based on an **evaluation criterion**.

Representation + Evaluation + Optimization

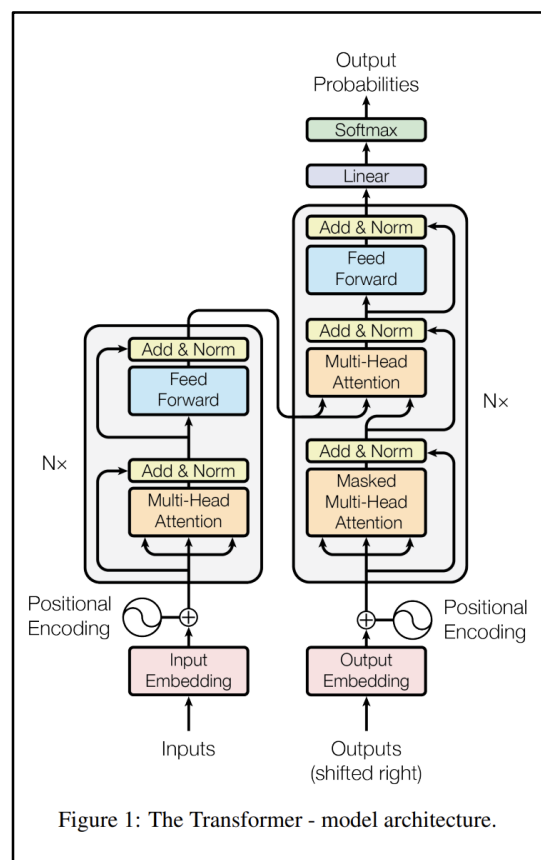
We focused mostly here.

$$\underset{\hat{q}}{\operatorname{argmin}} \sum_{\forall i \in \text{Data}} \operatorname{error}(q_i, \hat{q}(x_i))$$

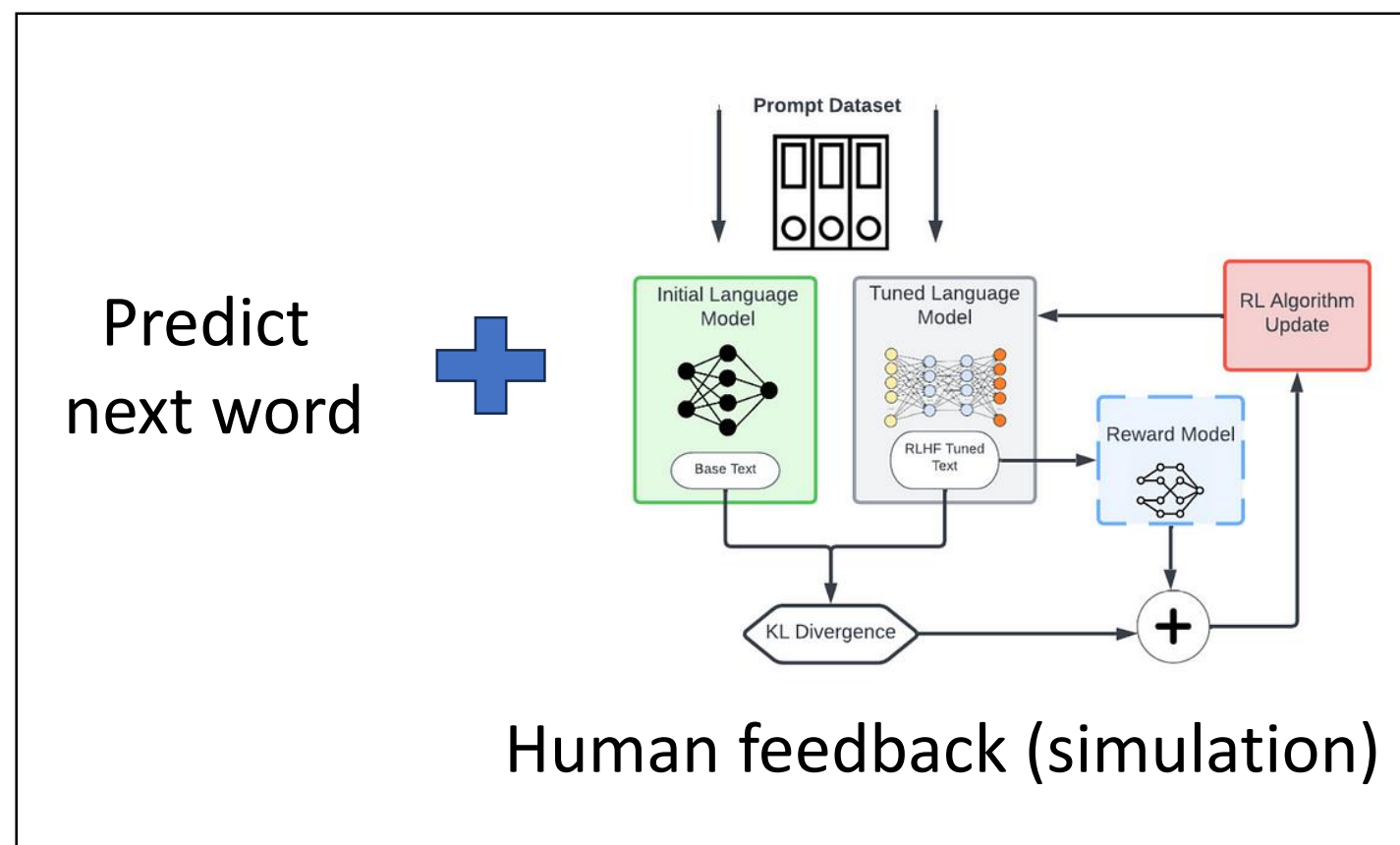
Learning = Representation + Evaluation + Optimization

GPT

Representation



Evaluation

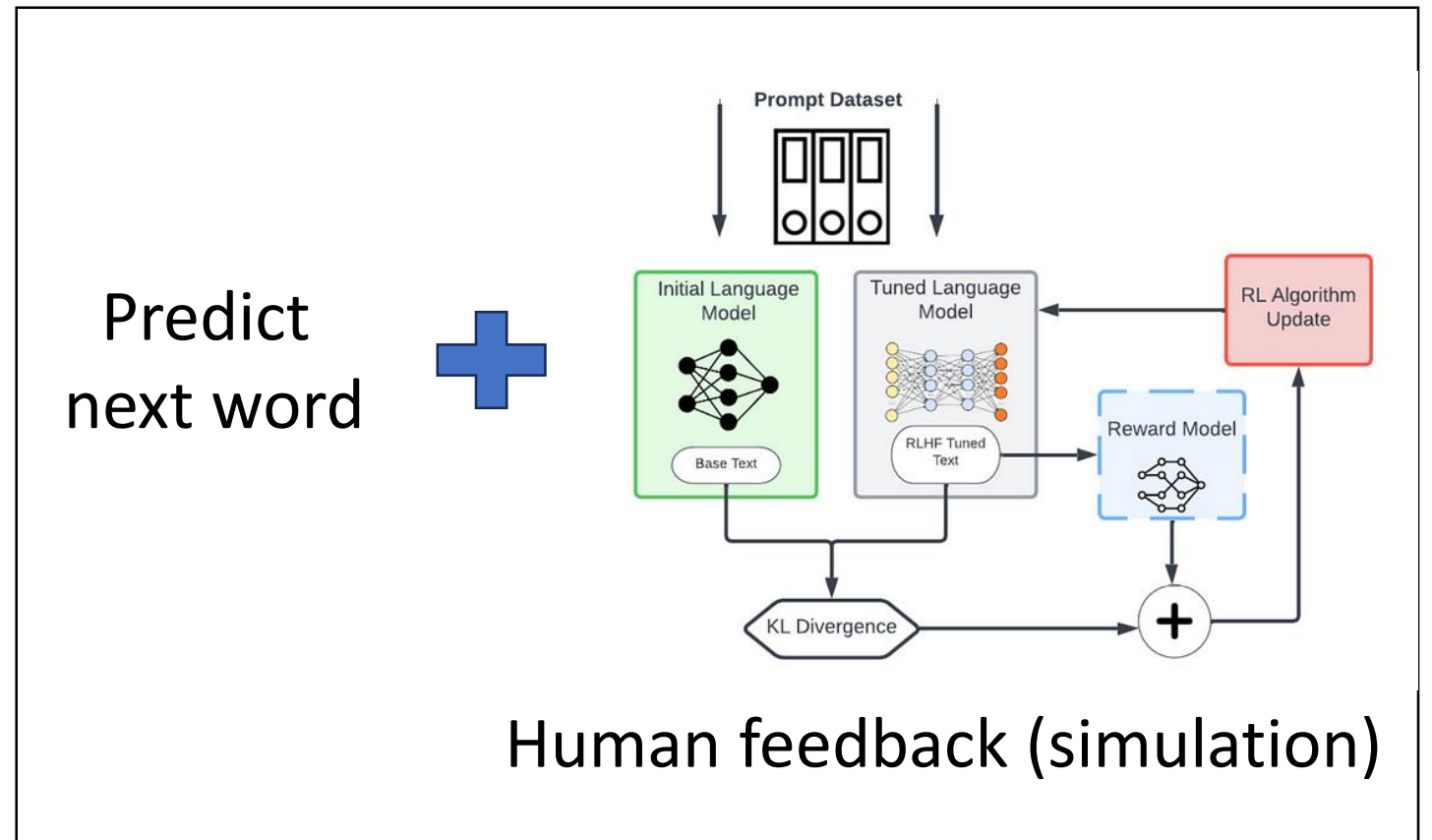
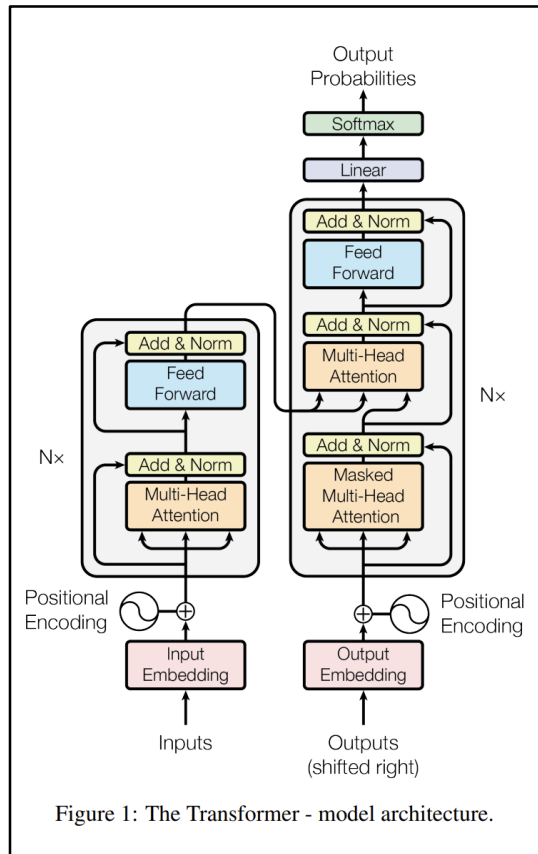


#1 Problem Formulation and #2 Data Engineering.

Inspired by

Representation

Evaluation

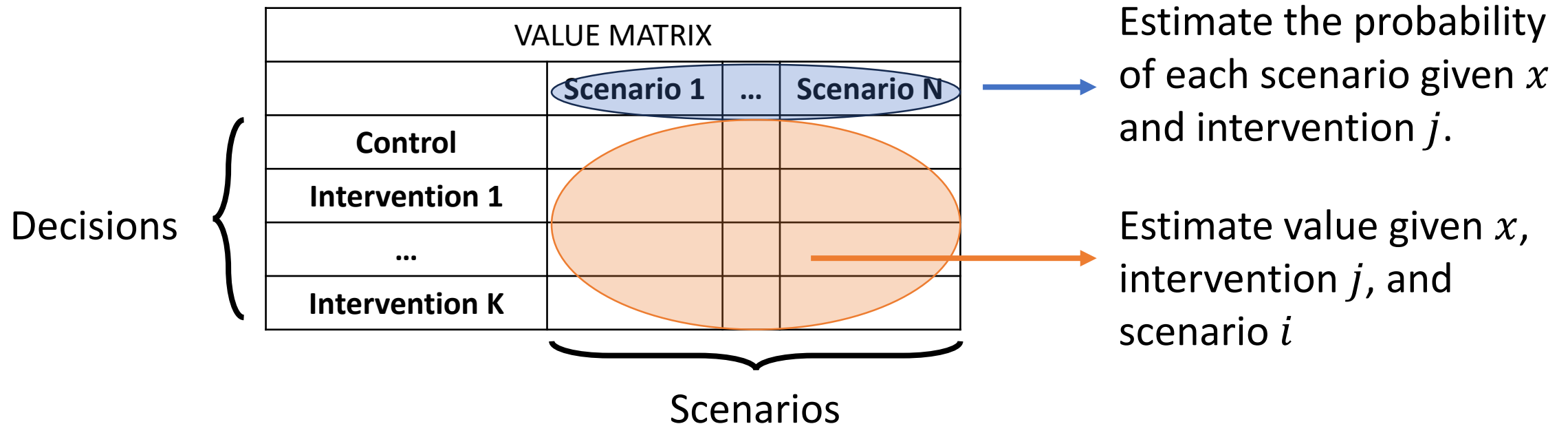


#4

When we reason about a business problem, the expected value framework helps us move from a unique problem to a data science solution.

The expected value framework divides the problem into probability estimates, value estimations, and cost-benefit estimations.

Decision Logic = Choose intervention with largest expected value.



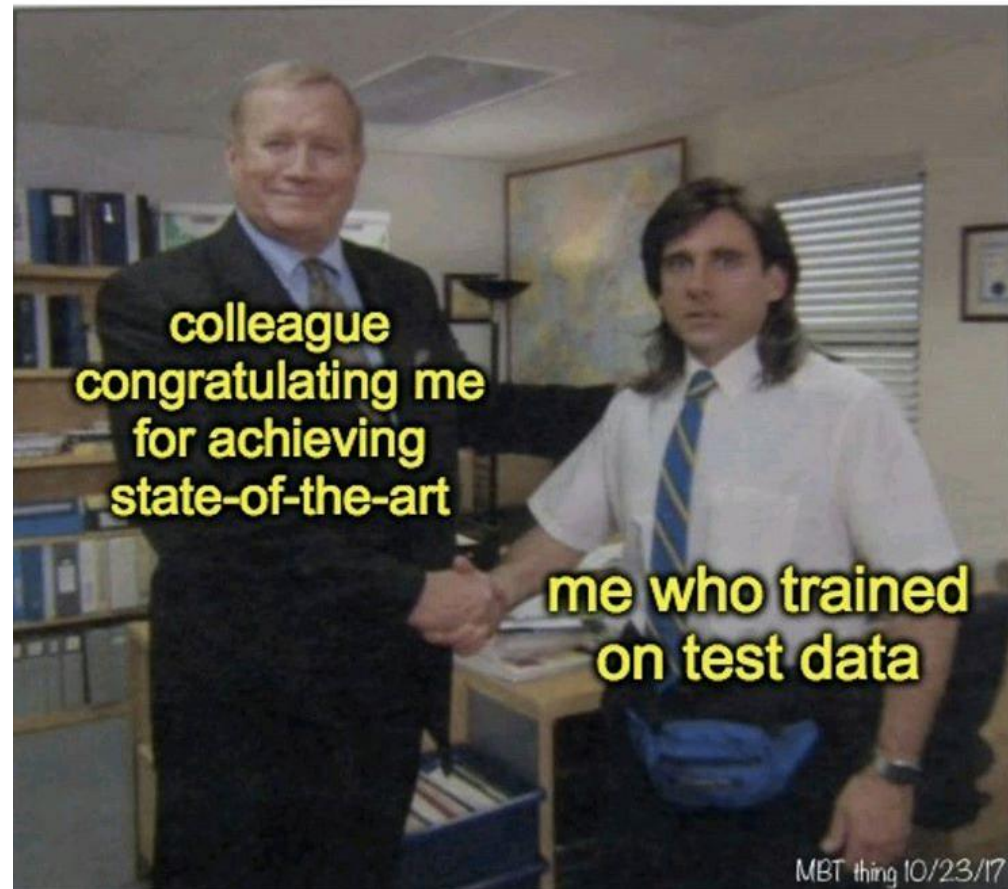
These are classic data science problems!

#5

Modern statistical modeling and machine learning is all about complexity control.



datascienceuniverse



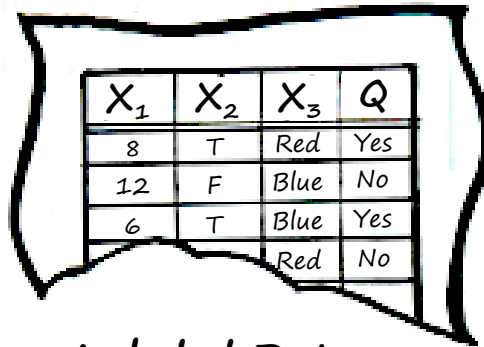
18 likes

datascienceuniverse Don't train on test data! 😏 ...
more

Framework #3: AI in Use vs. Machine Learning

#3 Complexity control and avoid overfitting.

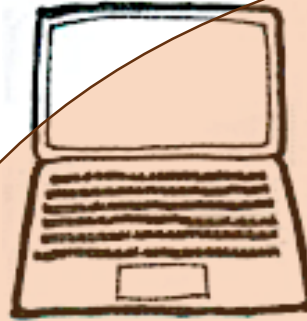
Machine Learning:



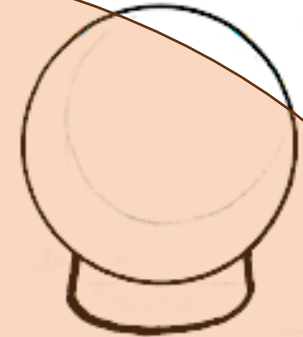
A hand-drawn table with four columns labeled X_1 , X_2 , X_3 , and Q . It contains four rows of data. The table is shown on a piece of paper with a torn bottom edge.

X_1	X_2	X_3	Q
8	T	Red	Yes
12	F	Blue	No
6	T	Blue	Yes
		Red	No

Labeled Data

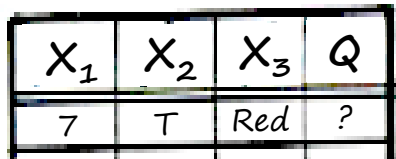


Machine Learning Algorithm



Learned Model

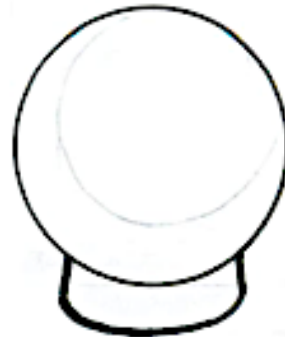
AI in Use (Inference):



A hand-drawn table with four columns labeled X_1 , X_2 , X_3 , and Q . It contains one row of data. The table is shown on a piece of paper with a torn bottom edge.

X_1	X_2	X_3	Q
7	T	Red	?

Data Instance



Model



Predicted Quantity



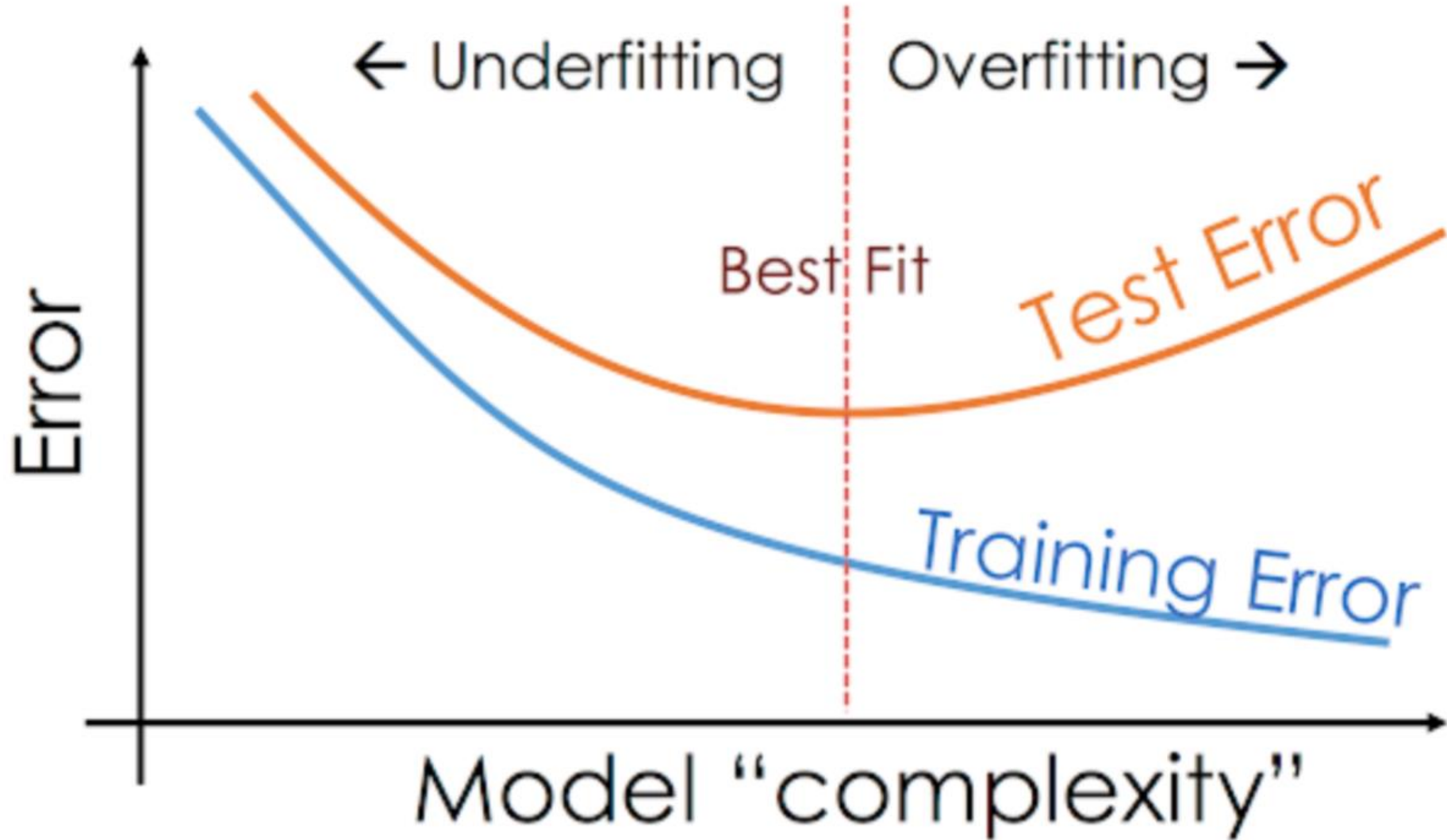
Decision Logic



Action



Fitting curves



Learning curves

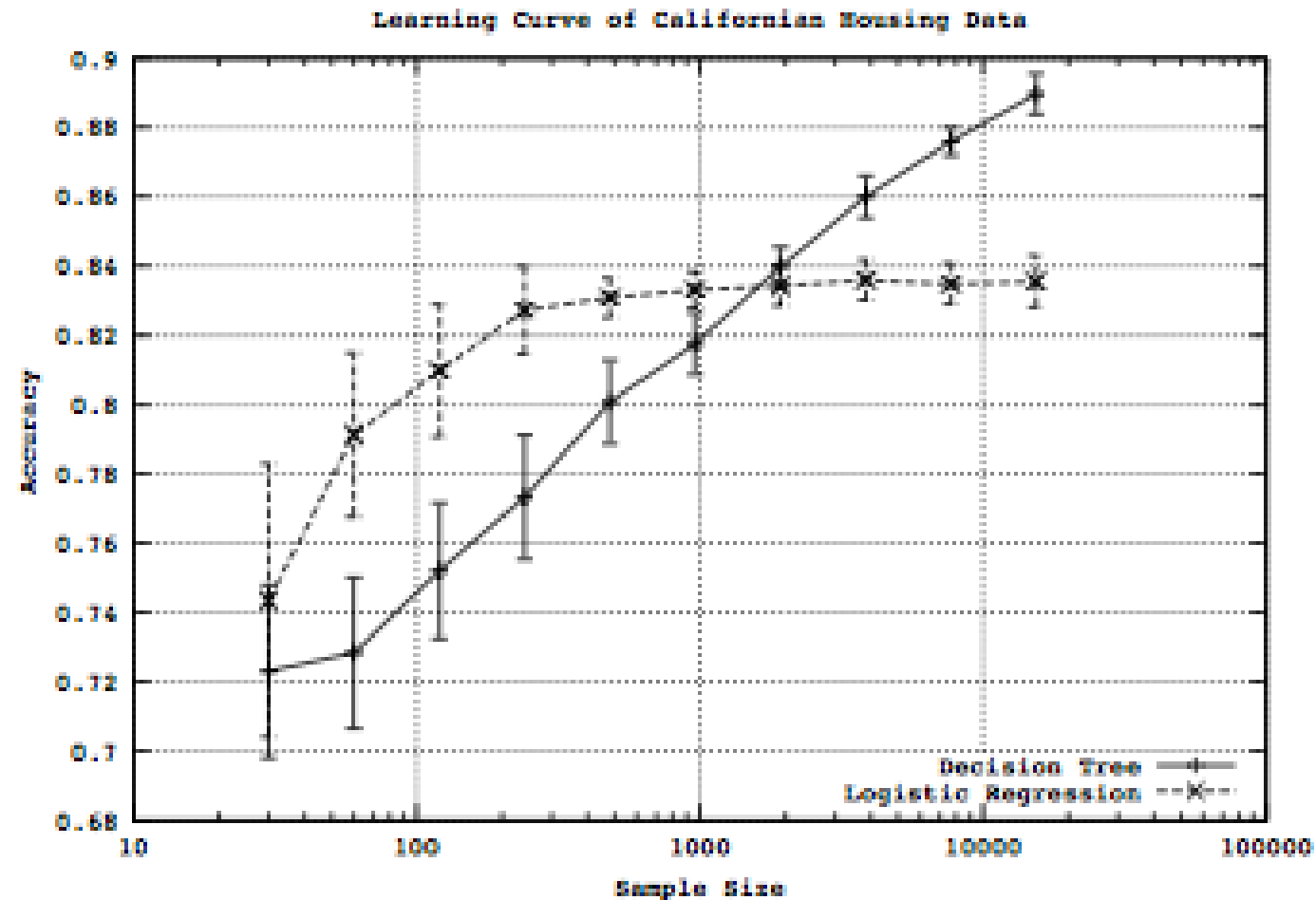
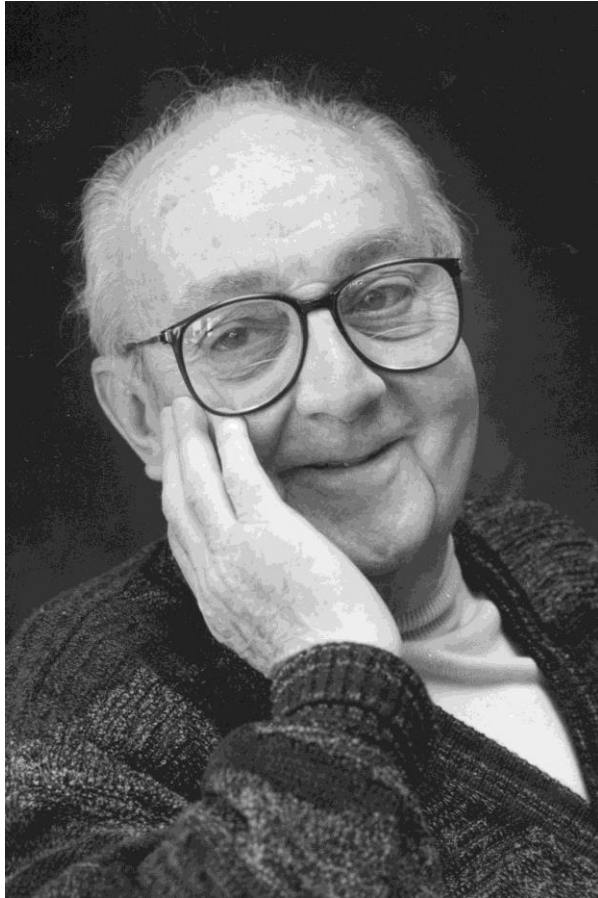


Figure 2: Log-scale learning curves

#6

Your models will be wrong!

Perhaps my favorite data science quote.



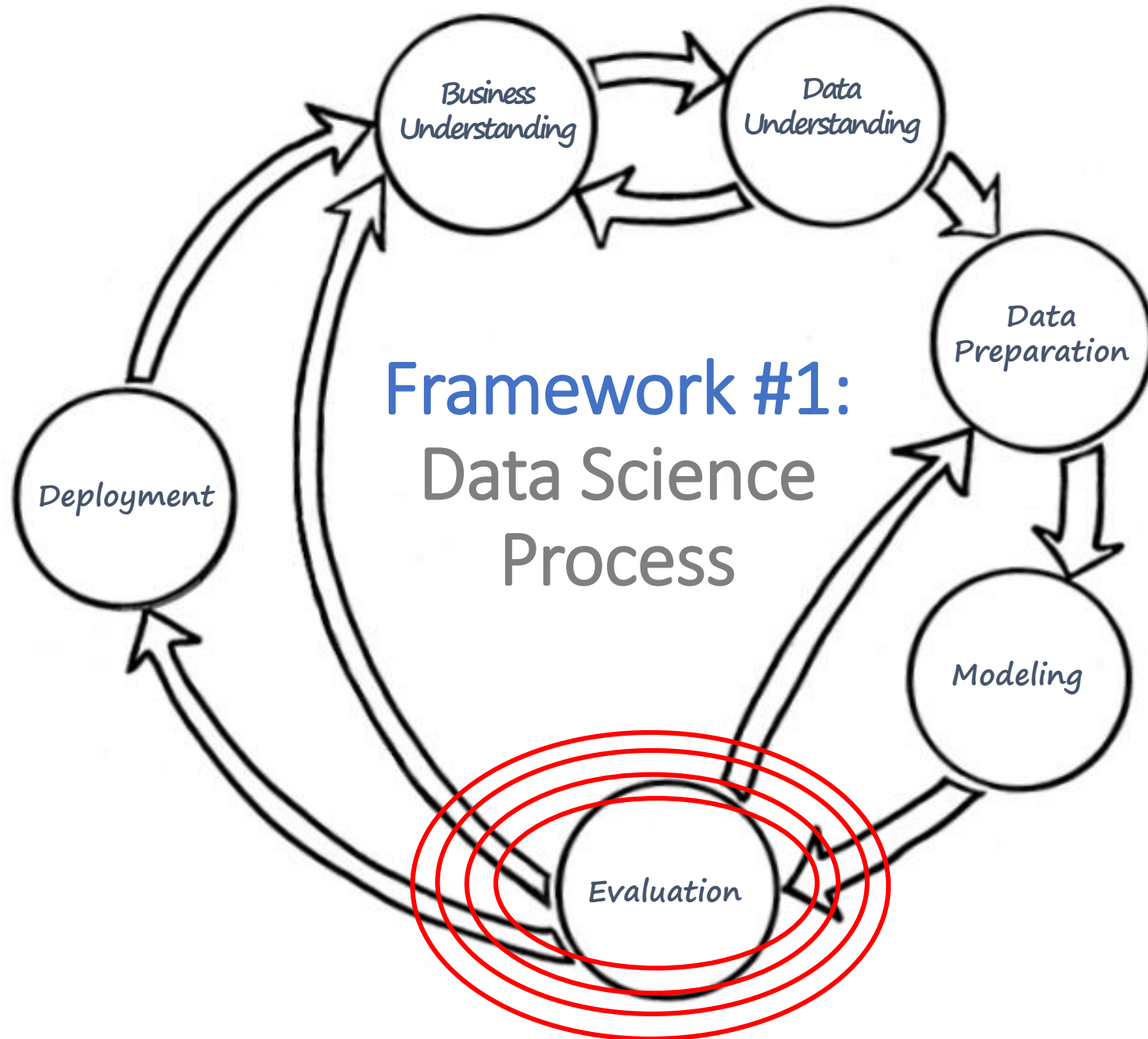
**“All models are wrong...
.... but some are useful.”**

— George E. P. Box

**George Box frees us to explore all types
of models, proxies, etc.**

Key to success:

Business Understanding ↔ Evaluation

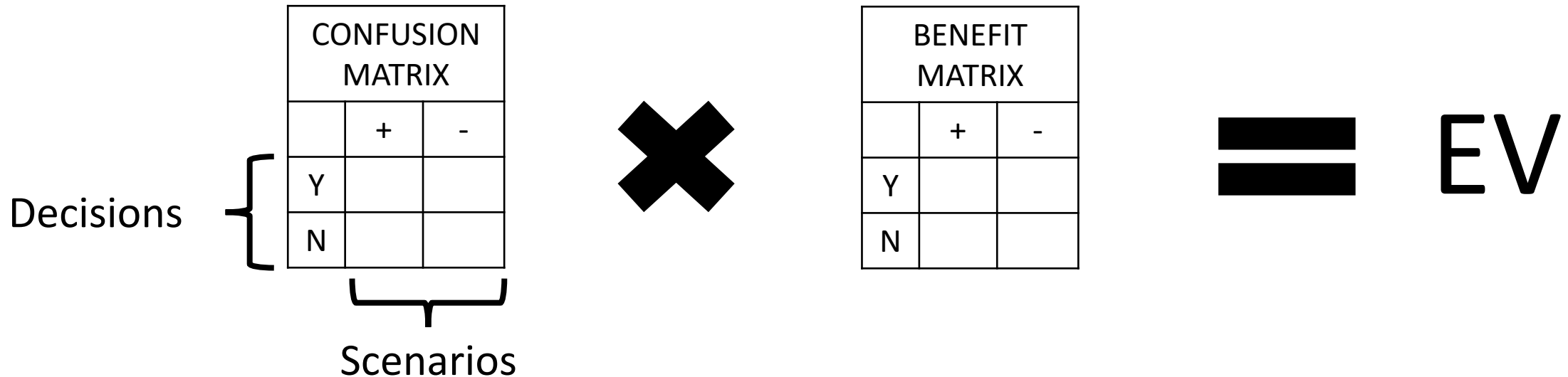


Framework #6: Expected Value

Also a data-driven framework to test solutions!

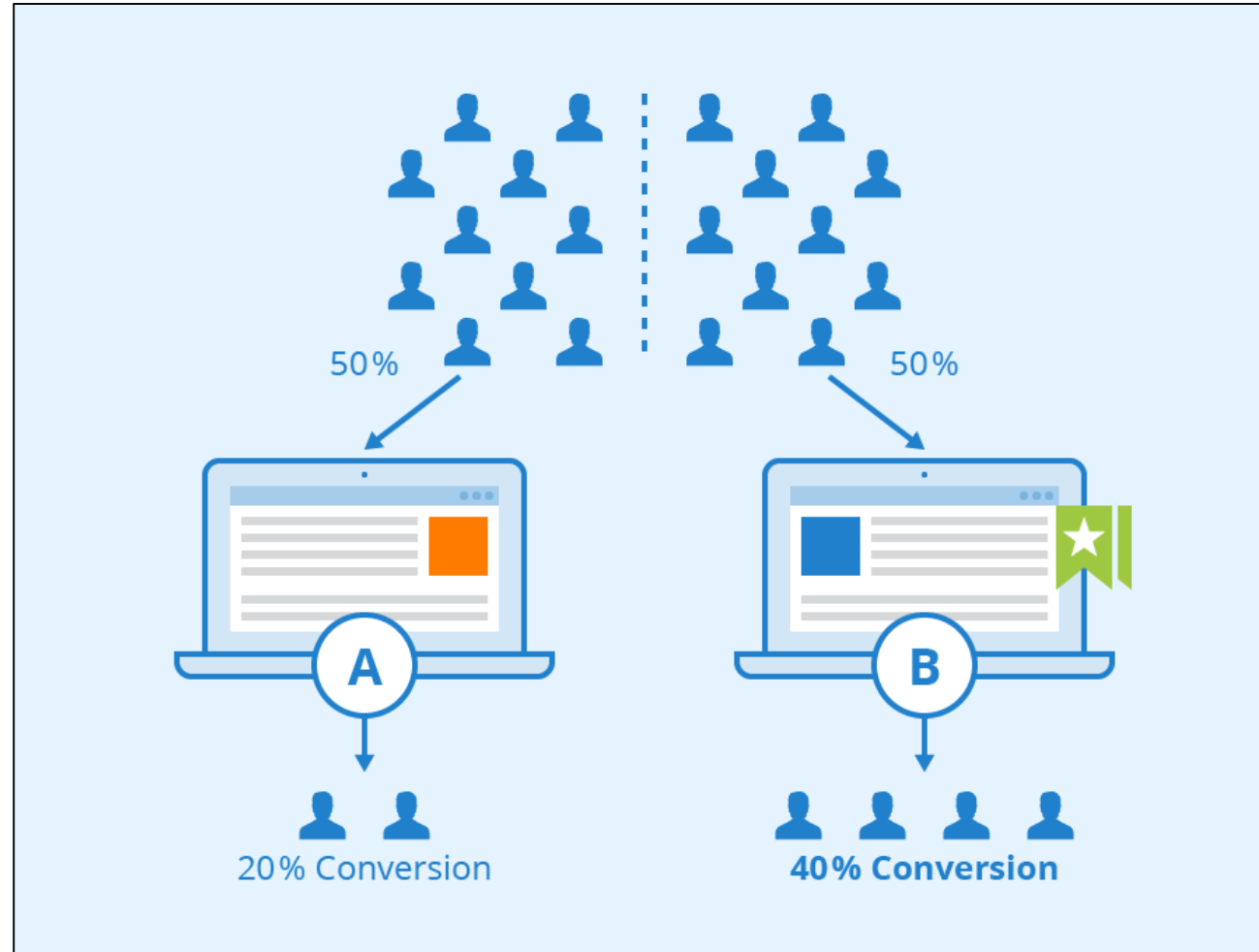
Predictive Power

Business Value



A/B testing

(for data collection and **evaluation**)



#7

But first... a fairy tale...

BUSINESS INSIDER

IBM's Watson Supercomputer May Soon Be The Best Doctor In The World

LAUREN F FRIEDMAN
APR. 22, 2014, 10:14 AM

Most people know Watson as [IBM's answer to Jeopardy star Ken Jennings](#). But IBM's aspirations for its artificially intelligent supercomputer are now less quiz show champion and more medical genius.

"Watson, the supercomputer that is now the world Jeopardy champion, basically went to med school after it won Jeopardy," MIT's [Andrew McAfee](#), coauthor of [The Second Machine Age](#), said recently in an [interview with Smart Planet](#). "I'm convinced that **if it's not already the world's best diagnostician, it will be soon.**"



Ben Hider / Getty Images



IBM's Watson Is Tackling Healthcare With Artificial Intelligence

The company has been making a splash in AI, and its investments in the healthcare industry are starting to produce results.

Danny Vena ([TMFLifelsGood](#))

Mar 19, 2017 at 10:49AM

International Business Machines ([NYSE:IBM](#)) has been [betting big on artificial intelligence](#) (AI). The company's AI-enabled *Jeopardy!*-winning cognitive supercomputer, Watson, has become the catch-all for the company's efforts in the area. Watson has been touted to revolutionize such diverse areas as [cybersecurity](#), customer service, and even [tax return preparation](#).

But nowhere is IBM's bet on Watson more evident than in the area of healthcare. The supercomputer's ability to analyze vast stores of data and recognize patterns make it a natural fit for medical applications.

Move your
old 401(k)



Get Started

Investing involves risk, including risk of loss.
Fidelity Brokerage Services, Member NYSE, SIPC.
© 2017 FMR LLC. All rights reserved. 608227.7.1





Matthew Herper, FORBES STAFF
I cover science and medicine, and believe this is biology's century.
FULL BIO



Virginia "Ginni" Rometty, chief executive officer of International Business Machines Corp. [+]

It was one of those amazing “we’re living in the future” moments. In an [October 2013 press release](#), IBM declared that MD Anderson, the cancer center that is part of the University of Texas, “is using the IBM Watson cognitive computing system for its mission to eradicate cancer.”



SAVE \$30
ON A NEW iPhone 8

After 24 monthly bill credits with finance ag & qual'g iPhone trade-i

LEARN MORE



iPhone 8

If you cancel wireless service, remaining smartphone may become due & you may lo us for details. Qual'g service & credit



Well, now that future is past. The partnership between IBM and one of the world's top cancer research institutions is falling apart. The project is on hold, MD Anderson confirms, and has been since late last year. MD Anderson is actively requesting bids from other contractors who might replace IBM in future efforts. And a scathing report from auditors at the University of Texas says the project cost MD Anderson more than \$62 million and yet did not meet its goals. The report, however, states: "Results stated herein should not be interpreted as an opinion on the scientific basis or functional capabilities of the system in its current state."

In February of this year, the University of Texas, which runs M.D. Anderson, announced it had shuttered the project, leaving the medical center out \$39 million in payments to IBM—for a project originally contracted at \$2.4 million. After four years it had not produced a tool for use with patients that was ready to go beyond pilot tests. M.D. Anderson wouldn't comment to me about Watson specifically, but it appears that the problems stemmed mainly from internal struggles over how the project was managed and funded.

“Health care has been an embarrassingly late adopter of technology,” says Manish Kohli, a physician and health-care informatics expert with the Cleveland Clinic.

JUL 2, 2017 @ 10:32 AM 94,478

The Little Black Book of Billionaire Secrets

Is IBM Watson A 'Joke'?



Jason Bloomberg, CONTRIBUTOR

I write and consult on digital transformation in the enterprise. [FULL BIO](#) ✓

Opinions expressed by Forbes Contributors are their own.

On the [May 8th](#) edition of *Closing Bell* on CNBC, venture capitalist Chamath Palihapitiya, founder and CEO of Social Capital, created quite a stir in enterprise artificial intelligence (AI) circles, when he took on [IBM](#) IBM +0.81% Watson, Big Blue's AI platform.

"Watson is a joke, just to be completely honest," Palihapitiya said.
"I think what IBM is excellent at is using their sales and

T-MOBILE
FOR BUSINESS

SAVE \$300
ON A NEW
iPhone 8

After 24 monthly bill credits with finance agmt & qual'g iPhone trade-in.

LEARN MORE

iPhone 8

If you cancel wireless service, remaining balance on your smartphone may become due & you may lose credits; contact us for details. Qual'g service & credit check req'd.



TECH • MASTERING DATA

Has IBM Watson's AI Technology Fallen Victim to Hype?



Early on, some large companies that were interested in Watson as portrayed by IBM, privately noted later that while they saw great promise, Watson in reality, was a set of technologies that needed to be stitched together at their site. That meant buying Watson—and then preparing data for Watson—was more a big integration project than a product purchase.

According to Tech Review, most of the issues with Watson have nothing to do with a technological problem. They are more about the availability and quality of data for Watson to process.

My former colleague Derrick Harris has great take on all this in his [Architech newsletter](#). Read the whole thing, but in his view, the biggest issue for business people is falling for the hype without understanding what's happening in the technology right now. He thinks the technologists in the trenches have a good handle on that. Too bad the marketers and PR maestros at their companies choose not to reflect it.

Business Impact

A Reality Check for IBM's AI Ambitions



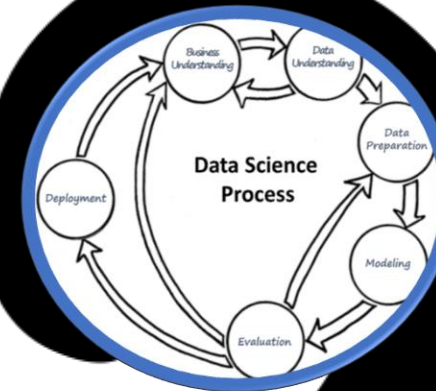
Read the article... but essentially...

*“Whoops! Where’s the labeled
training data?!”*

#7

Predictive analytics projects generally start with high uncertainty and therefore high risk. Start simple and keep iterating.

Investing in Predictive Analytics



“the biggest issue for business people is falling for the hype without understanding what’s happening.”

So remember...

Don't fall in love with the machine!

I hope you enjoyed the course as
much as I did!

