Building ML Products in Insurance/Finance

Reza Khorshidi, DPhil (Oxon)

Deep Medicine Program, University of Oxford

@RezaKhorshidi

What is insurance?

Insurance is the business of transferring the risk to its optimal owner.

Section I

A bit of history + intro

We have been over excited about AI since the ea

VOL. LIX. No. 236.] [October, 1950

MIND

A QUARTERLY REVIEW

OF

PSYCHOLOGY AND PHILOSOPHY

I.—COMPUTING MACHINERY AND INTELLIGENCE

BY A. M. TURING

1. *The Imitation Game.*

I PROPOSE to consider the question, 'Can machines think ?' This should begin with definitions of the meaning of the terms 'machine' and 'think'. The definitions might be framed so as to reflect so far as possible the normal use of the words, but this attitude is dangerous. If the meaning of the words 'machine' and 'think' are to be found by examining how they are commonly

It will simplify matters for the reader if I explain first my own beliefs in the matter. Consider first the more accurate form of the question. I believe that in about fifty years' time it will be possible, to programme computers, with a storage capacity of about 109, to make them play the imitation game so well that an average interrogator will not have more than 70 per cent chance of making the right identification after five minutes of questioning. The original question, "Can machines think?" I believe to be too meaningless to deserve discussion. Nevertheless I believe that at the end of the century

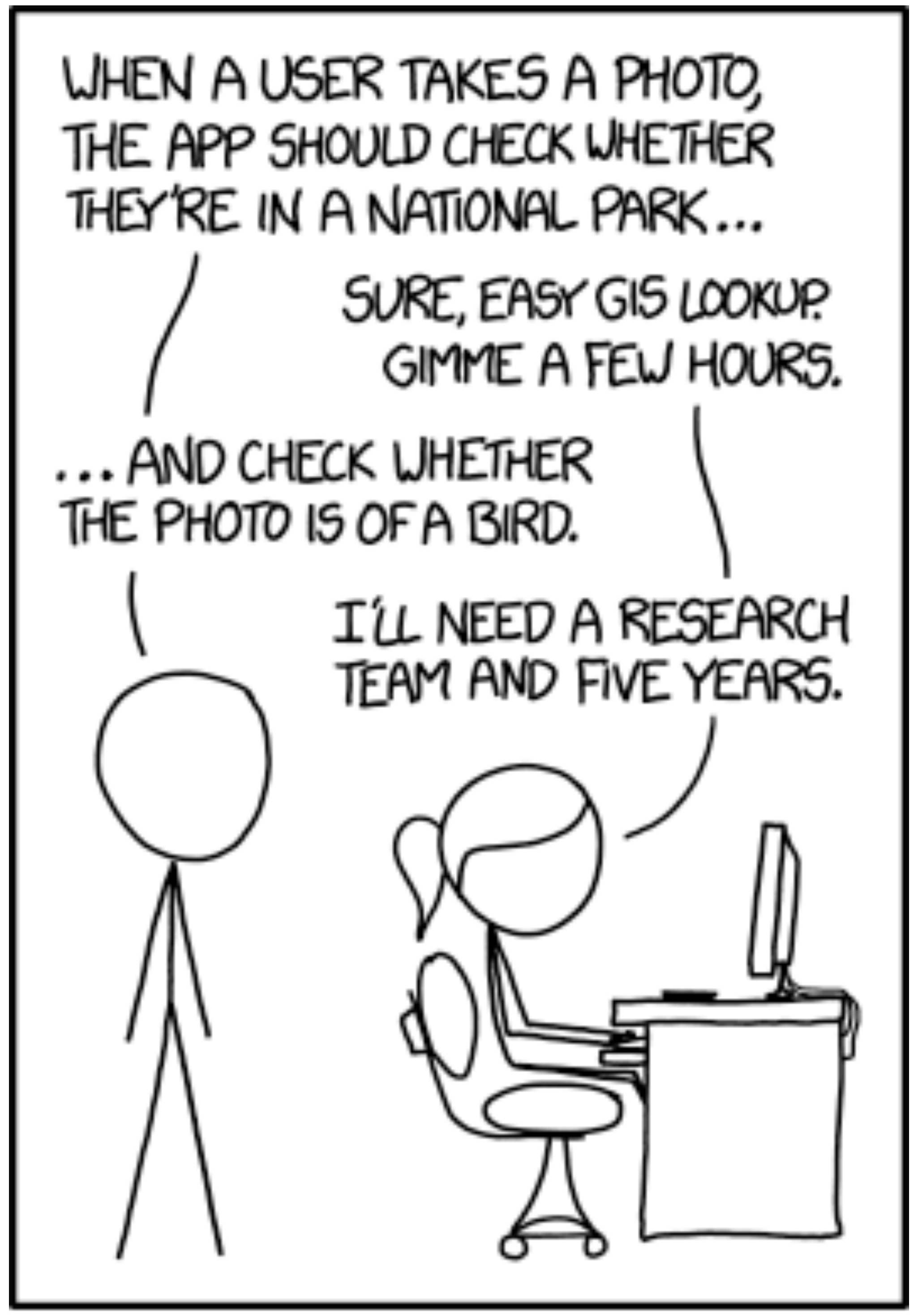
As I have explained, the problem is mainly one of programming. Advances in engineering will have to be made too, but it seems unlikely that these will not be adequate for the requirements. Estimates of the storage capacity of the brain vary from 10^{10} to 10^{15} binary digits. I incline to the lower values and believe that only a very small fraction is used for the higher types of thinking. Most of it is probably used for the retention of visual impressions, I should be surprised if more than 10^9 was required for satisfactory playing of the imitation game, at any rate against a blind man. (Note: The capacity of the *Encyclopaedia Britannica*, 11th edition, is 2 X 10^9) A storage capacity of 10^7, would be a very practicable possibility even by present techniques. It is probably not necessary to increase the speed of operations of the machines at all. Parts of modern machines which can be regarded as analogs of nerve cells work about a thousand times faster than the latter. This should provide a "margin of safety" which could cover losses

Making predictions is hard ...

Hype has created some unrealistic expectations about AI...

09’2014

https://xkcd.com/1425/



IN CS, IT CAN BE HARD TO EXPLAIN THE DIFFERENCE BETWEEN THE EASY AND THE VIRTUALLY IMPOSSIBLE.

Making predictions is hard ...

09’2014

https://xkcd.com/1425/

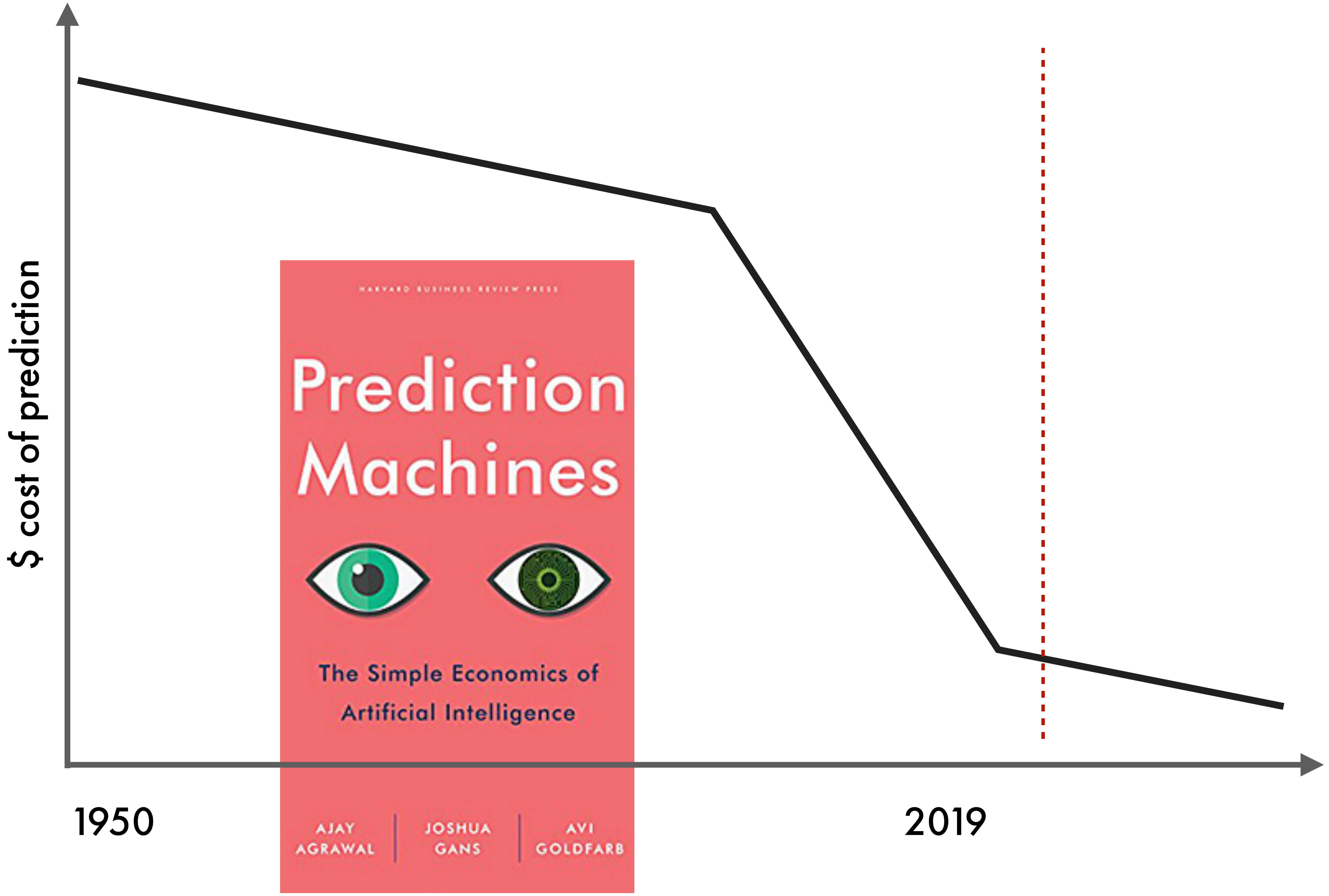
Hype has created some unrealistic expectations about AI...

...but less than two years after that cartoon was published, the image classification problem depicted in it could be solved with around 10 lines of Python code...

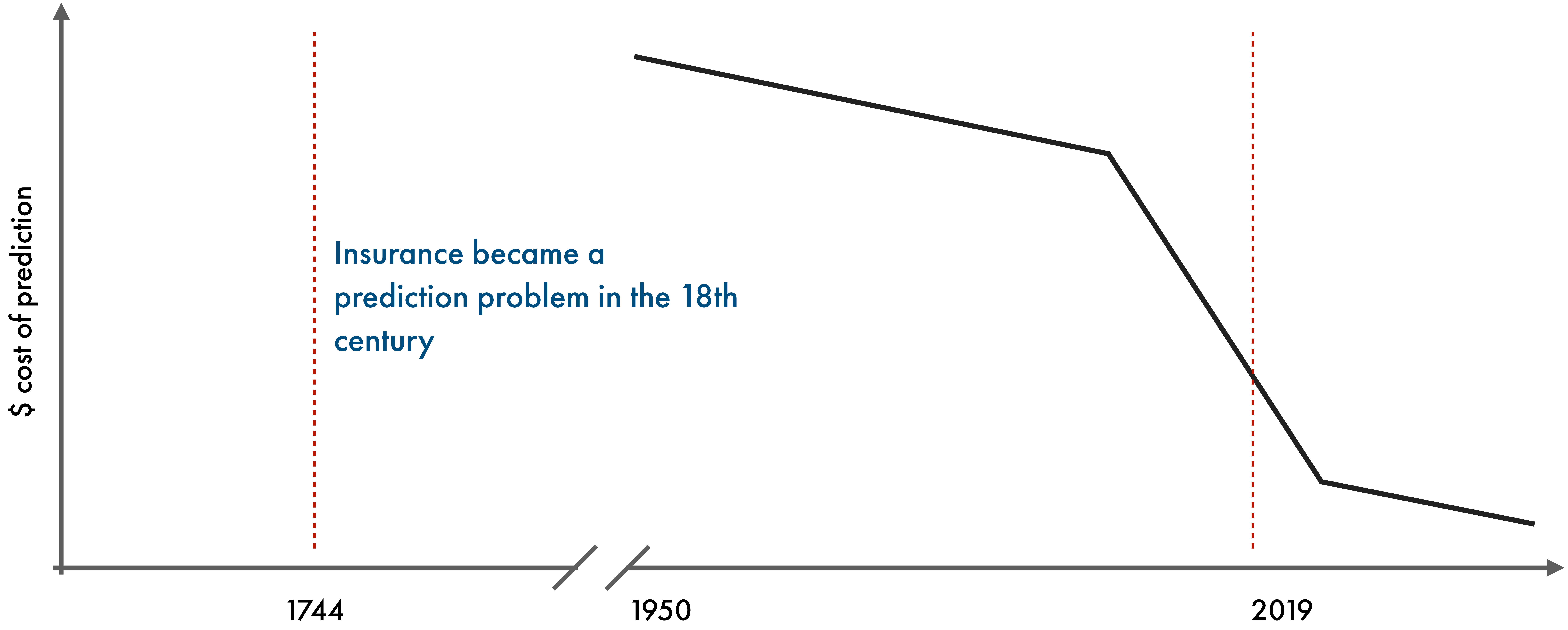
```
from keras.preprocessing.image import load_img
from keras.preprocessing.image import img_to_array
from keras.applications.vgg16 import preprocess_input
from keras.applications.vgg16 import decode_predictions
from keras.applications.vgg16 import VGG16

# load the model
model = VGG16()

# load an image from file
image = load_img('mug.jpg', target_size=(224, 224))

# predict the probability across all output classes
yhat = model.predict(image)

# convert the probabilities to class labels
label = decode_predictions(yhat)

# retrieve the most likely result, e.g. highest probability
label = label[0][0]

# print the classification
print('%s (%.2f%%)' % (label[1], label[2]*100))
```

The simple economics of AI/ML

AI can reduce the cost of prediction — this means that more businesses will aim to turn their existing problems into prediction problems.



The case for AI in insurance



The case for better general insurance

$10 tn

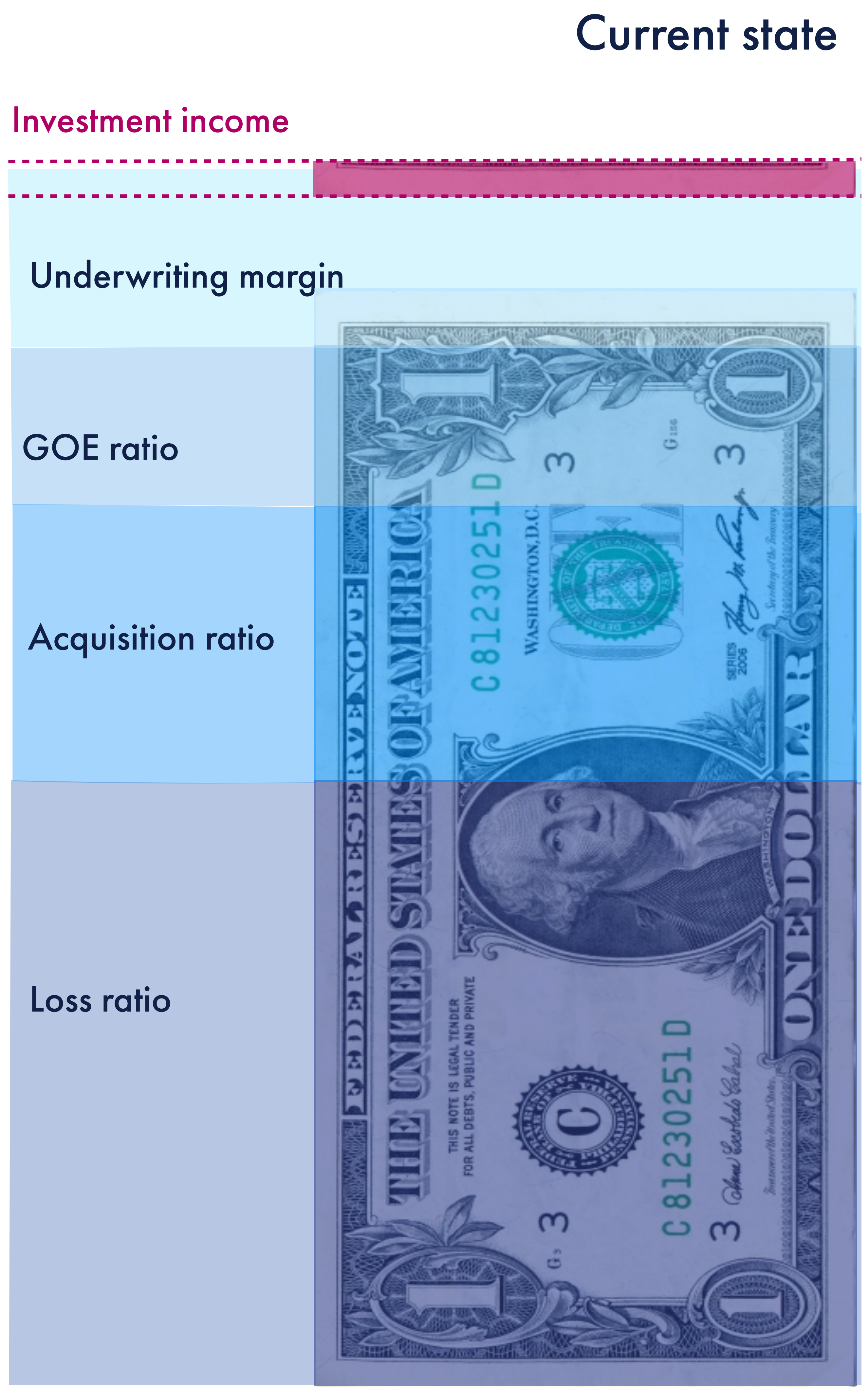
$1 tn

The case for better medical insurances

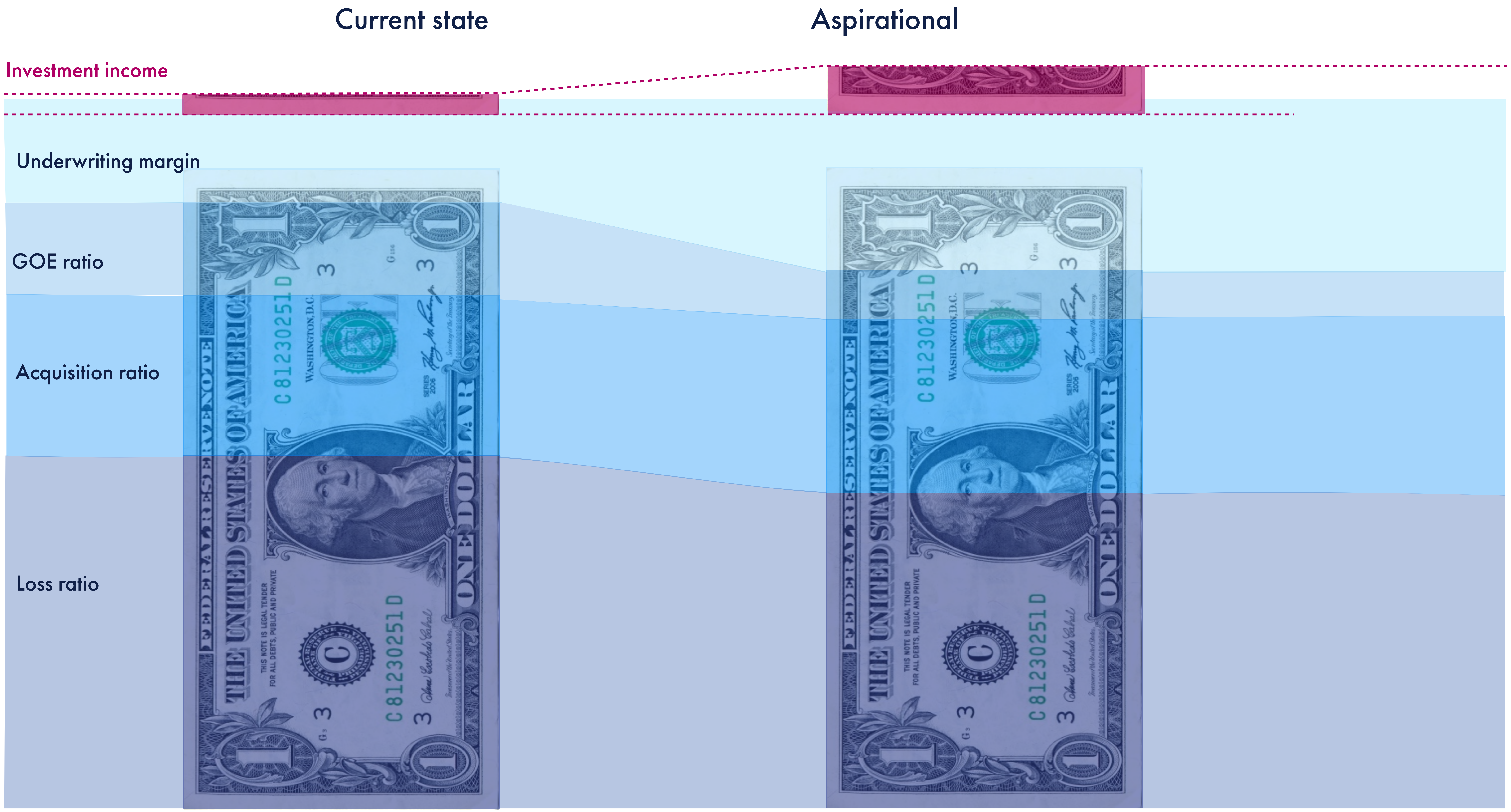
$10tn

$3.5tn

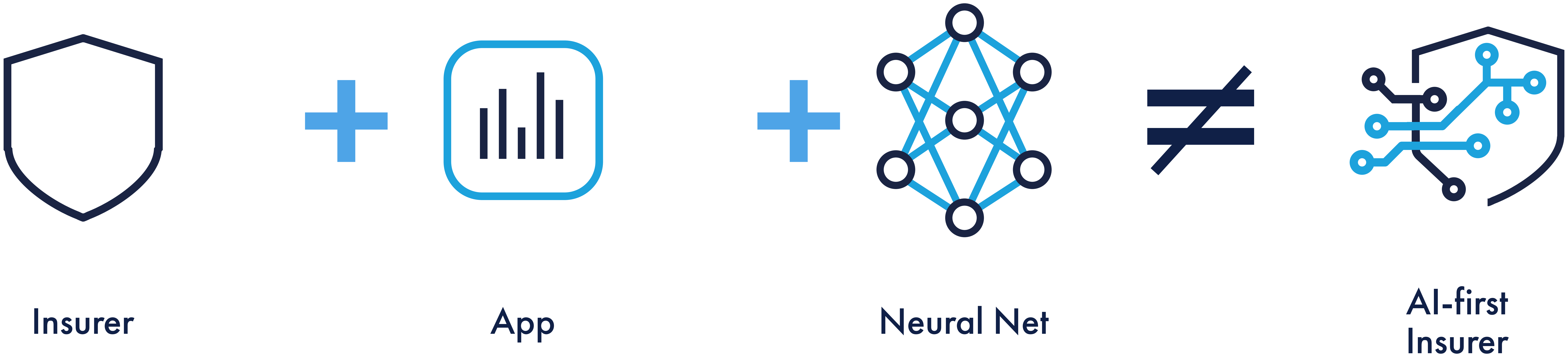
The simple economics of insurance



The simple economics of insurance

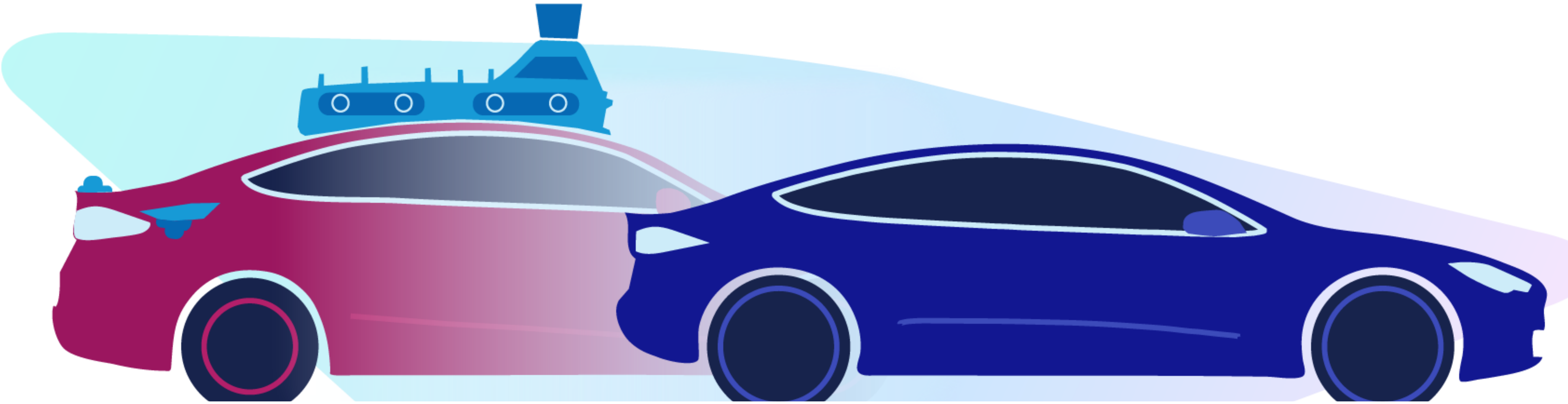


It will take more than ML models to improve the economics of insurance



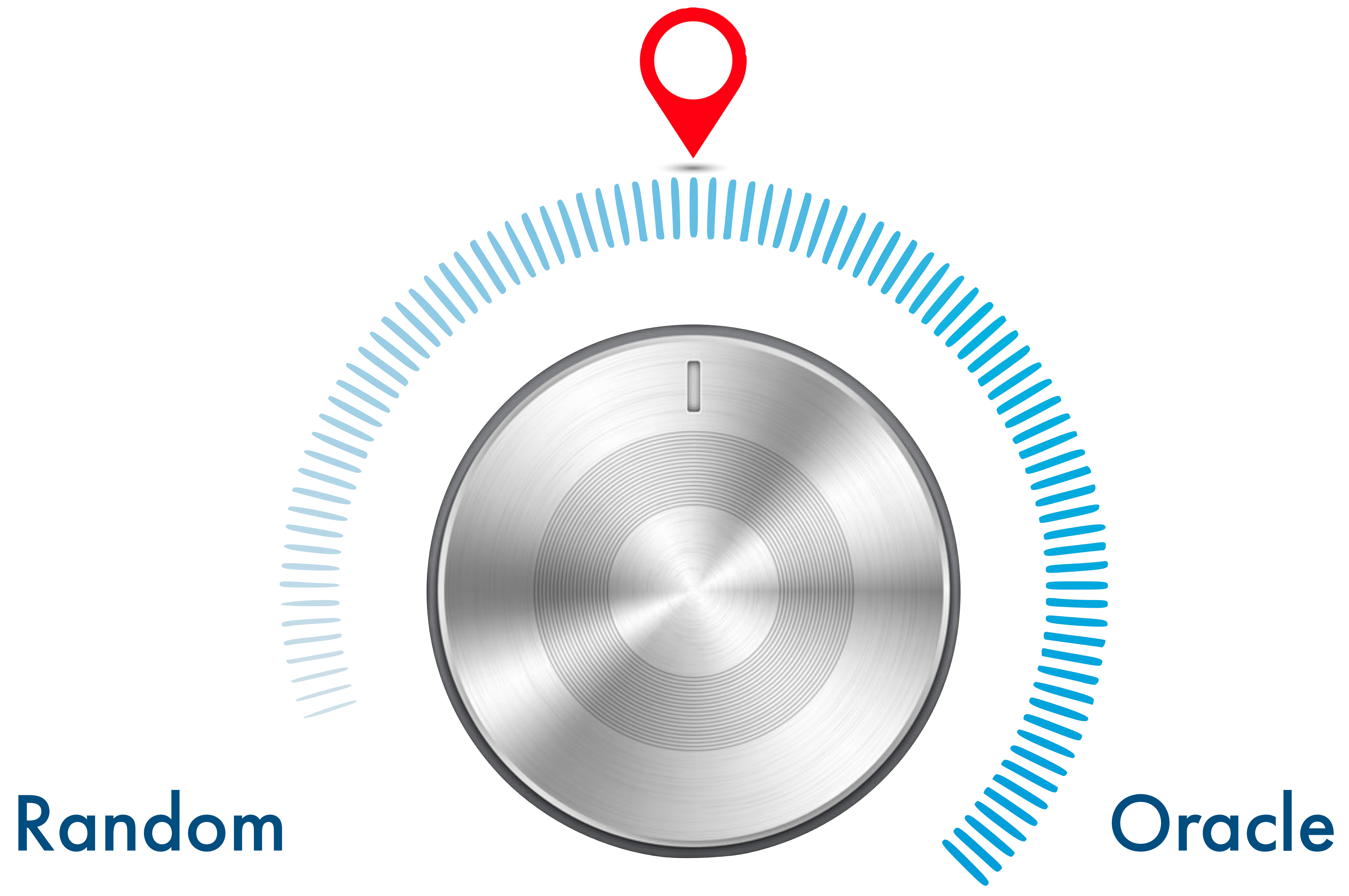
Example I: Tesla

Tesla is an AI-first company, with the ultimate goals of building the world's best AV. However, it began with a well-designed EV, equipped with the necessary array of sensor for autonomy. Over time, it can cruise control, automatically park, automatically break, do its own insurance, and so on. Eventually, however, it might reach full autonomy; there, the universe of possibilities for Tesla will be vast.



Example II: Amazon

Switching to anticipatory shipping, for instance, leads to a different company/product; not reliant on the same approach to customer acquisition and user experience, and a much stronger reliance on ML's ability to correctly anticipate.



Amazon Patents "Anticipatory" Shipping — To Start Sending Stuff Before You've Bought It

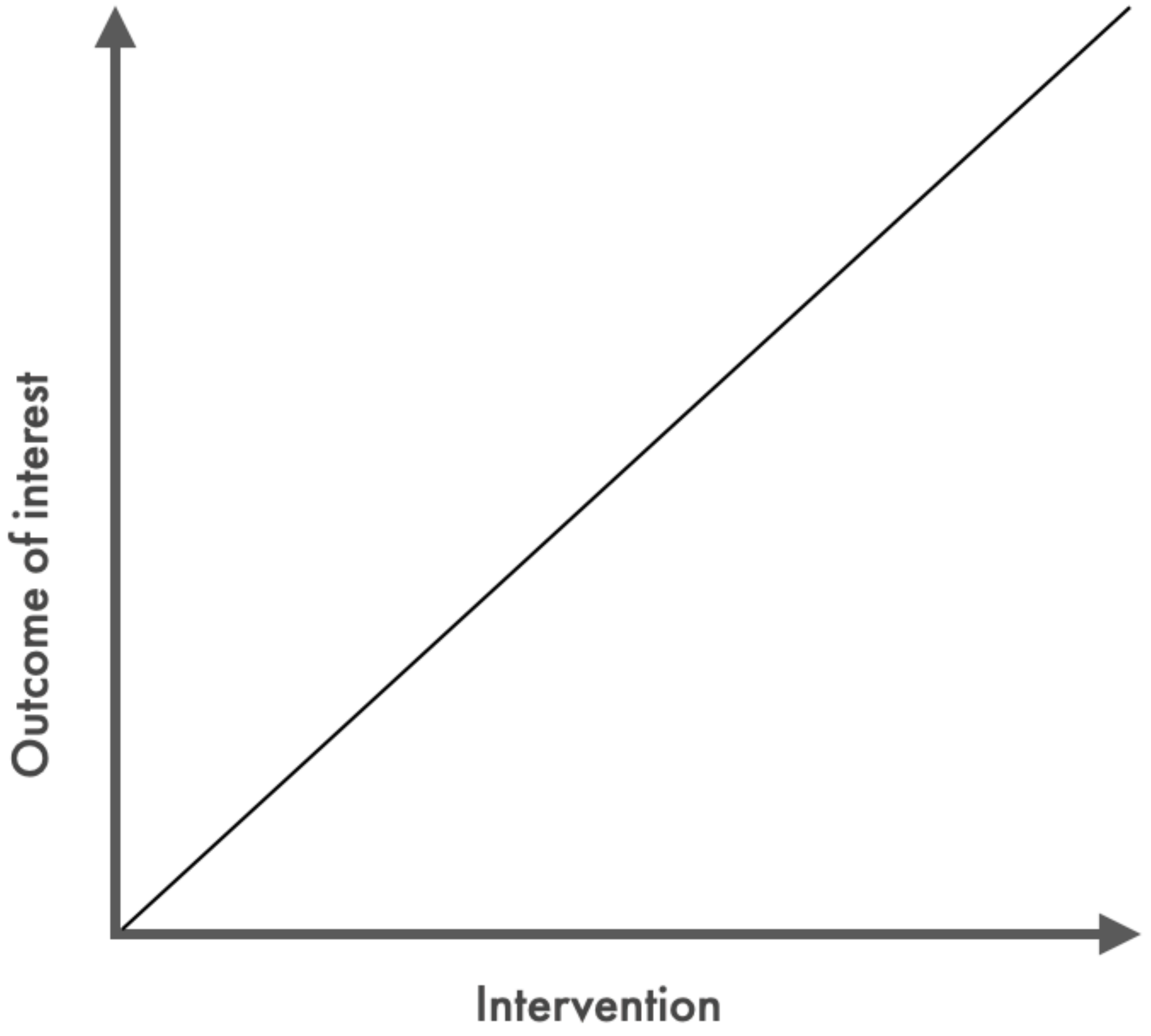
Natasha Lomas @riptari / 6 years ago

Comment

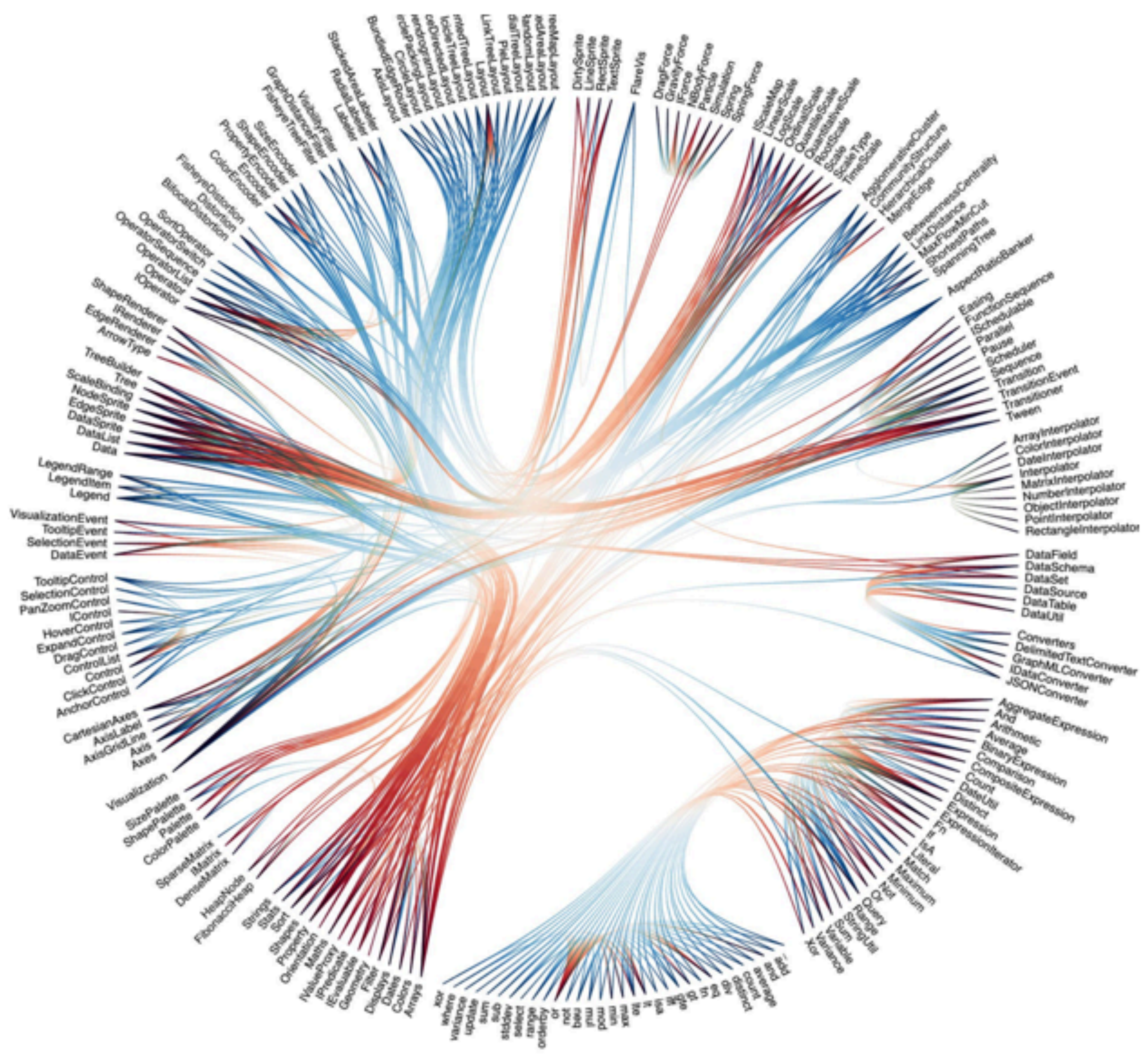
ML models in most applications become part of "computationally irreducible" systems.

The principle of computational irreducibility says that the only way to determine the answer to a computationally irreducible question is to perform, or simulate, the computation.

What we wish for ...



The reality of systems ...



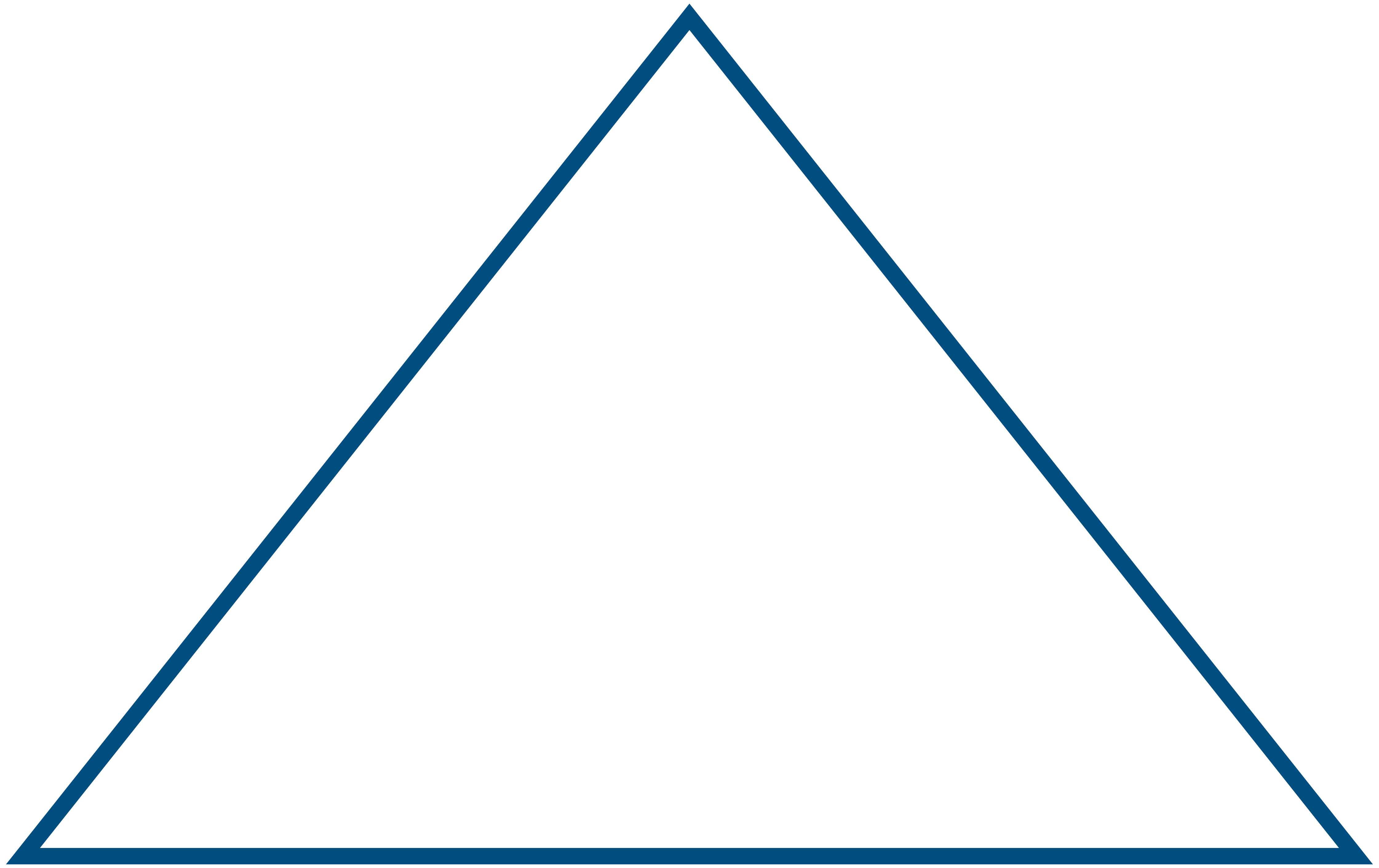
https://www.wolframscience.com/nks/

Section II

From model.fit() to market.fit()

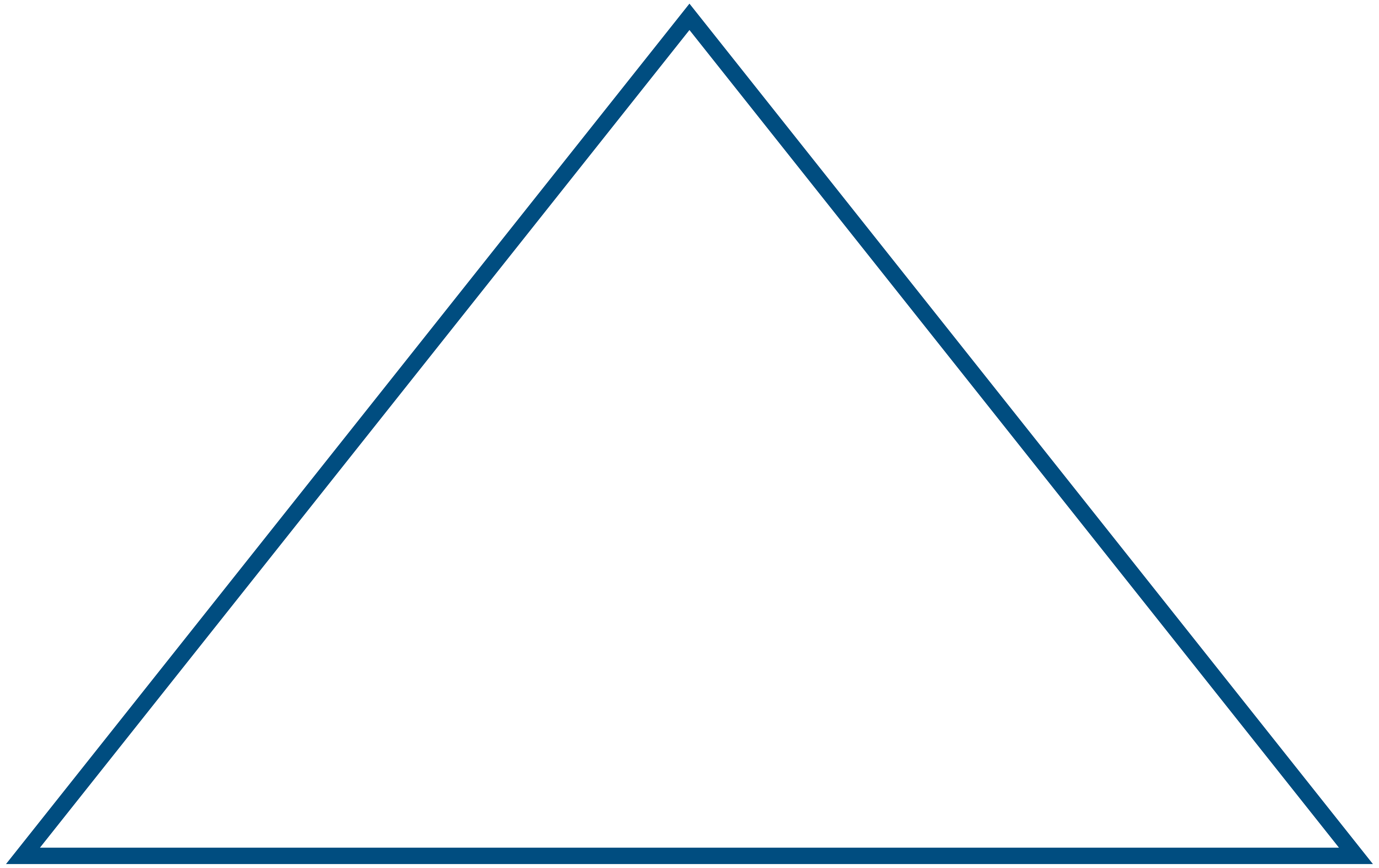
Let’s answer this question:

- What os the sum of the angles of a triangle?



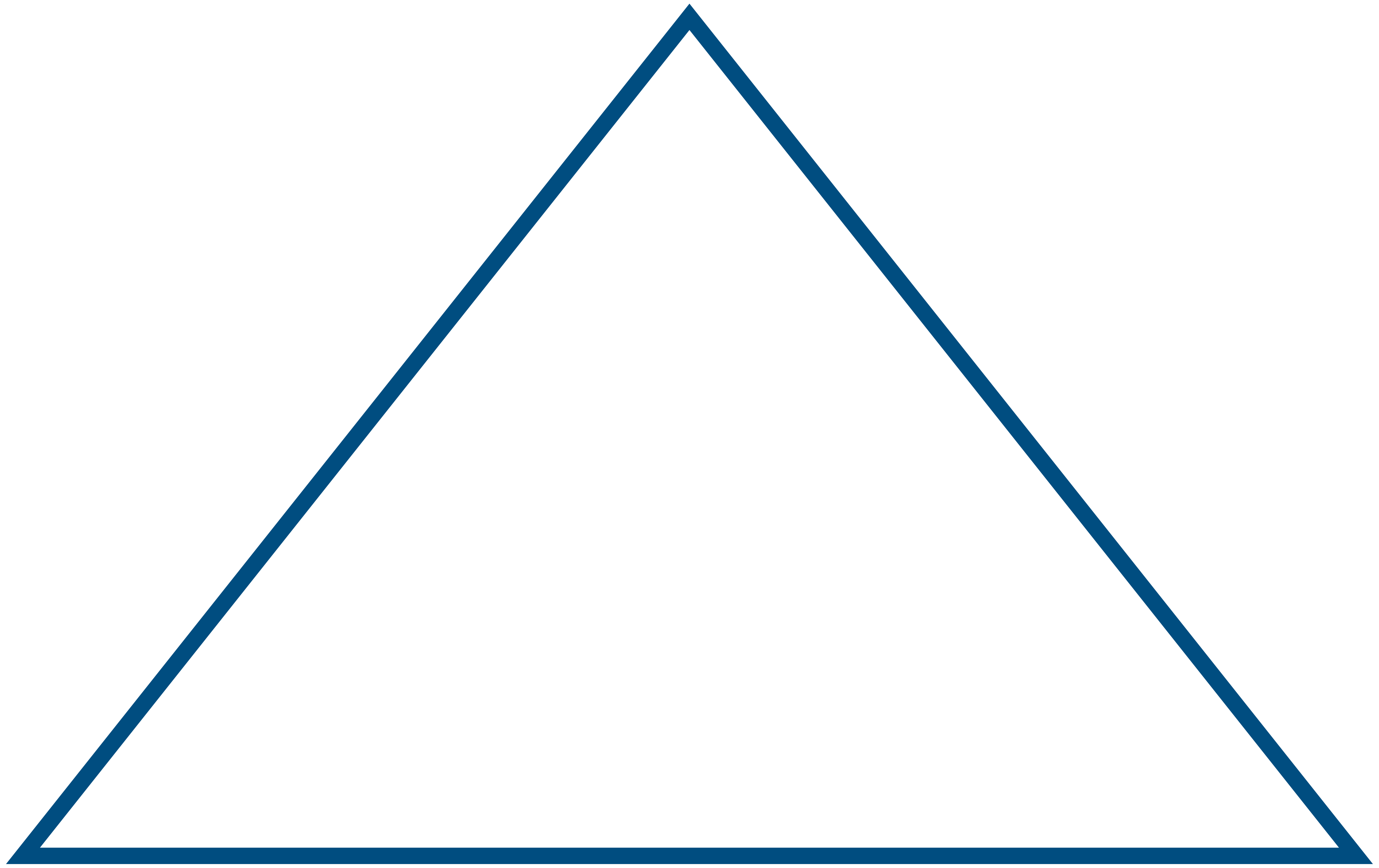
Let’s answer this question:

- What os the sum of the angles of a triangle?
- Are you sure?



Let's answer this question:

- What os the sum of the angles of a triangle?
- Are you sure?
- How do you know this is so?



Let’s answer another question:

- Does BioLight work?

“BioLight is bringing game-changing technology to the red light therapy market!”

“$1,099.00 to make it possible for someone to have a quality or ability again that they have not had for a long time; to bring back to a state of health, soundness, ...

Let's answer another question:

- Does BioLight work?
- Are you sure?

"BioLight is bringing game-changing technology to the red light therapy market!"

"$1,099.00 to make it possible for someone to have a quality or ability again that they have not had for a long time; to bring back to a state of health, soundness, ...

Let's answer another question:

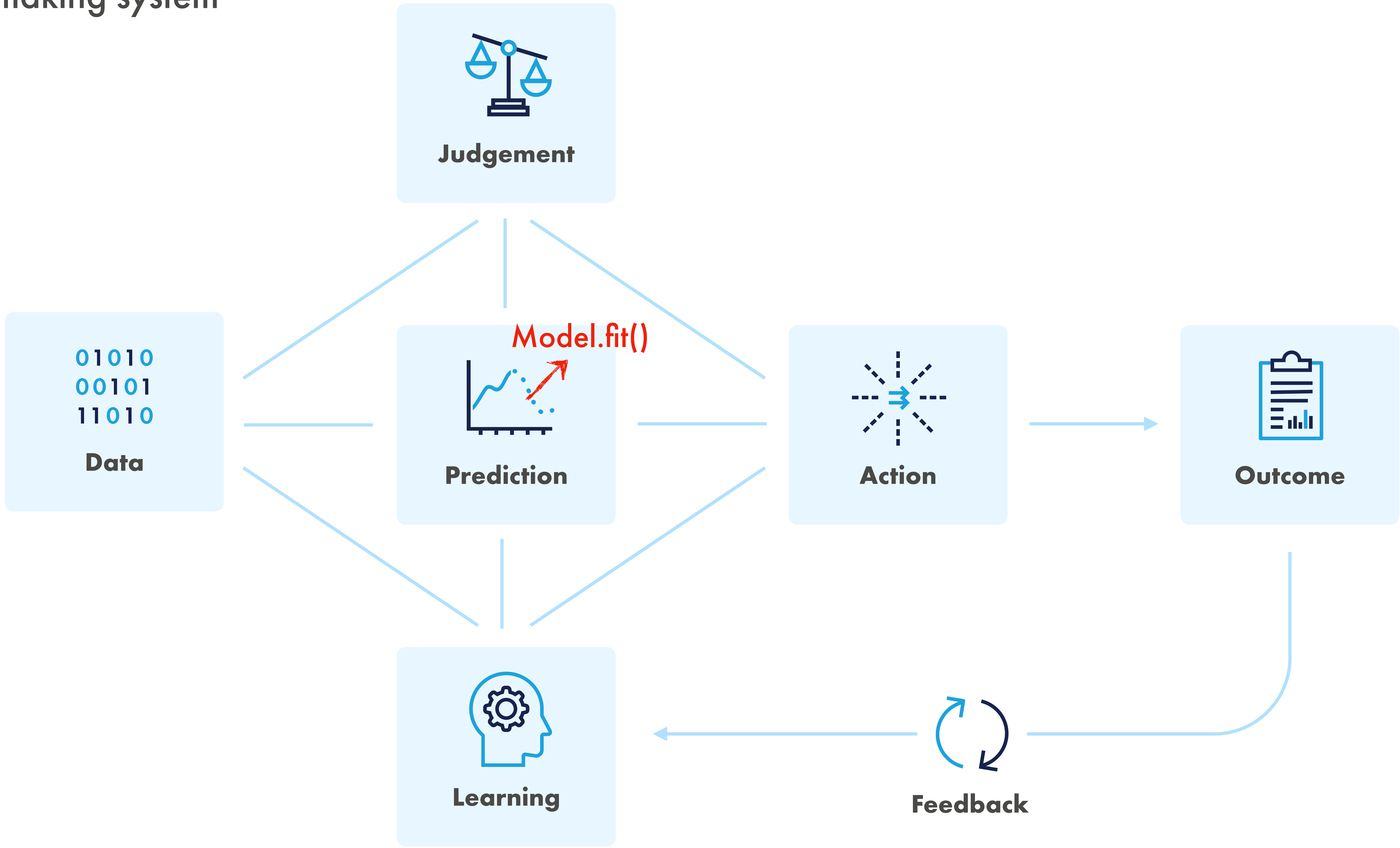
- Does BioLight work?
- Are you sure?
- How do you know this is so?

"BioLight is bringing game-changing technology to the red light therapy market!"

"$1,099.00 to make it possible for someone to have a quality or ability again that they have not had for a long time; to bring back to a state of health, soundness, ...

ML is a component in a decision-making system

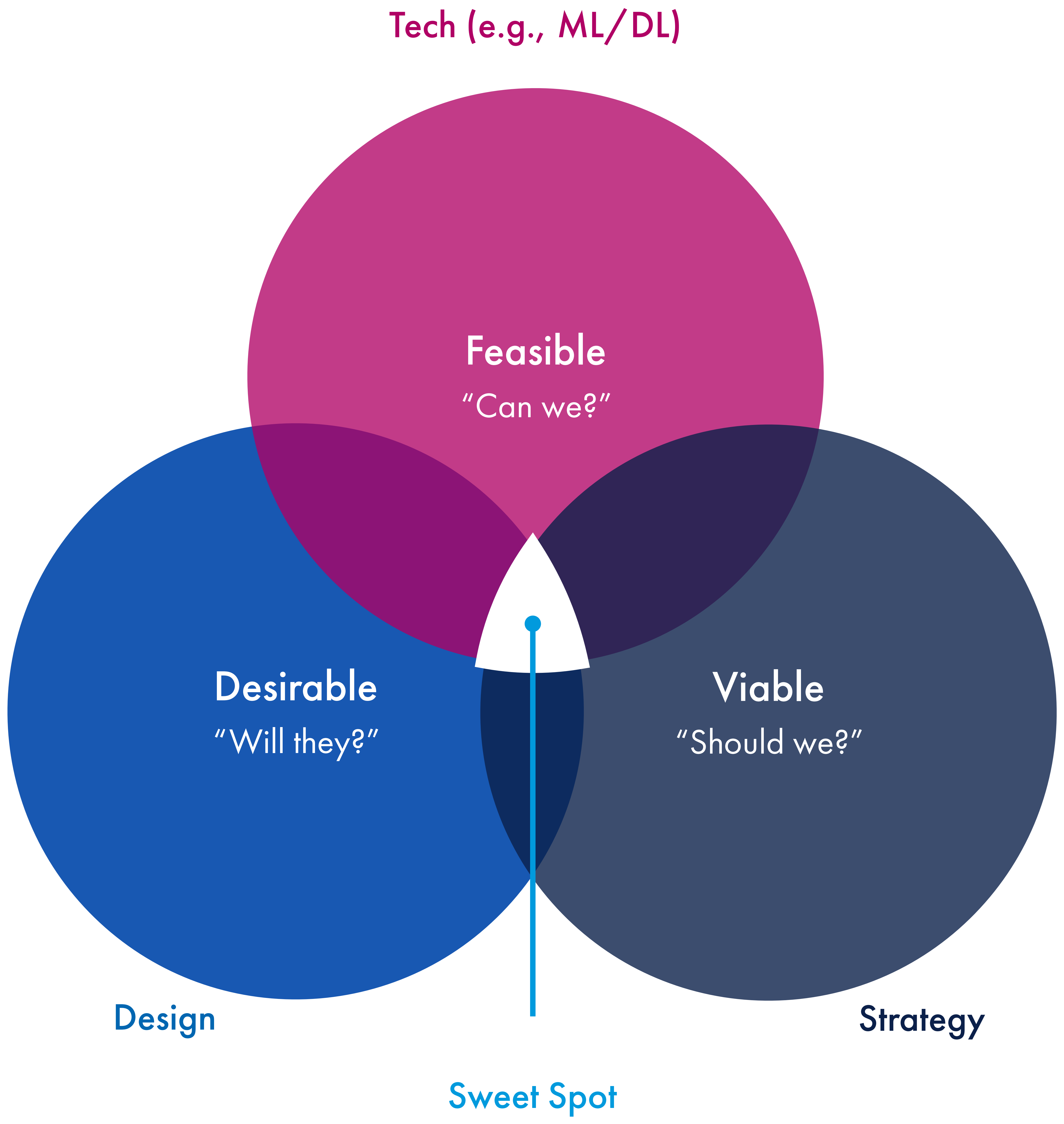
The new economics of prediction will have an impact on the economics of other components in the system. For instance, cheaper prediction is likely to lead to more expensive judgment, or data. Therefore, the impact of ML should not be considered in isolation, and rather as part of a system.



Achieving Market.fit() requires jointly optimising for multiple criteria.

A product needs to optimise for simultaneously being (1) desirable (i.e., users/customers really want it), (2) feasible (i.e., can be built on the current tech/operational capabilities), and (3) viable (i.e., lead to an economically sustainable solution / business model).

In most cases, Model.fit() relates to all three of these criteria; in AI-first products, however, it is usually more related to feasibility criteria.



A product is a specific offering intended to meet a set of customer needs.

The goal is to not leave things to chance and rather devise a systematic approach to discover, design and deliver.

When a great product meets a lousy market, market wins.
When a lousy product meets a great market, market wins.
When a great product meets a great market, something special happens.

$$P(fit|\mathcal{M}_o, \mathcal{M}_a) \approx P(fit|\mathcal{M}_a)$$

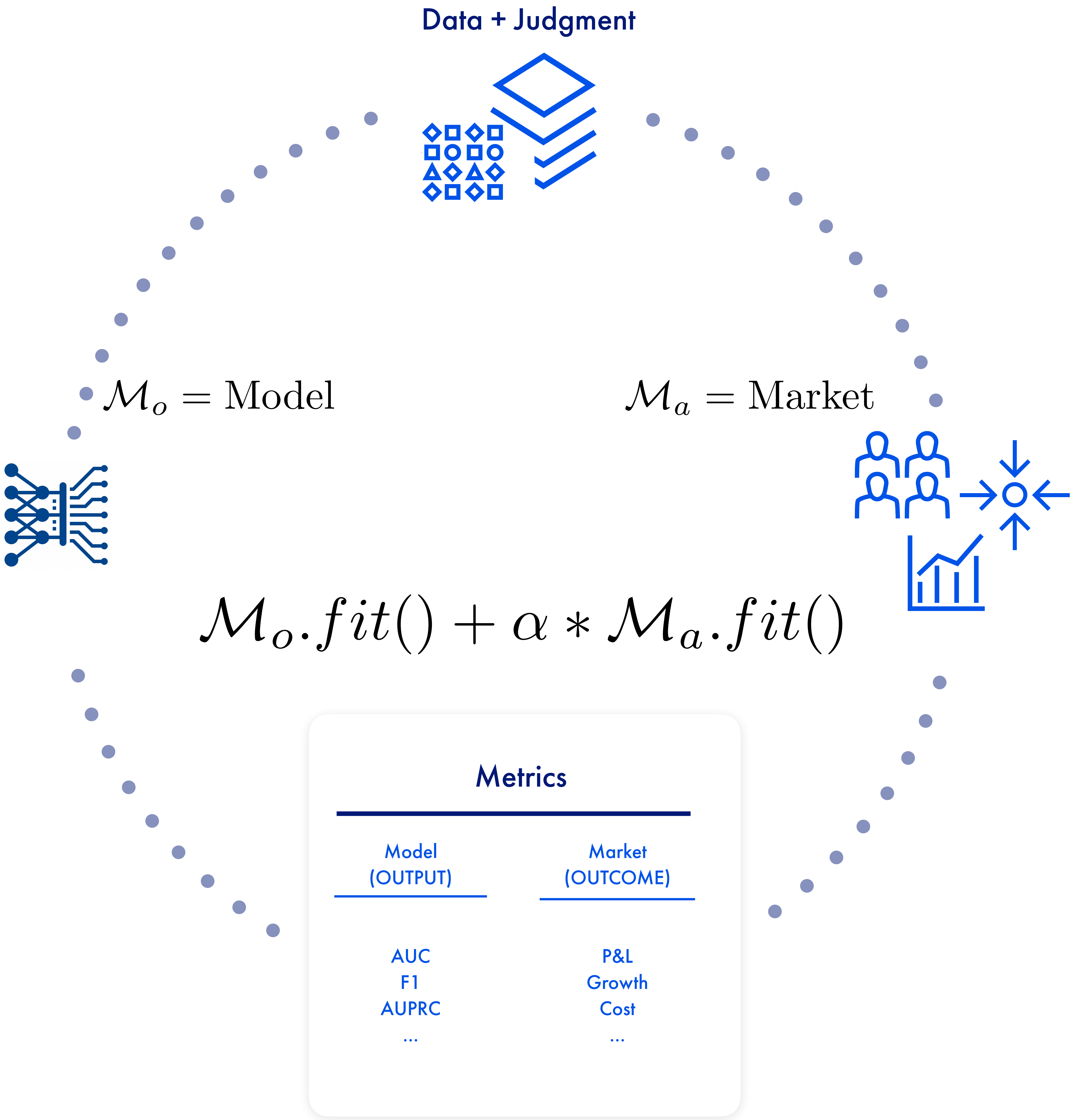
Another example: Twitter's image crop

Using AI to crop the images.

This brings us to the world of products and real-world evidence.

This diagram summarises the ecosystems in which ML systems operate. They consist of models, data, market/context, and metrics.

The key to this tutorial is model vs market, or output vs outcome mindset — the ability to separate these, and optimise for both.



Section III

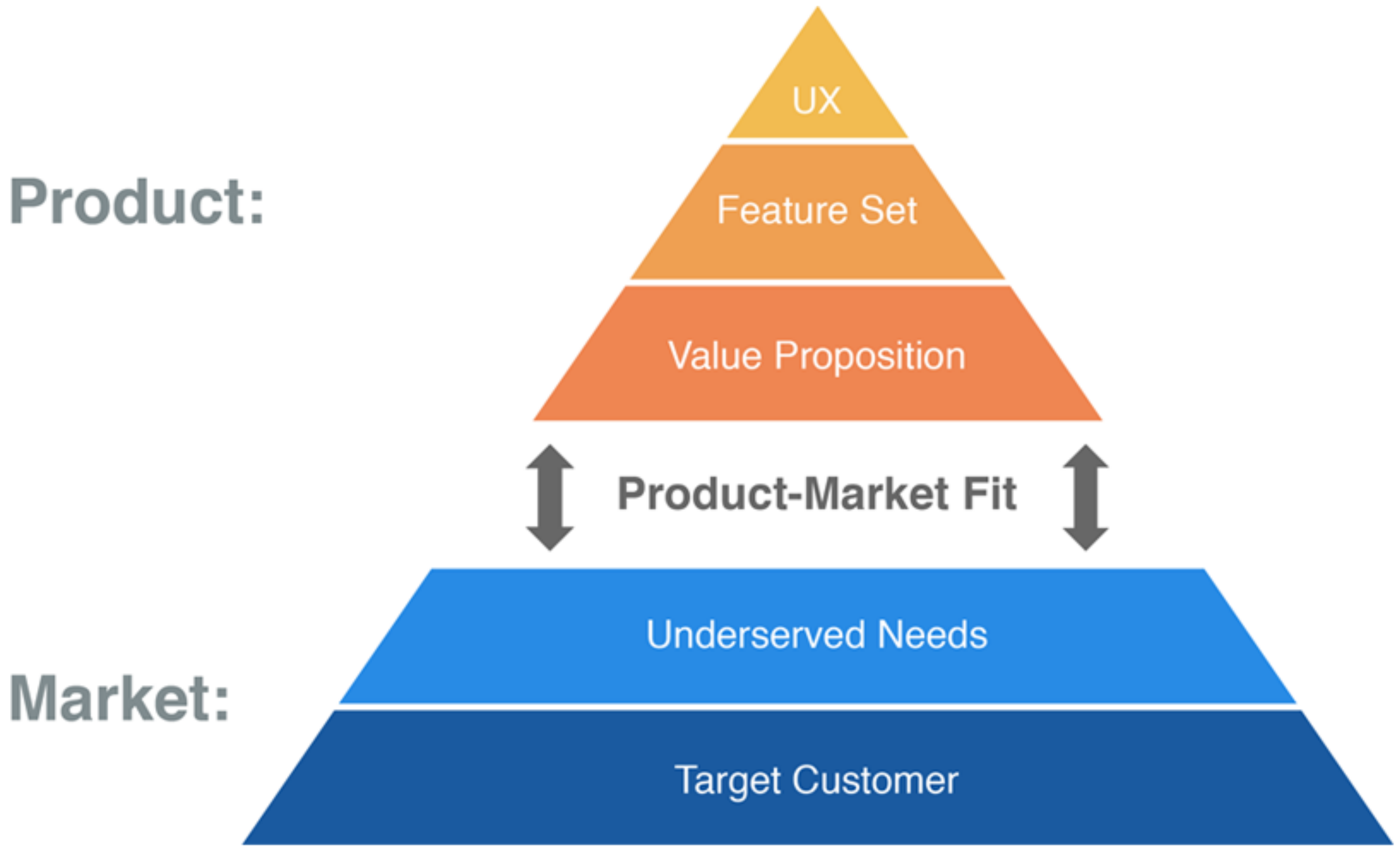
Product-market fit

What is product-market fit?

A market is a set of related customer needs.

A product is a specific offering intended to meet a set of customer needs.

Product-market fit is the measure of how well a product satisfies the market.



https://leanproductplaybook.com

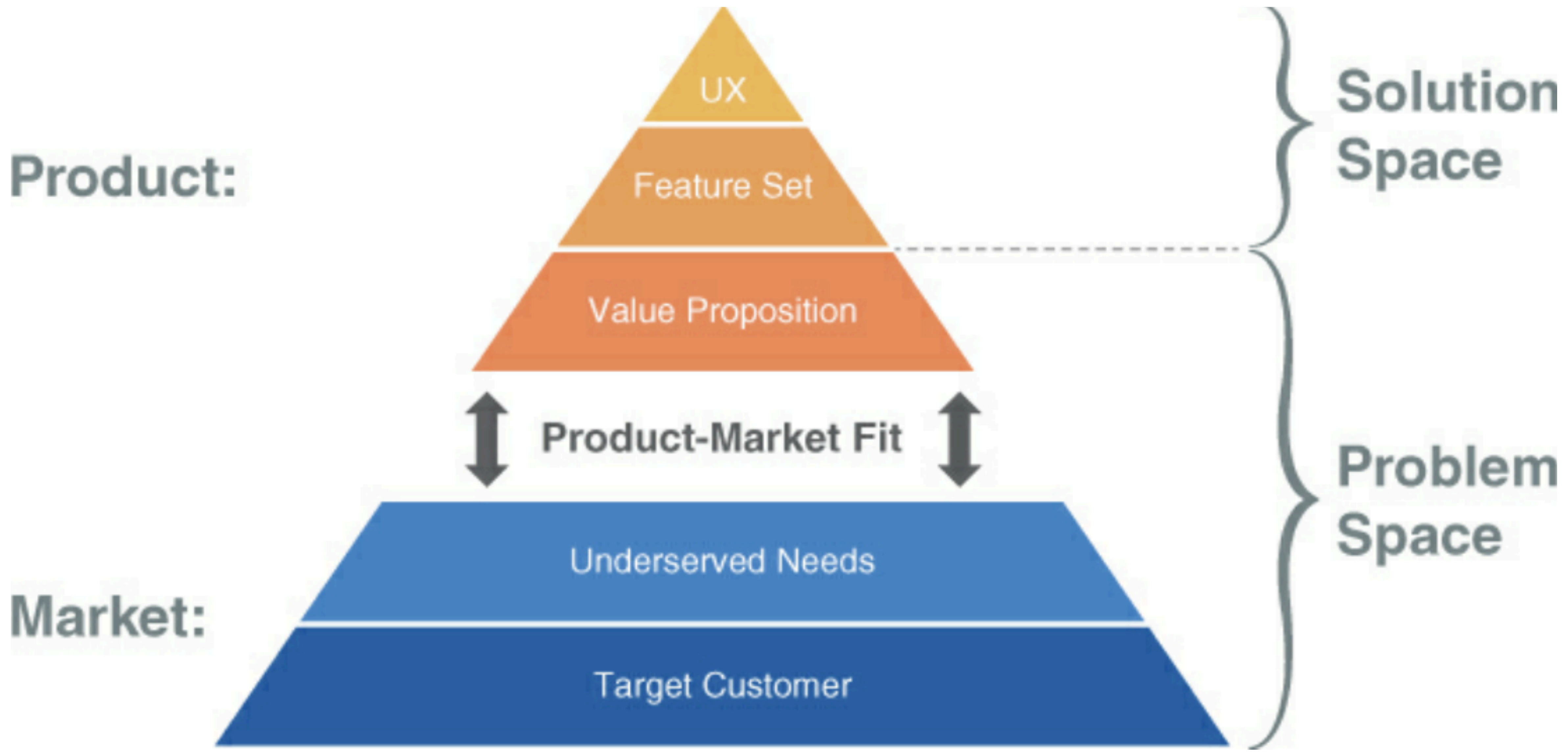
The main reason products fail is because they don't meet customer needs in a way that is better than other alternatives.

Products we build will exist in the solution space.

Problems, on the other hand, define the market.

A common discovery process is using the solution space to discover the problem space. That I, keeping problem space and solution space separate and alternating between them is one of the best ways to achieve product-market fit.

Iteratively testing and improving the product's hypotheses is preferred over prematurely jumping to the solution space.



"Customers don't care about your solution. They care about their problems." Dave McClure

First things first: Who is your customer?

There are many ways to define the customers; for instance, one can segment the users' market based on demographics, behaviour, needs, ...

Another approach is to define personas; this means making many assumptions in order to simply the overall market and hence has the risk of leading the decisions without GOOB.

Note that, there can be differences between users vs buyers. Therefore, it is important to make that distinction early in the process, when applicable.

ML Scientist

Great at ML maths,

good at programming

ML Engineer

Good at ML maths,

Great at programming

Data Scientist

Great at nothing

Good at everything (?)

Data Engineer

Great at data,

Good at system design

Identify underserved customer needs.

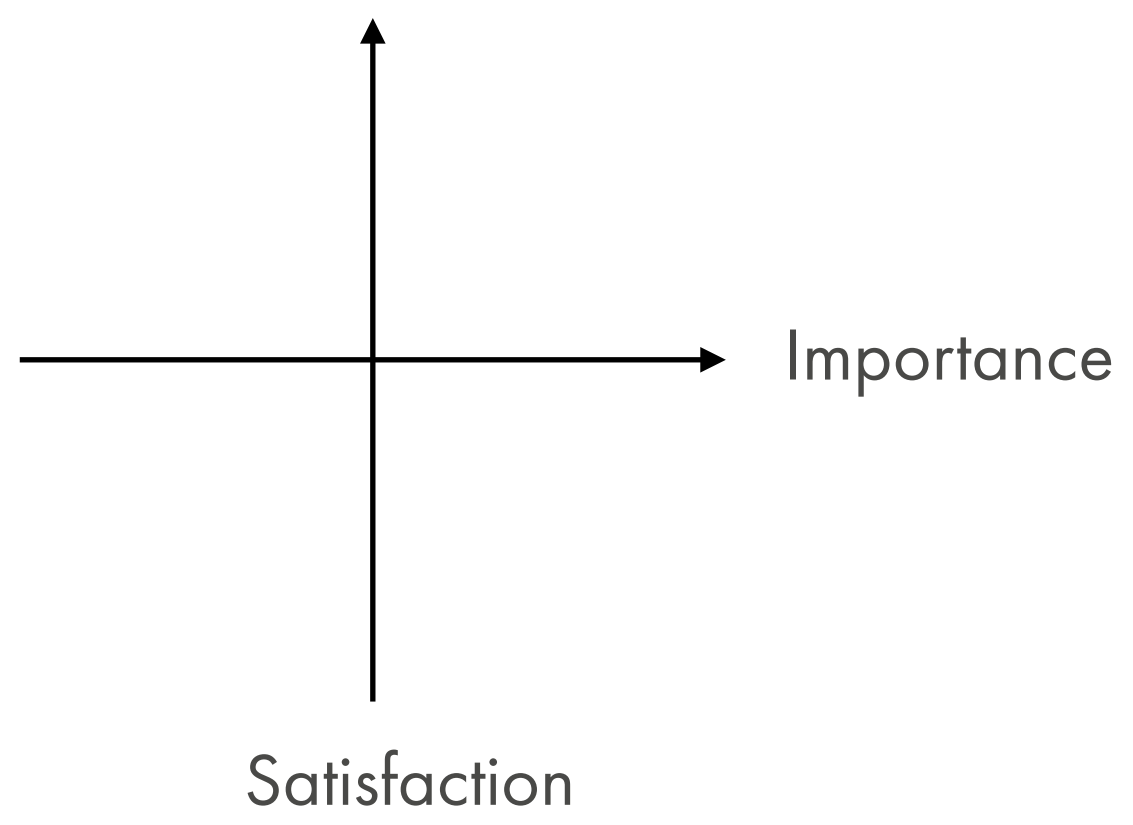
A good discovery process aims to illuminate the problem space, despite customers being much more aware, articulate and opinionated (and potentially helpful?) in the solution space.

There are frameworks, such as Gap = importance - satisfaction that can help decide how to prioritise the value proposition to the customers; the bigger the gap the more underserved the need/opportunity.

Customer Benefit	Typical Customer Comment
1. Help me prepare my tax return	"I don't really know much about taxes. I try to follow the instructions but they're confusing. I'm not sure which forms I should be filling out."
2. Check the accuracy of my tax return	"I'm not that great at math, so I know I'm probably making several mistakes when I'm adding and subtracting all those numbers on my tax forms."
3. Reduce my audit risk	"I'm worried about being audited but don't really know how risky my tax return is. It would be great to know if it would raise any yellow flags with the IRS so I could fix those parts."
4. Reduce the time it takes me to enter my tax information	"I spend lots of time each year entering data from all the tax forms I receive from my employer, bank, and brokerages."
5. Reduce the time it takes me to file my taxes	"I normally print my tax return and then go to the post office, wait in line, and mail it so I can get delivery confirmation. It would be great if I could avoid that hassle."
6. Maximize my tax deductions	"I don't know about all the deductions that I'm eligible to take. I'm probably leaving some money on the table."

Can ML predict the P(audit)?

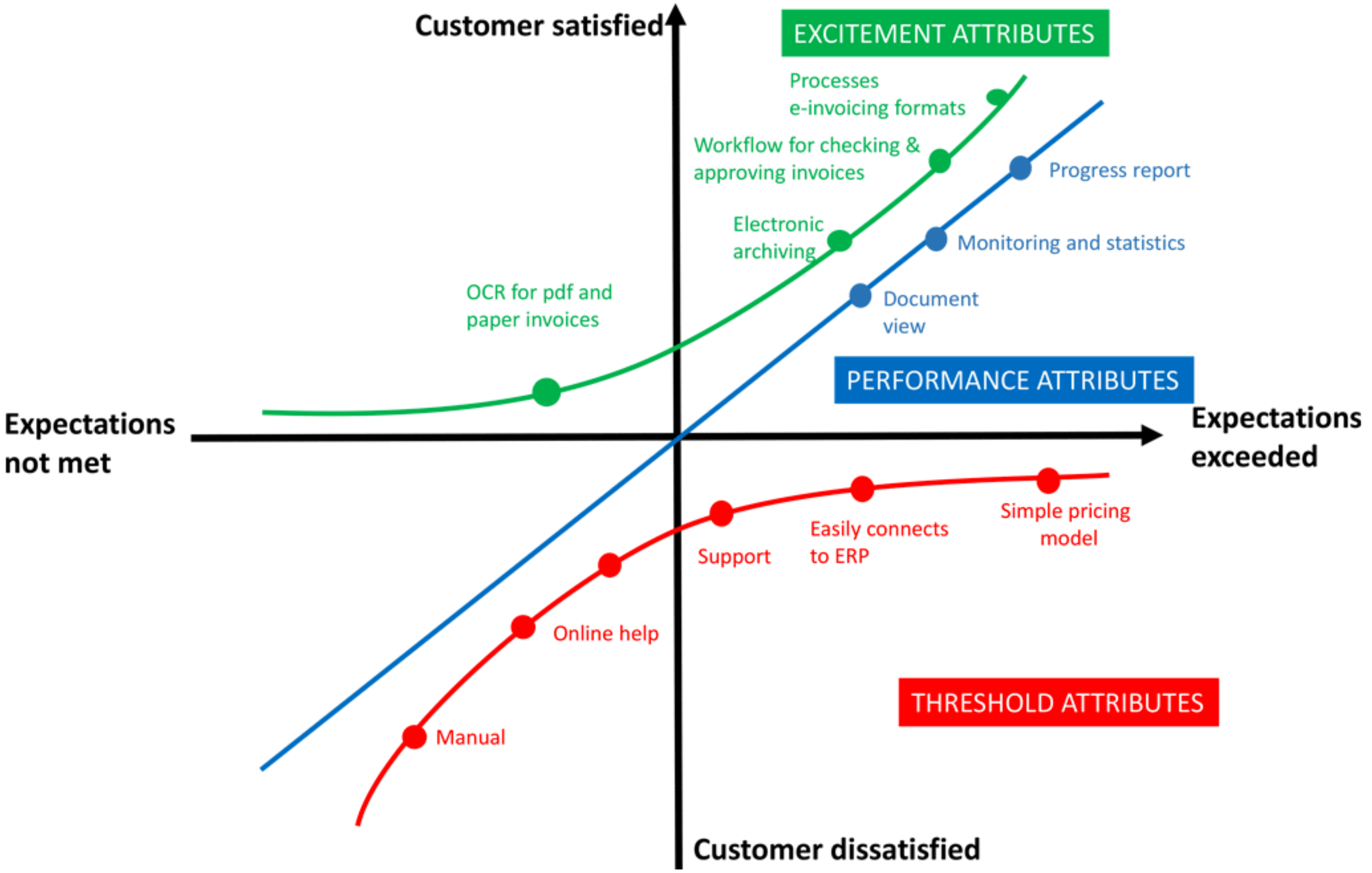
Can ML automate some tasks?



Given the problem-space insights, define the value proposition

Based on a clearer understanding of the users and their problems (that need to be solved), many value propositions and hence features can be ideated.

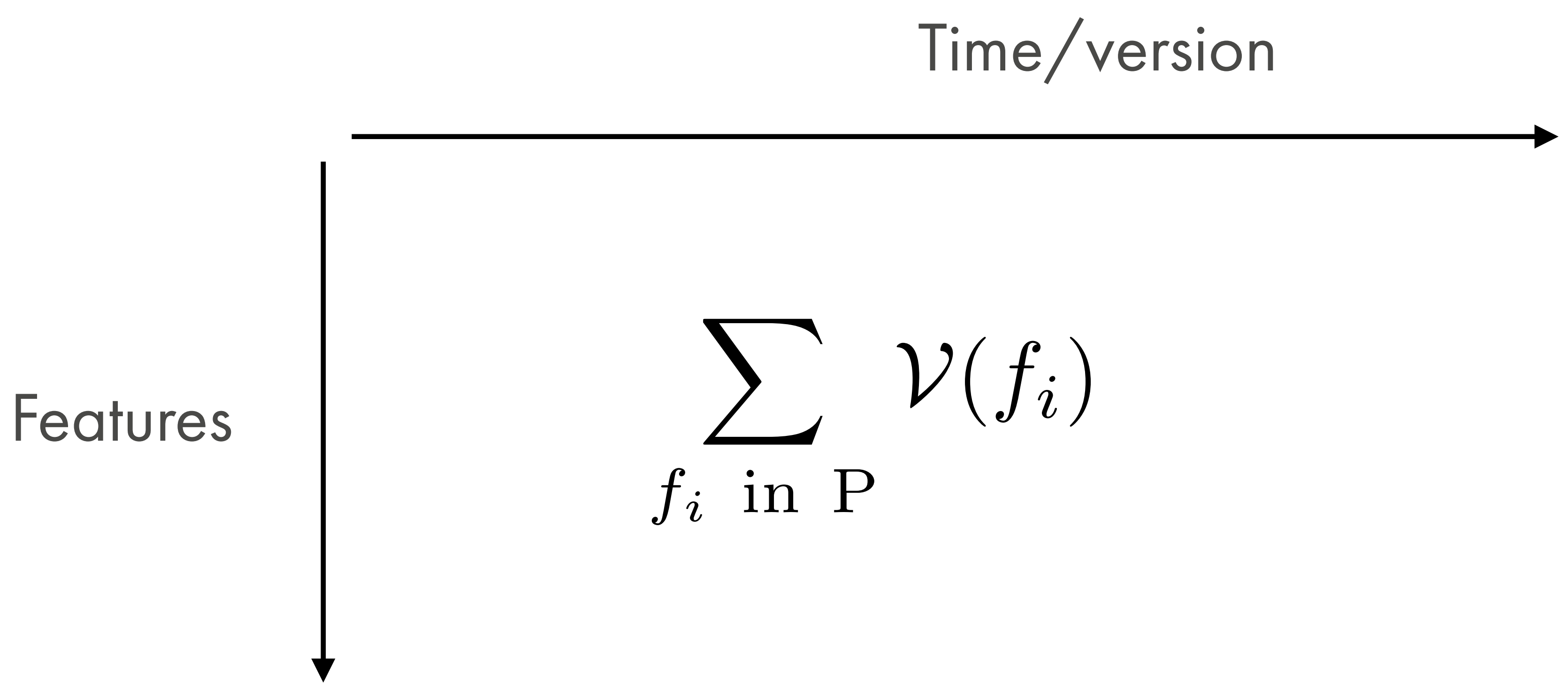
A useful structuring/organising framework for the prioritisation process can be the Kano model; this will lead to the needs being classified as must-haves, performance benefits, or delighters.



What are your MVP (i.e., minimum viable product) hypotheses? What features/value does it offer?

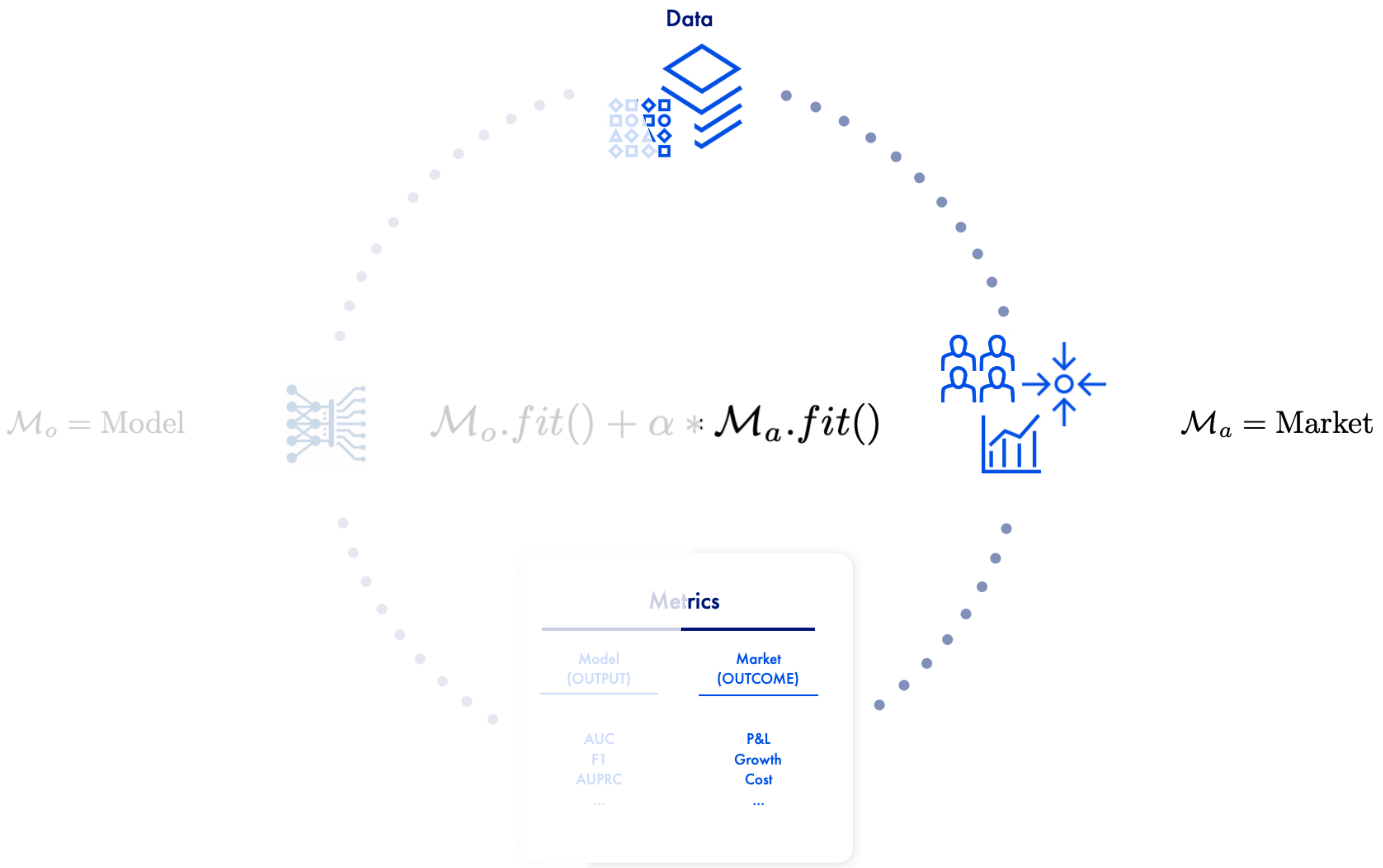
Given the list of value in mind, we can design a product value proposition template, which can summarise various MVP hypotheses, and competitors in one place — based on this, we can prioritise.

	Competitor A		**My Product**	
	Now	**In 1 Year**	**Now**	**In 1 Year**
Must-Haves				
Must-have 1	Y	Y	Y	Y
Must-have 2	Y	Y	Y	Y
Performance Benefits				
Performance benefit 1	High	High	Medium	**High**
Performance benefit 2	Medium	**High**	Low	Low
Performance benefit 3	Low	**Medium**	High	High
Delighters				
Delighter 1	Y	Y		
Delighter 2			Y	Y
Delighter 3		**Y**		
Delighter 4				**Y**



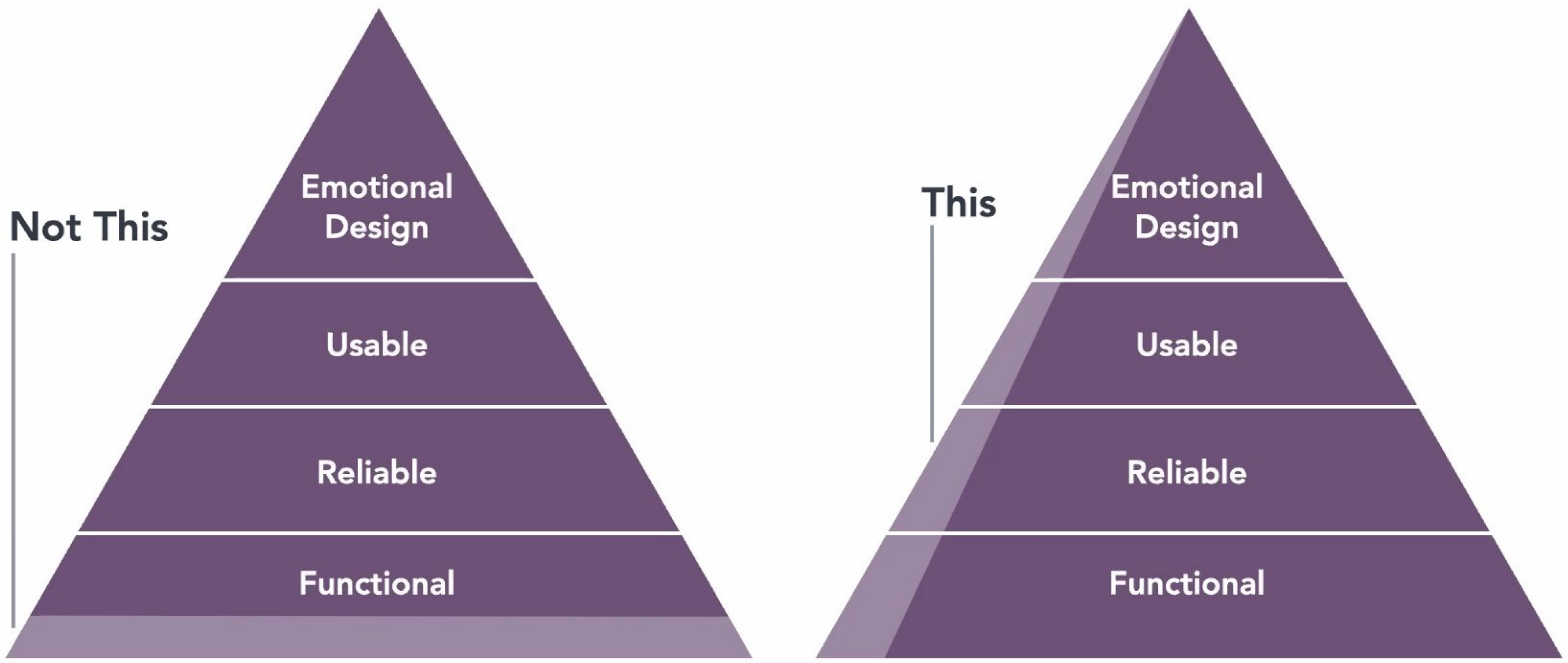
A quick recap: So far, you haven't built much and rather have been exploring the problem space and hypothesising potential solutions.

- Formed hypotheses about your target customersFormed hypotheses about their underserved needs
- Articulated the value proposition you plan to pursue so that your product is better and different
- Identified the top feature ideas you believe will address those needs and broken them down into smaller chunks
- Prioritised those feature chunks based on ROI
- Selected a set of those feature chunks for your MVP candidate/hypothesis, which you hypothesize customers will find valuable



It's now time to build — starting with the MVP (minimum viable product) prototype.

Many debate what does and does not qualify as MVP; while the terms minimum gets a lot of attention, it being a viable product is also important. An MVP should not be viewed as minimally functional parodic and rather a combination of all key attributes.



The process of testing MVP hypotheses (and lean mythology) is very close to scientific methodology.

Tests can be both qualitative and quantitative:

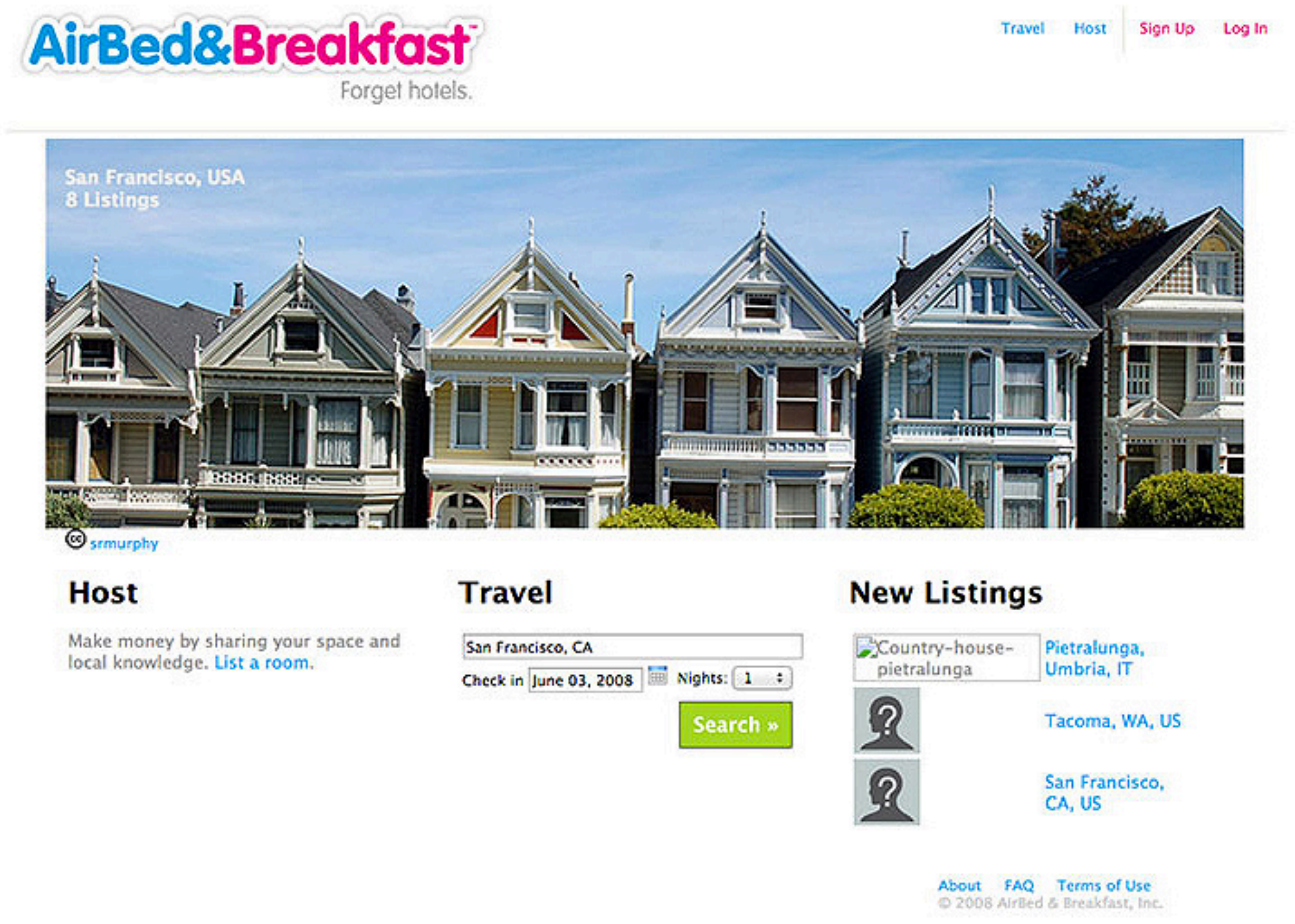
We will talk more about this.

e Wizard of Oz MVP and concierge MVP allow you to actually test your live product or service; but instead of the final version, you are using manual workarounds — common for AI

	Qualitative Tests	**Quantitative Tests**
Marketing Tests	Marketing materials	Landing page/Smoke test Explainer video Ad campaign Marketing A/B tests Crowdfunding
Product Tests	Wireframes Mockups Interactive prototype Wizard of Oz & Concierge Live product	Fake door/404 page Product analytics & A/B tests

An MVP example: Airbnb

One the hypotheses behind Airbnb idea was that property listings with professional photos would get more business. Here, they manually recruited the hosts, and photographers. Next, they manually uploaded them to the website, as if the real product was live.



Another MVP example: Uber

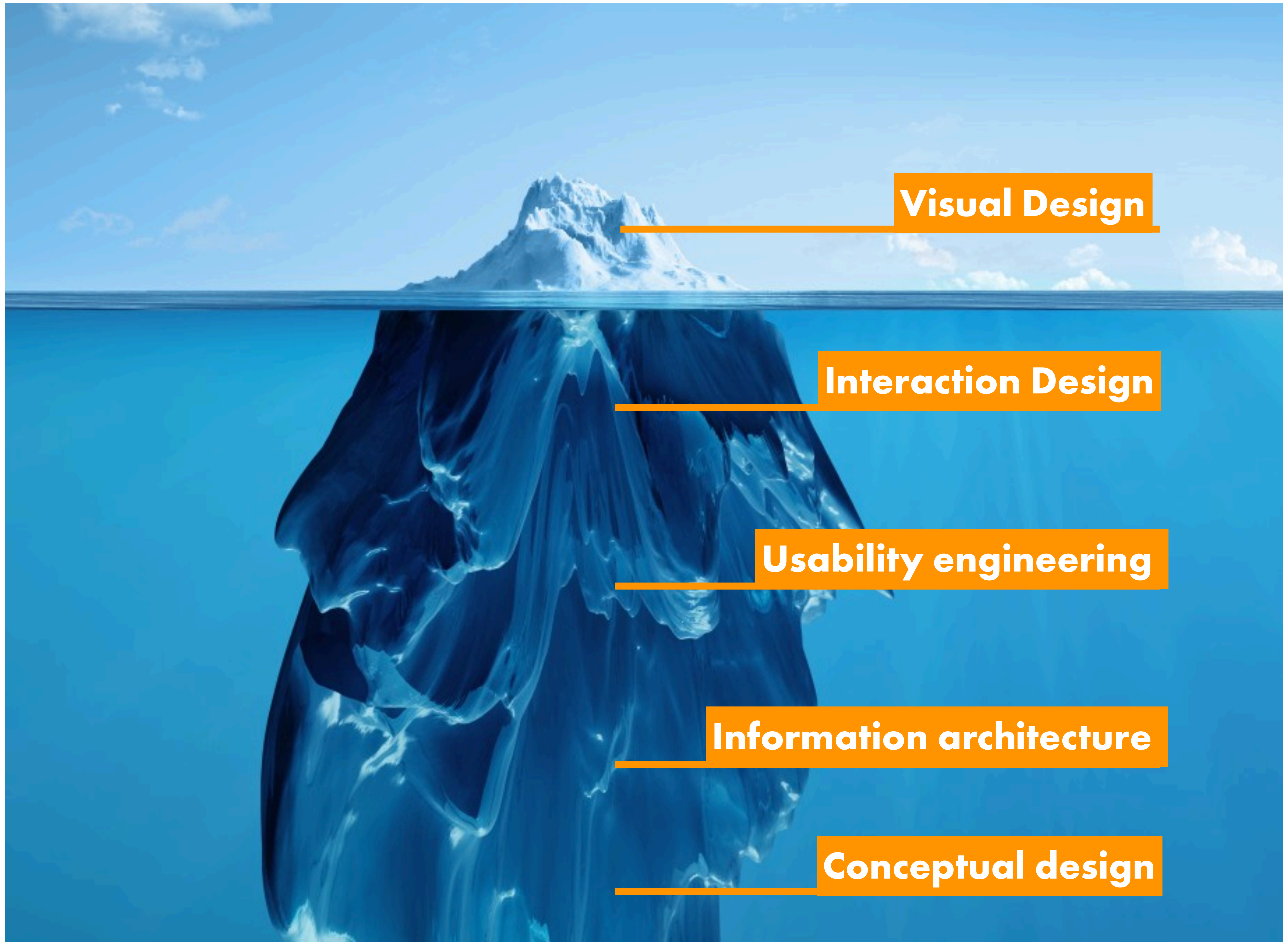
Uber's MVP, back in 2010, did one simple thing: It connected drivers with iPhone owners in San Francisco who were happy to allow the app to pay on their behalf, using their credit card.

It was enough to fulfil their main goal of offering: Taxi services where and when you need, with little to no friction.

Now, the product can be tested with actual customer — build, measure, and learn loop (+ UX design)

It's crucial as you conduct your user tests to differentiate between feedback on usability versus product-market fit. Feedback on usability has to do with how easy it is for customers to understand and use your product, whereas feedback on product-market fit has to do with how valuable they find your product.

You should consider pivoting if you just don't seem to be achieving gains in product-market fit after several rounds of trying to iterate. If, despite your best efforts, your target customers are only lukewarm on your MVP, you should consider a pivot.



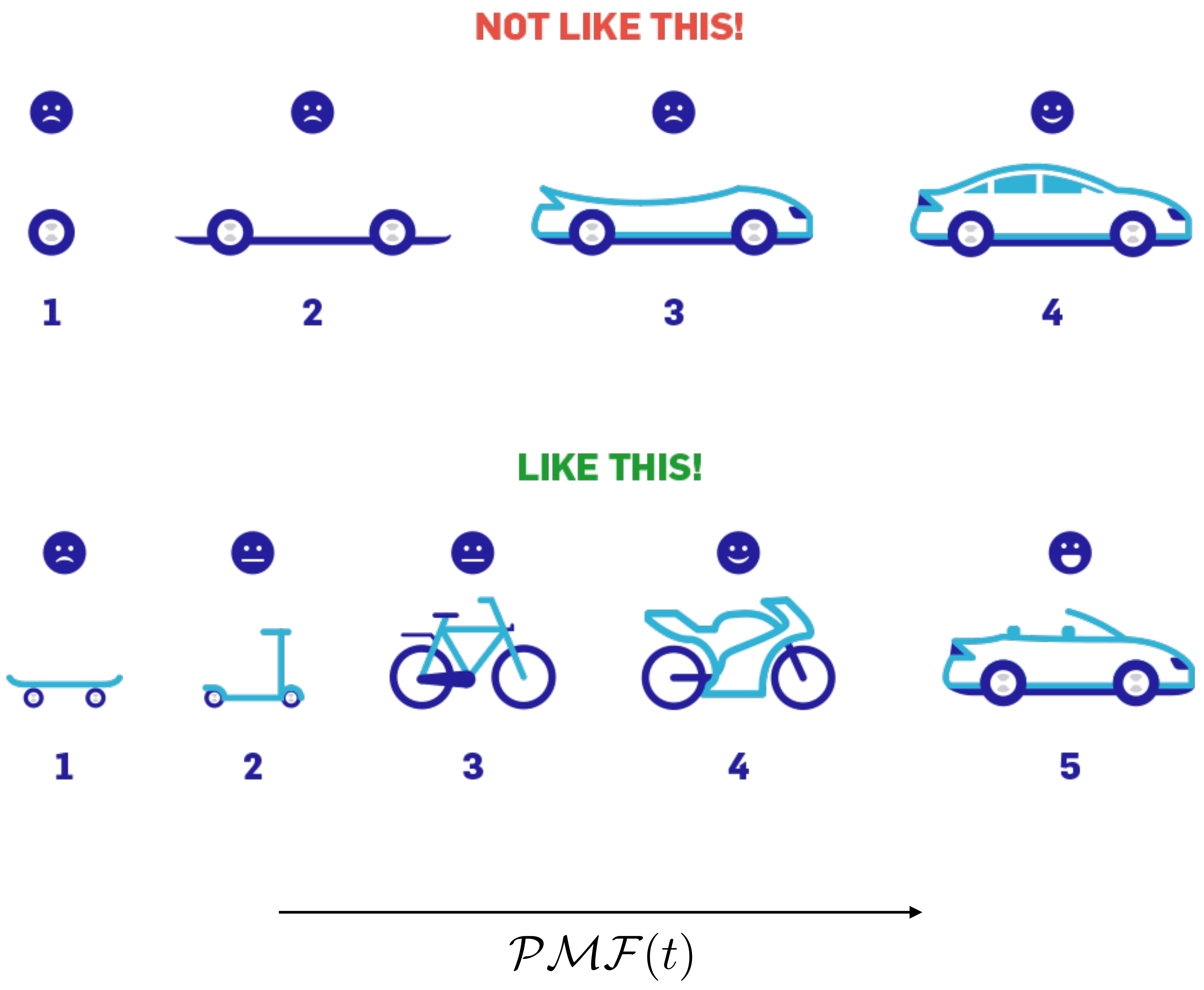
Build, measure, learn

In biomedicine, the "measure, model, perturb, repeat" loop is a commonly accepted norm.

Given the so-called market in most application domains of AI is not computationally irreducible, the only way to know its reaction to our idea/product (hence our product's fit to it) is to "perturb/intervene and measure".

"If you are not embarrassed by the first version of your product, you've launched too late."

Reid Hoffman

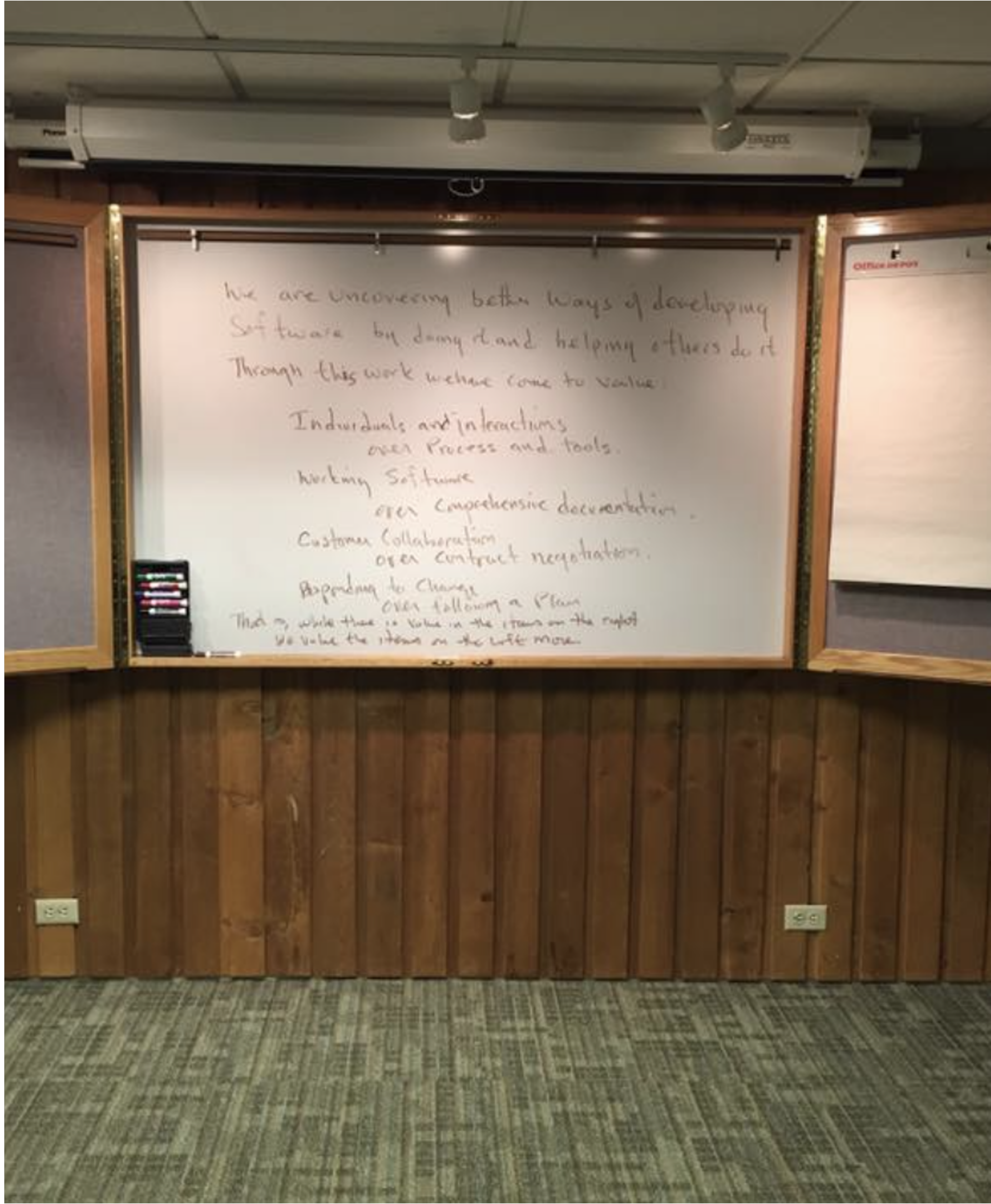


Agile encourages early and continuous delivery of working software that is valuable for customers.

"We are uncovering better ways of developing software by doing it and helping others do it. Through this work we have come to value:

- Individuals and interactions over processes and tools
- Working software over comprehensive documentation
- Customer collaboration over contract negotiation
- Responding to change over following a plan"

http://agilemanifesto.org



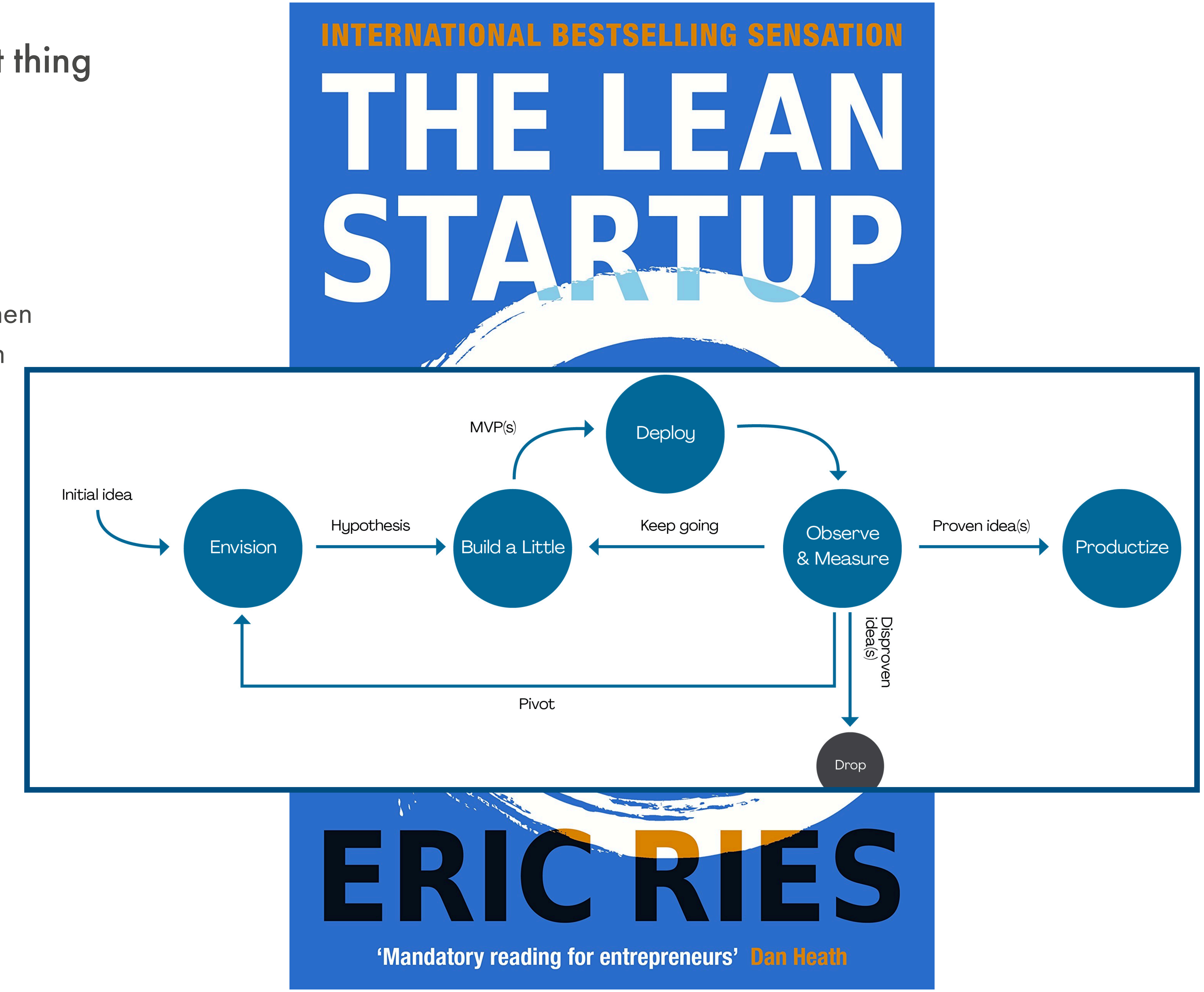
12 Principles of Agile

1. Our highest priority is to satisfy the customer through early and continuous delivery of valuable software.
2. Welcome changing requirements, even late in development. Agile processes harness change for the customer's competitive advantage.
3. Deliver working software frequently, from a couple of weeks to a couple of months, with a preference to the shorter timescale.
4. Business people and developers must work together daily throughout the project.
5. Build projects around motivated individuals. Give them the environment and support they need, and trust them to get the job done.
6. The most efficient and effective method of conveying information to and within a development team is face-to-face conversation.
7. Working software is the primary measure of progress.
8. Agile processes promote sustainable development. The sponsors, developers, and users should be able to maintain a constant pace indefinitely.
9. Continuous attention to technical excellence and good design enhances agility.
10. Simplicity--the art of maximising the amount of work not done--is essential.
11. The best architectures, requirements, and designs emerge from self-organising teams.
12. At regular intervals, the team reflects on how to become more effective, then tunes and adjusts its behaviour accordingly.

Experiment in order to discover/build the right thing that customers are willing to pay for.

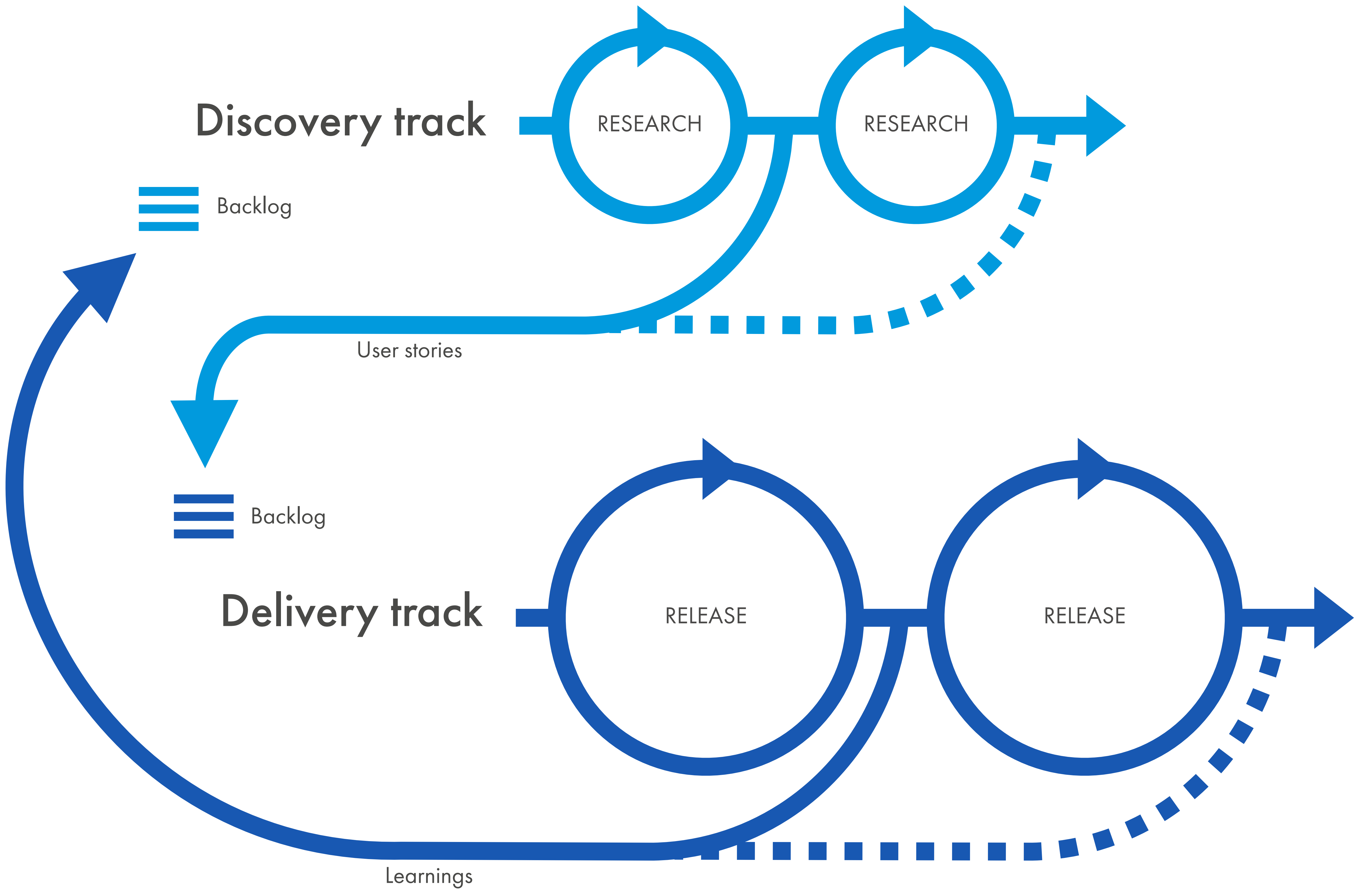
A project is not complete after you ship a version — as only then you receive user feedback from real-world interactions, which enables you to refine the goals for the next iteration. As a consequence — you should strive to execute very short iterations focused on learning what the user wants.

Everything is a grand master experiment.



Successful products benefit from vision, strategy, objectives, and discovery and delivery

- Discovery work focuses on fast learning and validation with the same agile and lean principles in mind.
- Discovery is a necessary part of product development; its progress should be visible to the whole team.
- If we're doing discovery right, we substantially change and kill lots of ideas.
- Keep measuring and learning even after you ship.



Section V

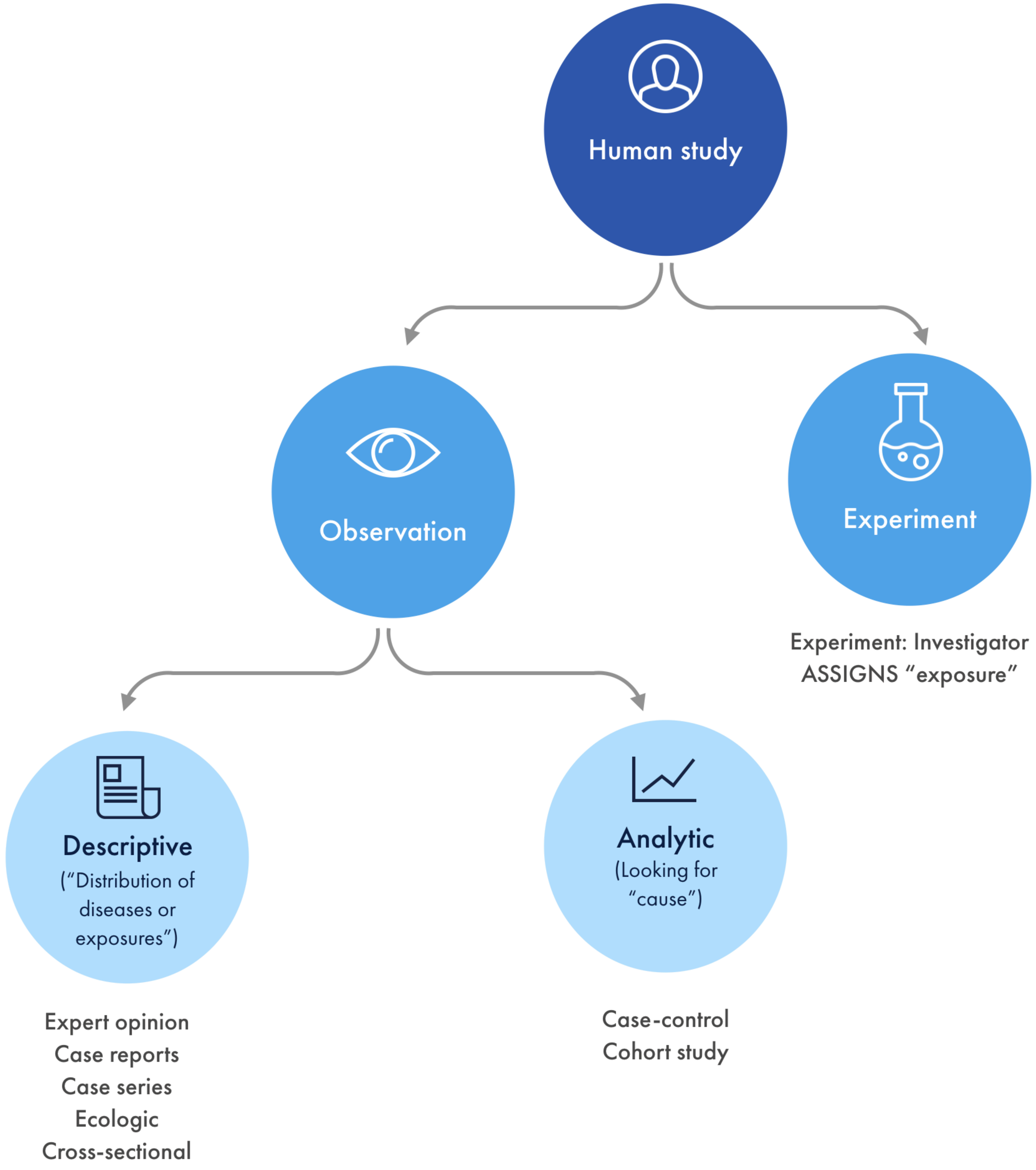
Experiments and metrics

The field of medicine has advanced the topic of measuring evidence in various real-world setting.

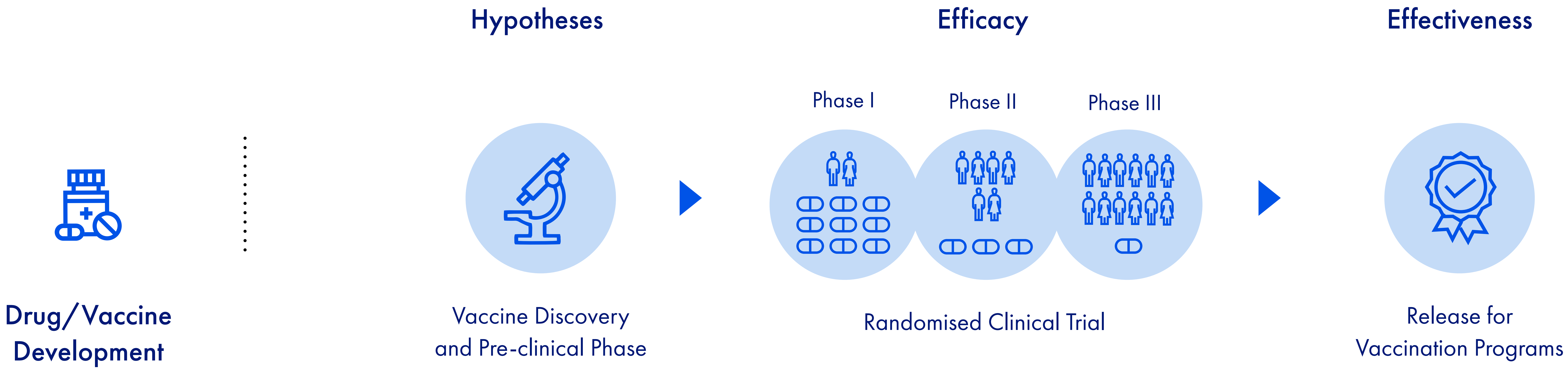
Scientific reasoning consist of a set of reasoning processes that permeate the field of science: induction, deduction, experimental design, causal reasoning, concept formation, hypothesis testing, and so on.

"To consult the statistician after an experiment is finished is often merely to ask him to conduct a post mortem examination. He can perhaps say what the experiment died of."

Ronald Fisher



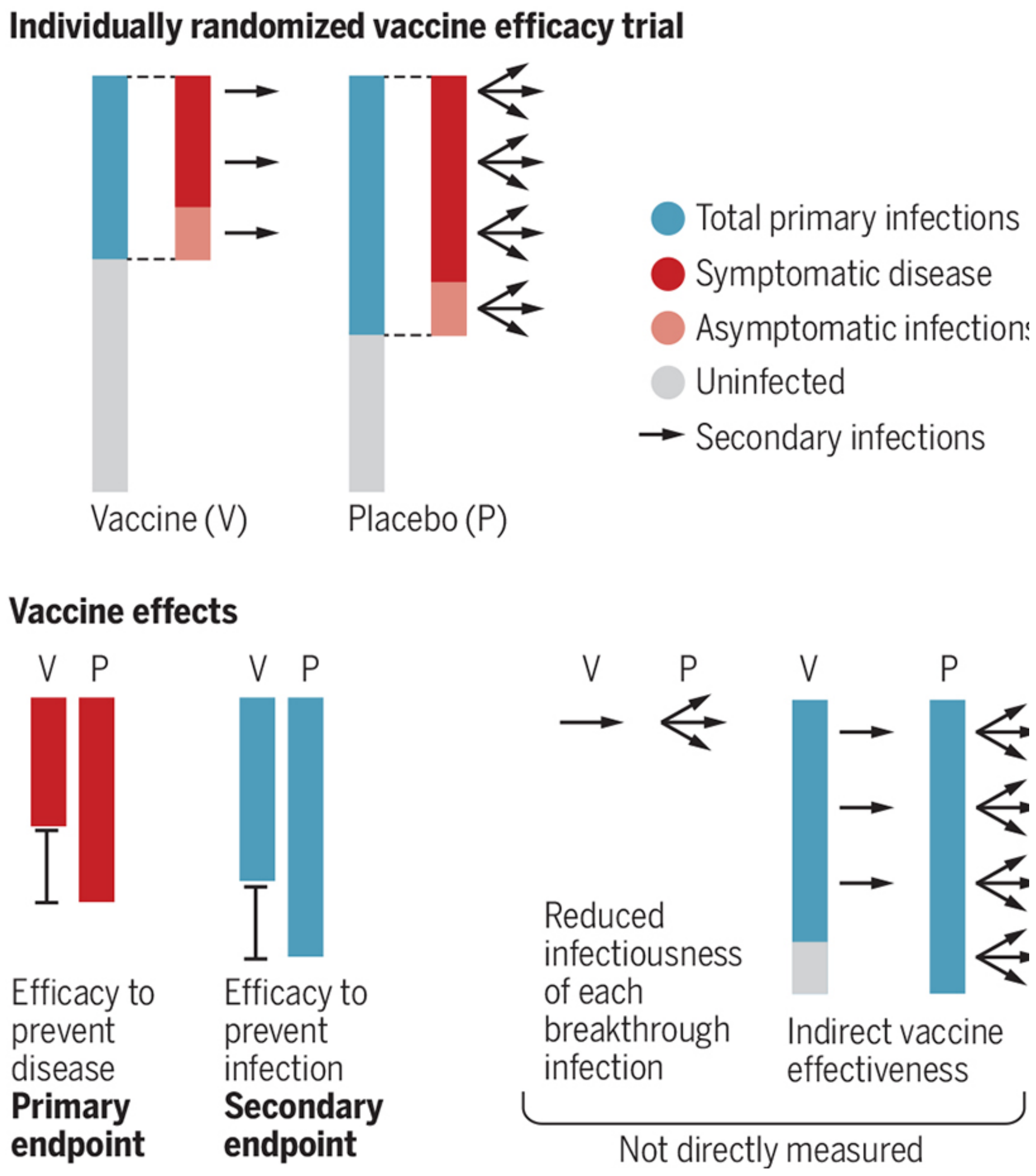
Different disciplines in science have devised different approaches to answer research questions.



In interventional studies, investigator assigns the exposure/intervention (e.g., COVID vaccines)

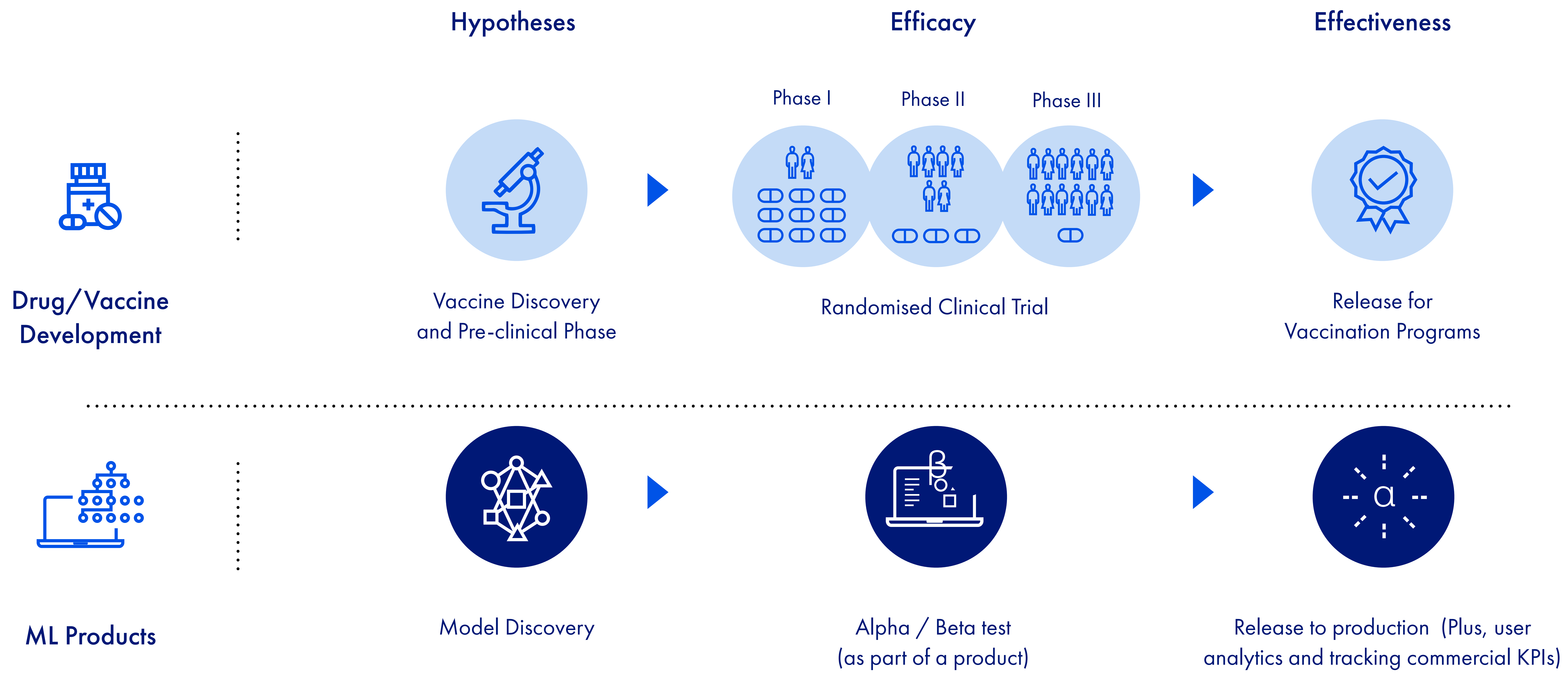
Randomised controlled trials (RCT), refers to a scenario, where a group of participants fulfilling certain inclusion and exclusion criteria is "randomly" (i.e., each participant has an equal chance of being allocated to the two groups) assigned to two separate groups, each receiving a different intervention.

There are additional interventional study designs such as nonrandomised controlled clinical trials, Interventional studies without concurrent controls, before–after (or, pre–post) studies, factorial study design, crossover study design, and cluster randomised trials.



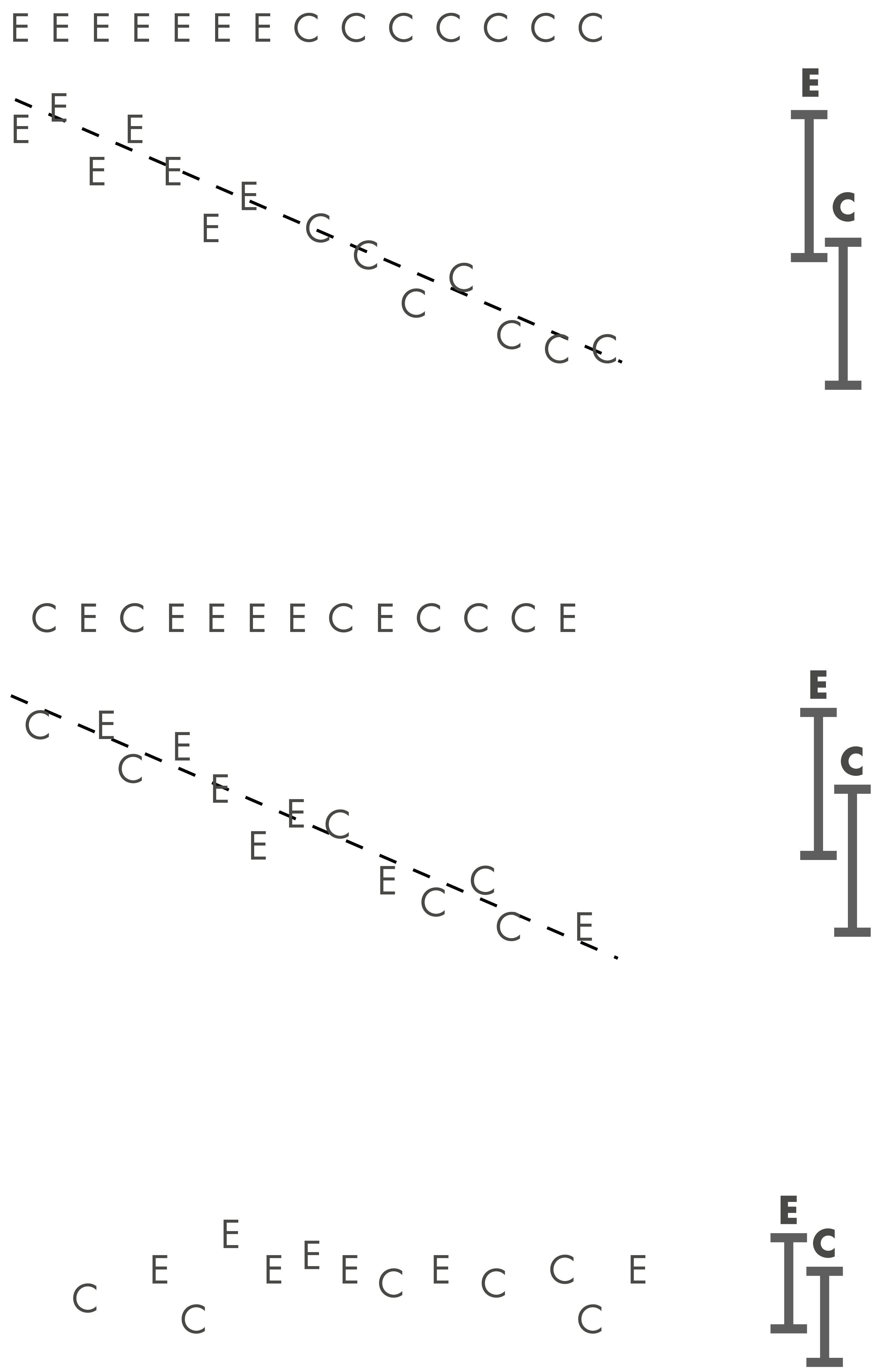
Lipsitch & Dean, 2020

Different disciplines in science have devised different approaches to answer research questions



We should ideally randomise; when not possible, we can introduce the covariates and control.

Imagine where we have an instrument drift that is a time-correlated confounder, which is common in ML and other domains. For instance, a trading signal loosing its edge over time, or a virus mutating over time, ...



Running an interventional study is not always an option; hence the need for observational studies.

Observational studies are those where the researcher is documenting a naturally occurring relationship between the exposure and the outcome that he/she is studying. The researcher does not do any active intervention in any individual, and the exposure has already been decided naturally or by some other factor; ; it is also referred to as natural experiment. Depending on the direction of enquiry, these studies can be directed forwards (cohort studies) or backwards (case–control studies).

Perspect Clin Res. 2019 Jan-Mar; 10(1): 34–36.
doi: 10.4103/picr.PICR_154_18
PMCID: PMC6371702
PMID: 30834206

Study designs: Part 2 – Descriptive studies

Rakesh Aggarwal and Priya Ranganathan[1]

Perspect Clin Res. 2019 Apr-Jun; 10(2): 91–94.
doi: 10.4103/picr.PICR_35_19
PMCID: PMC6463505
PMID: 31008076

Study designs: Part 3 - Analytical observational studies

Priya Ranganathan and Rakesh Aggarwal[1]

› Eur Heart J. 2017 Feb 1;38(5):326-333. doi: 10.1093/eurheartj/ehw411.

Influenza vaccination and risk of hospitalization in patients with heart failure: a self-controlled case series study

Hamid Mohseni [1], Amit Kiran [1], Reza Khorshidi [1], Kazem Rahimi [1] [2]

It is important to aim for an efficient design, which can capture the outcome of interest

While in some applications the use of more observations (i.e., larger N) is an option, in many others, the experiment needs to be run very efficiently. There are many axes along which an experiment can be optimised, in order to achieve reliable conclusions without necessarily increasing the sample size.

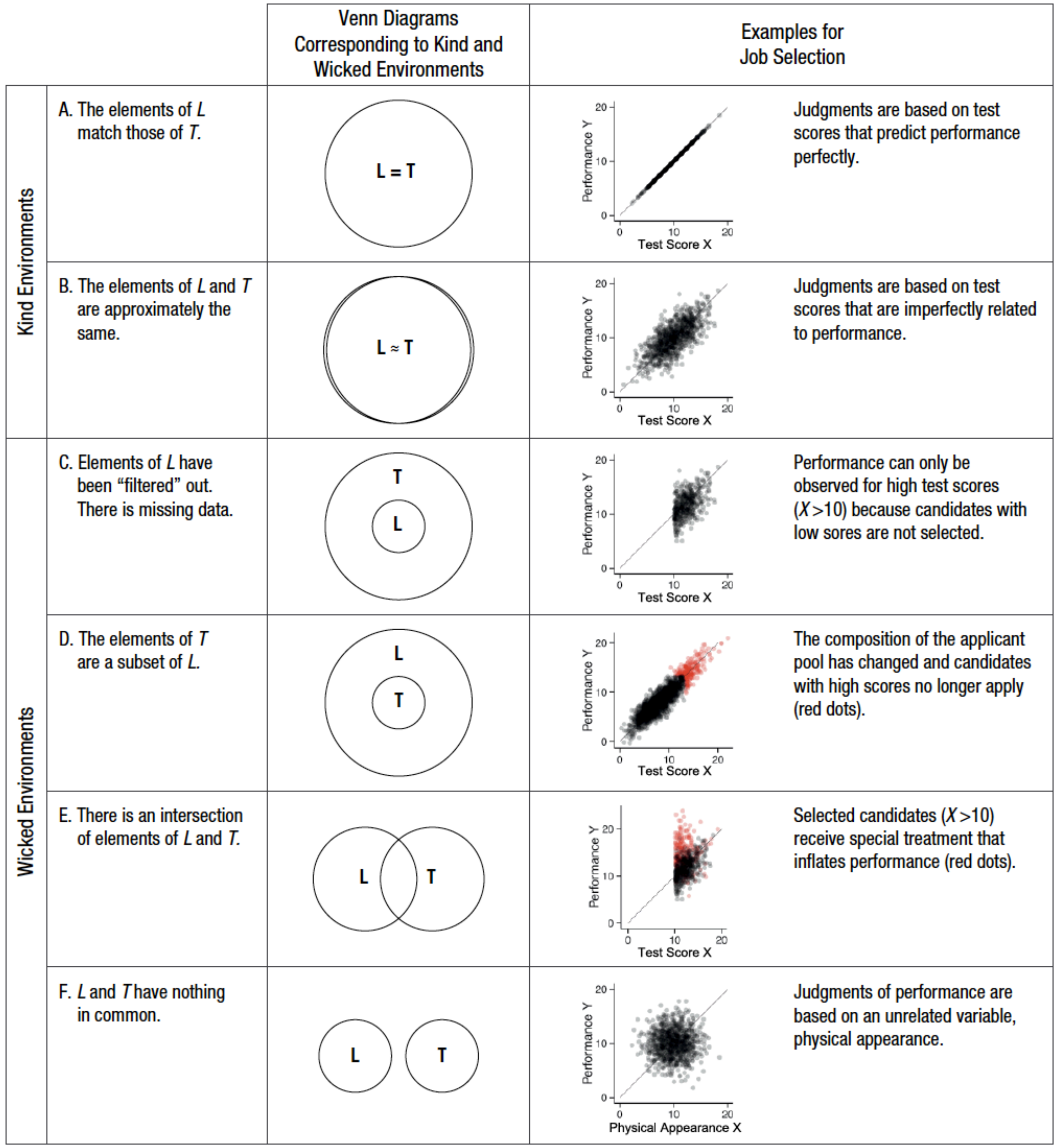
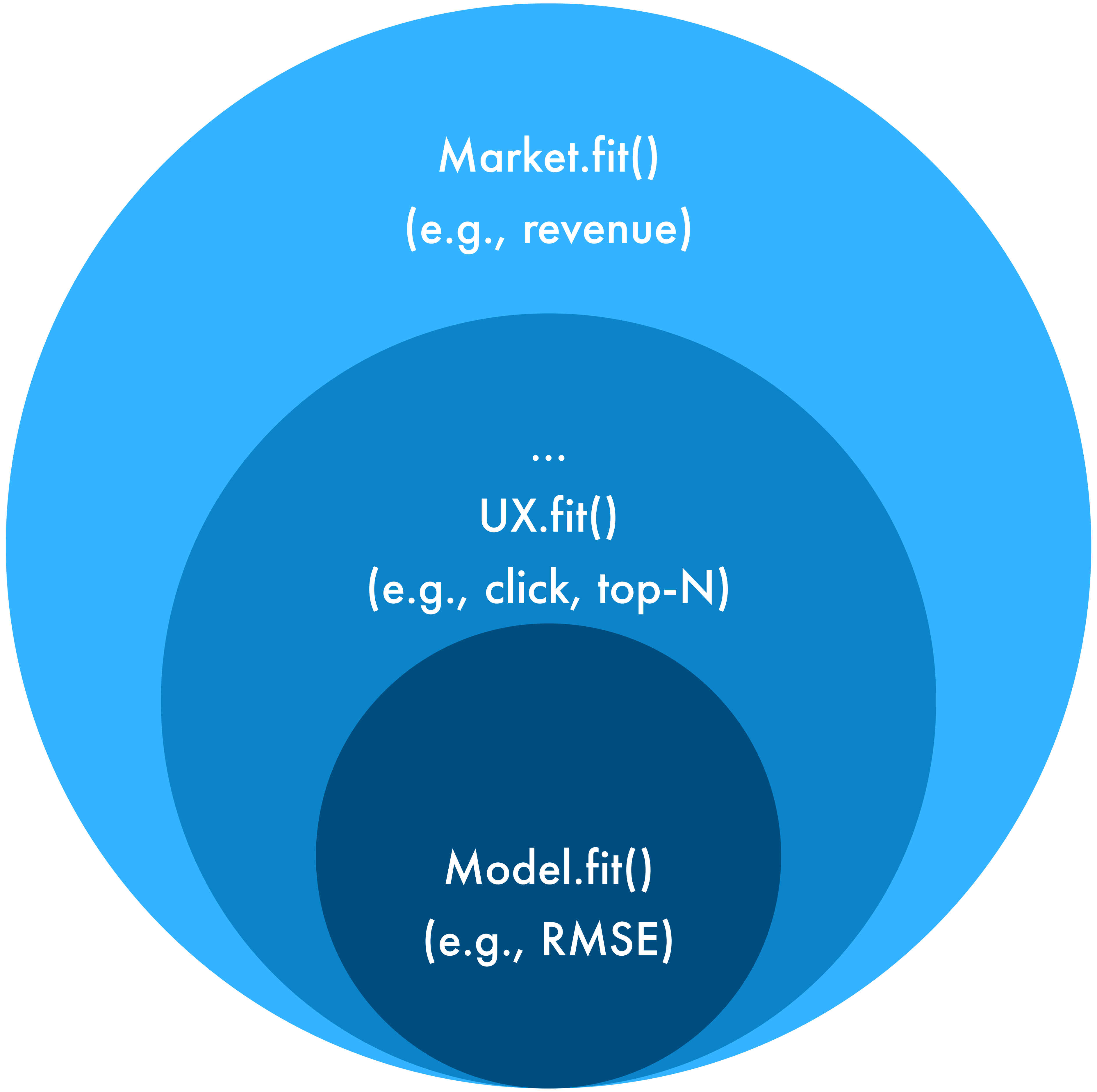


Fig. 1. The two-settings framework. On the left, we show six ways in which the elements of information in the learning setting (L) and the target setting (T) do or do not match. On the right, we show an example scenario involving job selection.

Hogarth et al 2015

Micro, meso and macro view on metrics — Model.fit(), UX.fit(), and Market.fit()

As discussed at the beginning of the tutorial, when doing ML for impact, one operates in a nested ecosystem of goals. For instance, when building a recommender system, model.fit() is done on a metric like KL divergence or RMSE. However, top-N accuracy, precision, recall, or FDR, for instance, are more likely to capture the quality of how accurate the results will feel to the users (depending on the choice of UX). On the other hand, the app/co might care about the click rate, revenue, LTV, or other metrics.



Correlation between ROUGE and Human Evaluation of Extractive Meeting Summaries

Feifan Liu, Yang Liu

Thanks!

@RezaKhorshidi