

KANTAR

Learning from the past as we head towards our AI driven futures

Jon Puleston
VP, Innovation
Kantar Profiles



Viral marketing

Media Research Group Conference 2002

“We won’t need to buy advertising in the future, you just need to work out how to make your coms go viral”



Protestations on AI...

“We won’t need to do traditional research anymore”

Good Judgement Project

Longitudinal forecasting tournament of global events

28,000 took test → 3,000 invited to participate → Top 100 = Superforecaster



Philip Tetlock

Experience of 3 years taking part in the Good Judgement Project, pitted against smartest future forecaster's in the world trying to predict things.

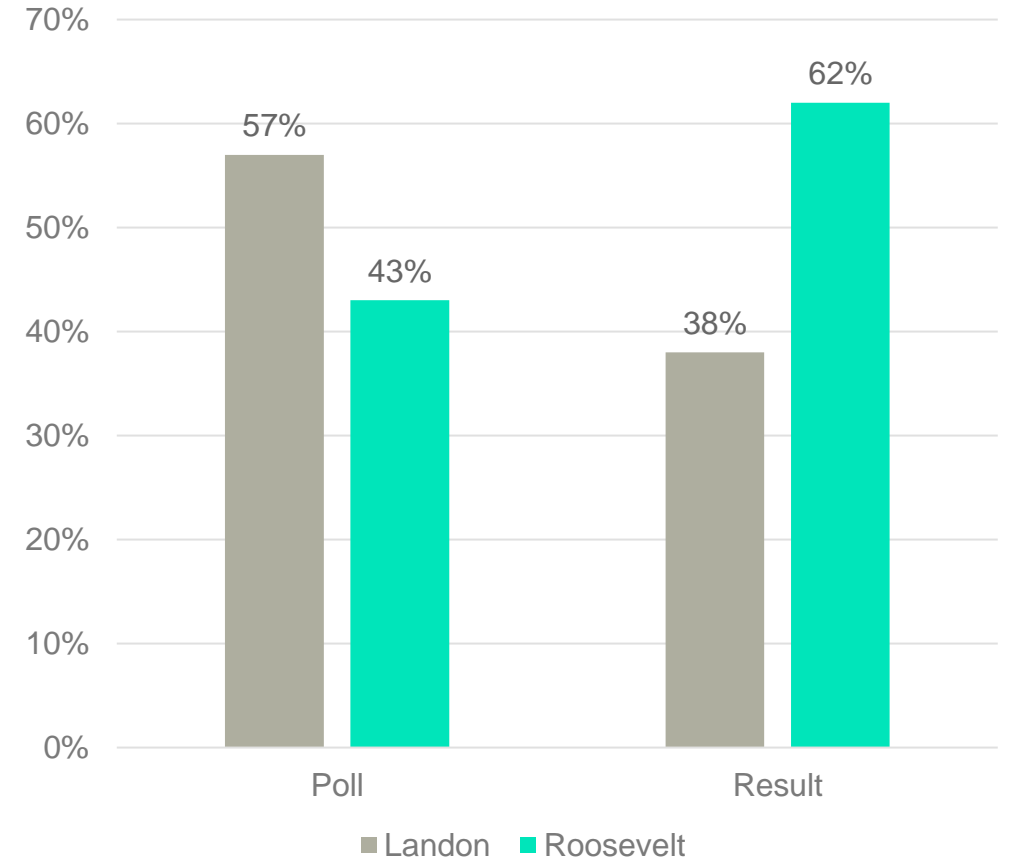
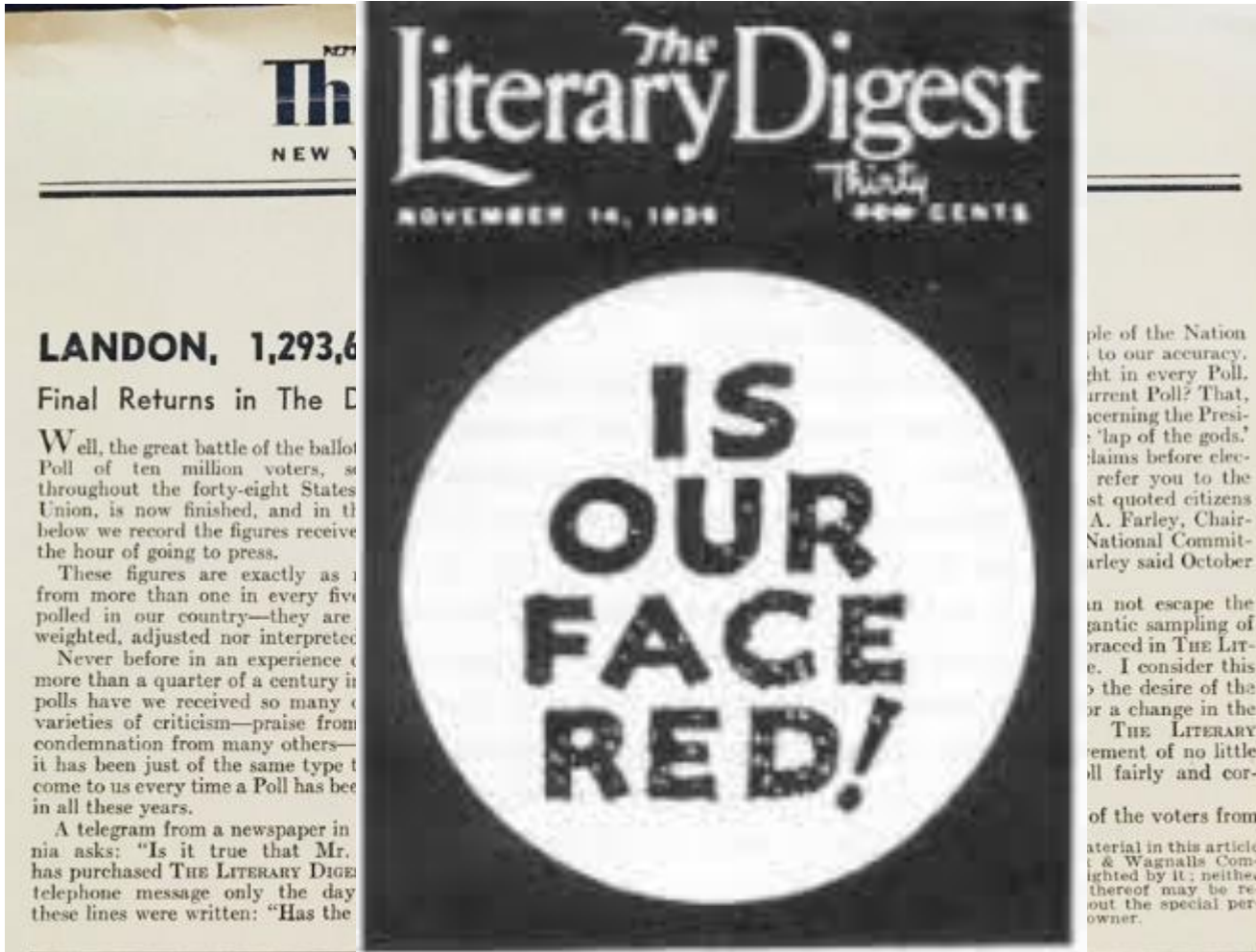
Came 102nd

It's not on impossible to predict certain things, rarely turns out how you expect & the best predictors turn out to be the ones that change their minds the most

**A common thought for nearly a century of research
that a sea change is in the air...**

1930's

We won't need to do traditional research anymore: just ask our readers



1950's

We won't need to do any traditional research anymore: just find out what consumers unconscious desires are



Ford invest £250m developing this car using a fashionable new methodology:
“Motivational Research”

Inspired by emerging psychoanalytical theories. The belief that consumers were motivated by unconscious desires and emotions, rather than rational decision-making.

Cars designed around consumers “sexual fantasies”

They could not

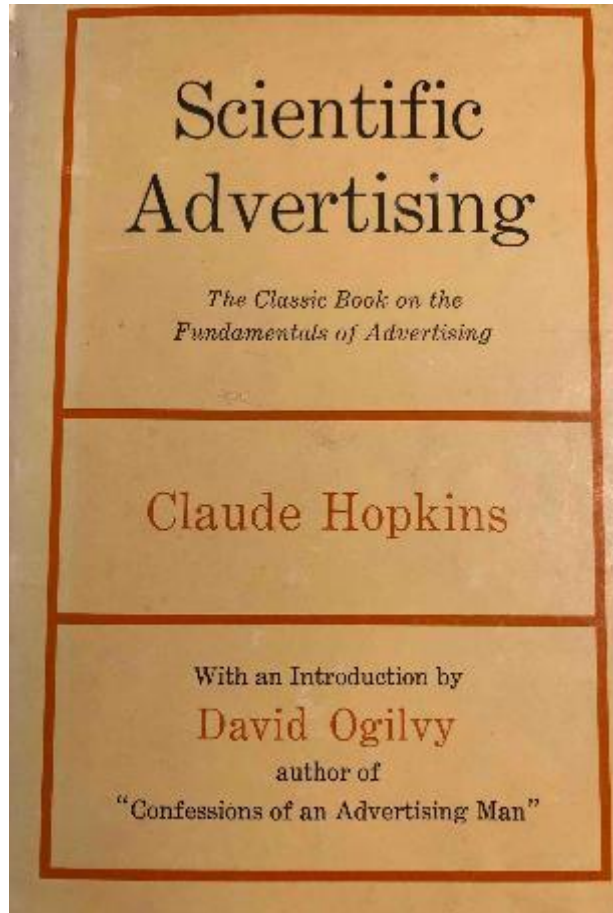
1950's

We won't need to do any traditional research anymore: just get some housewives in a room and conduct a focus group



1960's

We don't need to do any traditional research anymore: just do some ad testing



1910's Ad performance = Coupon response counting

1930's Ad performance = Noting and reading studies

1960's Ad performance = Ad copy testing

Short cut to predicting ad performance:
“How much do you like this ad?”

Yes worked quite well but lots of examples from the history of marketing of legendary advertising campaigns that failed conventional market research ad tests



1950's

We won't need to do any traditional research anymore: just put some blindfolds on and ask them if they like the taste



Mono-diet of sweet, salty & fatty food



The method excluded the emotions, memory associations and variant experience of eating food

2000s

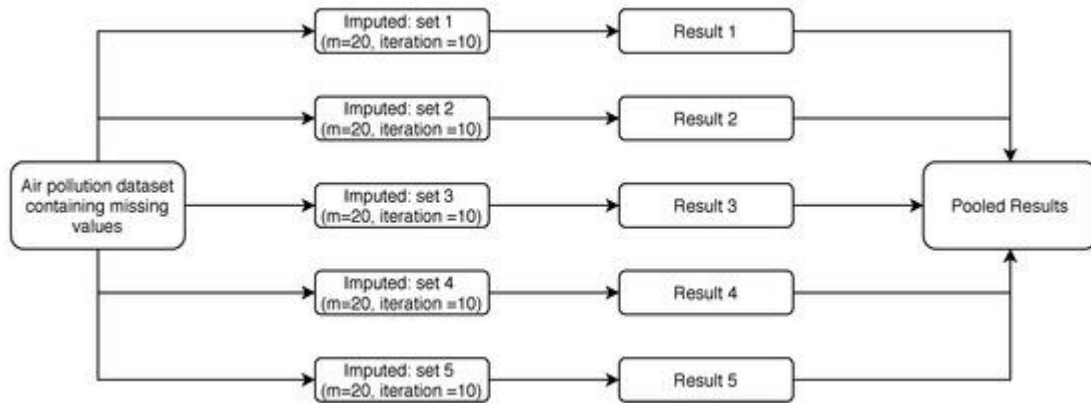
We won't need to do any traditional research anymore: we can do online surveys



2023 still not fully nationally representative of all age groups

2000s

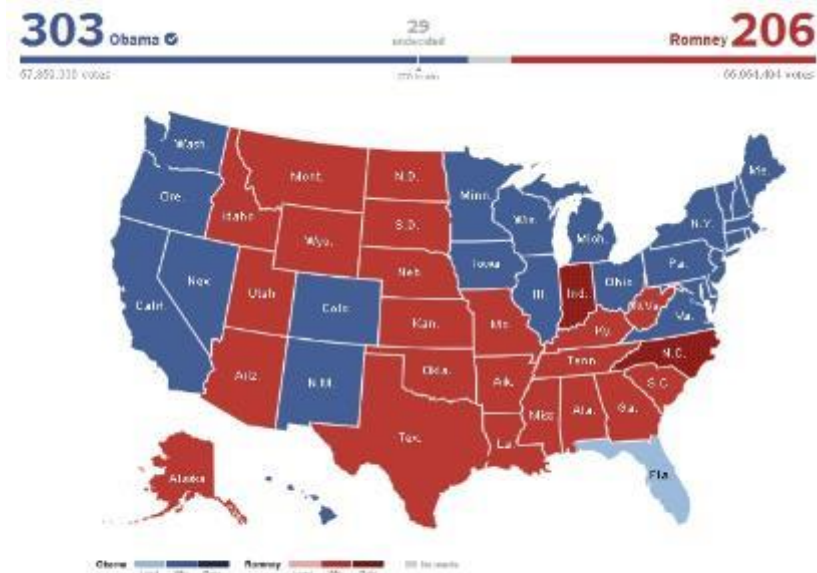
We won't need to do any traditional research anymore: we can model data using random forest ascription



5 weeks and £30k bill

2010s

We won't need to do any traditional research anymore : can use prediction markets and ensemble modelling



State-by-State Probabilities



Nate silver correctly predicted result of every state in the 2012 US election by aggregating poll predictions thousands of polls.

4 years later

Princeton Poll and economic aggregator forecast there was a “99% chance Clinton winning the 2016 election....

Princeton Election Consortium

A first draft of electoral history. Since 2004

As of November 8, 6:06AM EST: **Snapshot (191 state polls): Clinton 307, Trump 231 EV** **Meta-margin: Clinton +2.2%** RSS

Clinton Nov. win probability: random drift 99%, **Bayesian >99%**

Senate snapshot (48 polls): **Dem+Ind: 50, GOP: 50**, Meta-margin: D +0.9%, Nov. control probability: Dem. 82%


FRONT PAGE
ARCHIVES

HIGH-IMPACT RACES

[Find competitive House districts near you!](#)

Republicans:

NRSC



Democrats:

help us reach **\$ 400,000**

to date **1,587**

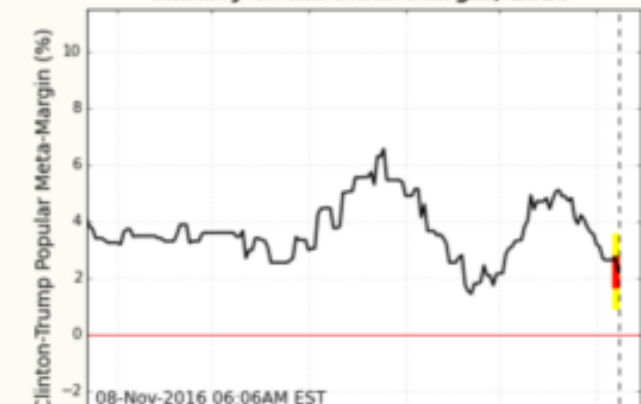
\$ 361,787

Contribute

History of the Meta-analysis

August 21, 2016: The forecast has been sharpened using information from this year's campaign. Also, the Meta-Margin to electoral votes relationship, which generates the November electoral vote "strike zone," has been updated. [See this explanation.](#)

History of the Meta-Margin, 2016



GET PEC ON YOUR MOBILE

Electoral College Clinton Nov win

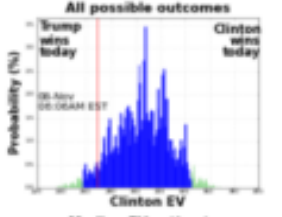
Clinton 83%

Trump


2016 PRESIDENTIAL RACE

Reducing state-level Presidential polls to a single high-precision snapshot. [more](#)

All possible outcomes

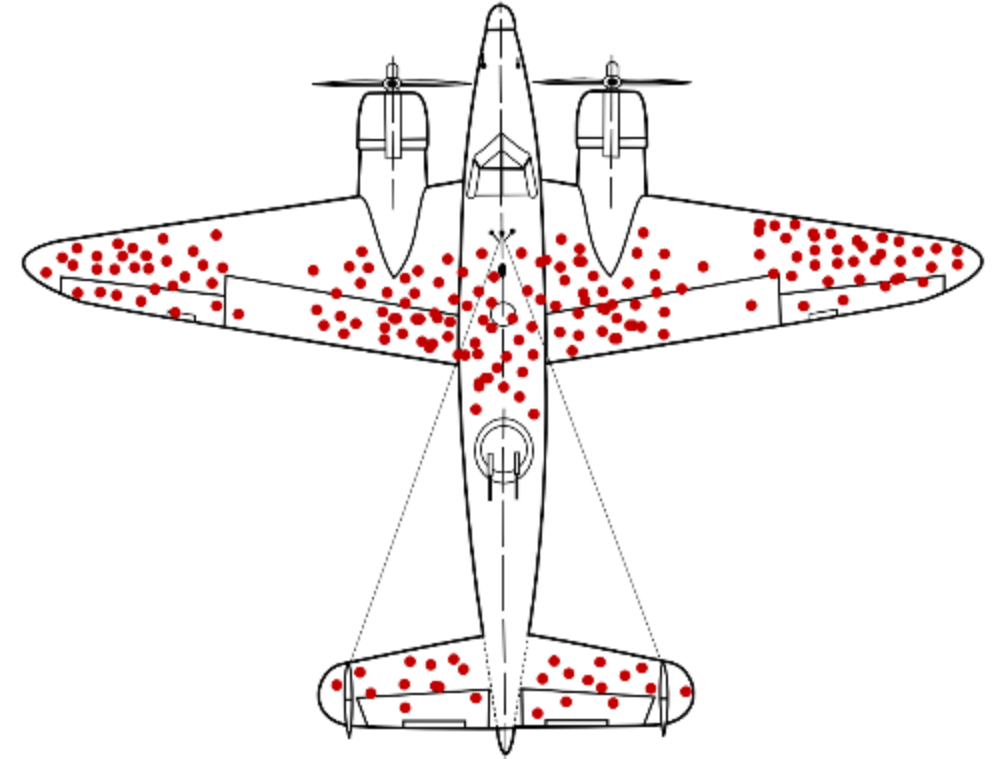
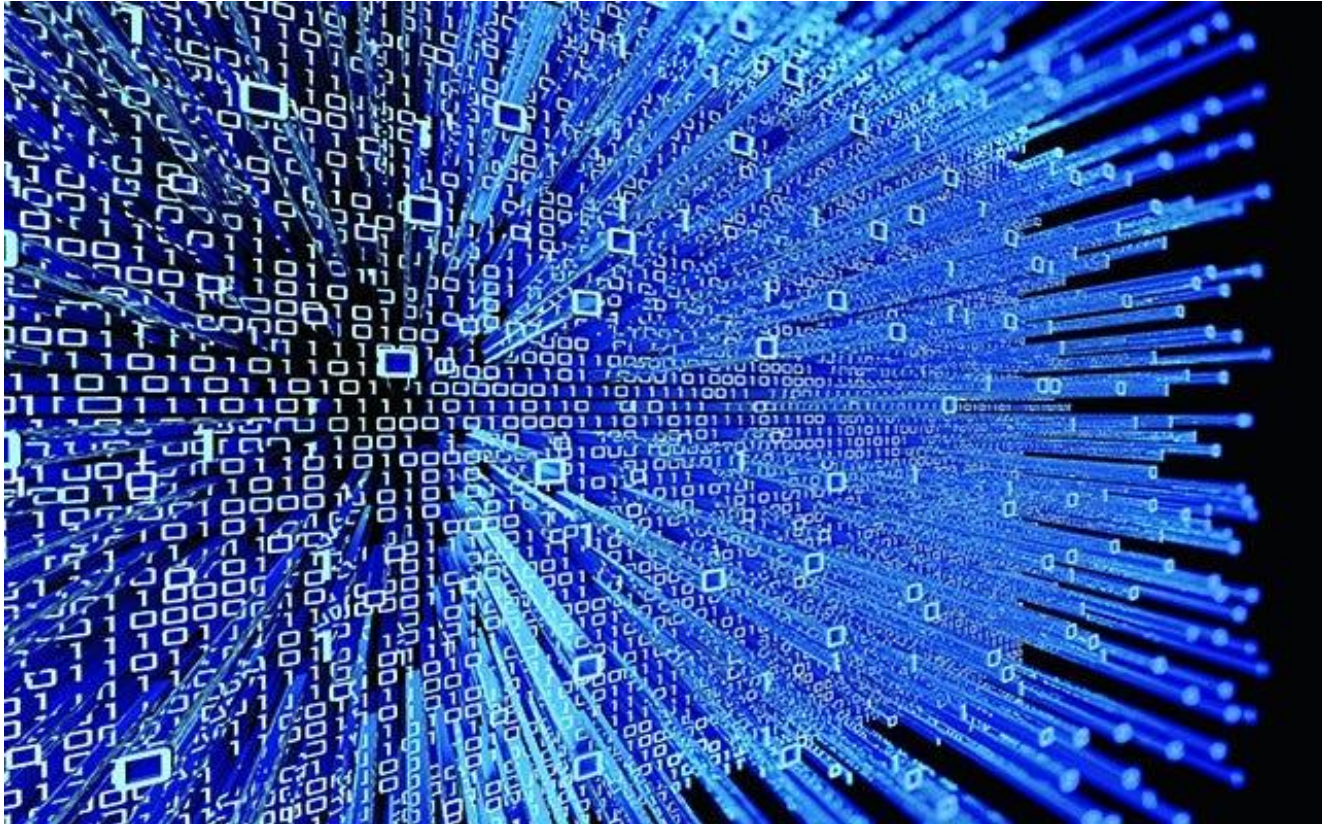


Median EV estimator



2010s

We won't need to do any traditional research anymore: We can use our own big data sources



Survivor bias

Some of these protestations have slowly come true others have not, and fatal flaws found in others

Looking back on the history of new innovations some do come along and wipe out the old way of doing things pretty quickly...

~~The horses~~ → The car

~~Photography~~ → digital photography

~~Neanderthals~~ → Homo Sapiens

But in most cases new innovations have grown alongside, rather than wipe out old ones

Radio → TV = Radio & TV

Shark → Human = Sharks & humans

Books → Kindle = Books

What will be the impact of AI on traditional market research?

Kill off or Live alongside?

To answer this helpful to think about issues held back and have hampered all these new techniques for conducting research over the years

1. Sample representivity
2. Reaching robust sample sizes
3. Practical limitations Reproducibility hurdles of new solutions e.g. finite usable use case scenarios
4. Techniques failing because they have closed & limited or one-dimensional perspectives

Sample representativity

Liberty polls, online research, social media text analytics & big data analytics all faced the same challenge of lack of sample representativity

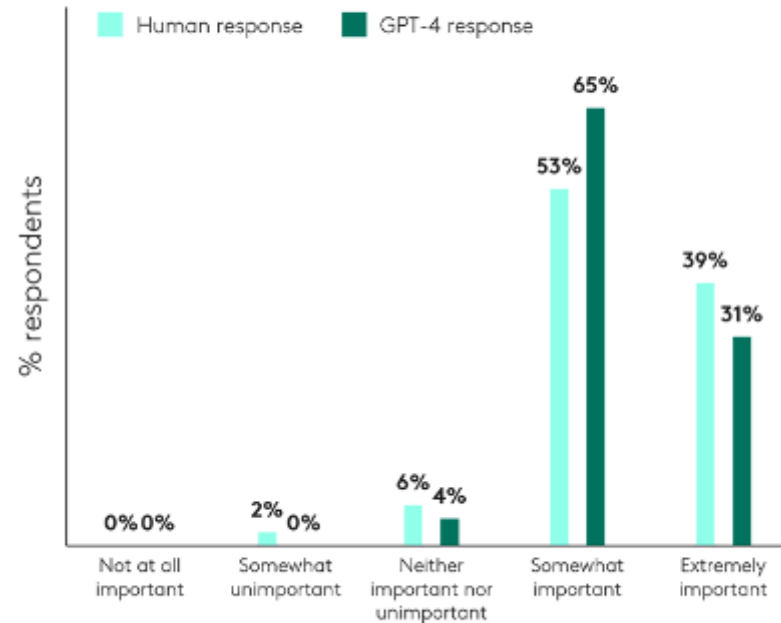
LLM's have the same challenge

Trained on the views of:

- Males
- White people
- Westerners
- English language speakers
- The well educated
- Those with a technological bent

Example 1

What is the importance of price while purchasing a luxury product?

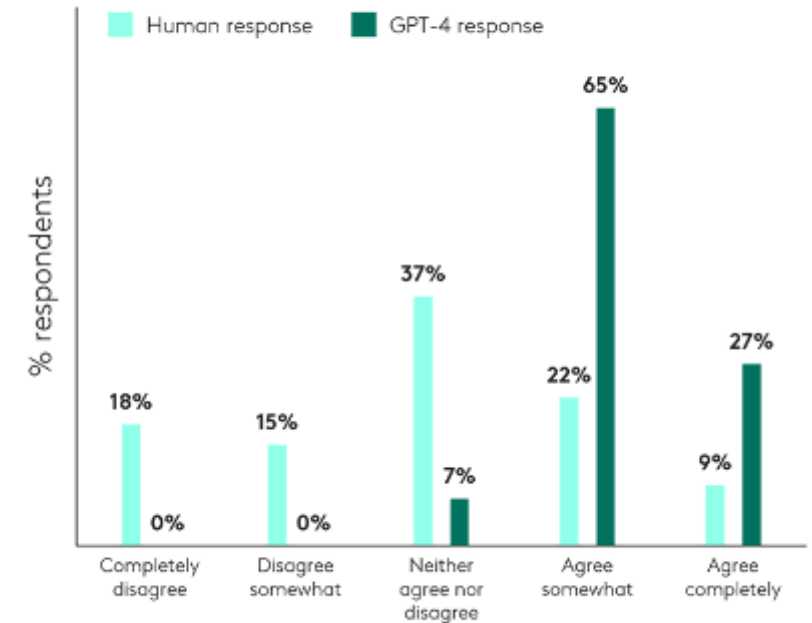


N = 5275

Source: Kantar's early stage R&D

Example 2

My product is a way for me to bond/connect with others who share my passion

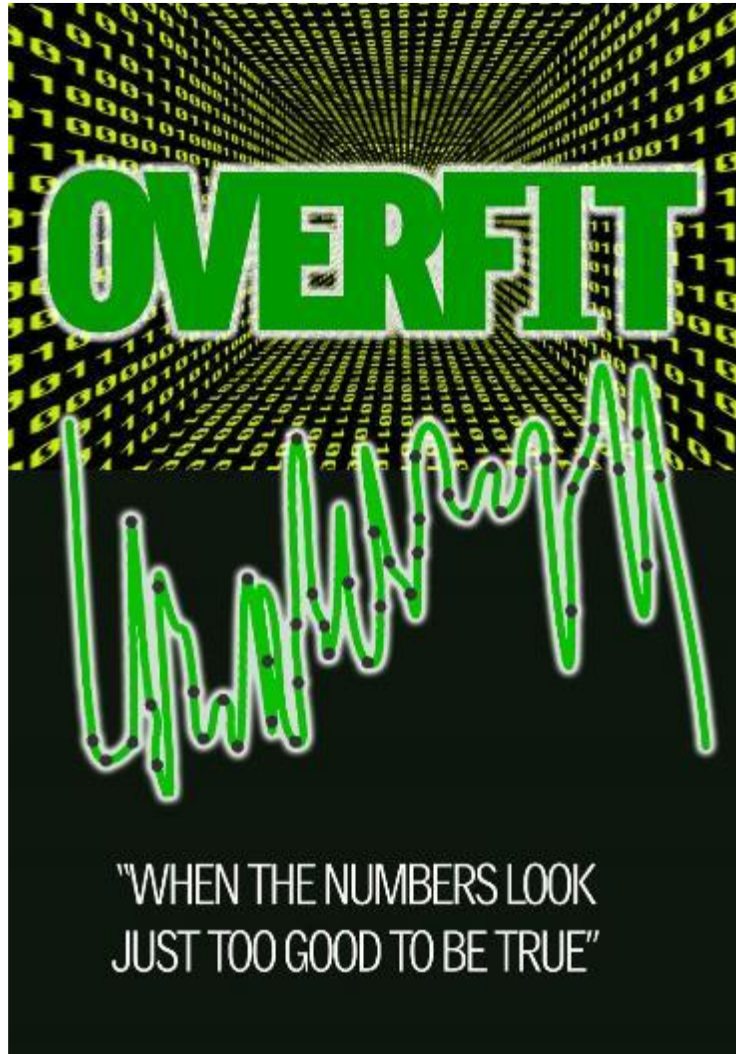


N = 5050

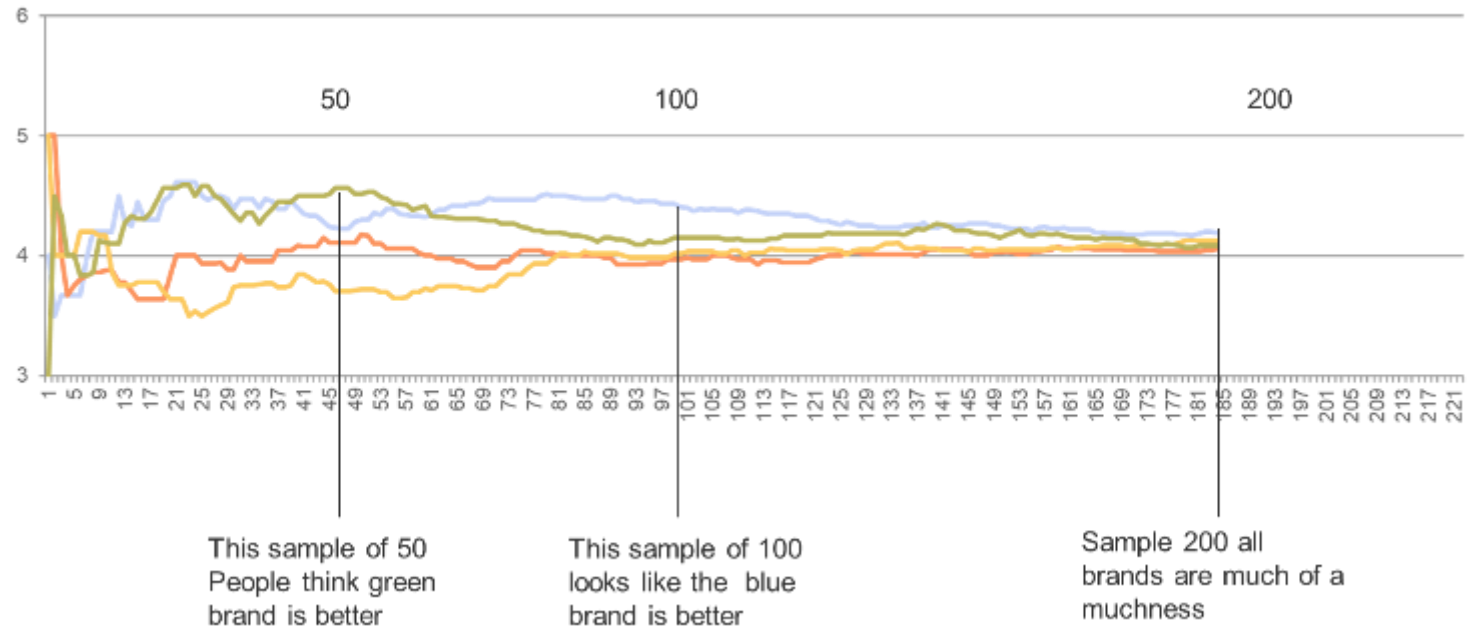
To handle sample bias researchers set quota and weight data, but to do both these you need to understand the make-up of the sample.

→ LLM's no easy way to find this out.

Sample size challenges



How many people do you need before you get a reliable answers



Case study from the early days exploring the potential of Text analytics

Project X

Understanding Consumer views of shampoo

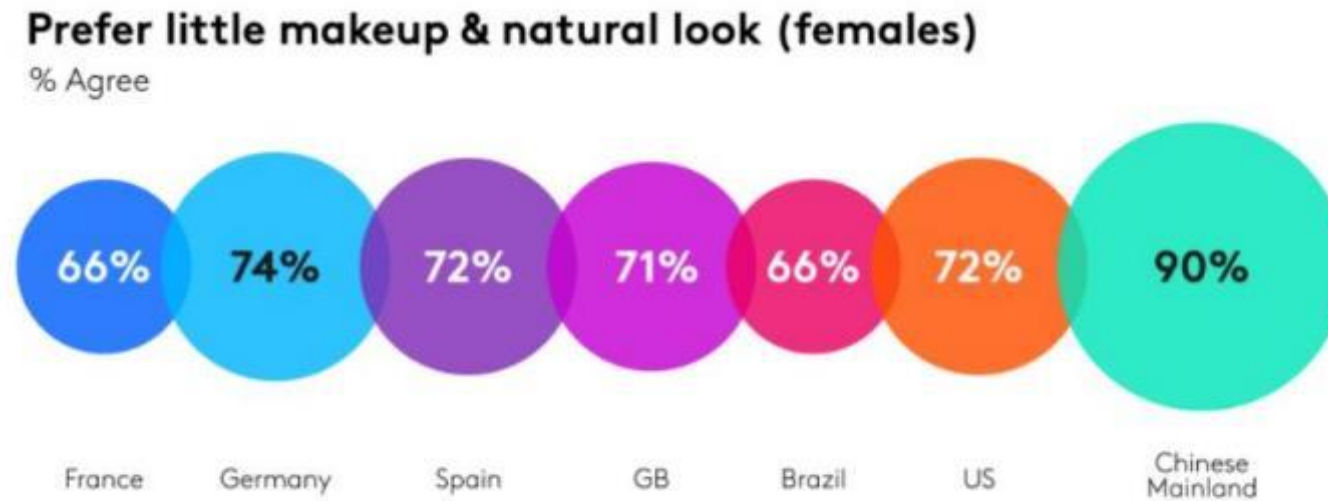
Scraped 200,000 comments from twitter referencing a shampoo brand → After 2 week of manual cleaning left with 200 usable ones

Big issues “Dove”: are they referring to bird watching, making song references, or talking about world peace

The majority of comments were about L’Oreal, no references to any of the more functional brands like VO5 or Voseine

Over the years the role of using text analytics has grown and evolved

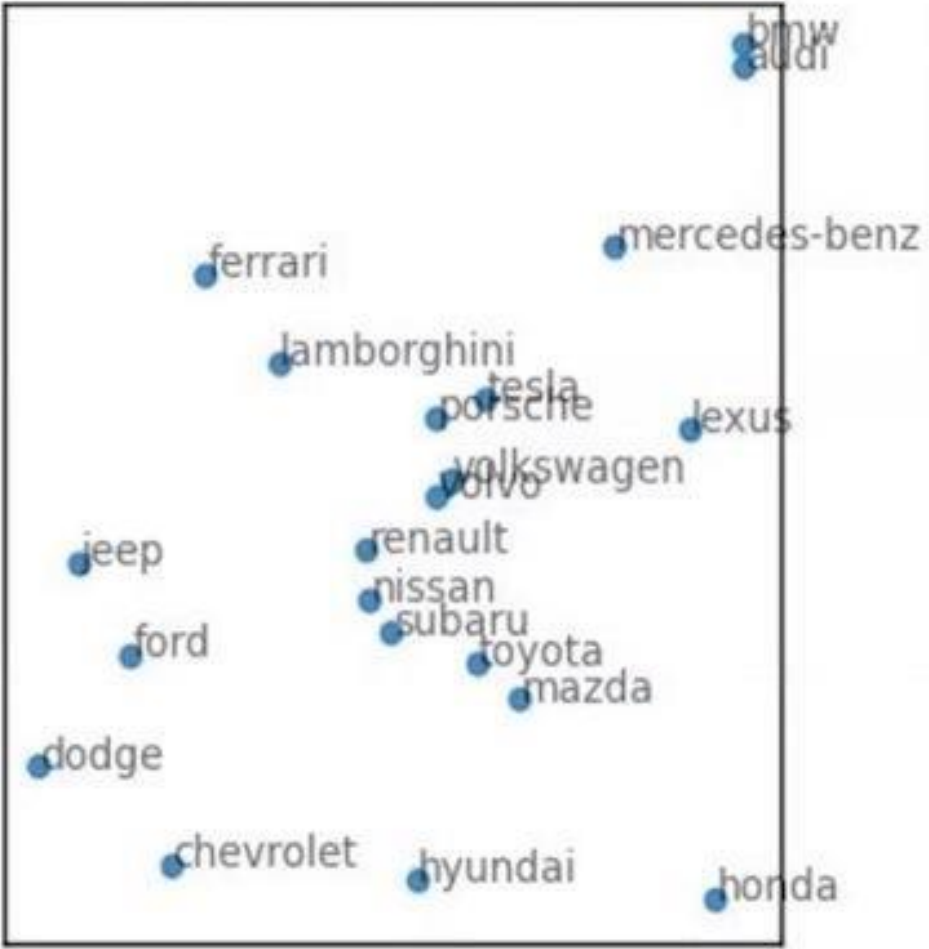
Kantar analyses for example over 2 billion open ended searches for beauty products every year to understand what is driving choices



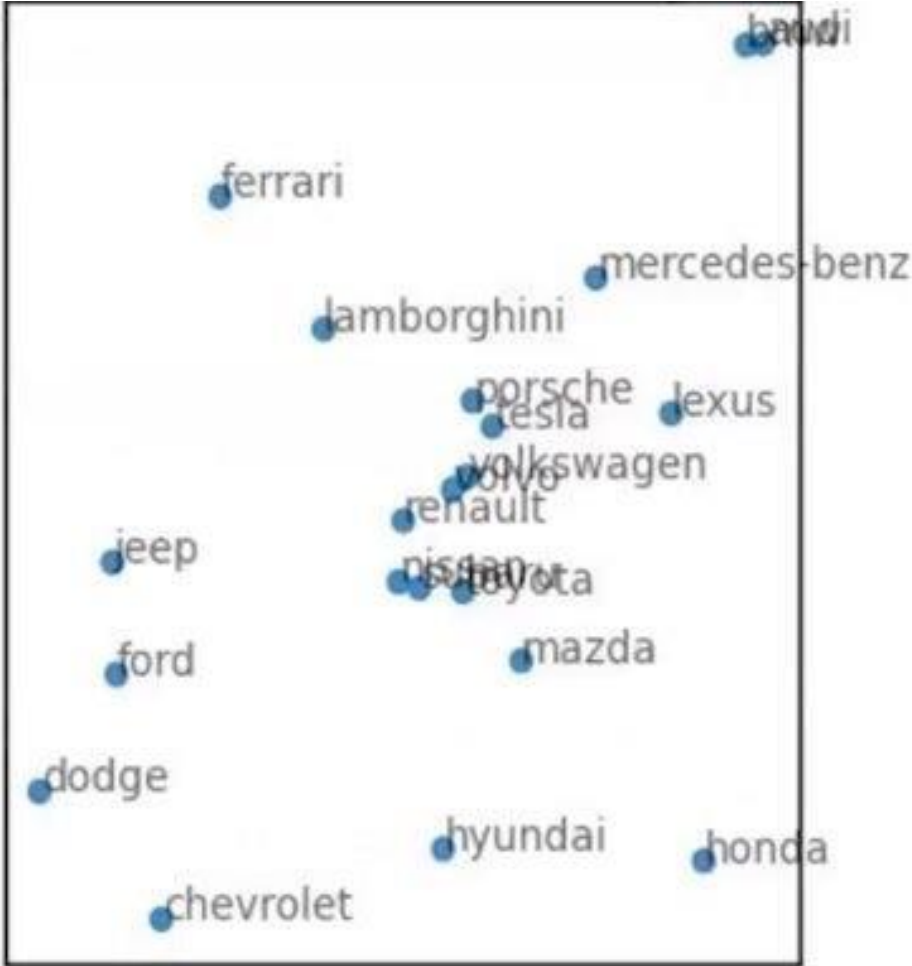
Source: Kantar, Usage Core Panels

...but this has not really encroached on traditional MR but expanded

Reproducibility hurdles: Using LLM's to produce perceptual maps the of the car market

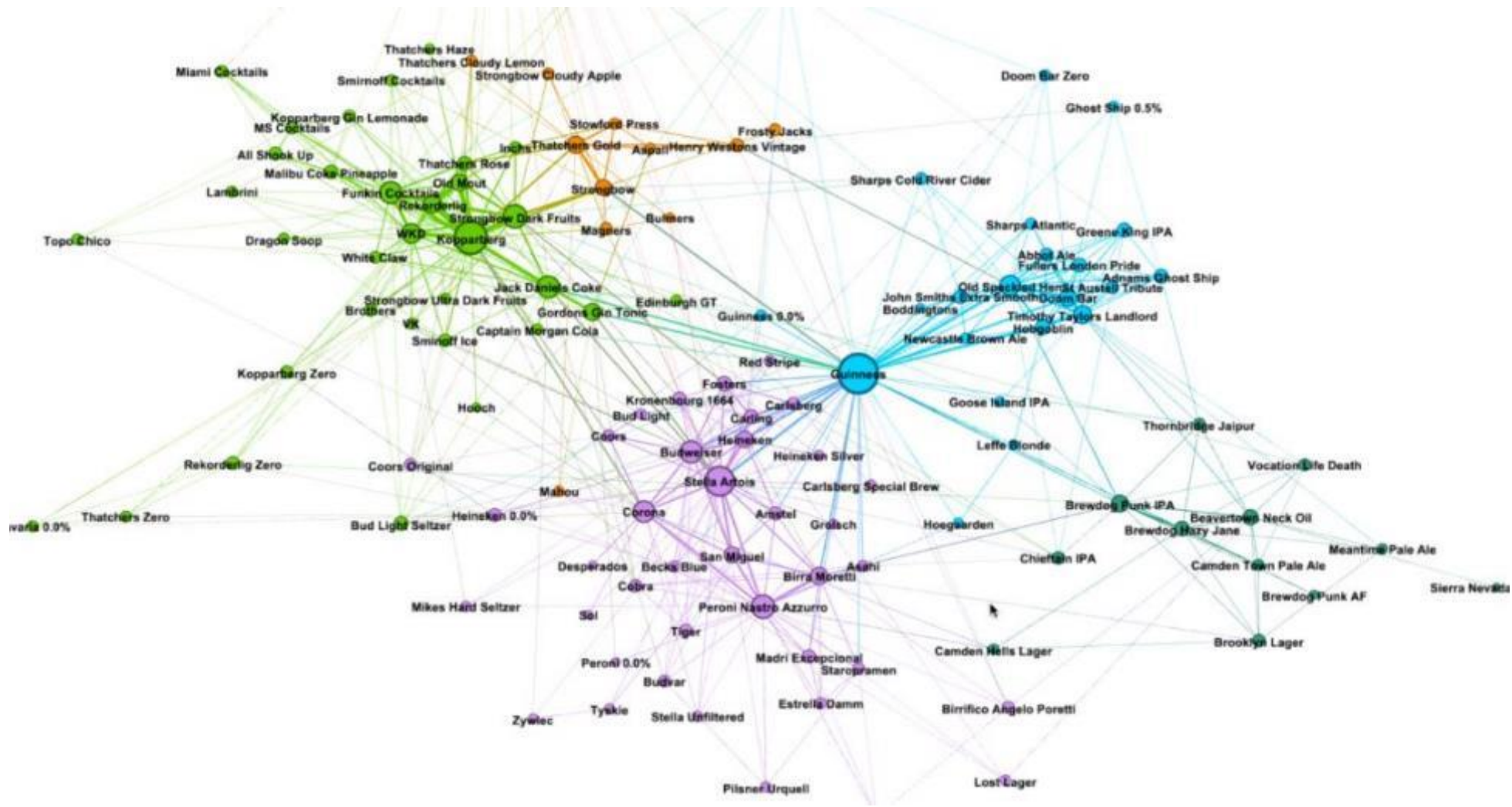


LLM



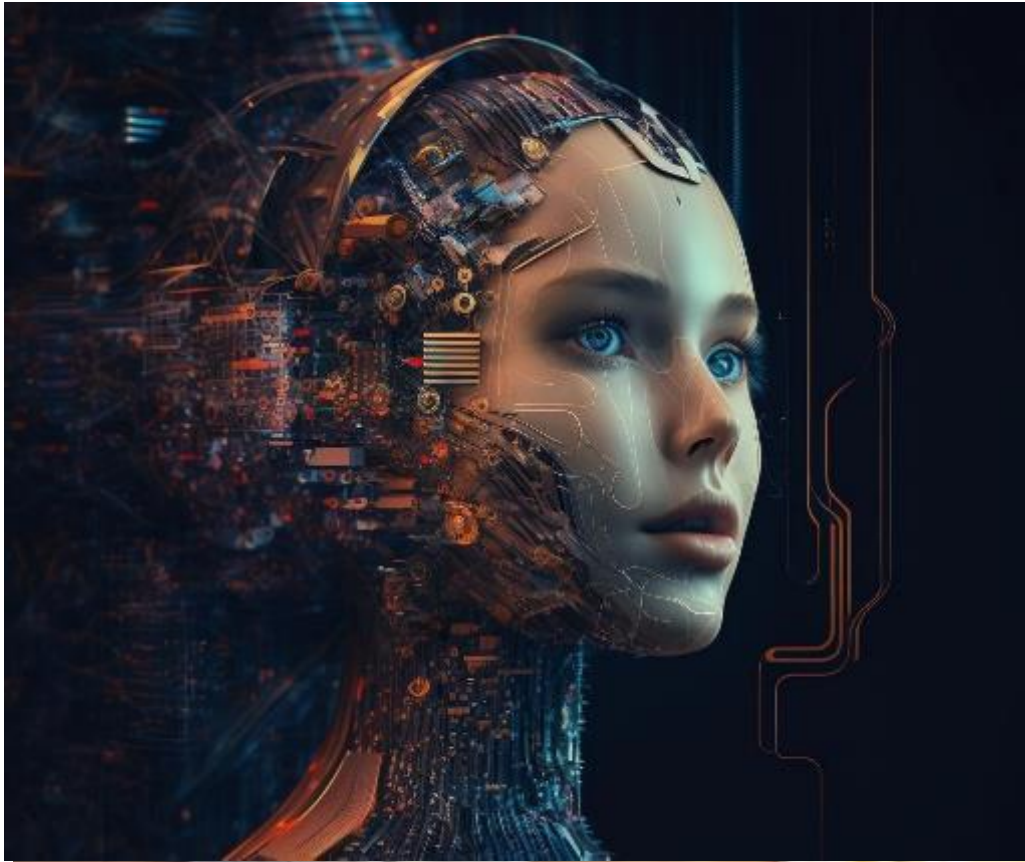
Market research

You trying doing that with less prominent brands and for categories where there is less natural discourse



Needed 6,000+ Nat rep respondents to map the top 100 brands with some accuracy

What are sold with Generative AI



What we get when trying to use it for real purposes “Market researcher”



Training sample of 1?

LLM's are probability models feeding off a huge amount of existing published research, weary of what fresh insights it can bring.

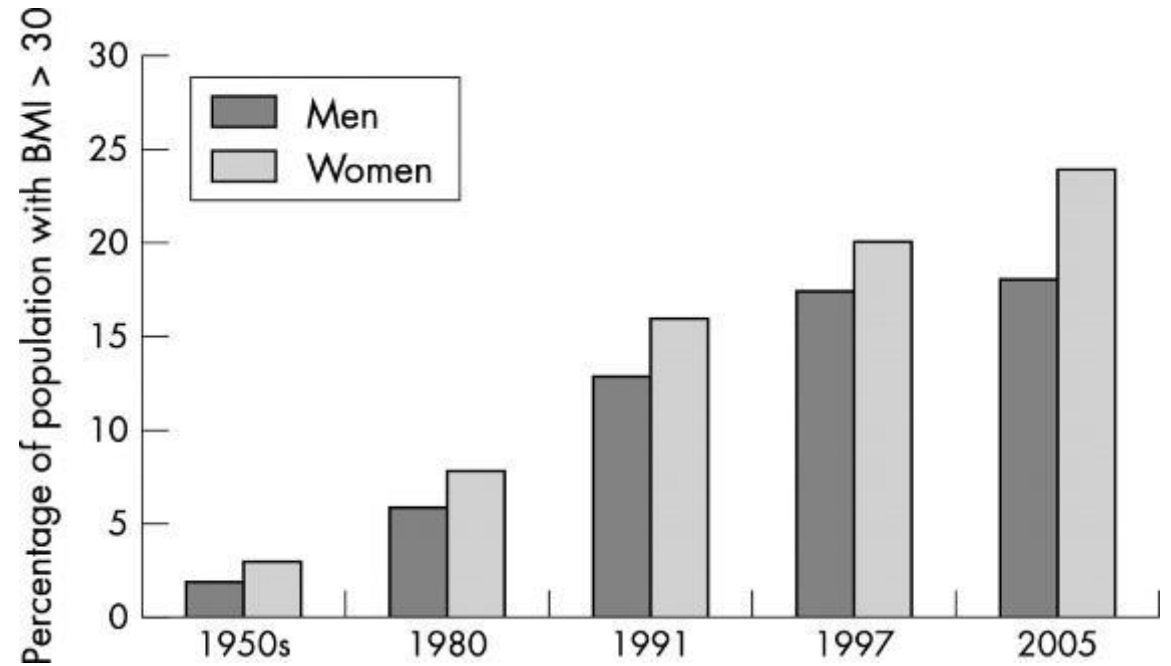
Remember it's still just **searching the internet with a fancy interface** that masks where its getting its info from.

The challenges of research techniques offering limited closed and over simplistic perspectives

Close perspective research

The impact of over simplest taste testing of the 1960s – it was not that it did not work, but it offered limited perspective

Which led so many companies down the same path of developing food full of Salt, Fat & Sugar & carbs: “because that’s what consumer liked”



I call this the 32 issue

What is the next number in this sequence?

1, 2, 4, 8, 16, ?

Marketing solution



1, 2, 4, 8, 16, 32



Research data

Some bright marketers may spot another marketing solution...

1, 2, 4, 8, 16, 64

There are in fact a huge number of solutions

0 (subtracting each number in the sequence from 17)

11 (adding the current number in the sequence to the product of the previous two numbers)

12 (subtracting the previous number in the sequence from the square of the current number)

128 (raising 2 to the power of each number in the sequence)

13 (alternating between adding and subtracting the previous number in the sequence)

15 (adding the digits of each number in the sequence together)

17 (adding 1 to each number in the sequence)

18 (adding 2 to each number in the sequence)

18 (adding the previous three numbers in the sequence)

19 (adding the product of the first two numbers to the sum of the next two numbers)

23 (multiplying each number in the sequence by 3 and subtracting 5)

24 (adding the current number in the sequence to twice the previous number)

24 (adding the first two numbers, then adding 2 to each subsequent sum)

25 (adding the product of the first two numbers to the cube of the current number)

31 (subtracting 1 from each number in the sequence)

32 (continuing the pattern of doubling each number)

32 (multiplying each number in the sequence by 2 and adding 1)

32 (multiplying each number in the sequence by 2, then subtracting 2)

64 (squaring each number in the sequence)

10000 (writing each number in the sequence in binary and concatenating them)

Any number (they are actually random numbers that just happen to be in order)

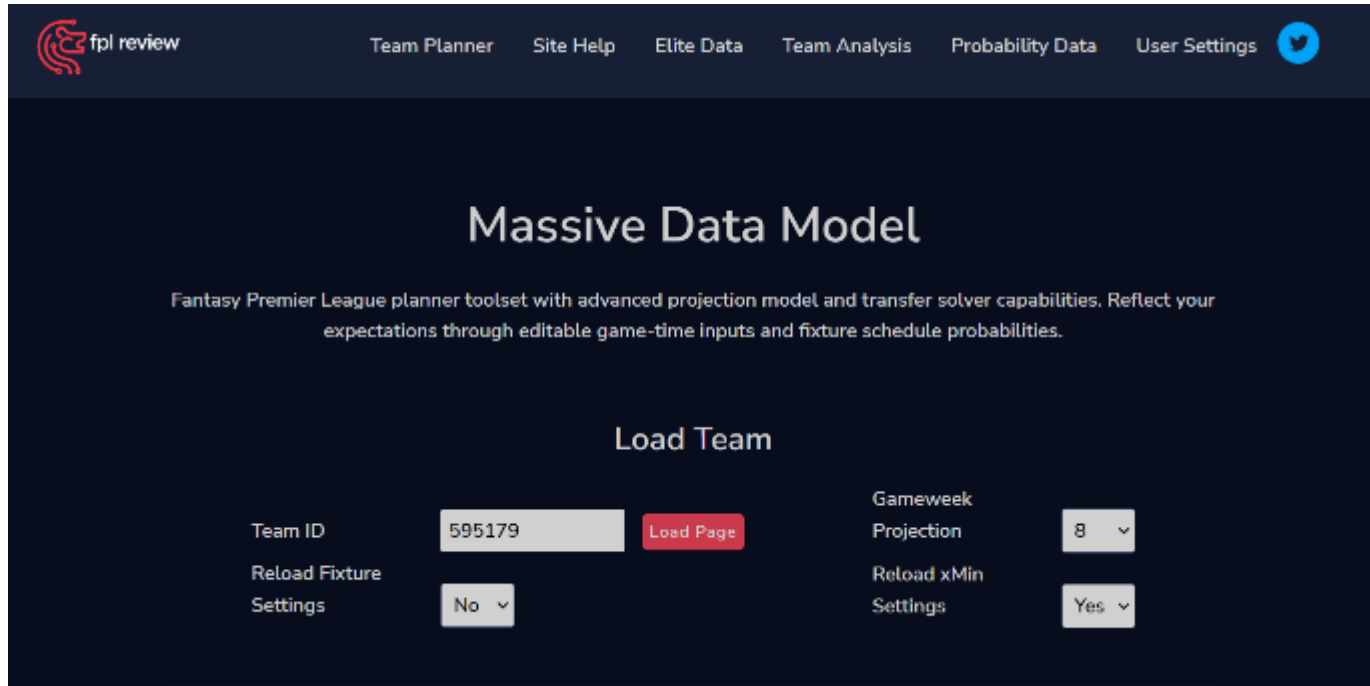
But everyone looking at the data from the same perspective: doing the same simplistic analysis and all coming to the same conclusions and all missing the bigger picture....

Ad testing faces the same challenge: If all you are measuring is “liking”, “memorability”, “branding” the 32 solution are car ads like this....



Here is where machine learning and AI potentially opens up opportunities to find the 64+ solutions...

Case study: Fantasy football AI models



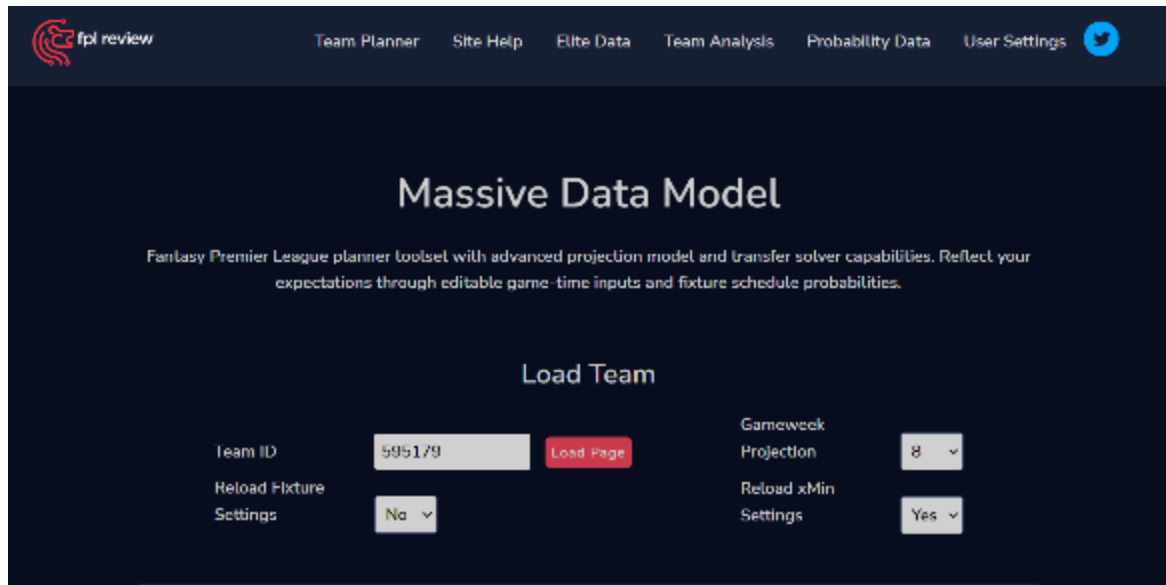
The screenshot shows a dark-themed web interface for 'fpl review'. At the top, there is a navigation bar with links: Team Planner, Site Help, Elite Data, Team Analysis, Probability Data, and User Settings. The main content area is titled 'Massive Data Model' and includes a subtitle: 'Fantasy Premier League planner toolset with advanced projection model and transfer solver capabilities. Reflect your expectations through editable game-time inputs and fixture schedule probabilities.' Below this is a 'Load Team' section with several input fields: 'Team ID' (text input with '595179'), 'Reload Fixture Settings' (dropdown menu with 'No'), 'Gameweek Projection' (dropdown menu with '8'), and 'Reload xMin Settings' (dropdown menu with 'Yes'). A red 'Load Page' button is positioned between the Team ID and Gameweek Projection fields.

Famed machine learning model which as Ingested every kick in football used to predict the performance of players.

If you followed its suggestions, it would guarantee you a top 10,000 Fantasy Football finish last year (out of 10 million)

Machine learning and AI has the potential to find better solutions, but can only be as good as the training data they are primed with.

Case study: Fantasy football AI models



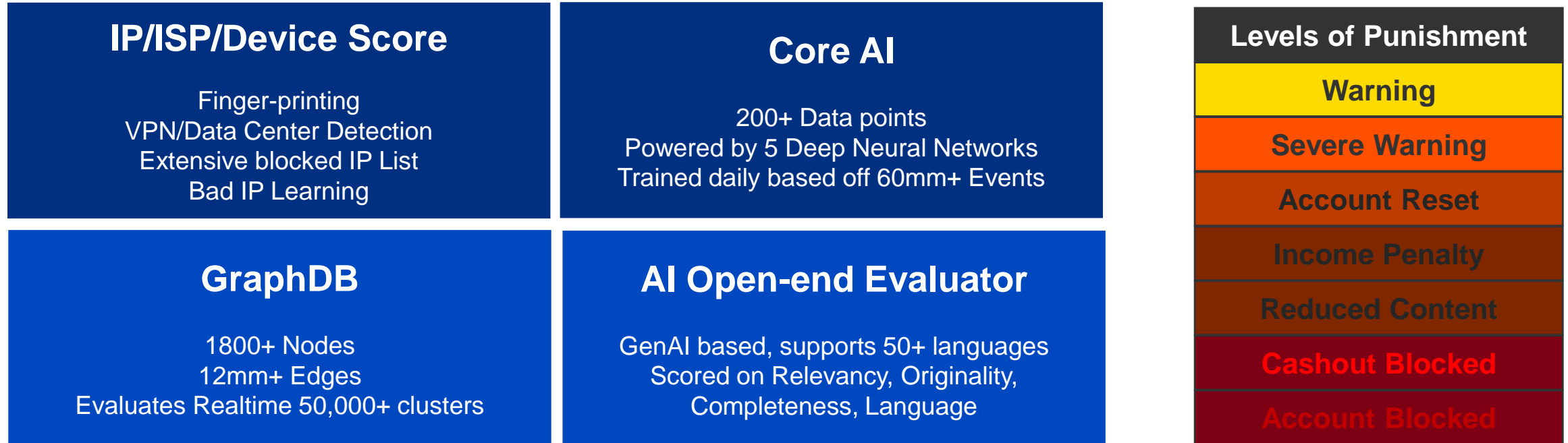
The screenshot shows a dark-themed web interface for 'fpl review'. At the top, there is a navigation bar with links for 'Team Planner', 'Site Help', 'Elite Data', 'Team Analysis', 'Probability Data', and 'User Settings', along with a Twitter icon. The main content area is titled 'Massive Data Model' and includes a subtitle: 'Fantasy Premier League planner toolset with advanced projection model and transfer solver capabilities. Reflect your expectations through editable game-time inputs and fixture schedule probabilities.' Below this is a 'Load Team' section with several input fields: 'Team ID' (text input with '595179'), 'Reload Fixture Settings' (dropdown menu with 'No'), 'Gameweek Projection' (dropdown menu with '8'), and 'Reload xMin Settings' (dropdown menu with 'Yes'). A red 'Load Page' button is positioned between the 'Team ID' and 'Gameweek Projection' fields.

If you did exactly what the AI models said this year...

After 10 games you would be languishing in 4 millionth

Model suffered catastrophic failure as a result of rule change resulting in 6 minutes of extra time

Case study: Kantar Profiles CUBE fraud detection



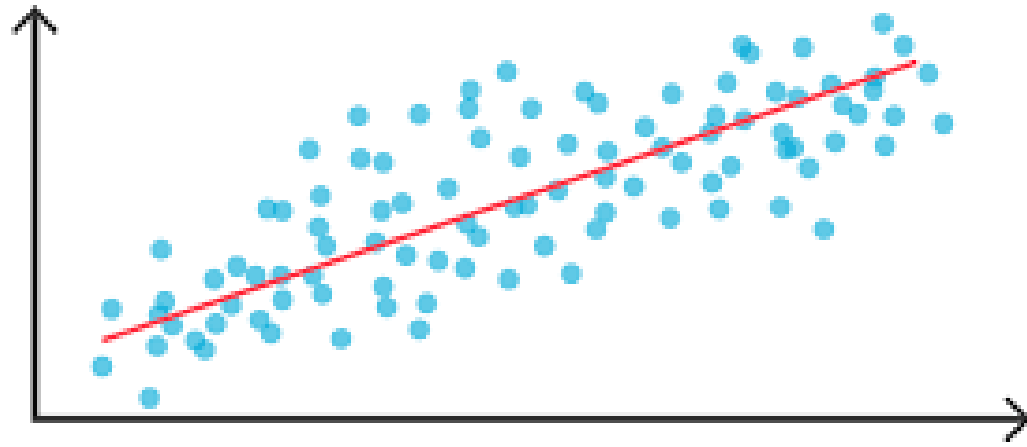
Real time AI fraud detection system had to be trained on over **1 trillion of data points** collected over 8 years in the UK & USA

Applied in Germany

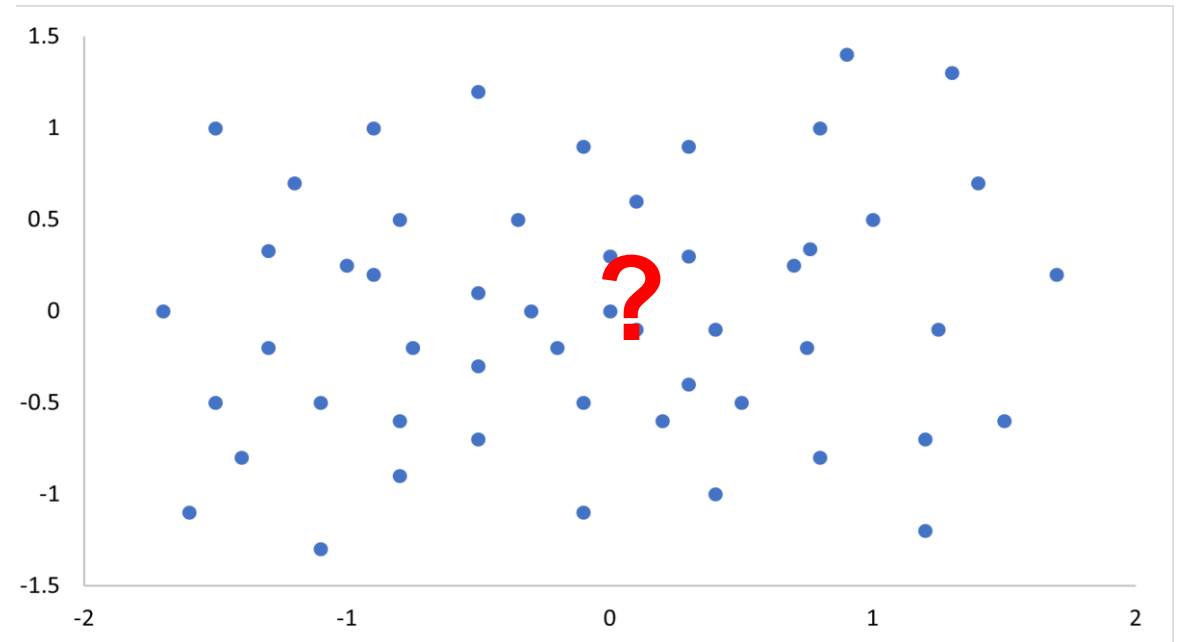
Germans more honest = 20% miss screens
needed another 100 million data points to calibrate

AI models need reliable outcome data to be able to connect the dot

What AI models hoping for

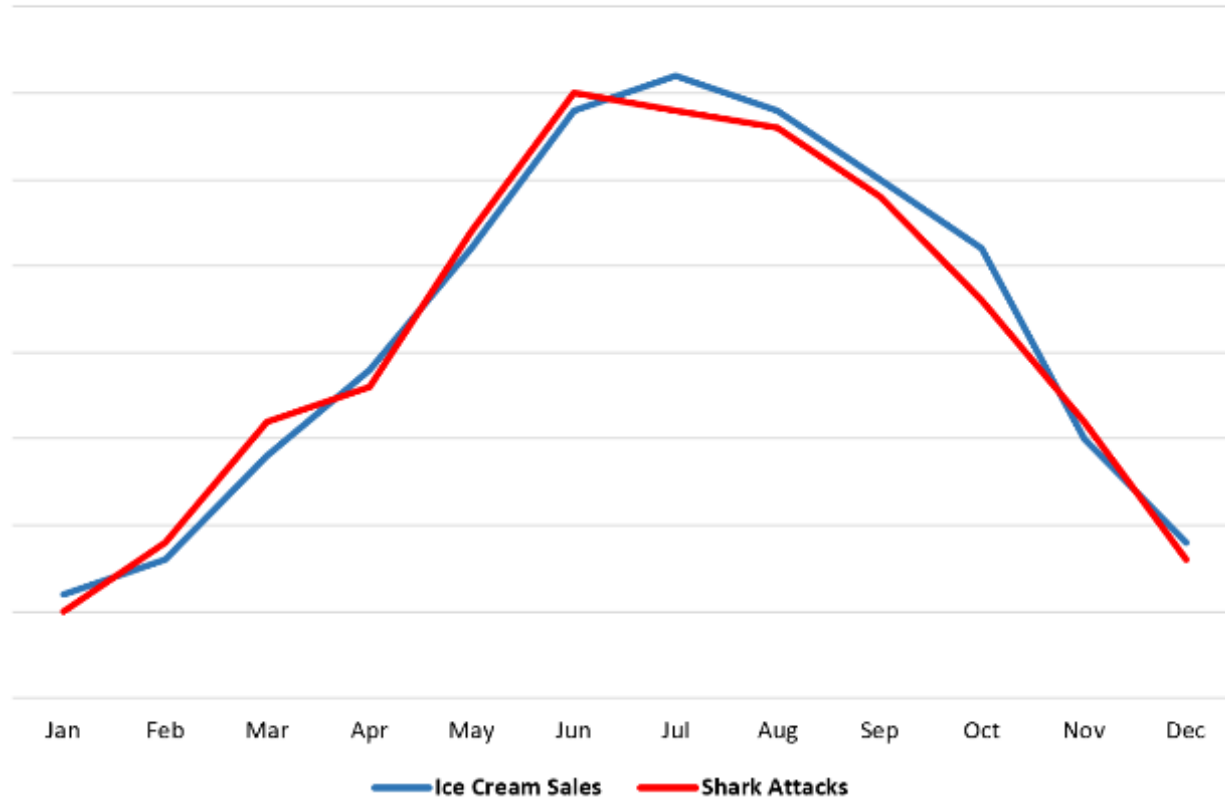


What data market research data can be like



Ai can also struggle to differentiate between Correlation & Causation

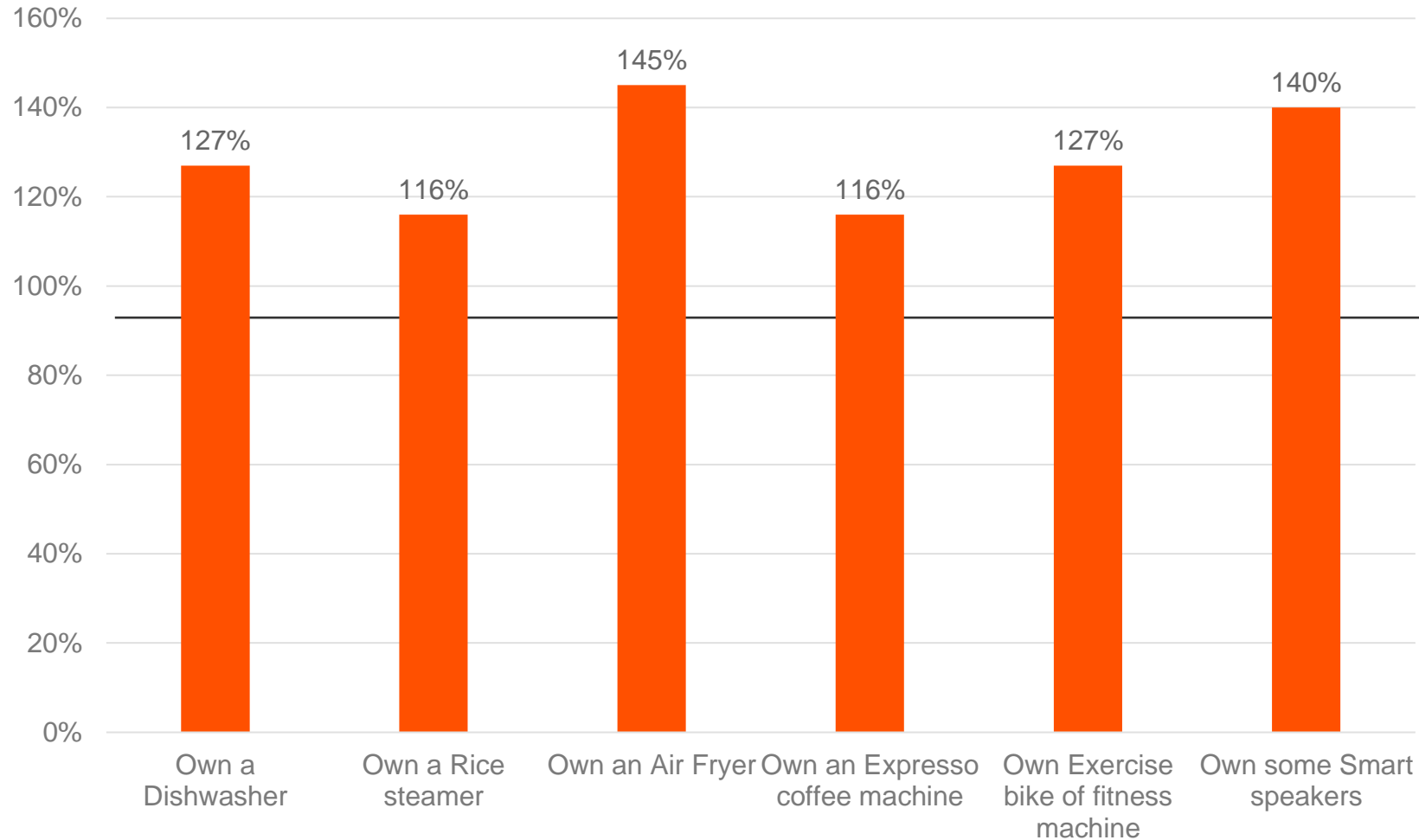
Ice Cream Sales vs. Shark Attacks



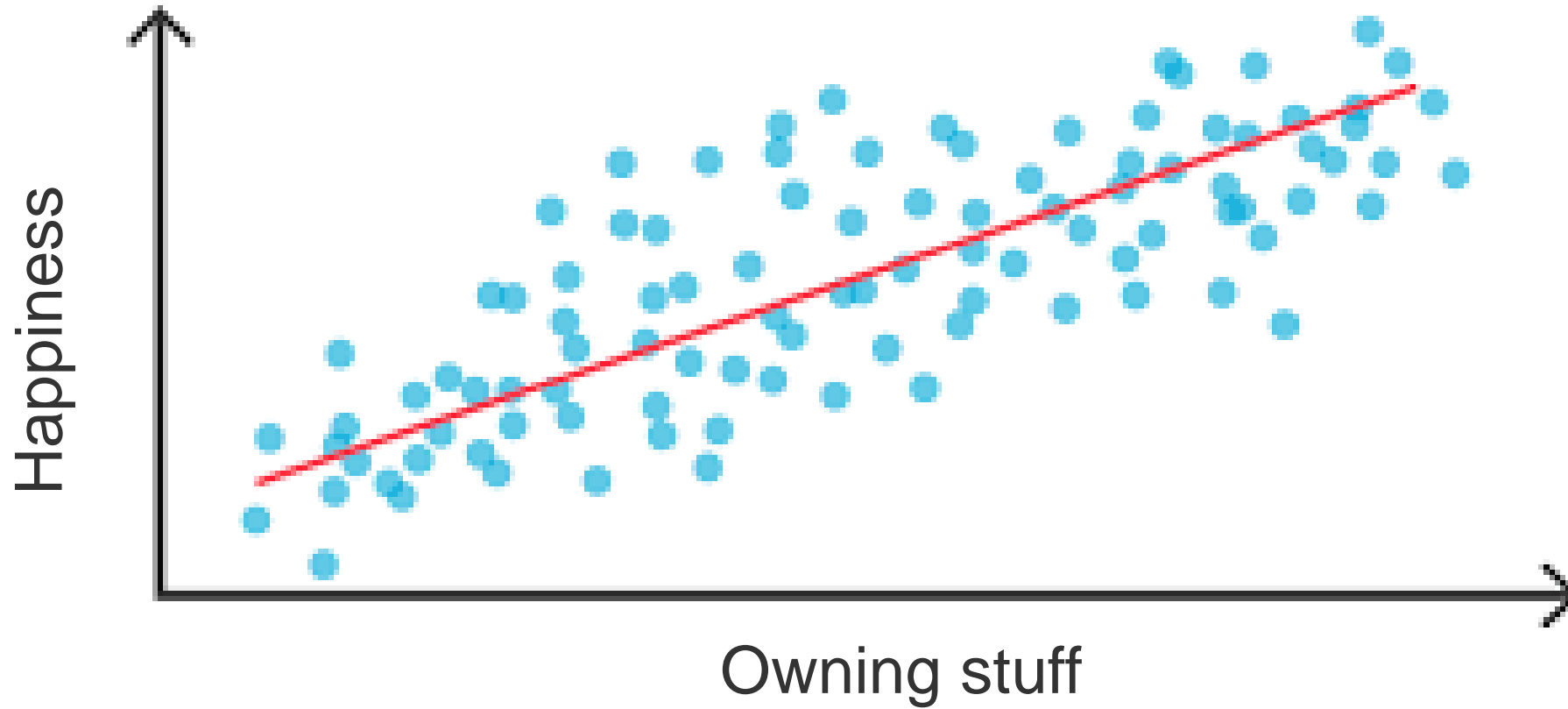
Poor input data and no outcome data the result can be hallucinations...

Case study: quest to find the secret to happiness-based poll of 20,000 people from around the world about their lifestyle and attitudes

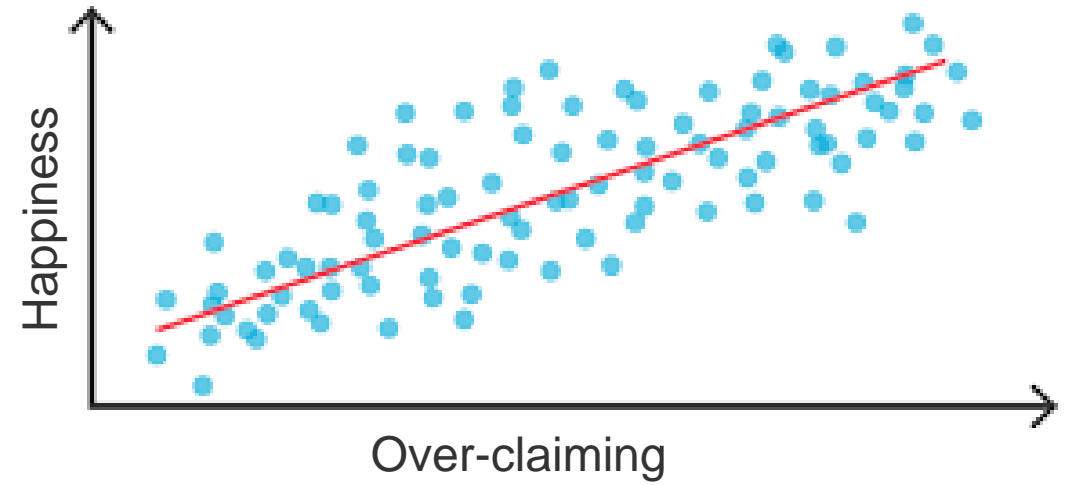
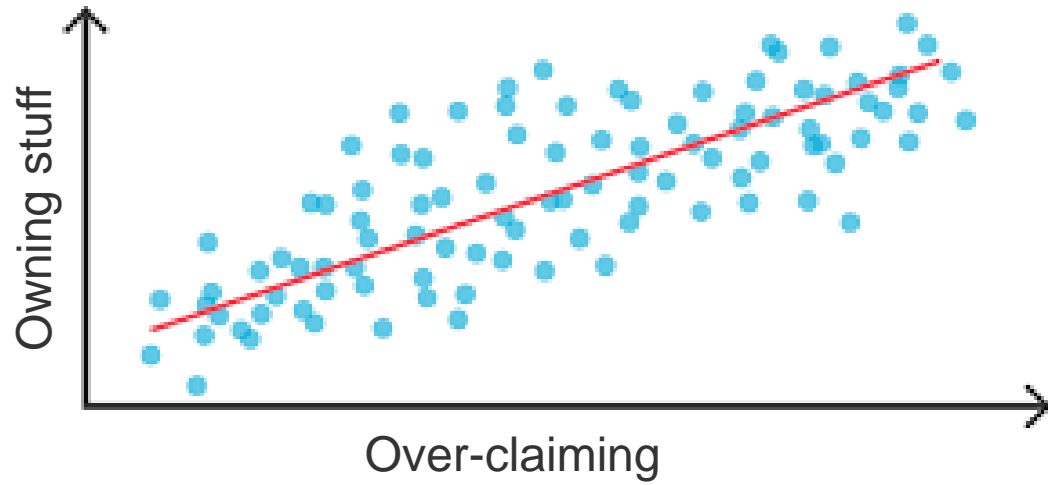
The secret to happiness = Owning an Air Fryer



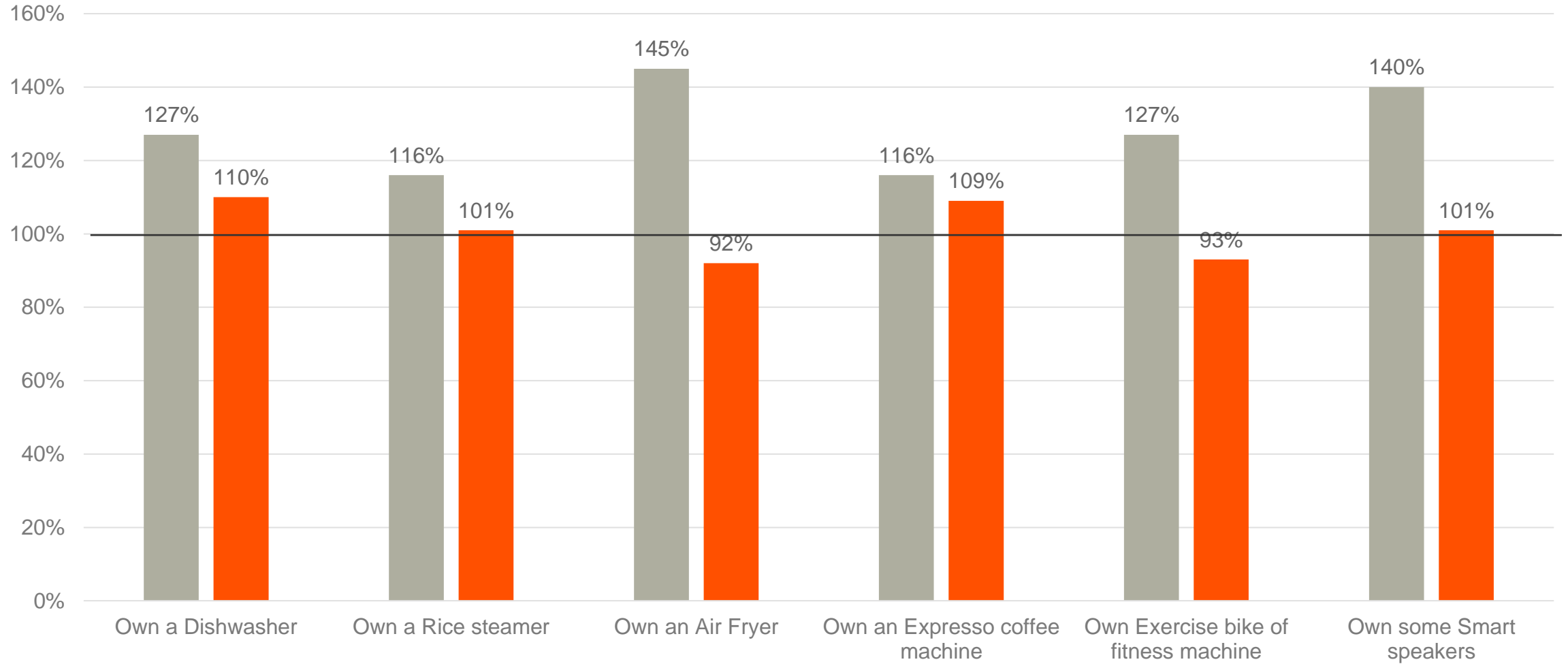
Correlation v Causation?



Correlation v Causation?



The secret to happiness – weighted to take account of overclaim



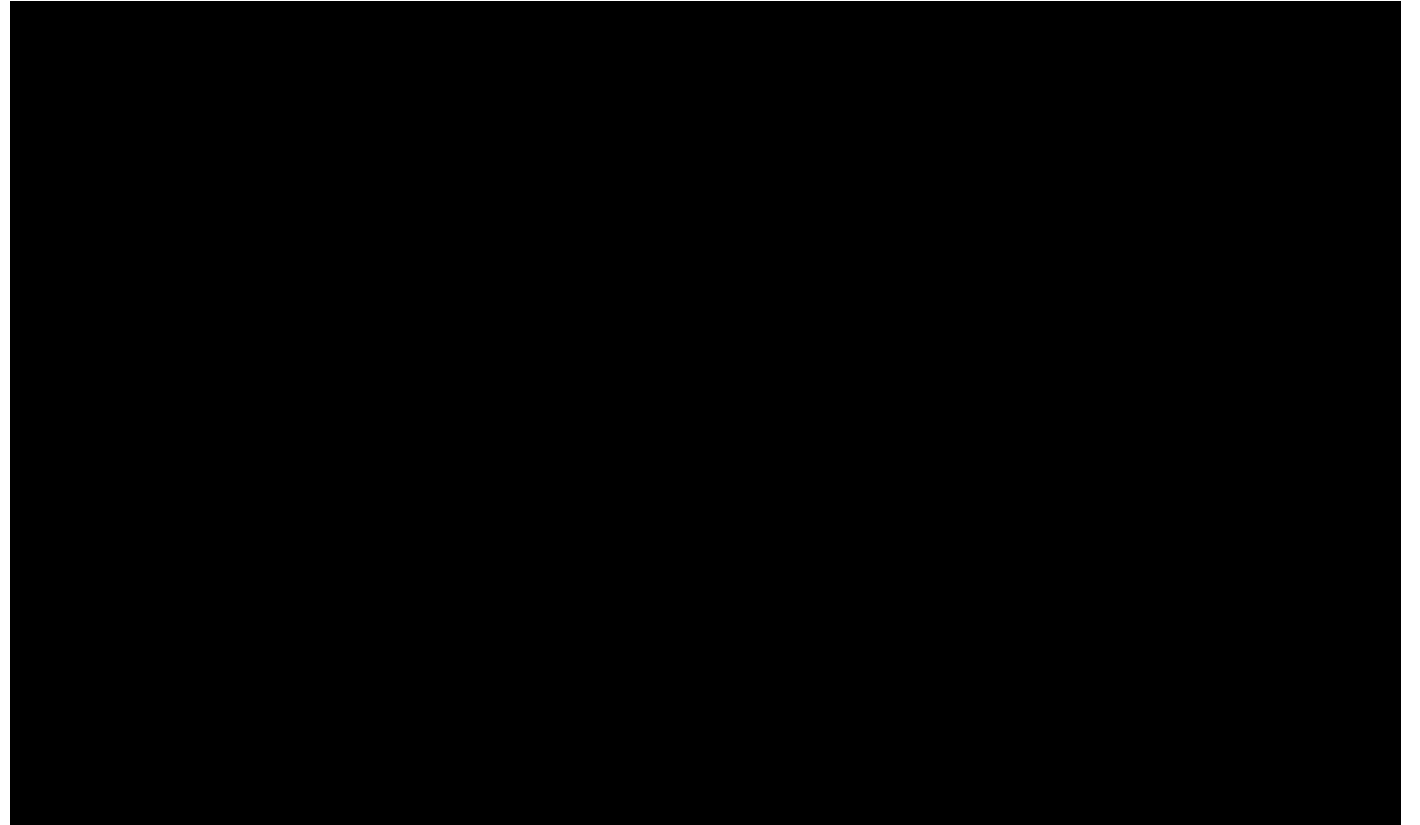
Case study

the evolution of Ad copy testing

- For years we have been working with a very **limited set of lowest common denominator input measures** show a clear relationships with ad performance: “liking”, “memorability”, “surprise” & “branding”
- **Going beyond this is very difficult using traditional research technique**
- **Lots of weaker signals many of which are non-linear and intertwined**

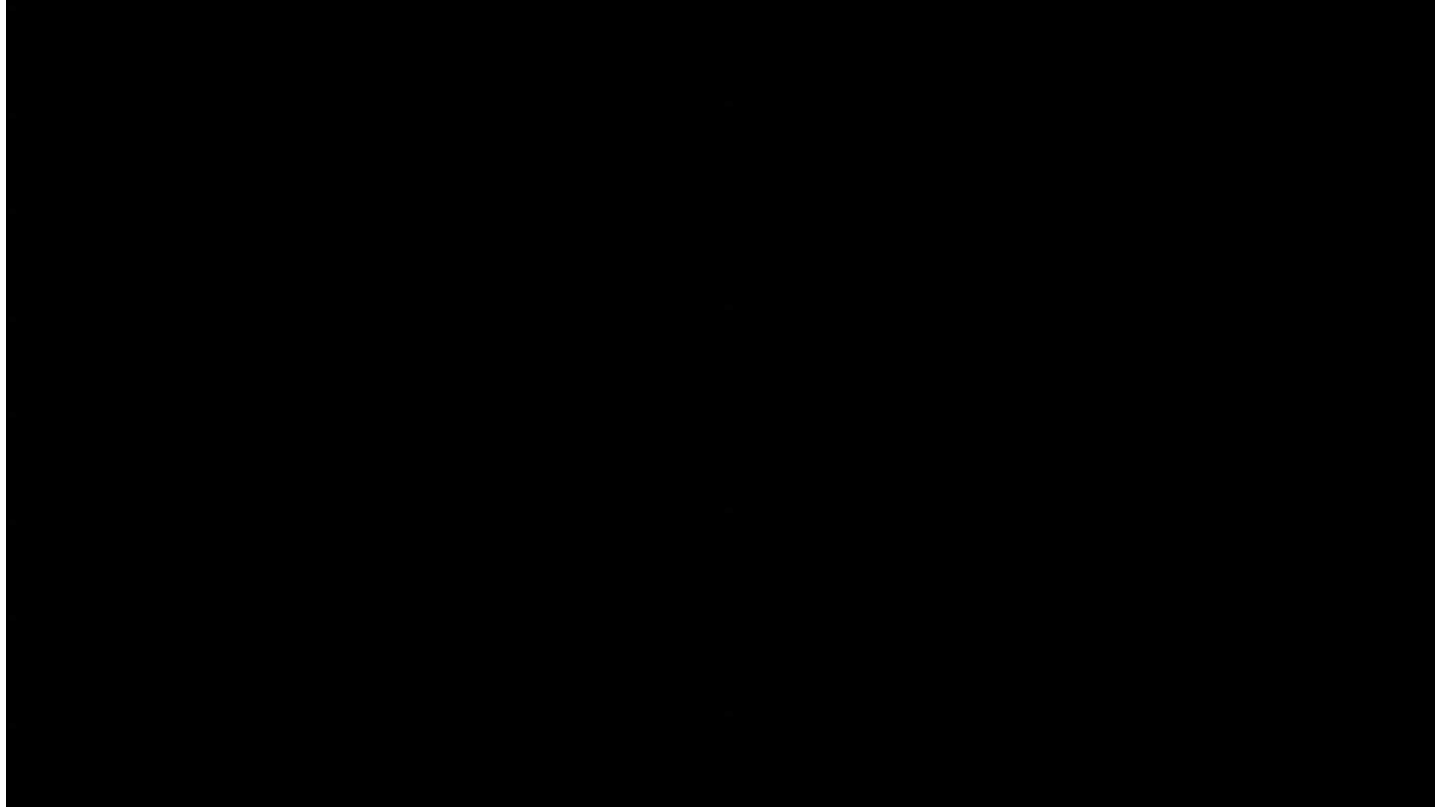
Ads that you know will do well in a traditional ad copy tests are of a type...

- Funny
- Emotional
- An element of surprise
- Strong characterisation



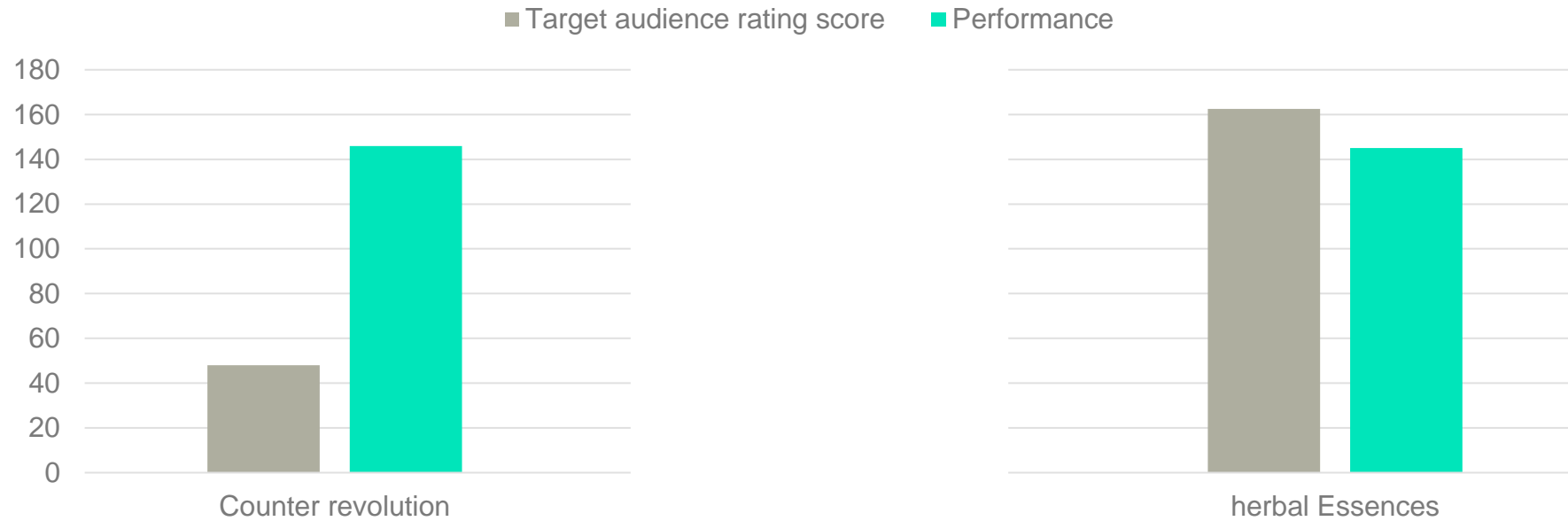
But there many ads that don't
do well in ad traditional tests

That can actually be really
effective



How these ads compared in test and reality

Ad effectiveness



Success Drivers breakthrough research study in 2018

..by expanding out the range of input measures from 5 → profiling 200+ different factors and using machine learning enabled a much richer understanding of the underlying strength of a wider range of advertising

Liking
Emotion
Relevance
Branding
Purchase intent



Level of characterisation
Message delivery
Message confidence
Human Face to camera time
Eye contact
Brand integration
Type of sound and voice over
Quality of information delivery
Emotional tone
Product benefits



Linear regression $R^2=0.30$

ML Universal structural modelling $R^2=0.62$

KANTAR

2023 where we are at now...

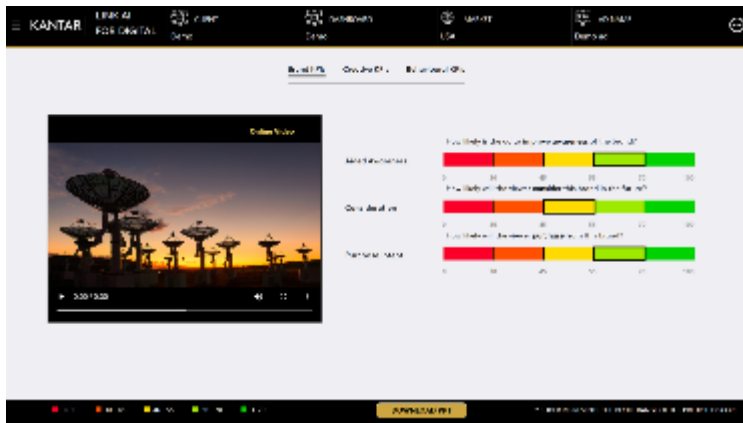
Link AI



Copy tests combining

Brand metrics

The ad's likely contribution to campaign effectiveness with media weight behind it



- Aided Awareness
- Consideration
- Purchase Intent
- Ad Recall (TikTok only)

Creative metrics

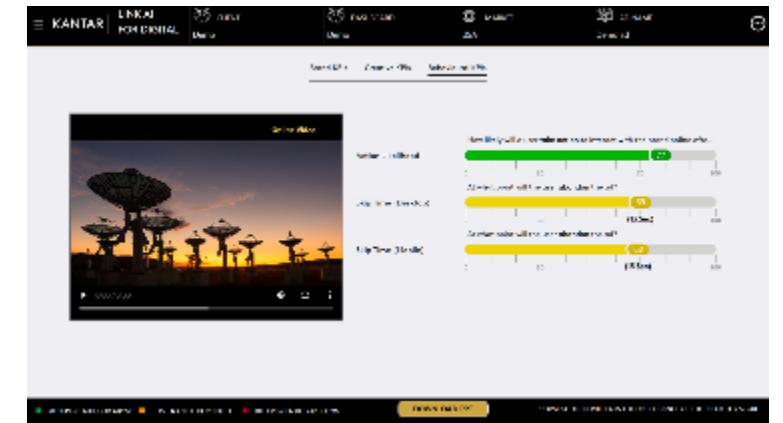
What people think/feel about the ad/brand



- Impact
- Power
- Branding
- Enjoyment
- Persuasion
- Affinity

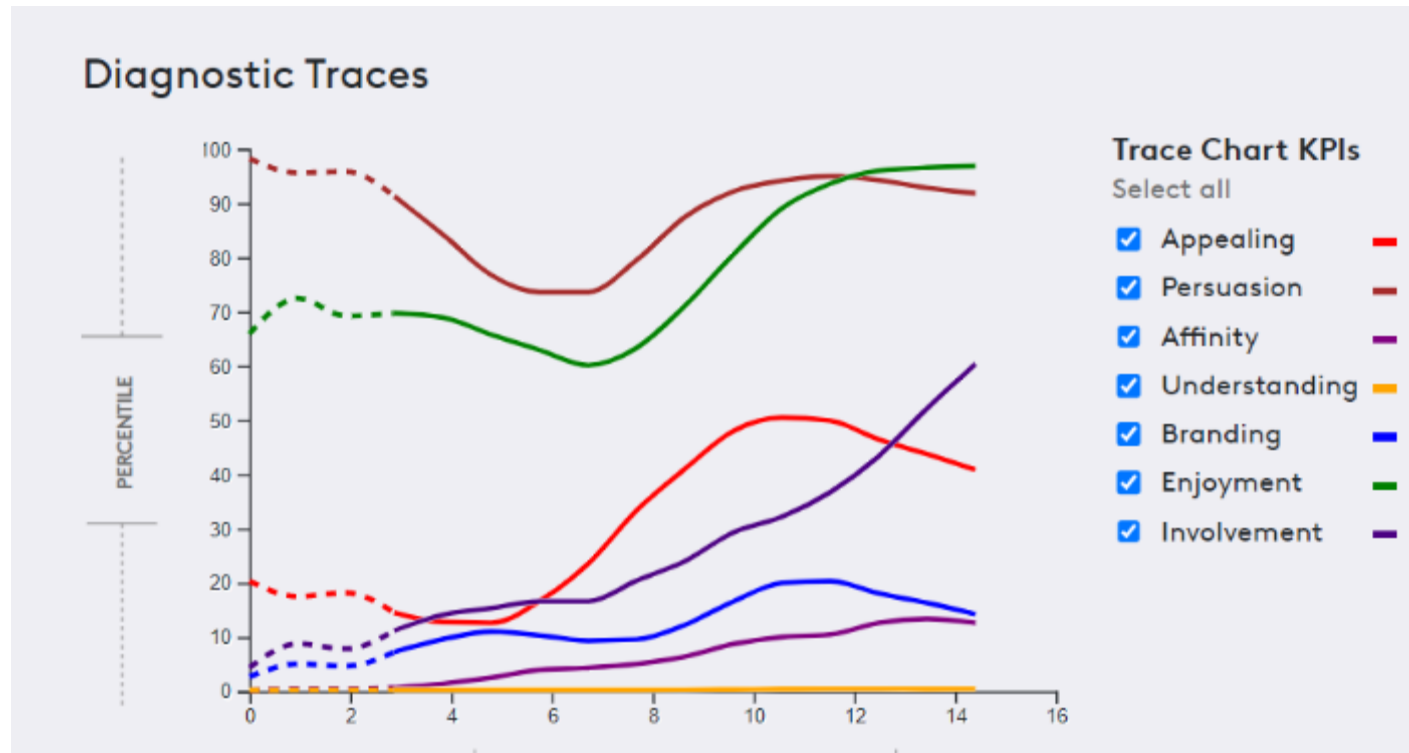
Behavioral metrics

How people interact with the ad



- Interaction
- Skip Time
- Action Likelihood
- Watchability
- Likeability

Undertaking Second by second analysis of ads



Using machine learning technology to extract **20,000 ad features** from a single ad

Performance prediction power improvements



R2=0.85+ without even doing any research

Will AI wipe out traditional research?

Some elements of what we do that is highly reproducible - believe it will create a hunger for research in other areas of research

A symbiotic relationship

AI needs to feed off enormous amounts of quality training data to get smart

Chat GTP 3 = Billions → Chat GTP 4 = Trillions

Challenges faced when creating large scale models

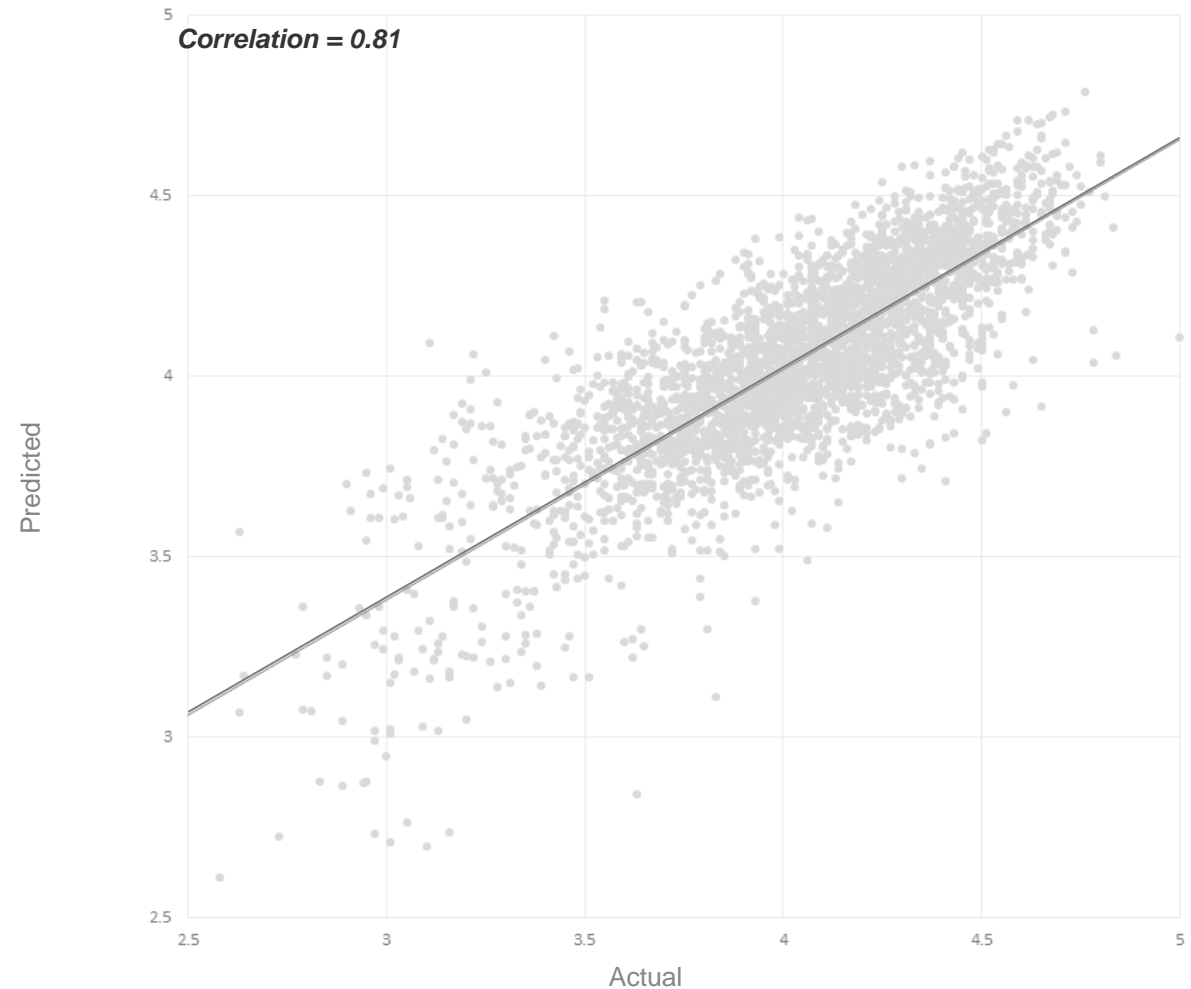
Catastrophic Forgetting: When many of the weights (where "knowledge is stored") change, it's improbable that previous knowledge remains intact. During sequential learning, newer inputs can overshadow older ones.

Model Drift and Decay: This happens when a model's performance deteriorates on fresh, unseen data, or when foundational assumptions about the data alter.

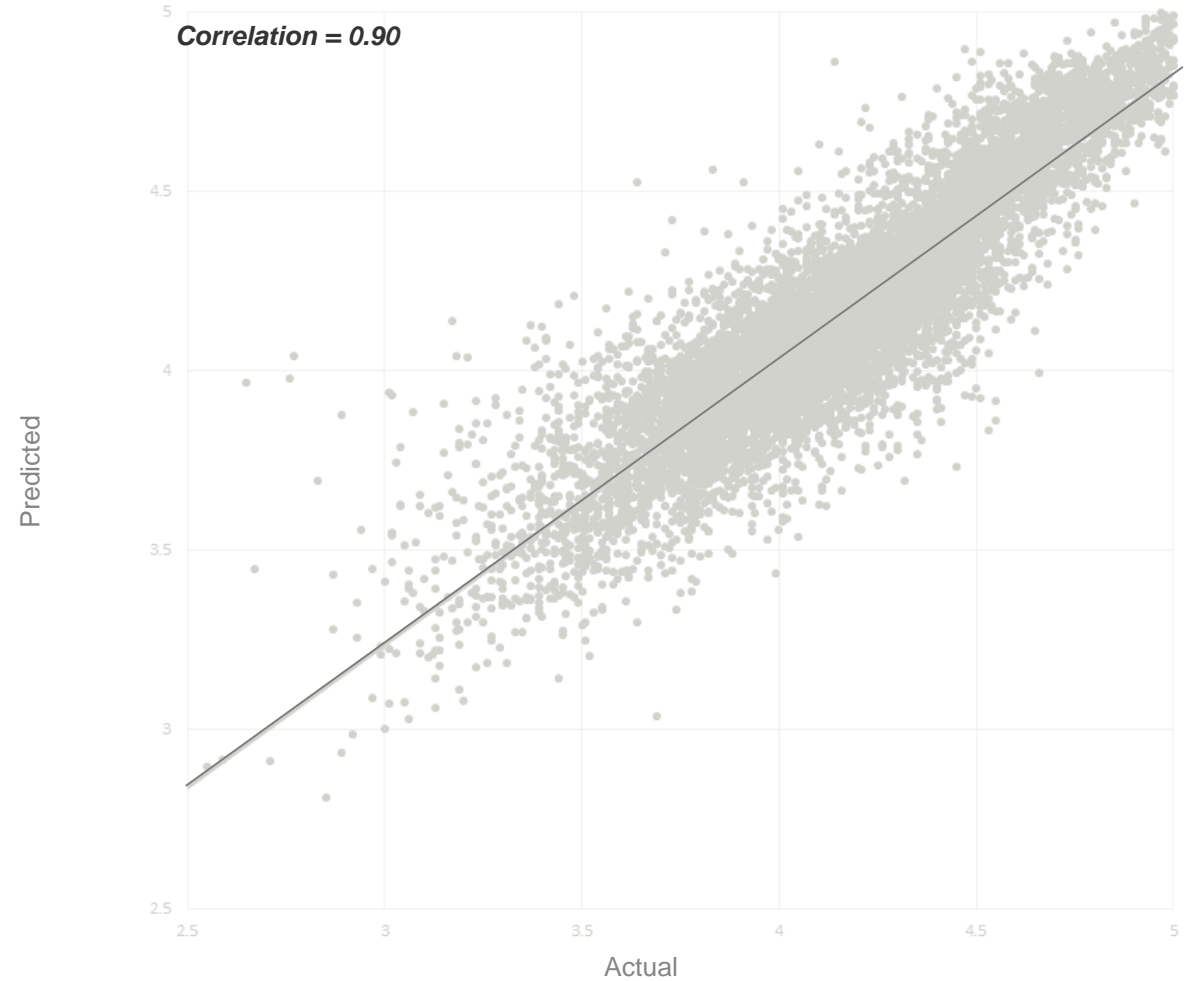
Diversity Collapse: This is when models begin to lose track of the authentic data distribution, even if that distribution remains stable.

Branding

**Link Ai Needed 20,000
ads to reliably
predicting one single
link measure
“Branding” with
correlation of 0.80**



To reach $c=0.90$
...needed to prime
with 200,000 ads



For each Incremental improvement in predictability AI requires an order of magnitude increases volume of data...

2,000 = c0.6

20,000 = c0.8

200,000 = c0.9

Thinking more broadly...

Will AI replace traditional data analysis techniques?

Pretty sure it will

Growth area will be using AI to analyse more everyday data...

**LLM's already in mainstream use for Qual Analysis
and text coding**

..but now can used for general data analysis



Personalitygpt.xlsx
Spreadsheet

can you you analyse this data and describe the key characteristics of poeple who Get annoyed by spelling mistakes

The top 10 characteristics that would be best to use if you were creating a test to predict if someone gets annoyed by spelling mistakes are:

1. Like my own company
2. Intelligent
3. Like to stick with things that I know
4. Don't like to draw attention to myself
5. Planner
6. Can get stressed out
7. Open to change
8. Organized
9. Prepared to compromise
10. Quiet

Using only these 10 characteristics, the predictive model (Random Forest Classifier) achieves an accuracy of approximately 74.4%. This means that the model can correctly predict whether someone gets annoyed by spelling mistakes in 74.4% of cases when using these 10 characteristics.



personalitygpt2.xlsx
Spreadsheet

Can you have ago at the same task with this data. Isolate the 10 characteristics the combined best predict if someone gets annoyed by spelling mistakes and indicate how acurate the preduction is

Principal component analysis

The top 10 characteristics that would be best to use if you were creating a test to predict if someone gets annoyed by spelling mistakes are:

1. Like my own company
2. Intelligent
3. Like to stick with things that I know
4. Don't like to draw attention to myself
5. Planner
6. Can get stressed out
7. Open to change
8. Organized
9. Prepared to compromise
10. Quiet

Using only these 10 characteristics, the predictive model (Random Forest Classifier) achieves an accuracy of approximately 74.4%. This means that the model can correctly predict whether someone gets annoyed by spelling mistakes in 74.4% of cases when using these 10 characteristics.



fpldata2.xlsx
Spreadsheet

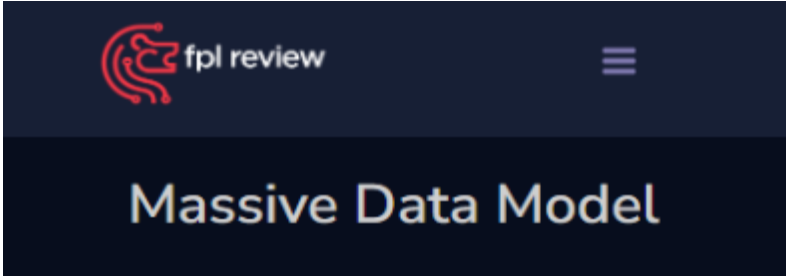
Can you work out which 15 player would have the highest overall 7_week_points_total with these conditions. 1. Total Buy_value must not be more that 100. 2. you must choose a total of 15 players but you can only pick a maximum of 3 players from each Team, exactly 2 from positions G, 5 from positions D, 5 from position M and 3 from positions F. Tip start by thinking of optimising the Points per million and iteratively replace players until you reach 100 of Buy_value



Certainly! We can approach this by iteratively replacing the player with the lowest "Points_per_million" value in the optimized team with the next available player who would score more points and help spend closer to the 100 budget, without violating the other constraints.

Here's the recommended selection of 15 players based on the new dataset and given constraints:

Position	Player Name	Buy Value	Team	7-week Points Total
G	Areola	4.2	West Ham	24.38
D	Tsimikas	4.5	Liverpool	24.26
D	Gabriel	4.7	Arsenal	24.96
D	Zinchenko	4.9	Arsenal	24.41
D	Burn	4.7	Newcastle	22.13
M	Salah	12.8	Liverpool	65.58
M	Mbeumo	6.7	Brentford	32.27
M	Palmer	5.0	Chelsea	24.00
M	Gross	6.3	Brighton	28.33
M	J.Ayew	5.4	Crystal Palace	23.59
F	Haaland	14.0	Man City	63.975
F	Wilson	7.8	Newcastle	22.33
F	Isak	7.5	Newcastle	25.64
G	A.Becker	5.5	Liverpool	27.60
D	White	5.6	Arsenal	24.68



Ascription...



are you able to do random forest ascription?



Yes, I can perform Random Forest-based attribution, which is often termed "feature importance" in the context of Random Forest models.



In a Random Forest model, feature importance is typically computed using one or more of the following methods:

1. **Mean Decrease Impurity (MDI):** This method calculates the total decrease in node impurity (typically measured using Gini impurity) averaged over all trees in the forest for each feature. Features that tend to split nodes on more "informative" criteria will have a higher importance.
2. **Mean Decrease Accuracy (MDA):** This method involves permuting the values of each feature one by one and measuring the decrease in the model's accuracy. A significant decrease in accuracy upon permuting a feature indicates its importance.
3. **TreeSHAP:** SHAP (SHapley Additive exPlanations) values are a more recent and advanced way to interpret Random Forest models. TreeSHAP provides a consistent way to distribute the prediction value across features.

If you have a dataset and a specific task in mind, I can guide you through the process of training a Random Forest model and extracting feature importance using one of these methods.

Needing the DP team to help analyse your data and the skill involved in doing this has always held back the types of research we can do.

Easier DIY analytics has to drive up demand for research?

Summary assessment

Using LLM's to creating synthetic data ? – I would be weary

Using LLM's to analyse data – We will be doing this every day this time next year

Using AI to fill in data gaps/scale data – Yes but not beyond the limits of statistics

Using AI to predict things outside of model scope – not in foreseeable future
(need a lot of training data to be reliable and very vulnerable to network errors)

Will AI replace traditional market research – No, AI can't live without it

LLMs role in research will clearly be significant to develop research more efficiently, summarise and analyse and facilitate doing research faster and more efficiency

AI will grow demand for ever more accurate training data.

If market research is to be use as the source of truth for AI solutions...

We will need to get our own house in order...

Ensuring quality and representative sample

Ensuring our research is delivering reliable unbiased answer