



BIT Review 2021-22





Data science and artificial intelligence	03
Predictiv	05
<u>Qualitative research</u>	08
Synthetic Data	10
TESTS	13





# Data science and artificial intelligence

Evidence is built on data. Data analysis has always been a core element of what BIT does, even if it's often behind the scenes. In 2016, we took a bet that data science – techniques now often referred to as artificial intelligence (AI) – could find high impact applications across the UK government. We were pleased to find that they could,<sup>1</sup> particularly in prioritising scarce resources, and in understanding complex behavioural patterns that standard statistics struggled with. Since then, government departments have created their own data science teams, and we've shifted our focus: How can data science extend and improve our work as behavioural scientists?



- We uncovered age-biased language in more than 11 million job ads using new techniques in natural language processing.<sup>2</sup> By using word embeddings, we were able to find problematic words beyond those identified in the academic literature, plus detect synonyms and other phrases used in the same way. This work created a fuller picture of the problem of age-biased language in job descriptions and laid the foundation for automatic tools to reduce it.
- In our larger trials, we can now see whether particular groups of people respond in different ways by using machine learning. This enables us to find patterns in how people respond that would be difficult or impossible to detect through traditional statistics.
- Using regional and national data, we've examined how crime clusters over space and time. We've also considered the role of anonymised data on violent incidents from emergency departments in identifying patterns to inform policy and prevention.

What's next? We're set to begin a large programme exploring how AI techniques can provide insights on suppliers and vendors to support social impact goals, such as net zero. Supplier information on social outcomes is critical, but often messy, complex and hard to access at the right time – the perfect challenge for us to tackle.

Forecasting negative outcomes for individuals – and trying to prevent them – is another area we are very excited to explore. How far in advance can a machine learning model predict when a student is heading towards dips in attendance, behavioural incidents or even exclusion? Are there cases where, if alarm bells were raised a little earlier, we could support people to get back on track?

It's an exciting time to be a behavioural scientist focused on data science. New tools and techniques are emerging every year, and with so many of them open and widely accessible from the outset, the time between invention and application is shorter than ever. We built our synthetic data tool<sup>3</sup> to accelerate high impact data projects across the public sector. It's already being used by Ofqual and the Ministry of Justice, and a similar approach is now being taken by the Office for National Statistics itself. We're happy to see the approach expand the scope and creativity of data science and research projects by connecting policymakers and analysts more closely to the data they need to innovate.

3 https://www.bi.team/blogs/accelerating-public-policy-research-with-easier-safer-synthetic-data/



<sup>2</sup> https://ageing-better.org.uk/resources/ads-for-all-ages

5

## Predictiv

Predictiv is BIT's policy testing lab. A one-stop shop for designing, running, analysing, and presenting the results of online experiments and surveys. Since its creation in 2016, Predictiv has run over 200 projects with BIT's public and private-sector clients.

Predictiv can be used to run any type of online survey or experimental design but really comes into its own with online randomised controlled trials (RCTs, also known as A/B tests), considered the gold standard of impact evaluations. In an online RCT, participants from an online panel are randomly assigned to two (or more) groups and shown different versions of communication materials, products, or services.

Predictiv has access to an extensive pool of participants. The size of this pool in the UK is over 500,000 people, in the US nearly 2,000,000, and Predictiv can also reach participants virtually anywhere in the world if required.

The fact that the assignment is random means that test results that show a certain test worked better in increasing a target behaviour or in decreasing an undesirable one can be stated with confidence. This means policy-makers and other decision-makers working with BIT can rigorously devise and, crucially, test the most effective intervention in every context before its actual deployment. Thanks to its bespoke methodology and reach, Predictiv is a powerful tool in tackling a multitude of challenges. Its most common uses are:

### 1. Identifying best performers

It is incredibly powerful to know with confidence the best ways of communicating, or the best policy, or the best new product. Predictiv provides this by comparing and measuring the performance of different versions of content or media against each other to determine which one is best.

For example, in the current climate change crisis, it is crucial to find the best way to encourage buyers to make eco-friendlier choices such as buying used items instead of new. Predictiv, together with one of our clients, designed a set of messages for a well-known marketplace and tested them in a simulated shopping environment. The aim was to see whether any of the messages led to an increased number of used items (hypothetically) purchased compared to no message (ie business as usual).

The result of the experiment was clear: messages emphasising cost-saving and environmental benefits associated with used items significantly increased the number of second-hand items selected compared to no message.



n = 3,003 \*\* p < .01 \*p < .05, +p < 0.1

Primary analysis, with covariates (age, gender, income, ethnicity, education, employment) Corrected for multiple comparisons



Both the 'Cost' and 'Environmental' messages increased the number of used items selected, compared to a scenario where participants were not shown a message.

### 2. Answer 'what if' scenarios

Sometimes the uptake of a scheme or initiative is lower than originally expected and stakeholders start wondering what changes might be most effective at increasing its attractiveness. Should eligibility criteria be changed? Can the sign-up process be improved? What about the words used to communicate and explain what it is? How do you know what changes would actually make a difference? (And which would only make things worse). Testing changes in this way in the real world entails a big gamble but with Predictiv it is possible to scope the impact of changes in a sandbox environment before any are implemented in real life.

### 3. Investigate public opinion

Opinion surveys are a useful way of investigating the general public's or a specific sector of the population's attitude or sentiment towards a certain topic.

Predictiv provides a streamlined and efficient way to conduct large scale surveys with quick turnaround times.

### 4. Measure behaviour in a simulated environment

By observing people's behaviour in a simulated immersive environment that closely mirrors real life, Predictiv allows the collection and analysis of valuable insights that are more likely to be replicated in the real world than can be produced with survey-style research and testing.

For example, Take A BITe is Predictiv's own simulated food delivery platform. It has been designed so elements of the page, such as the order in which options are shown, or the amount of information given for each choice, can be easily changed and altered. In this way Predictiv can measure precisely how different ways food is presented affect the choices people make.

This can play an important role in tackling a major public health issue in many countries today - obesity. Great efforts are being put into tackling it and the involvement of food delivery platforms is crucial to success. Predictiv's Take A BITe has explored how changes in the choice architecture could nudge people towards healthier food choices.

Predictiv is not only a flexible platform well-suited to give answers to wide-ranging policy challenges, it is also a time and cost-effective way of doing so. Over the next year, we plan to expand our capabilities so that we have ever more functionality for our clients available at low cost. We're really excited about applying our current capabilities in new ways as well as seeking out new opportunities for innovation and challenge. Please get in touch if you have any questions, ideas or projects - predictiv@bi.team.

### **Qualitative research**

A core part of understanding problems and creating solutions is to appreciate that everyone has a different perspective of the world. We seek to understand these perspectives better by observing, engaging and co-creating with people to get to know their circumstances, constraints and priorities.

We weave qualitative research across projects, using a wide variety of methods, including card sorting, sketching, mapping, testing, deliberative assemblies, observational and ethnographic techniques, traditional interviews and focus groups to name a few. Our approach enables innovation and creativity, and challenges us to find respectful and meaningful ways to engage with people.

As part of our early problem exploration work we make significant efforts to incorporate qualitative research and insights to inform our work. This is because these activities help us ask and answer different questions - notably 'What?' 'How?' and 'Why?". For example, we might be interested in why a particular group of people makes certain sexual health choices, how a multi-site intervention aimed at helping workers build relationships with young people at risk of exploitation has been implemented, or *what*, from the perspectives of domestic violence survivors, impacts womens' decisions to engage with an NGO.



By identifying contextual and environmental factors, and using them to inform and develop solutions, we are able to better inform policies, services and products. Qualitative research also lives within our RCTs and evaluations. When seeking to understand whether policies and interventions work, it's the contextual work we do that helps us get at what's really going on and contribute these insights to the evidence base. We also know that sometimes what we need to land a point isn't an effect size, but a story.

For instance, one of the main reasons why we were able to design and evaluate interventions empowering adolescent girls in Kenya to make more positive choices about their sexual and reproductive health was because we found non-verbal engagement methods that addressed the taboo of open discussion of such topics. Traditional research techniques would not have done this.

Meeting people where they are upends the power dynamics between the researcher and the researched as we enter into their worlds. Exciting work from our Gambling Policy & Research Unit demonstrates this. An online gambling audit project meant that our team literally stood in the shoes of players as they assessed real-world gambling websites for harmful practices.

### What is around the corner?

We plan to grow and refine our design-influenced work. Recent collaborations with the design team at UK innovation foundation, Nesta, are helping us to incorporate an even wider range of qualitative methods, including human-centred design and testing, taking a more holistic design-informed approach throughout the research process and using creative approaches to co-design research.

Global

10

# Synthetic Data

Rigorously analysing good data is critical for BIT's work. We often share and receive data from our partners and other collaborators. This helps give us a broader picture of the issues we're trying to solve and allows us to look for deeper patterns and insights that can deliver even more socially impactful results.

Unfortunately, how the process of sharing data usually goes is far from efficient. Think of it like a dysfunctional car market.

Imagine that before you buy your vehicle, you can't see it, can't test drive it, can't read reviews of it (because there aren't any), and will have to wait up to six months for delivery. When your car arrives, it might be quite different from what you were expecting - and missing lots of parts. Not to mention that before you're allowed to buy it, you have to justify exactly why you want the car, how you will use it, and when you will destroy it.

Of course, such a market wouldn't last for long.

These data sharing challenges exist because of strict governance and protections required for data handling. But that doesn't mean the approach to sharing data needs to come with all those delays and issues. The solution is 'synthetic data', which can dramatically reduce these obstacles without sacrificing safeguards.

Sounds too good to be true? How does it work?

Synthetic data is a new copy of a data set that is generated at random, but made to follow the structure and some of the patterns of an original data set. It can be made at low or high fidelity, that is, to be more or less accurate to the original data source.

Because synthetic data doesn't contain any information from the original data set - and is completely different in its own right - it doesn't compromise data security. But it is close enough to the original so that it can be loaded into normal tools to see if it will work as intended, and code can be written and debugged with the synthetic set in advance of the original set being available. All of this helps to speed up the process, potentially quite considerably.

While making progress with the synthetic data, partners can decide on appropriate protection protocols for the actual data set. As soon as those are in place and the original data set is shared, researchers can start working with it immediately. The preparation that previously could only begin once the data was transferred has already been completed. Returning to the car analogy, synthetic data allows you to test drive a car that isn't the one you're contemplating buying, but is so similar that you'd never be able to know the difference.

BIT advocates for expanded understanding and use of low fidelity synthetic data sharing. The UK public sector collects a wealth of data. Embracing synthetic data will result in greater innovation and use of data science to accelerate and enhance policy research without compromising data security.

No method existed for creating safe, low fidelity synthetic data, until our BITx team built one. We have since been working with the Ministry of Justice, Department for Education and others, and with our partners in ADR UK (Administrative Data Research UK) to advance this exciting initiative.

These data sharing challenges exist because of strict governance and protections required for data handling.

### 2016 Olympics: Low-fidelity synthetic data



Each point is randomly generated from the general statistics of height and the general statistics of weight at the 2016 Olympics.

No points correspond to a real person. Weight **increases** with height, on average.





Each point is randomly generated from the **general relationship between height** weight at the 2016 Olympics

No points correspond to a real person. Weight increases with height, on average



A quick example: the 2016 Olympics

Height and weight of every athlete at the 2016 Olympics.

**Every point corresponds to a real person.** Weight **increases** with height, on average

Global



# TESTS

TESTS is BIT's methodology for running projects: **Target, Explore, Solution, Trial and Scale.** 

The methodology was developed by BIT and has been refined through use in over 1,000 Behavioural Insights (BI) projects around the world over the past 10 years.

While not every BI project will go through all five stages, it has proven to be a helpful structure for guiding much of BIT's work - whether it's exploratory research that results in policy advice (and ends with the 'Solution' stage) or a project that goes all the way to recommendations for scale-up of a rigorously tested intervention.

Importantly, the TESTS framework itself has evolved alongside our work: When we speak about Trial today, we don't mean this to only refer to randomised controlled trials (RCTs), but a whole range of evaluation methods. Similarly Scale was added to the framework a few years ago as we started to think more about how our solutions could be embedded in systems more long term.



### 1. Target

Encouraging or maintaining positive behaviours is the ultimate aim of most public policies. However, we have found that policy or programme objectives can often initially be too broad and not explicit about which behaviours they target. Target is the crucial first step of a BI project to determine which behaviour or behaviours to focus on and how to measure it.

We work with our partners to narrow down the broader policy challenge to specific populations and behaviours of interest. Ideally, we determine a short list of target behaviours at this stage - but in some cases we might also conduct some Explore work first.

### 2. Explore

In order to understand how any proposed interventions may change target behaviours, it's essential to understand the context in which they take place or the behaviours themselves more deeply. No amount of reviewing literature can replace speaking to and observing people and those around them. During this phase, we combine qualitative and quantitative research methods with our expertise in behavioural science.

Our approaches range from analysing large datasets to understand at what stage of a process end users drop out to using observations and interviews to understand experiences in depth. We then use insights from the behavioural sciences to interpret our findings and identify barriers to target behaviours.

### 3. Solution

This stage involves designing an intervention that is grounded in the Explore findings, insights from the behavioural science literature and feedback from the people we design for. This is also where we may use our EAST framework to generate ideas (see page \*\*\*). These solutions can consist of 'nudges', the light touch and low cost interventions that BIT was initially known for, as well as a wide range of other tools and techniques from the world of applied behavioural science.

We may start with a long list of ideas, which we narrow down gradually to a couple of potential solutions after assessing feasibility and potential impact of each and gathering a first round of feedback from stakeholders. We then regularly user test the most promising ideas, to get feedback and to weed out as many problems as possible before we take our idea to larger-scale testing - because you really don't want to set up a RCT and then realise your solution didn't have the desired impact just because of a very fundamental flaw, which any end user could have pointed out in 5 minutes.

The end goal of this stage is one intervention that has been fully designed, user tested, approved, and is ready for implementation, for trial in the real world.

14

### 4. Trial

Evidence-based policy requires knowing what works and what doesn't. We therefore run impact evaluations to fully understand the effects of intervention on the target outcome.

The most rigorous way of establishing the impact of the solution on the behaviours we try to change is an RCT and we continue to use them where possible and where they generate the evidence we need. But especially for more complex solutions and settings, it is not always feasible or desirable to rely (only) on an RCT. In that case, we have experience in a number of alternative evaluation approaches - ranging from quasi-experimental designs to qualitative implementation and process evaluations.

Importantly, we know that data collection can often be costly. This is why we try to rely on administrative data as much as possible - if we work to improve educational attainment, we try to rely on routine exam results where feasible rather than administering a new test; if we aim to increase cancer screening, we use GP data rather than a survey.

This penultimate stage will provide evidence that enables a decision to be made on whether or not to scale up the intervention being tested.

#### 5. Scale

This is the point where the results of the project are reported and successful solutions are put forward to be scaled for maximum social impact. Scaling can take a lot of different forms: changing policy so that a recommendation is included in a new law or regulation; rolling out a solution to the whole target population (rather than just a small sample) or implementing a solution again in the future.

If the intervention did not have the intended impact then we aim to understand the reasons behind this and discuss implications with our partners. We believe strongly that not only those projects that lead to significant improvements in outcomes are successes: firstly, it is as important to know what doesn't work as it is to know what does - after all, governments and organisations want to spend money where it has the biggest potential for impact.

Secondly, each phase of our projects generates valuable insights in its own right - so a better understanding of barriers gained during the Explore phase can be just as valuable as a significant Trial result.



Global

15