Department of Natural Sciences

Data Science Research Project (DATA40160)

Can Natural Language Processing capture the "Significance" and "Reach" of Impact Case Studies from the Research Excellence Framework

Mateusz Kopaczewski

September 2024

Abstract

As part of the Research Excellence Framework (REF), each Higher Education Institution (HEI) and each discipline (also known as "Unit of Assessment") must submit an impact case study which demonstrates the impact their research has had outside of academia on the economy, society, culture, policy and more. This research aims to identify the patterns between the class 4* and 3* rating of impact by employing the increasingly complex machine learning models, including XGBoost, BERT and Longformer, to classify impact case studies and interpret the predictive features behind the models' decision-making through the topic modelling and interpretability techniques.

The study found 40 topics in Panel A impact submission, particularly related to healthcare and medicine, with certain topics associated with cancer research with consistently higher scores. However, as seen in other pioneer literature, this has no significant predictive power. The simple models, such as Naive Bayes with Bag-of-words, demonstrate substantial institutional bias and capture only surface-level predictive features, like "estimating" and "guideline". In contrast, the best-performing model of XGBoost with Longformer embedding and the integration of handcrafted features could capture the semantic context, providing more information in the decision-making of the model according to the macro F1-score.

Introduction

A number of countries have introduced some sort of performance-based system that is responsible for the allocation of public funds for research. These nationwide evaluations vary greatly with different emphasis on expert judgement and peer review and more quantitative indicators of research quality (Siversten, 2017). Despite the peer-reviewing process being overly expensive and laborious for both academics and expert evaluators, it remains the most robust method for evaluating research quality because it can capture broader context of research quality that cannot be captured by metrics alone (Harbach and Halffman, 2020; Guthrie, Ghiga and Wooding, 2017; Heesen and Bright, 2021).

The United Kingdom has a long history of using expert peer-reviews to evaluate research quality dating back to the first Research Assessment Exercise (RAE) in 1986. The RAE was introduced as part of the complex restructuring of policy which required universities to justify their use of public expenditure by demonstrating high-quality research (Robinson, 1986). Subsequent assessments occurred until 2014 when it was replaced by the Research Excellence Framework (REF). The REF was largely unchanged to its predecessor continuing to evaluate research quality based on expert peer reviews of the submitted outputs and environment narrative (Pinar and Horne, 2022). However, with the REF HEIs were also asked to submit a narrative of how the impact of their research outside of academia.

HEIs are required to submit three differently weighted elements: outputs (60%), impacts (25%) and environments (15%) for each UoA. This weighing for impacts was around half of the weighting given to the outputs in the overall quality profiles. However, the value of impact case studies should not be underestimated. According to the publicly available data on QR research funding allocation to HEIs, the median value for a single 4* impact case study is around £880k for panel A (Medicine, Life and Health Sciences) where the actual value is dependent on the number of full-time equivalent (FTE) staff (Collet, 2024). There is a higher ratio of impact case studies to the number of outputs meaning that a single impact case study can contribute more to the overall quality profile than an output. This ratio increases sharply with higher number of full-time equivalent (FTE) staff, as the requirement for number of impact case studies per several FTE staff decreases sharply (Kerridge, 2021). This means that achieving a 4* impact case study could be disproportionally more valuable than a 4* output especially for large HEIs.

This has led to a change in strategic and research focus by HEIs as they are required to reconceptualise their perception of high-quality research to focus on the broad societal impact of research (Watermeyer and Chubb, 2019). The main criteria for classifying the impact score are based on the significance and reach of the research papers. The significance describes the extent to which the impact of research alters and benefits society, whether it is in a positive shift in performance, awareness, or practices of the "beneficially". Reach points to the range of the research's influence across various groups or sectors (Pollitt et al., 2023). Leading to HEIs focusing on fostering more internal structures and dedicated funding avenues that can support researchers at planning, measuring and demonstrating societal impact (Grant, 2012). Consequently, there is a growing body of literature into aspects of how impact is evaluated to get insight into ways that impact case studies can be improved. Whilst quantitative indicators can be used by these analyse to measure impact, e.g. bibliometrics like citation counts or altimetric to capture citations in policy documents (Williams et al, 2018) , the actual score of impact case studies is subjective because it is based on how well the narrative is constructed and interpreted by the board of panellists.

There is a range of research that suggest the subjectivity of impact evaluation including suggestions that a “strong narrative” can distinguish between high and low scoring impact case studies (Pidd and Broadbent, 2015). These scholars emphasise the influence of rhetoric over the content (Watermeyer and Chubb, 2018), with compelling and well-crafted impact case studies being unconsciously favoured over less well-written impact case studies with the same impact (including Watermeyer (2019), Gow and Redwood (2020) and McKenna (2021). Others highlight the lack of evidence for assessing rhetoric over substance, suggesting there are difference in language and narrative, but they only have marginal influence on the REF outcome (Reichard et al, 2024). Scholars have also highlighted the subjective nature of assessing interpretations of impact. With individual panel members and the dynamics within the group having huge influence on what is considered worthwhile impact (Oancea, 2013). A panellist from an interview conducted by Watermeyer and Chubb (2018) agreed, noting that "the panel had quite an influence on the criteria", admitting that certain types of impact that are more easily quantifiable could be favoured (Weinstein et al, 2019). Especially when dealing with unfamiliar impact when panellists often resorted to their own experiences and biases to assess impact (Reichard et al, 2020).

However, there has been limited research exploring how Natural Language Understanding (NLU) can be leveraged to uncover intricate patterns between high- and low- scoring impact case studies. NLU, a field mostly saturated by computer scientists, is now attracting attention from a wide range of disciplines, including psychology, sociology and linguistic. NLU allows machines to understand the meaning behind text by extracting a variety of features that each uncovering dimensions of language, such as syntactic and semantic relationships.

With the recent breakthrough in deep learning have revolutionised Natural Language Processing (NLP) sparking excitement among researchers about the advancements in the capabilities of machines to further understand language and text generation, highlighting a progressive step in NLU. The pinnacle of advancements can be traced to the two fundamental components: the introduction of the transformer neutral network architecture, presenting in the "Attention is all you Need" paper published by a group of Google researchers, and the application of transfer learning, a well-established technique in computer vision, into NLP. The compound benefits of these techniques have led to a shift from using lightweight models for NLP tasks, to leveraging the deep contextualised understanding of large language models. These models can capture deep syntactic, semantic and discourse-level features. By fine-tuning them on specific downstream tasks, they can achieve better results in many NLP tasks with fewer labels (Sebastian Ruder et al, 2019).

This study aims to leverage topic modelling to uncover underlying themes and structures within impact case studies. However, a key challenge is that deep learning methods, including large language models, inherently lack interpretability. To address this, the research proposes an empirical approach to evaluate the patterns captured by these models and assess their effectiveness in capturing the significance and reach of impact case studies. This approach involves comparing the behaviours of complex models like BERT with simpler, more interpretable supervised and unsupervised machine learning models, while also using topic modelling to identify key themes within the case studies. By understanding how these themes correlate with high and low impact scores, we can provide clearer insights into the influences of topics and their associated score toward the predictive features of the advanced model. These insights will be integrated into predictive models utilizing advanced NLP techniques, such as BERT and transformers, to explore whether the semantic and contextual features captured by these models can effectively differentiate between high- and low-scoring impact case studies. Therefore, the research involves investigating the patterns captured in impact case studies and aims to answer the following research questions:

Research question 1: What implicit and explicit aspects of the impact criteria can be captured by machine learning models?

Research question 2: To what extent can machine learning models capture the significance and reach of impact case studies?

The research project is structured as follows: Chapter 1: Introduction provides an overview of the research area and motivation behind it, along with the main research questions. Following by Chapter 2: Literature Review details the background of Research Excellence Framework (REF) and the significance of impact case studies (ICS). It highlights the importance of high ICSs and discusses potential biases and limitations of expert evaluation, particularly from the distinct cultural and disciplinary influences as well as ethical concerns. This section also covers the existing literature on natural language processing (NLP) and automated essay scoring (AES) as a potential alternative tool for impact assessment. In Chapter 3, Methodology provides information on data collection and preprocessing, the model utilised, and their evaluation methods. Chapter 4 presents and discusses the key findings, including essential features, topic modelling and model performance. This will follow with the last chapter 6 of the conclusion, which includes limitations and suggestions for future research.

Literature Review

Section Introduction

The literature review section will begin by formalising the Research Excellence Framework (REF) and its focus on the impact agenda. It describes the historical development of the currently employed REF assessment while exploring the mechanisms used for evaluating the impact of case studies with the essential criteria of reach and significance. The process involved in evaluating and classifying the quality of research impact is detailed, with the information addressing criticism of the peer-review process and the influence of the linguistic features according to pioneer literature. The section will further highlight the key studies on machine learning, narrative style and topic modelling in the context of impact case studies, identifying the present gap in the existing literature that requires further exploration.

Background of the Research Excellence Framework

In the United Kingdom, the Thatcher Government undertook a series of initiatives as part of a broader policy restructuring aimed at reducing public expenditure (Robinson, 1986). As a result, publicly funded organisations were required to demonstrate they were operating with economy, efficiency, and effectiveness (Rhodes, 1994). This shift significantly impacted higher education institutions (HEIs), leading to the introduction of the Research Excellence Exercise (RAE). The RAE fundamentally changed the way that Higher Education Institutions (HEIs) were allocated funding by introducing a systematic and standardised method for evaluating research quality across various institutions and disciplines. The government's objective with the RAE was to provide accountability for public investment by asking institutions to prove the quality of their research to secure funding.

Over the coming years, RAE assessments would be carried out periodically, with each iteration being refined to address the criticism of the previous one, making the process increasingly sophisticated and robust at assessing the research quality. However, as the RAE became more sophisticated, it also became more disruptive, burdensome and expensive. This led the treasury to abolish the RAE and propose a cheaper, less onerous system that relied more heavily on metric-based evaluation. However, fierce opposition from the academic community about the abolition of peer review resulted in the rejection of the metric-based system (Macilwain, 2009). Instead, the new treasury replaced the RAE with an equally controversial Research Excellence Framework (RAE), which remained largely the same but included a new "impact agenda". This required HEIs to demonstrate the effect of research outside academia (Silverstein, 2017). The treasury's initial motive for the new impact agenda, especially following the repeated economic downturns, was to encourage research with higher economic rationale (Barker, 2007), although over time the understanding of the criteria for impact broadened to include cultural, social, and environmental impacts to provide a more holistic outlook on impact.

The REF and Impact evaluation

This study has introduced the background behind the inception of the Research Excellence Framework and the impact agenda. This section will now focus on formalising the mechanism of evaluating impact case studies in the REF. The REF evaluates research quality at the disciplinary level through Units of Assessment (UoA). These UoAs are organised into four broad panels, each responsible for evaluating specific Units of Assessments (UoA) within their remit. Higher education institutions have the flexibility to choose to which UoA they wish to submit their research to. For each UoA, HEIs are required to submit three differently weighted elements: outputs (60%), impacts (25%) and environments (15%). The number of submissions to each element is determined in accordance with the number of Full-Time Equivalent (FTE) staff members within each UoA. FTE is a standardised metric that reflects the proportion of full-time work. For example, a part-time staff member working 10 hours per week roughly equals 0.25 FTE. Focusing specifically on the impact aspect of the Research Excellence Framework. Each HEI must submit at least one impact case study, rising to two if it has between 20 and 24.99 FTE staff and an extra case study for every 15 FTE staff beyond that point.

These impact case studies must provide a compelling narrative that provides a link between the research conducted by the institutions and the impact. The REF guidelines define impact as: "an effect on, change or

benefit to the economy, society, culture, public policy or services, health, the environment, or quality of life, beyond academia" (REF, 2021). However, in practice, each discipline interprets and defines impact somewhat differently within its respective guidelines, as the norms and traditions of each discipline shape both the definition of impact and what is deemed to be a worthwhile impact (Oancea, 2013). The narrative of impact case studies often draws from multiple research outputs over a pre-defined research period.

The impact case study narrative must follow a predefined format, which has five sections:

1) Summary of Impact
2) Underpinning Research
3) Details of Impact
4) Sources to corroborate impact
5) References to research

The evaluation consists of a dual peer review where panel members assess the narrative of impact case studies, followed by a moderation stage by panel and sub-panel chairs. The explicit criteria from the REF guidelines emphasise assessing both the “reach” and “significance” of the impact that HEI has achieved outside of academia. Despite this, some studies suggest that implicit factors and subjectivity influence the outcome of impact evaluation (this will be discussed further in sections 1.6 and 1.7).

The importance of good impact for Higher Education Institutions

Higher Education Institutions (HEIs) have been subjected to increased accountability for their research through various assessments. In recent years, this accountability has taken the form of the impact agenda within the Research Excellence Framework (REF), which encourages HEIs to consider the societal impact of their research and forces them to reconceptualize their perception of what constitutes high-quality research (Watermeyer and Chubb, 2019). This shift is not only philosophical but also carries significant practical implications for HEIs.

Firstly, based on the 2023-2024 financial year, the allocation of mainstream QR funding for impact case studies reveals that a 4-star impact case study is estimated to be worth approximately £110,000 for Panels A and B and around £60,000 for Panels C and D, based on the REF2021 cycle. However, the actual value can vary significantly between institutions depending on the number of FTE staff and impact case studies submitted to each Unit of Assessment (UoA). This means that for larger submissions, the number of 4-star impact case studies can disproportionately be more valuable than individual research outputs (Collet, 2023).

In addition, the overall quality profiles generated by the REF are used to inform various national and global league tables for universities. These profiles contribute to the reputation of HEIs, affecting how prospective students, staff, and collaborators perceive the institution. As Duarte, Alves and Raposo (2010) note, “The image of HEIs is critical for their competitiveness”.

Subjectivity and Bias in Impact Evaluation

The UK REF 2021 evaluated approximately 6361 impact case studies at the unit of assessment level, giving each institution a score on a scale ranging from 4* (world-leading), 3* (Internationally Excellent), 2* (Recognised Internationally) and 1* (recognised). An extra category, Unclassified, also exists from research that cannot be assessed against the REF criteria. The impact case studies and their distribution of scores are in the public domain, providing a unique opportunity for researchers to analyse the construction of impact case studies.

However, it is essential to consider how data used in these evaluations is never entirely objective. The data is not simply the "empirical stuff" (Dourish, 2022), but it is "imagined and articulated" by humans who may perpetuate their existing subjectivities and biases (Bigo, Isin & Ruppert, 2019). This subjectivity is present even in processes like the REF, where, despite peer-reviewing being widely viewed as the best alternative to other metric-based systems, such as systems that rely on bibliometrics, which tend to disadvantage certain groups of people (Thelwell, 2023). The process of evaluating significance and reach is a complex process that can be assumed to require complex domain knowledge. Therefore, while the REF provides transparent guidelines, the extent to which conscious or subconscious bias of expert panellists can affect their interpretations is unclear. This uncertainty raises concerns regarding the potential for different biases to influence the evaluation process's

outcome, leading to the deviation between the expected outcome and the actual assessment of the significance and reach of the data.

Despite this, many predictive models operate under the assumption that the target variable is unbiased. To this extent, most studies assume that the expert panel members are correct in their assessment of the significance and reach of impact case studies. As such, any deviation from the REF outcomes and the results of any predictive model that disadvantages a certain group alludes to the presence of bias. Nevertheless, even if efforts are made to correct any existing bias by the expert panel members, studies should still aim identify potential entries for bias and mitigate the consequences.

One type of bias that can occur in human and predictive evaluation is institutional bias. Institutional bias can be defined as a situation where an institution is systematically disadvantaged over another institution, not because of any differences between quantifiable impacts but because of existing prejudices. Another type of bias can be disciplinary, where a certain type of impact from the specific discipline that is more easily quantifiable (e.g., medicine or engineering) is favoured as opposed to disciplines where the impact is less quantifiable. Alternatively, certain types of impact could be undervalued in specific disciplines. This can skew the generalisability of our model. However, in practice, each discipline interprets and defines impact somewhat differently within its respective guidelines, as the norms and traditions of each discipline shape both the definition of impact and what is deemed to be a worthwhile impact (Oancea, 2013)

Criticism of the impact agenda

There is a debate about the extent to which higher education institutions should be held accountable for the broader implications of their research. Some scholars emphasise the social responsibility that researchers must contribute towards society in return for the public funding they receive, encouraging research that has long-term societal benefits (Oceana, 2018). However, other scholars scrutinise the impact agenda for applying market logic to higher education institutions. If researchers are required to pursue specific types of more easily quantifiable impact, this could stifle their academic freedom (Weinstein et al., 2019). It presents the undesirable academic movement of competition, subjecting the learning process and outcomes to being perceived as 'commodities,' which contrasts with the original objective of learning (Fairclough, 1995; Brown & Carasso, 2013)."

In addition, the REF is a laborious process for 1) the institutions that select the content and craft the impact and environment statements, 2) the thousands of experts in sub-panels peer-review the submissions. This takes a substantial amount of time; consequently, there has been a broad range of research that has tried to replace some of the decision-making of the REF panel (Gilles, 2008; Brassington 2022: Pinar and Home 2022: Watermyer and Derrick and Borras Batalla, 2022).

Descriptive and Predictive analysis on Impact Case Studies

A large body of literature has performed text-mining techniques to identify words and phrases that occur in high-scoring impact case studies. Some studies used the distribution of words and phrases to identify topics and themes across impact case studies. Thorpe et al. (2018) highlighted distinct linguistic variations between high and low-scoring environmental statements, a similar methodology was later employed by Reichard et al. (2020) in their analysis of impact case studies. Their studies identified the importance of language choice in crafting impact case studies. Thorpe et al used computer-aided analysis tools to analyse the differences between ranked environment statements in the Business and Management UoA. Thorpe et al.'s study concluded that lower-scoring submissions were less likely to use passive voice, more prone to incoherence, tended to adopt a "finished article" style of discourse, and “cited specific details rather than generalities”. Reichard et al supported these findings in impact case studies by performing thematic analysis and identifying differences in the frequencies of words and phrases between higher versus lower-scoring impact case studies. Reichard et al identified that writing styles and clarity might influence the evaluation outcomes of impact case studies. Their study found that high-scoring case studies conform to clear, direct writing which was often simplified in how it represents the connection between research and impact contrasting with the more complex ones. In contrast, lower-scoring case studies tended to have less distinctive and more cautious language which was less confident in its assertion of impact. Lower-scoring impact case studies also tended to have more “filler phrases” that could be associated with “academese” writing and were less likely to have a clear structure using sub-headings and paragraphs. This

lack of clarity and organization further contributed to their lower evaluations. Thorpe et al. (2018) and Richard et al. (2020) findings suggest that outside the explicit criteria, there may be some implicit factors that affect the quality of impact case studies. However the analysis has been done on a small subset of impact case studies, we cannot be certain how the result generalises to the whole dataset.

On the other hand, in her later study, Reichard et al. (2022) clarified that there are not enough statistically significant differences between high and low-scoring impact case studies to suggest that language has enough influence on the score. She claimed there are fewer differences than similarities between the linguistic aspects of low and high-scoring impact case studies. Nevertheless, the concept that well-crafted case studies score better than poorly written case studies with equal impact was mirrored by many scholars (including Watermeyer (2019), Gow and Redwood (2020) and McKenna (2021). Therefore, further analysis is needed to determine the precise extent to which language affects the evaluation of impact case studies.

Several studies have shifted their focus away from specific stylistic differences in language, instead, they utilise topic modelling techniques to identify tangible topics and themes within impact case studies (see section 2.1 for more detail). A notable example is a study commissioned by the UK Research Funding Council (King's College London and Digital Sciences, 2015), which aimed to identify and categorise the societal benefits that arise from public-funded research. The study utilised computer-aided topic modelling techniques to identify 60 unique impact topics and 3709 unique pathways to impact. The research concluded that UK HEIs had a diverse wide range of impacts, including health, public policy, environment and culture with a wide range of beneficiaries spanning multiple countries. However, the diversity and complexity of impact interpretations, coupled with the sparsity of certain impacts, present significant challenges in quantifying impact, given the varied and often limited nature of the evidence supporting these impacts.

Terama et al, (2016) built on the King's study by using an alternative computer-assisted topic modelling technique. The study identified less granular but more interpretable topics allowing for a more detailed analysis of variations between topics. The study corroborated King's study claims of diverse interpretations of impact but added that the interpretations of impact varied across disciplines and institutions, with broader institutions having a larger variety of impacts. Most importantly, the study found no specific interpretation of impact consistently received higher scores based on the Research Excellence Framework (REF) criteria. This reinforces the complexity of impact evaluation; the topic alone is not sufficient to determine the significance and reach of research, suggesting a multi-dimensional approach which considers more nuanced aspects of significance and reach is required. However, the study had low granular categories which might mask specific trends or patterns.

The University of Oxford study, commissioned by the British Academy and the Academy of Social Sciences (hereafter referred to as the Oxford Study), represents the most comprehensive analysis of impact case studies to date. This study is especially significant due to the considerable resources dedicated to thoroughly understanding the effects of SHAPE research on individuals, the economy, policy, and society. The study used state-of-the-art BERT topic modelling which unlike the previously employed topic modelling techniques can capture semantic meaning by considering the context of surrounding words. This approach allows the study to identify 84 more nuanced topics across almost 4000 SHAPE impact case studies from social sciences (Panel C, N=2,146), arts and humanities (Panel D, N=1,528) and Psychology, Psychiatry and Neuroscience (Unit of Assessment 4, N=326). The study further reassigned 1,366 Impact Case Studies using human reviewers because they couldn't be effectively categorising into any specific topic. Furthermore, the study extracted various types of linguistic features and broad-level features on the sources of income, environment and geography of the impacts. These features and topics were used to identify beneficiaries of SHAPE research, tracing funding sources, quantifying grant income, and mapping the impacts on specific groups, organizations, and sectors.

The Oxford study did not examine the relationship between these features and impact scores, but it has made the underlying research data publicly accessible. Therefore, this research has conducted a regression analysis with these features to get further insight. The result collaborates with the findings of Terama et al. (2016) which indicate no correlation between thematic topic and Research Excellence Framework (REF) outcomes. As anticipated, the analysis reveals that grant income is a significant predictor of research impact. However, the use of this feature might introduce institutional bias, making it unsuitable for predictive analysis. More intriguingly, aligning with Thorpe et al. (2018) and Reichard et al. (2020) linguistic features, including part of speech tags, sentiment and readability appear to be moderate predictors of REF outcomes.

To the best of our knowledge, only one study, conducted by Williams et al. (2018), has employed a predictive rather than descriptive approach to analysing impact case studies. In their study, they performed a binary classification of impact case studies, distinguishing between "high quality" and "low quality" based on their placement within the percentile rankings, specifically categorizing those in the upper 20th percentile as "high quality" and those in the lower 20th percentile as "low quality." The study identified various features to evaluate the quality of impact case studies. The explicit textual features included narrative style analysis using TF-IDF to extract significant n-grams, reflecting the thematic focus and complexity of the research narratives. Implicit textual features were assessed through various readability metrics (e.g., Flesch Reading Ease, SMOG, etc). The study introduced a novel method for determining the reach of impact case studies by using the Digital Object Identifiers (DOIs) of referenced research in impact case studies to extract bibliometric and policy indicators. The study performed backward elimination with Support Vector Machines, Random Forests, and Neural Networks to identify features contributing to high-scoring impact case studies. The results indicated that policy and bibliometric indicators of referenced research were strong predictors of research quality.

Gaps in the Literature

The thematic analysis of impact case studies reveals the significant diversity in how the impact is interpreted across disciplines. These distinct interpretations of impact arise because each discipline's understanding of impact is shaped by the field's norms, traditions and methodologies, leading to slightly different criteria for what is considered valuable impact toward society, the economy, the environment and others (Oancea, 2013). these differences in interpreting and measuring research impact lead to a multidimensional dataset. As the number of different dimensions grows, the dataset becomes increasingly complex and distinct impact case studies become more thinly spread across dimensions (Hastie, Tibshirani and Friedman, 2009). The large number of impact dimensions and the sparsity of certain types of impact can pose significant challenges in identifying patterns and relationships across data, both by humans and computational models. The work of Tarama et al. (2016) produces six distinct impact types, including education, public engagement, environmental and energy solutions, and enterprise and clinical use. Using these themes, the study claimed that no interpretation of impact is more highly rewarded than the other. However, these topics have low granularity because they ignore many of the impact nuances. Therefore, it remains unclear if specific subtopics within these categories could be more highly rewarded than others.

On the other hands, The King's study identified 60 more granular topics across 6679 non-redacted impact case studies from all disciplines. But because of the diversity of interpretations of impact across different disciplines each with varying subject-specific terminology, the model struggled to capture all the unique interpretations of impacts by disciplines. Instead, it led to ambiguous topics like "Asia" and "cancer" which were difficult to categorise as any specific types of impact. These topics seemed more likely broad areas or fields than specific types of impact. The Oxford study was able to identify 84 more granular topics, by narrowing down the focus on certain more closely related disciplines within the SHAPE (Social Sciences, Humanities, and the Arts for People and the Economy). This research builds on the existing research by focusing on Medical and Life Sciences (Panel A, N=1928) to produce even more granular topic categories. This research will further identify different topics of research from Medical and Life Sciences to determine if these topic variations can be used to predict the quality ratings of impact case studies.

This is fed to the second part of this research which is aimed at understanding the extent to which machine learning models can capture significance and reach of impact case studies. Although the present peer review biases are well-documented with implemented mitigation approaches are accounted for, various research have pointed toward the potential biases with limited research utilising the natural language processing (NLP) and neural network in order to uncover and asses factors driving high-quality impact case and the role of narrative style, coherence and clarity that have been discovered by Thorpe et al. (2018) and Reichard et al. (2020) in influencing the impact case study quality rating. Most existing studies, such as Williams et al. (2018), primarily focus on utilising the bibliometrics or policy citation in predicting the rating of impact case studies, highlighting the attention on descriptive analysis of impact case studies. Therefore, there is existing gap in the utilising combination of external text-driven data, NLP and handcrafted linguistic features of readability, sentiment and other to provide better understanding of underlying feature providing quality impact case studies.

Methodology

This research aims to identify to what extent machine learning can capture the significance and reach from the qualitative narrative of impact case studies. This process will involve comparing different machine learning models and scrutinizing the way the models make decisions to classify low, medium and high-quality impact case studies. To improve the interpretability of the models, we designed a set of handcrafted features based on the preliminary primary data obtained from qualitative solicitations with experts in the field. The byproduct is a deeper understanding of implicit and explicit aspects of impact evaluation and the degree to which they contribute to assessing impact case studies. This feeds and directly contributes to the research's second objective: to understand the differences between high- and low-scoring impact case studies. The research and the utilised code are provided at (https://github.com/mateuszkopaczewski/mds).

Data collection

Primary data

This research gathered important domain-specific information through expert elicitation from panellists involved in the Research Excellence Framework (REF). The panellists conducted a public discussion and Q&A session to share their experiences as REF panel members, with our questions helping guide parts of the debate.
The research designed the questions to address the challenge of building a predictive model to predict the quality rating of impact case studies from the impact case studies narratives alone. It focuses on aspects of impact narratives that could distinguish between high-versus low-scoring impact case studies. A quick personal interview followed the discussion to follow up on the answers to the questions discussed in the debate.

The questions and the formalised overview answers are displayed below:

1. What are the most critical aspects of the impact case studies?

It is crucial to address the REF criteria directly and identify a problem the research is solving to prevent failing into the biggest pitfall of impact case studies: losing the connection between the research and impact. A well-crafted narrative that addresses a specific problem can anchor the reader's attention to particular parts of the story, effectively guiding them through the various contexts of the impact case study. This approach helps to clearly illustrate the link between research (cause) and impact (effect). Simply presenting findings is less effective compared to providing a cohesive narrative.

2. Does linguistics play a role in how you view an impact case study? For instance, grammar, sentence structuring, and language quality (such as using more technical or sophisticated language). 3. If so, which aspect of linguistics do you think is the most important in viewing an impact case study (i.e., technical language, sophisticated language, grammar, etc.)?

Clearly state the impact in the first sentence and paragraph to ensure the reader understands the significance without ambiguity. Scientists evaluate the impact of case studies in Panel A, so they must follow the principles of good scientific writing. Scientific writing requires clarity, conciseness, and factual language, avoiding unnecessary jargon and overly complex language. This approach ensures that the key points are easily understood and accessible to non-technical readers.

Nevertheless, impact case studies are ultimately assessed on their significance and reach. Therefore, a well-written impact case study is unlikely to receive a higher score than a poorly written one if both demonstrate the same quantifiable impact. However, it may unconsciously influence the score if the research impact cannot be found.

3. What are the critical distinguishing areas between 3-star and 4-star impact case studies – for example, presentation, structure, linguistics, or significance and reach of the Impact case study?

It is crucial to include any measurable improvement in outcomes directly tied to the research to demonstrate the tangible effect of research. Providing factual or quantitative data can distinguish between achieving a 4-star rating for outstanding impact and a 3-star rating for good impact in case studies.

4. Do the quantitative features of an Impact Case Study matter? For example, document length, number of citations, number of words, number of visualisations or figures to back up the arguments, number of countries benefiting from the paper, etc.

Word and citation limits prevent overly complex explanations and encourage writers of impact case studies to focus on the point at hand.

5. Do different kinds of impact matter (Health, Tech, Environment, Society)?

No single type of impact will consistently result in higher-scoring impact case studies. Instead, the focus should be on providing evidence that the specific interpretation of impact has significance and reach. However, certain types of impact might have a more tangible impact that is easier to prove.

6. Do Impact Case Studies focusing on the UK matter more than concentrating on other countries or globally as "The UK funding bodies expect that many impacts will contribute to the economy, society and culture within the UK, but equally value the international contribution of UK research." (from a REF report)

The impact's geographical location, whether within the UK or internationally, is not as crucial as the overall reach and significance. The critical consideration is how far-reaching the effect is regardless of where it occurs. For instance, research that affects a small, remote community in the middle of nowhere might have a more profound and significant impact than research that impacts a major city in the UK. The accurate measure of impact lies in the change it brings about, how widely it is felt, and its importance in the relevant context.

REF data

REF 2021 required higher education institutions (HEIs) to submit a certain number of impact case studies for each Unit of Assessment (UoA) depending on the number of Full Time Equivalent (FTE) staff in that UoA. These impact case studies describe the impact of research conducted between 1 August 2013 and 31 December 2020, extending beyond academia. The case studies followed a predefined structure containing five main sections, each having a word or reference limit, including Summary of Impact (100 words), Underpinning Research (500 words), Reference to Research (maximum six references), Details of Impact (750 words) and sources to corroborate the impact (maximum ten references). Expert panellists evaluated each impact case study based on the reach and significance of the stated effects, and the results and submissions were later published on the official REF website.

Formatting

This public data is available in a spreadsheet, which provides raw data extracted from the REFs impact case study database, including the geography of impact, funding beneficiaries, associated funding and the five main sections from the predefined format. Although REF states that the data is formatted in Markdown, this research analysis has found a mixture of Markdown, LaTeX and HTML formatting, including some erroneous formatting that did not conform to any clear formatting. The research also identified that the case studies had slightly different citation styles and varied use of bullet points, numbered lists, and headings. Non-text items, such as photographs, tables and institutional shields, were removed from the text. All other text was included in the submissions except for text that was deemed commercially sensitive or required restriction, which was redacted and marked by the notation "/[text removed from publication]/" (REF, 2021).

To mitigate the adverse effects of noise introduced through inconsitent formatting, this research leveraged the capabilities of a pre-trained large language model, specifically GPT-4o mini, to reformat the impact case studies into a suitable format. ChatGPT has been proven effective in reformatting data and improving input data quality, thereby enhancing the performance of downstream NLP tasks (Qin et al., 2023). This effectiveness is primarily due to the model's extensive training on a diverse corpus of text data, which gives them a deep understanding of language patterns, grammar, syntax and semantics—allowing them to easily capture and filter out irrelevant

data that might introduce noise (Alawida et al., 2023). This was achieved through simple prompt engineering, using direct and unambiguous 'instructive' prompts to guide the model in completing specific tasks within the given context (Giray, 2023). This prompt was designed to ensure the text preserves the semantic, stylistic, structural and syntactic elements of the text but removes erroneous formatting (irregular or non-standard formatting) that traditional methods cannot capture.

Constructing the target variable

This research also uses the result spreadsheets containing the quality profiles of all 157 UK higher education institutions (HEIs), including impact, environment, and output submissions. These quality profiles indicate the proportion of submissions in each institution at the unit of assessment level which received the following ratings: 4-star (World-leading), 3-star (Internationally Excellent), 2-star (Recognised Internationally), 1-star (Recognised Nationally), and Unclassified (REF, 2021).

The distributions provide information about the overall performance of impact case studies within a specific institution and unit of assessment (UoA). Still, they do not reveal the exact score of any individual impact case study, making the data not suitable for classification or regression tasks by default. To address this problem, we calculate the weighted average of impact case studies, termed the Grade point average (GPA), to allow for univariate comparison between impact case studies. For each assessment unit in a specific institution, the GPA is given by equation 1.

$$GPA = \sum_{i=0}^{n=4} \frac{1}{100}\left(P_i \times \ i\right) \tag{1}$$

Where, P_i represents the percentage of impact case studies with the i-star quality rating. It is also valuable to frame this problem in the classification lens because of the categorical nature of the ratings. Equations 2 and 3 show the methods for deriving 5 and 9 class labels, respectively.

$$\text{Class Label}_5 = \text{round}(\ GPA) \tag{2}$$

$$\text{Class Label}_9 = \text{round}(\ 2 \times \ GPA) \tag{3}$$

The small number of impact case studies ranked unclassified, 1-star or 2-star, means keeping these categories separate does not make sense as this can contribute to the class imbalance. Therefore, in the 9-class label configuration, classes 0 through 5 are combined into a single class, while classes 6 and 7 are grouped into another class and class 8 forms its separate class. This is reasonable because unclassified, 1-star and 2-star classes all receive the same amount of funding (zero). Consequently, we create a 3 Class Label to reflect this logical grouping.

Variance inside GPA

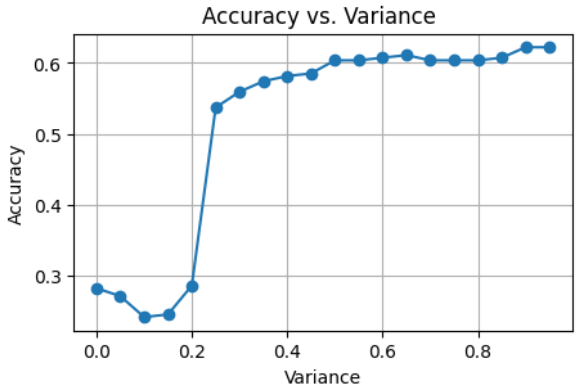


Figure 1: Accuracy under different thresholds of variance

When we use GPA as the target variable, we aggregate the range of possible scores the HEI receives in a UoA to a single number. It is impossible to distinguish between impact case studies that may have received varying individual ratings with this single score, leading to a loss of crucial information about individual impact case study variability and nuances.

Zhang, Watson, and Hodgson (2022) proposed training the model on a subset of impact case studies, which had a variance below a specific threshold in the scores aggregated to calculate GPA. However, setting the threshold too low can result in too few eligible impact case studies for training, which will be too little to train a robust machine learning model, while setting the threshold too high would cause the GPA to be overly distorted by the variability in scores.

Contrary to the findings of Zhang, Watson, and Hodgson (2022), when experimenting with 20 different thresholds of variance on the training set, we found that decreasing the threshold led to no improvement in accuracy on the testing set. This research would also suggest against lowering the threshold as doing so could introduce bias towards larger institutions. Smaller institutions typically have higher variance due to having fewer impact case studies on average.

Imbalance in dataset

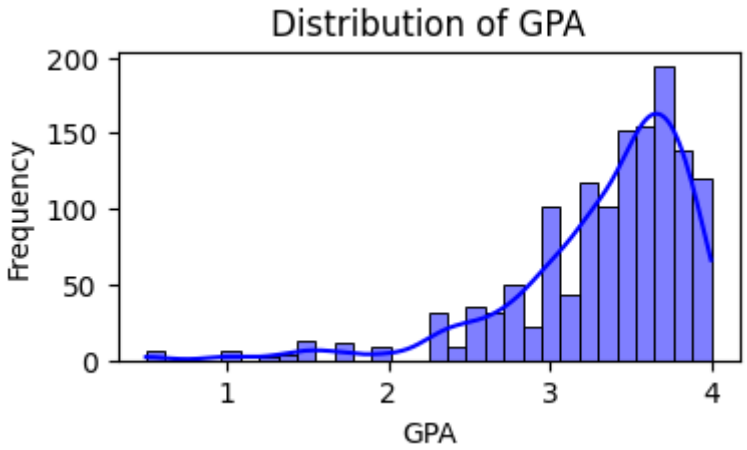


Figure 2: Distribution of GPA across panel A

Most machine learning models work with the assumption that the underlying data trained on is balanced. When the data is imbalanced, the model is trained primarily on the overrepresented classes, leading to low predictive

performance on the minority set. This research uses various resampling methods to minimise the implications of imbalanced data.

Traditional methods of dealing with imbalanced data are oversampling by duplicating random data instances and undersampling by removing data instances. Excessive oversampling leads to excessive duplication of data instances, increasing the chance of the model overfitting to specific examples and leading to poor generalisation of unseen data. Similarly, excessive undersampling can lead to the excessive removal of data instances, causing the loss of valuable cases that the model could have learned from, increasing the chance of underfitting. A suitable solution could be a combination of oversampling and undersampling, which can help balance retaining the majority class information and increasing minority class representation.

Whilst traditional oversampling and undersampling methods introduce the risk of overfitting or underfitting, this research utilises more advanced techniques with lower associated risk of underfitting and overfitting. The following section explores SMOTE and data augmentation, offering a more robust way to handle imbalanced datasets.

SMOTE

SMOTE (Synthetic Minority Over-Sampling Technique) identifies random data points from the minority class. Rather than just duplicating the sample, SMOTE finds the k-nearest neighbours and creates a synthetic data point by interpolating between the chosen data points and the k-nearest neighbours. Repeating this process until the minority class is sufficiently balanced. Interpolating between the k-nearest neighbours introduces variation to the new data point reducing the risk of overfitting.

Whilst this may be effective when dealing with frequency-based methods that rely on word co-occurrence, it struggles with more complex word representations that encode semantic and syntactic relationships. Where interpolating between these vectors could produce meaningless or nonsensical relationship degrading the model's performance.

Augmentation

Augmentation is a type of oversampling technique which increases the proportion of the minority class by creating new text samples from randomly selected instances of the minority class. Some common ways to augment data are synonym replacement, replacing words with similar word embeddings, back translation and generative models. These augmentations preserve the semantic meaning of the text whilst changing the data slightly to introduce variation within the instance, exposing the model to different forms of the same data. Through this variation, augmentation reduced the chance of overfitting

By creating slight variations in the original data without changing meaning, data augmentation can broaden the local space around the decision boundary, meaning the model can distinguish between different classes more effectively.

Words and Phrases

By default, the text in impact case studies is unstructured because it does not follow a predefined format or scheme, making it unsuitable for machine learning. It necessitates feature extraction methods to extract meaningful features from impact case studies and represent them in a numerical format (Kowsari et al., 2019).

Frequency-based methods

The simplest way to represent this text in a structured format is by using Bag of Words (BoW), frequency-based representations. Let us consider a document d_j which is constructed by concatenating the sections of impact case studies that do not include references. Namely, the "Summary of Impact", "Underpinning Research" and "Details of Impact" sections. Then the vocabulary V is all the m unique words $w_1, w_2, \dots, w_m$ in this concatenated text.

In the Bag of Words (BoW) approach, each document d_j is represented as a vector $v_j \in R^m$, where m is the size of the vocabulary. The vector v_{ij} has m elements, each corresponding to a unique word in the vocabulary V. For each element in the i-th position of vector v_j corresponds to the frequency of the word w_i from the vocabulary in the document d_j. The vector essentially captured the Term Frequency of each word from the vocabulary in the document.

An alternative frequency-based method is (TF-IDF) Term Frequency Inverse Document Frequency, similar to BoW, d_j is converted into a vector $v_j \in R^m$ where the i-th element v_{ij}represents the frequency of each word in the vocabulary V. However, TF-IDF also scales the computed Term Frequency by a constant c, depending on how often it appears in the document. Where the constant c is equivalent to the value of $IDF(w_i)$ given by equation 1.

$$IDF(w_i) = \log\left(\frac{N}{n_i}\right) \tag{4}$$

Where n_i is the number of documents the word $w_i$appears in.

These methods suffer from several limitations that need to be considered for use as features. By ignoring the position of words, various essential semantic, syntactic and discourse-level text elements are abstracted away (see section 2.4). This loss of nuanced aspects of text can lead to losses in performance on downstream tasks (Devlin et al., 2019). However, incorporating n-grams (i.e. sequences of n consecutive words) into frequency-based methods can provide improved context by capturing words that commonly appear together.

However, each unique word in the corpus adds a new dimension to the vector space, leading to highly high-dimensional vectors in corpora with an extensive vocabulary. As these dimensions increase, the data needed to fill or represent the vector spaces adequately grows exponentially (Thudumu, 2020). This exponential growth leads to sparsity, as most documents only use a small subset of the total vocabulary, resulting in many sparse vectors with primary zeros. This issue is exacerbated by n-grams, which increase the number of possible combinations of words and phrases, resulting in even higher dimensionality.

Despite their simplicity, frequency-based methods have been effective in various NLP tasks (Turney & Pantel, 2010). Using frequency-based and machine-learning methods would likely capture words and phrases associated with higher-impact case studies. This could expose indirect impact indicators, including specific stylistic differences between impact case studies (Reichard et al., 2020), identify specific types of thematic differences, etc. However, this research suspects that frequency-based methods cannot capture the significance and reach of impact case studies. Counting the frequency of different words is likely only to identify surface level features and not capture a deeper understanding of the importance of specific outcomes or recognise the scope or scale of the impact. Nevertheless, understanding what frequency-based methods can capture can deepen our understanding of the differences between impact case studies.

The high dimensionality of the frequency-based embeddings for each document makes it intuitively tricky for humans to interpret. To address this, this research proposes using the Chi-squared test to identify the words most strongly associated with each of the three class labels. The Chi-squared test measures whether two categorical variables are independent (Meesad, Boonrawd and Nuipian, 2011).

Given the contingency table that represents the count of occurrences of a particular word in the document for each class, we can determine the chi-squared statistic by comparing the observed frequency to the expected frequency in the contingency table, formally given by equation 5.

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \tag{5}$$

$$E_{ij} = \frac{\text{Row Total} \times \text{Column Total}}{\text{Grand Total}} \tag{6}$$

Where O_{ij} is the observed frequency of feature f in class C_j and E_{ij} is the excpected frequency of feature f for class C_j

Topics and Themes

Identifying topics associated with case studies can help infer the potential significance of research based on its alignment with critical social needs or strategic priorities. A study by Terama et al (2016) has failed to capture a relationship between different interpretations of impact and the quality profiles of impact case studies. However, the topics they have identified were extremely broad, which might have obscured more nuanced association. This research performs topic analysis at higher granularity without compromising on interpretability, by narrowing the scope of topic analysis to impact case studies from Medical and Health Sciences.

Transformers

This section will introduce the pioneering transformers and how they have been used to build contextualised embeddings.

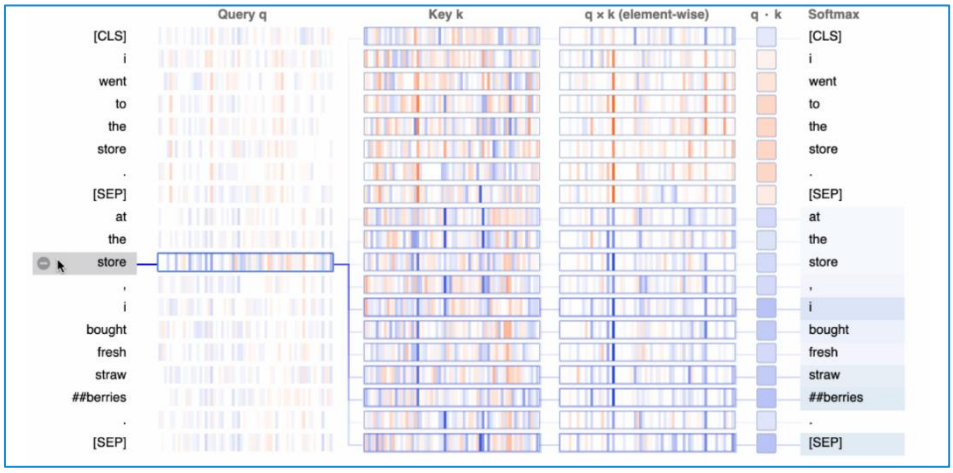


Figure 3: Example of attention visualization in transformer model for the word "store"

This brief section will formalise a type of deep learning architecture called a transformer which will be referenced throughout the paper. The transformer architecture was first introduced by Vaswani et al (2017) as an encoder-decoder architecture. The encoding layer consists of many encoders' layers stack next to each other which iteratively process the input tokens one layer after another, while the decoder consists of a stack of decoding layers which iteratively process the encoder's output as well as the decoders output so far. The function of the encoder layer is to generate contextualised representations of the tokens by producing a linear combination of the token representations from other input tokens via the self-attention mechanism. In each decoder layer, there are two distinct attention mechanisms: (1) cross-attention, which merges the encoder's output (the enhanced representations of the input tokens), and (2) self-attention, which blends information from the tokens that have been produced so far during the decoding phase. The core building block of transformers is scaled dot-product attention. Each attention unit computes predictions using the following process:

For each token i , the input representation x_i is multiplied with each of the three weight matrices to produce a query vector $q_i = x_i W_Q$, a key vector $k_i = x_i W_K$, and a value vector $v_i = x_i W_V$. The attention weights a_{ij} from token i and j are calculated by the dot product between q_i and the key vector k_j. The more similar the query and key vectors, the larger the dot product, whilst those that don't share much in common will have no overlap, resulting in smaller vectors.

The attention weights are scaled by diving them by the root of the dimension of the key vector $\sqrt{d_k}$ and passed through a softmax function to normalise the weights. The result of the attention mechanism for token i determined by a_{ij} which represents the attention score from token i to each of the other tokens. The attention mechanism can be expressed as the matrix multiplication in equation 4.

$$Attention(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V \quad (7)$$

Where Q , K and V represent matrices where the i-th rows are vectors q_i , k_i and v_i respectively, and d is the dimension of the keys.

The encoder generates different representations for each input feature using several attention heads, each with distinct query, key, and value vectors. These representations are then combined, given appropriate weights, and passed through a feed-forward layer to produce the final output.

Contextualised Embeddings

In traditional word embeddings, such as Word2Vec and GloVe, represent each word with a fixed vector that does not account for the context in which the word appears. This limitation arises because these embeddings do not consider the position of words within a sequence and instead treat them in isolation. This word embedding generates static representations for each word irrespective of its contextual usage. In contrast, contextual embeddings capture the meaning of words based on their context. There are various types of contextual embeddings; some are based on LSTM architecture with an attention layer, such as Context Vectors (CoVe) (McCann et al., 2017) or bidirectional LTSM-like Embeddings from Large Language Models (ELMOs). However, this study focuses on Bidirectional Encoder Representations from Transformers (BERT).

BERT stands out from other contextualised embeddings, like ELMo, because it uses a multi-head transformer encoder architecture that learns bidirectional context from text. This is achieved through a training method called Masked Language Modelling (MLM), where BERT topics randomly predict masked words in the text. This unique training method forces BERT to consider both the preceding and following words in a sentence to understand how they contribute to the meaning of the masked word.

Bert takes the input sequence of words and subwords and maps them to a predefined ID using the Word Piece tokeniser. These predefined IDs are derived from a vocabulary built during BERT's pre-training on a large corpus of text (e.g. "play" can have the id 2456). The Word Piece tokeniser also uses special tokens [CLS] and [SEP] to distinguish sequences' start and end, respectively. For words not found in the tokeniser, the tokeniser breaks them down into smaller subwords. For example, "unhappiness" might be split into "un" and "##happiness", where the prefix ## means that the preceding string is not a whitespace. The Word Piece tokeniser also creates attention weights these are binary vectors [0 and 1] that indicate the token should be attended to and which are just padding and should be ignored. These embedding vectors and their corresponding attention masks are passed through BERT's encoder layers, generating a hidden state for each token.

The previous section explains BERT's underlying architecture for creating token-level embeddings. This is visualised in Figure 7 where the token "store" is aggregated by its attention scores with all the other tokens in the sequence. The tokens "strawberry", "fresh", and "bought" have higher attention scores, which means they exert more influence on the contextual representation of the word "store" at that position. Although there is no explicit sentence-level embedding, the standard practice is to use the first token [CLS], which, unlike the other tokens, is specifically designed to be a representative summary of the entire input sequence (Choi et al., 2021). While taking the [CLS] token output can generate sentence embeddings, it is not explicitly optimised for sentence-level semantic similarity. Instead, for downstream tasks like some topic models that need to capture the semantic meaning of documents, sentence transformers that are explicitly fine-tuned for sentence-similarity tasks should be used. However, these Bert-based embeddings come with certain limitations on their memory requirements, which can be constrained by available resources, such as the 10 GB memory limit.

BERT topic modelling

Traditional topic modelling techniques like the ones employed by Terama et al. (2016) and King's Study (2015) utilise frequency-based methods to identify themes and topics across impact case studies. These frequency-based methods fail to capture the context of sentences, causing them to capture text semantics inadequately (Das et al., 2015). Grootendorst (2020) addresses this limitation by employing the contextualised word representations from BERT. Their research proposes dimensionality by reducing these embeddings, clustering them and extracting topics within each cluster by looking at the most common words in each cluster with TF-IDF.

Before implementing the BERT topic, our research performs a series of preprocessing steps on impact case studies to reduce potential noise. Scholars have defined noise as non-informative features of text that could obscure meaningful patterns (Jain, Murty and Flynn, 1999). Despite the advancements in transformer-based models like BERT topic are still not resistant to the effects of noise (Zhang, Li and Liu, 2021). To remove noise, this research has performed various preprocessing steps in the "Details of Impact" section of the impact case studies. These steps are taken in addition to removing the formatting detailed in section 2.1.

Given that the BERT topic strictly focuses on identifying semantic features, features that do not contribute to the meaning of the text become redundant, introducing unnecessary information, and steps should be taken to remove these features. This is especially important for the last stage of BERT topic where the topics are extracted using TF-IDF, which is extremely sensitive to noise. The first steps taken were to remove unnecessary, lowercase all the text, remove stop words ("the", "in", "is", etc.) and remove punctuation marks (e.g. "!", "?", ".", etc.). Subsequently, words in the text were lemmatised, breaking them down into their lowest base form whilst still providing meaning (e.g. "walking" to "walk") using sentence-level Part of Speech (POS) tags and the WordNet dictionary.

This research considered converting the pre-processed documents into contextualised vector representations using sentence transformers. Having been fine-tuned explicitly to encode for semantic similarity, these type of embeddings are specifically well-suited for clustering tasks because of their understanding of broader semantic connections between text. However, these models are not suited for our particular task as they have a length context window of 512 tokens, below the average token length of impact case studies, meaning key parts of impact case studies are likely to be truncated, leading to loss of information. Therefore, this research considered embedding models with larger context windows more suitable. In particular, this research leveraged OpenAI embedding. These embeddings are designed to capture the meaning of individual words and the larger context in which those words appear. This means the embeddings can reflect the relationships and dependencies between words across the document, allowing for a more comprehensive understanding of the text. By capturing both local and global contexts, this approach provided high-quality representations of impact case studies while remaining computationally efficient.

This research dimensionally reduces the embeddings by using the UMAP (Uniform Manifold Approximation and Projection) algorithm modified to find the nearest neighbours using cosine distance (similarity) between word embeddings (Leland et al., 2018). UMAP was used compared to other methods like PCA (Principal Component Analysis) because it preserves the non-linear relationship and can capture both local and global structure, which is crucial for capturing intricate aspects of the text (McInnes and Healy, 2018). These reduced UMAP data were then clustered using HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise). Figure 4 shows the results where each HDBSCAN cluster represents a unique topic.

Finally, a modified TF-IDF measures the word frequency for all the documents within a cluster to identify high-frequency words (Term Frequency) distinct from other clusters (High Inverse Document Frequency).

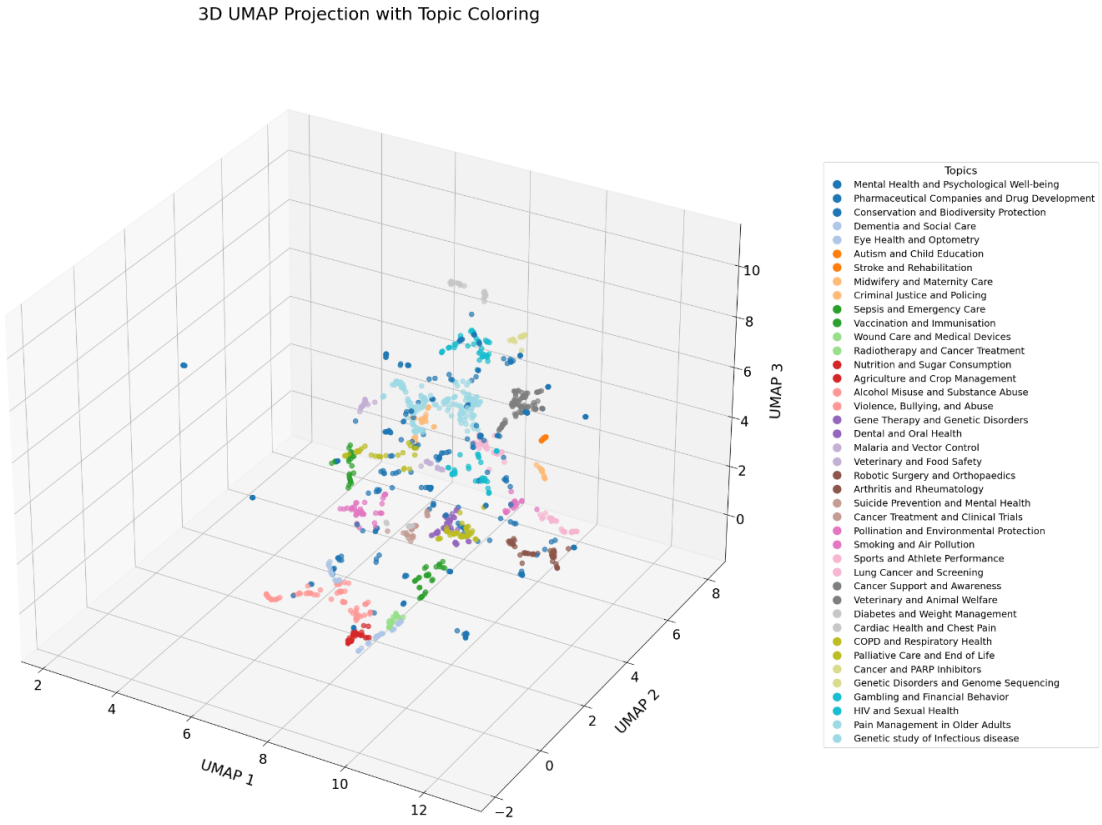


Figure 4: 3D umap projection with topics defined in impact case studies

Hyperparameter tuning BERT topic

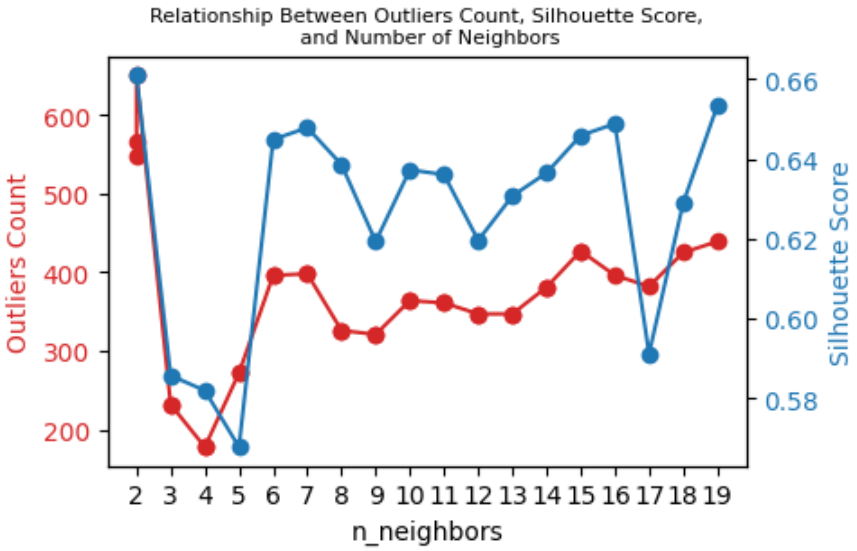


Figure 5: relationship between outliers count, silhouette score, and number of neighbours.

To evaluate the model's effectiveness at capturing topics and themes across text metrics, the similarity between the words associated with each cluster or topic was used to measure the coherence of the topic. The assumption is that the more coherent the topics are, the more meaningful and interpretable they are (Rosner et al., 2014). The coherence score is used in the hyperparameter tuning of the Bert Topic model, where we assess how the number of neighbours in UMAP dimensional reduction, the number of components (or dimensions), and the minimum cluster size in HDBSCAN can influence the coherence of topics (Röder et al., 2015) (see appendix for Bayesian optimisation of Bert topic).

Borrowing from the Oxford Study (2024) the model was also evaluated using the Silhouette Score (Shahapure and Nicholas, 2020). The Silhouette Score measures how well the documents are clustered, and a high score means that the documents are clustered close to other documents in the same cluster but far away from documents not in the same cluster. The number of neighbours influences whether the model captures more local or global data structures. A lower number will capture more of the local structure, capturing finer details and potentially leading to smaller clusters. A higher number of neighbours will capture more of the global structure, leading to a broader overview of the data (McInnes et al., 2018).

Figure 5 displays the relationship between several neighbours, the silhouette score, the number of unclassified topics, and outliers. There's a trade-off between outliers and silhouette score at specific points; for example, for a low number of neighbours (2-4), the silhouette score is high, but the outlier count fluctuates. Both metrics appear to become more stable at a higher number of neighbour's (e.g., 16-19). The Silhouette score is highest at neighbours = 2, but this comes with a very high outlier count, which is undesirable. Neighbours between 5 and 9 provide a relatively stable balance between outliers and silhouette scores. Therefore, we chose n neighbours to be 7, where the outlier count is moderate and the silhouette score is around 0.62, which seems to be a reasonable balance.

Linguistic analysis

Linguistic features have long been used to estimate the proficiency of writing. Dating back to Page (1966), who produced the first implementation of Automated Essay scoring utilising the concept of “trin” and “prox,” where "trin" represents the intrinsic variable of interest and "prox" is a variable that estimates "trin. Page highlighted how "Trin" features such as word length, essay length and the use of uncommon words can be used to train machine learning models to evaluate text proficiency. More recently, advancements in machine learning and natural language processing have allowed for more complex “trin” features by considering dimensions of essay quality such as relevance to the prompt, development of ideas, cohesion and coherence (Ramesh and Sanampud, 2021). These AES systems have been shown to demonstrate high accuracy comparable to human judgement in evaluating writing proficiency.

This section builds on recent Automated Essay Scoring (AES) advancements to create custom, handcrafted features. Insights from domain-specific research inspire these features, gathered through consultations with experts, to explore how various linguistic elements might influence the quality of impact case studies. This section aims to uncover how language can affect the REF outcome.

Discourse Features

The REF panellists have highlighted the importance of a well-crafted narrative to convey the link between research (cause) and impact (effect). The extent to which the narrative can effectively communicate its meaning is closely tied to its coherence. The coherence of narrative refers to how various discourse segments are interconnected in a unified and logical way to create meaningful text. Without coherence, all sentences are independent and meaningless; with coherence, all sentences are connected and flow, allowing for inference and comprehension. Therefore, coherence modelling has intuitive applications in estimating how well impact case studies have provided a clear link between research and impact throughout the text.

To understand the methods for modelling coherence a distinction must be made between coherence and cohesion. Cohesion refers to the use of lexical and grammatical devices to link elements within text. Halliday and Hasan (2014) categories five main devices used to make cohesive sentences; reference, which uses pronouns (e.g., she, he) and comparatives (e.g., this and that) to link back to previously mentioned entities; Substitution,

where elements (e.g., a person) are replaced with equivalents of the same grammatical function (e.g., they); Ellipsis, which omits understood parts of sentences (e.g., "She enjoys playing tennis, and so does her brother"—with "enjoy playing tennis" omitted in the second part); Conjunction, which connects sentences or clauses with words indicating addition, contrast, cause, or time (e.g., "additionally," "but," "because," and "afterwards"); and Lexical Cohesion, which links ideas through vocabulary, using methods like reiteration (repeating or substituting e.g., "expanding" and "expansion") and collocation (using related words together, e.g., "exciting" and "action").

On the other hand, refers to the logical connections and consistency between propositions that make a text logically organized and easy to follow. Although coherence and cohesion are related, a text can be cohesive—linked by grammatical and lexical means—without being coherent if it lacks meaningful or logical connections. In other words, cohesion alone does not ensure that a text makes sense or is meaningful. Coherence involves underlying logical connections between propositions, known as Coherence Relations, which might not always be explicitly marked. These relations can be semantic (linking underlying meanings) or pragmatic (relying on context and world knowledge) (Farag, 2020).

This research models the aforementioned aspects of cohesion in the following ways:

Lexical Cohesion

This research estimates the lexical cohesion by computing the similarity between segments (sentences or paragraphs) (Crossley, Kyle and Dascula, 2018). Given a document d_j with segments $s_1, s_2, \dots, s_n$,each segment is mapped into a vector $v_1, v_2, \dots, v_n$ in the semantic latent space using S-BERT (Reimers, 2019). In this latent space, vectors that are more semantically similar are positioned closer together. The similarity between vectors can be computed using cosine similarity as described in equation 4.

We capture the local coherence of the text by measuring the mean cosine similarity between all adjacent segments s_i and $s_{i+1}$for all $i \in [1, n-1]$where n is the total number of segments. The assumption is that a higher lexical cohesion is reflected by greater mean similarity between consecutive segments.

We also capture the global cohesion by measuring the mean similarity between all pairs of segments. This is the mean of all segments s_i and s_j where $i \neq j$. The global coherence captures the consistency of themes and topic across the topics. The higher the global coherence the more consistent the topics and themes across the text.

$$Sim(S_i, S_j) = \frac{v_i \cdot v_j}{||v_i|| \cdot ||v_j||} \quad (8)$$

The variance in the means of the segments calculated for the global coherence and the local coherence is also calculated to measure the consistency in cohesion. Lower variance suggests that some parts of the documents are less cohesive than others.

Referent Cohesion

The research identified entitles (including people, places and organisation) in each sentence using Named Entity Recognition (NER) techniques. The pronoun and nouns were then resolve between each sentence using coreference resolution. The total number of times that each entity was mentioned in adjacent segments normalised by the total number of entities was calculated. The assumption is that the higher the number of coreferential links the more continuity there is across sentences (Pogorilyy and Kramow, 2020).

Conjunctive Cohesion

The conjunctive cohesion was modelled by calculation the proportion of conjunctive adverbs and conjunctions (including “However”, “Therefore”, “Nevertheless” and more) compared to the total number of words. The assumption is that a higher proportion indicates higher conjunctive cohesion because there are more logical connections between clauses, sentence and paragraphs.

Sentiment

In exploring language differences between high and low scoring environment statements, Thorpe et al (2018) made the claim that impression management techniques could play a strong role in manipulating how the reader

perceives the environment. Research conducted by Pidd and Broadbent (2015) and Watermeyer and Chubb (2018) echoed these concerns in the narratives of impact case studies, suggesting that the framing the impact of research in an overly favourable light by emphasising the most compelling outcomes whilst downplaying the limitations could unfairly lead to a more favourable REF outcome.

To understand the importance of fostering a positive tone and the influence it has on manipulating the readers perception of the impact. This research utilised an existing transformer-based model, based on RoBERTa that has been fine-tuned for sentiment analysis (Barbieri, Collados and Neves, 2020). Although these are limited by their max sequence length, to account for this this research used the sliding window technique to break the text down into smaller chunks, that are within the model's max sequence length and aggregating the mean score for each chunk (Jaiswal and Milios, 2023).

Shallow linguistic features

This research has used word and sentence tokenizers from the NLTK library to extract lexical features, such as the number of words, stop words, sentences, sentence lengths, punctuation marks, and word lengths. Part-of-Speech (POS) tagging has been applied to identify grammatical roles within sentences, facilitating the extraction of syntactic features like the number of nouns, verbs, adjectives, adverbs, and conjunctions (Kamawat and Jain, 2015). Additionally, this research will extract lemmas by utilizing sentence-level POS tags and synsets from WordNet to lemmatize text into its base dictionary form, known as lemma (Yogish, Manjunath, and Hegadi, 2019). This combination of features has enabled the extraction of deeper linguistic characteristics, including syntactic complexity, lexical sophistication, lexical complexity, and lexical diversity (Mizumato and Eguchi, 2023). To model the influence of conciseness and clarity this research has also calculated various readability metrics, these include Flesch-Kincaid Grade Level (Kincaid et al., 1975), SMOG (McLaughlin, 1969) and Gunning FOG (Gunning, 1952) (Uto, Xie and Ueno, 2020) were calculated for each impact case study.

Predictive analysis

Loss function

This research frames the problem of predicting impact case studies as a multi-class classification task, where the objective of the models is to learn patterns from the data to accurately predict the quality profiles of unseen impact case studies.

Given this scope, the models trained in this section will utilize the categorical cross entropy loss function. This loss function takes the raw outputs of the models $(z_1, z_2, \dots, z_C)$ where C is the number of classes and converts them into probabilities $(p_1, p_2, \dots, p_C)$by applying softmax defined in equation 4. The multi-class loss is then computed using equation 5, where y_i is a binary indicator (1 if the class is correct, otherwise 0).

$$p_i = \frac{\exp(z_i)}{\sum_{j=1}^{C} \exp(z_j)} \tag{9}$$

$$\text{Loss} = -\sum_{i=1}^{C} y_i \log(p_i) \tag{10}$$

This loss is optimised using backward propagation with respect to the models' weights.

Models

Naive Bayes

Multinominal Naive bayes for text classification traditionally uses frequency-based methods as inputs where the discrete count of each word is treated as an individual feature. It is a probabilistic classifier based on the Bayes' Theorem, with the assumption that all the features are conditionally independent given each class. Although this relationship is rarely met in practice, Naive Bayes classifiers have performed surprisingly well in various problem domains including text classification (Raschka, 2014).

Given a document containing words $w_1, w_2, \dots, w_n$ the probability that the document belongs to the class C_k is derived from probability-based function based on Bayes Theorem shown in equation 5.

$$P(C_k \mid w_1, w_2, \dots, w_n) = \frac{P(C_k) \cdot P(w_1, w_2, \dots, w_n \mid C_k)}{P(w_1, w_2, \dots, w_n)} \tag{11}$$

Where $P(C_k)$ is the prior probability of class C_k, initialised as the proportion of document that belong to class C_k

Calculating the probability of the entire sequence of words together $P(w_1, w_2, \dots, w_n \mid C_k)$ given a class can be computationally intensive. However, under the assumption that each word is conditionally independent of others given the class, Naive Bayes simplifies the process by using the chain rule of probability. To reduce the computation to multiplying all the individual probabilities of each word occurring, as if they were independent.

$$P(C_k \mid w_1, w_2, \dots, w_n) \propto P(C_k) \cdot P(w_1 \mid C_k) \cdot P(w_2 \mid C_k) \cdot \dots \cdot P(w_n \mid C_k) \tag{12}$$

In multinominal naive bayes the model assumes that the words follow a multinominal distribution. This means the count of how often a word appears in a document is used to estimate the likelihood of each word.

$$P(w_i \mid C_k) = \frac{\text{Count}(w_{i \text{ in }} C_k) + \alpha}{\text{Total words in } C_k + \alpha \cdot V}$$

(13)

Where V is the size of the vocabulary and α is a constant used for Laplace smoothing to prevent previously unseen words not found in class C_k leading to zero probability.

During training, the model stores the prior probabilities $P(C_k)$and the likelihood probabilities $P(w_i \mid C_k)$for all classes and words.

During classification the model computes the log-probabilities for each class. The class with the highest log-probability is chosen as predicted label.

$$\log P(C_k \mid w_1, w_2, \ldots, w_n) \propto \log P(C_k) + \log P(w_1 \mid C_k) + \log P(w_2 \mid C_k) + \cdots + \log P(w_n \mid C_k)$$

(14)

The probabilities are used to predict classes of new documents during classification.

Extreme Gradient Boosting

XGBoost is an ensemble algorithm designed to optimize a tree-based algorithm's objective function using a gradient boosting framework. In this framework, multiple weak decision trees are built sequentially, and each tree is trained to reduce the classification error of the previous tree. The key notion is adjusting the gradient of the loss function to provide and guide the model in the direction as well as the step sized needed to adjust the objective function in the direction that decrease the error, thereby providing the better performance. By iteratively updating the model to minimize the loss, each new tree further enhances the model's predictive power. This type of algorithm is suitable for sparse data such as text data with high dimensional feature spaces, where multiple features input may have zero values due to the missing words or information in the document. In the traditional tree-algorithm with its greedy tree selection approach, it necessitates nearly complete and numerous sets of trees for comparison, which are time-consuming and computationally expensive (Zhao et al., 2019). This algorithm on the other hand can update and select the best tree iteratively from the subset of splits, chosen proportionally from the distribution of features.

In addition, this algorithm is equipped with the shrinkage and feature selection mechanisms, where the shrinkage aims to reduce the proportion of weight update by applying a learning rate, which helps control overfitting. While the feature selection involves limiting the number of features in consideration for splitting, improving efficiency. The algorithm aims to search for the estimator that minimizes the objective function over the dataset, $F_m(x)$, ensuring the optimal performance in each iteration update.

The regularized objective function which the algorithm aims to minimise is shown below:

$$L\big(y, F(x)\big) = l\big(y, F_m(x)\big) + \sum_{m=1}^{M} \Omega(f_m)$$

(15)

This first term in the equation represents the loss function which measures the differences between the true label, target value, and the predicted output of $F_m(x)$.

Where the penalize term for the complexity of the trees. The XGBoost incorporates the L1 and L2 regularization which helps deal with overfitting (Heuer, 2021). In this context, T denotes the amount of leaves while ω is referred to the vector weight. This is computed below:

$$\Omega(f_m) = \gamma T + \frac{1}{2}\lambda\left|\left|\omega\right|\right|^2 \tag{16}$$

Transformer-based predictive models

Until now, this research has focused on handcrafted features designed using specific domain knowledge of impact evaluation which were concatenated with frequency-based methods. However, modern techniques reduce the need for human intervention in the intricate and laborious design of features, instead they utilize deep learning techniques to automatically extract deep level representations and abstractions of words, phrases and sentences. As a result, they can capture intricate patterns across text yielding superior performance to traditional machine learning models (Young et al, 2018). In consequence, the field of Natural Language Processing has been dominated by deep learning methods, including Recurrent Neutral Networks (RNNs) and their variants Long-Short Term Memory (LTSM) (Hochreiter and Schmidhuber, 1997). Particularly because of their proficiency at handling sequential data and maintaining long-term dependencies over a long-time for the latter.

More recently, amongst the natural language processing community recurrent methods have been replaced by new state-of-the art transformer models like Bidirectional Encoder Representations from Transformers (BERT). If you recall BERT has been pre-trained on a large dataset to learn intricacies of human language. By leveraging transfer learning these models can be fine-tuned on specific tasks with relatively small amount of task specific data (Vaswani et al, 2017). Fine tuning BERT for specific tasks has been shown to have impressive results in various natural language processing tasks and received leading scores in various benchmarks like Super GLUE (Super General Language Understanding Evaluation) (c).The rich understanding of language possessed by transformer models means they can be fine-tuned for a specific task with relatively little data compared to other deep learning methods and still achieve similar levels of performance on specific tasks (Vaswani, 2017).

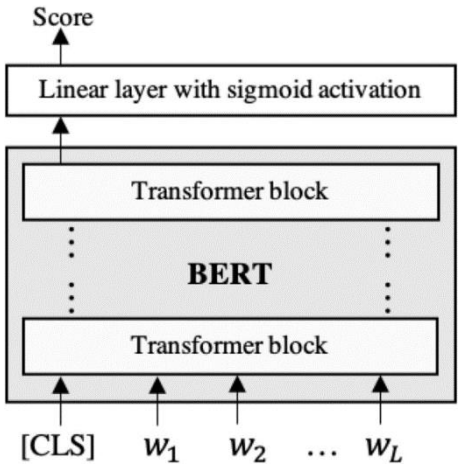


Figure 6: BERT architecture with transformer blocks and a linear layer with sigmoid activation for classification

Scholars have shown that early layers of BERT learn shallow linguistic features and the later layers learn more specific patterns for downstream tasks (Merchant et al, 2020). By leveraging this layered structure by placing a classification head on top of the BERT model (see Figure 5), one can effectively fine-tune the model for a specific task without needing to retrain the model from scratch. During the fine-tuning process, the weights of the

classification head and the later layers of the BERT model are optimized based on the model's predictions on the input data, using multiclass cross-entropy loss. The loss is then propagated backward through the network during backpropagation, which updates the weights to minimise the prediction error.

The multi-layer self-attention mechanism of BERT means that in every layer each token can attend to every other token preceding and following it in the sequence. This gives BERT a deeper understanding of language, this deeper understanding means features that have been previously removed in uncontextualized methods, including various syntactical elements of text, are no longer considered noise but insightful features that may improve the performance of the model on downstream tasks (Devlin et al, 2018). For this reason, BERT does not need as much preprocessing steps as the previously mentioned methods in this paper. Instead, we aim to preserve a lot of the features removed by these preprocessing methods. Except for the erroneous formatting (as detailed in section 1.2). Due to BERT being able to automatically capture deep linguistic aspects of text, including deeper syntactic, semantic and discourse level features without need to explicitly engineer them we do not pass our handcrafted features to the model.

While BERT is effective in various Natural Language Processing tasks it still struggles with imbalanced data. Several studies have explored methods to solve this problem. Recall one of these methods is by oversampling the minority classes with data augmentation. This involves slightly distorting the data by applying augmentation techniques like replacing tokens or using synonym replacement. The idea is that to slightly distort the text without changing the meaning. Which means the model should still make the same predictions for a slightly distorted text. This distortion also adds variance to the dataset reducing the risk associated with oversampling of the model overfitting to the data (Feng et al., 2021)

Although theoretically BERT and other variants like RoBERTa can have longer context windows, there is often a limit imposed by their designers. This is because the process of attending to every pair of tokens in the sequence has a quadratic time complexity. Therefore, for large sequence lengths the computational demands become unfeasible. The longformer variant of BERT is designed to handle longer sequences unlike BERT and RoBERTa which have a context window of around 512 tokens. It does this by with their novel attention mechanism that scales linearly with time (Beltagy, Peters and Cohan, 2020). Allowing it to process much larger context windows efficiently. This makes the Longformer more suitable for tasks requiring long-range dependencies without incurring the computational cost associated with traditional transformer models.

As illustrated in the figure 7, BERT-base uses full attention (a), where every token attends to every other token, leading to quadratic time complexity. To reduce this, strategies like sliding window attention (b) are often employed, where the text is split into smaller chunks, preserving local context but limiting global understanding. A more advanced method, dilated sliding window attention (c), allows tokens to attend to every few tokens, extending the model's range while maintaining lower complexity. The Longformer employs a combination of sliding window and global attention (d), where some tokens attend globally, and others locally, balancing efficiency with the ability to capture long-range dependencies in the sequence, making it ideal for tasks requiring both local and global contextual understanding.

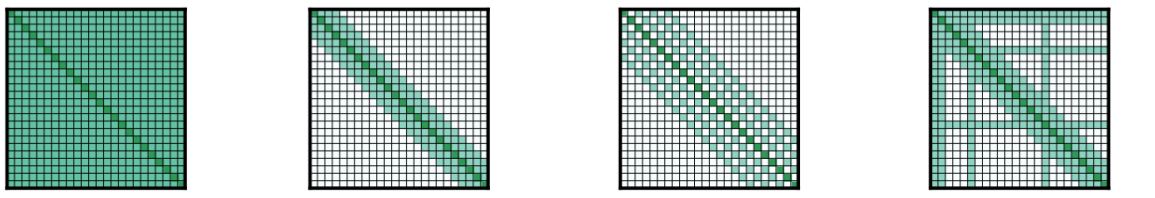


Figure 7: comparison of different attention mechanisms: (a) full $n2n\hat{}2$n2 attention, (b) sliding window attention, (c) dilated sliding window attention, and (d) global+sliding window attention

Even with these advancements this research found that training the longformer variant of BERT on the subset of the REF requires specialised hardware that is not accessible to most. This research was forced to use the

limited and in demand Nvidia A100 GPU with 80GB of VRAM available to us which imposed considerable limitations on our research.

Model Evaluation

By assuming that expert panellists are correct in their assessment of impact case studies, we can evaluate the extent of which machine learning models capture significance and reach of impact case studies by comparing the predictions of our model with the expected results. However, the findings suggest that this approach alone is insufficient and more rigorous evaluation methods that consider the decisions made by the model are required. Although, this task is inherently difficult because of ...

Metrics

This section begins by introducing the metrics that will be heavily relied upon to evaluate the performance of the proposed classification. These metrics are all associated with confusion matrices (see Figure 7). This is a $N \times N$ matrix where N is the number of classes in the target variable. Each entry a_{ij} in this matrix represents the count of samples predicted to the j-th class that belong to the i-th class. In this case the entries a_{ii} represent the frequency when the model has correctly predicted the i-th class.

The accuracy of the model is defined by equation 5 which represents the number of classes predicted correctly as a proportion of total number of predictions. Although in practice this metric can be misleading with non-uniform distributions. This is because the metric does not account for the underlying distribution of the data. Therefore, the model can simply predict the overrepresented classes in the imbalanced dataset and receive unfairly high scores.

$$\text{Accuracy} = \frac{\sum_{i=1}^{N} a_{ii}}{\sum_{i=1}^{N} \sum_{j=1}^{N} a_{ij}} \tag{17}$$

In the cases where class imbalance is a problem alternative metrics that consider the performance of the model on minority classes are important. This gives rise to metrics like Recall, Precision and F1 score which focus on the model's ability to correctly identify and predict minority classes.

$$\text{Recall}_i = \frac{a_{ii}}{\sum_{j=1}^{N} a_{ij}} \tag{18}$$

$$\text{Precision}_i = \frac{a_{ii}}{\sum_{j=1}^{N} a_{ji}} \tag{19}$$

$$F1_i = \frac{Precision_i \times Recall_i}{Precision_i + Recall_i} \tag{20}$$

The F1 score provides a harmonic balance between the proportion of correctly identified positive cases (precision) and the proportion of positive cases that were identified by the model (recall). This metric is particularly valuable in imbalanced datasets, where one class dominates another, as it prevents misleading results when accuracy alone is used. This is crucial because misclassifications could have significant implications for the allocation of funding and brand reputation of Higher Education Institutions. Specifically, this research will utilise the F1-macro score which ensures that each class is treated equally regardless of how often they appear in the dataset. This gives a more balanced view of the model's performance on each class; this is suitable when class imbalance is a problem to again prevent misleading results.

Generalisation

The generasibility of the model to unseen data is important. The model can achieve high accuracy by memorising the intricates and biases of the data without learning anything meaningful. It is crucial to measure the model ability to generalise on unseen data throughout the evaluating process to reduce the size of overfitting. One method to diagnose poor generalisation is by splitting the dataset into two unequally sized subsets, the larger subset is used to train the data, and the smaller subset is used to evaluate the model on the metrics mentioned above. This way we can see how well the models generalises to previously unseen data.

Another common method is K-Fold Cross Validation, where the training set is divided into K folds. In each iteration, K-1 folds are used for training, and the remaining fold serves as the validation set. This process is repeated K times so that each fold is used as the validation set once. K-Fold Cross Validation reduces the risk of variance and bias because it avoids relying on a single validation set, which might not represent the full dataset. However, running K-iterations on models with a lot of parameters, such as transformer-based models can be computationally expensive.

Hyperparameter Tuning

Hyperparameter tuning is crucial for finding the right balance between a model's complexity and its ability to generalize well to new data. In this research, we use Bayesian Optimization for each model and its variants to ensure that they are evaluated at their optimal performance, preventing any model from being unfairly compared due to suboptimal hyperparameter settings.

Bayesian Optimization is an advanced method for hyperparameter tuning that builds a probabilistic model of the objective function, allowing it to efficiently search for the best hyperparameters. Unlike exhaustive methods like grid search, which test every possible combination (Brochu, Cora, and Freitas, 2010), Bayesian Optimization finds optimal hyperparameters much faster. It achieves this by maximizing (or minimizing) the output of an objective function. In our case, this function takes a set of hyperparameters, trains the model with them, and outputs the cross-validated F1 macro average. This setup is specifically designed to account for class imbalance and improve the model's generalizability.

We can assume that every model that is mentioned has been fined tuned with visualisation of hyperparameter tuning for particularly important models in the Appendix.

Ethics and bias considerations in evaluation

During the model evaluation, the research acknowledges the inherent challenges related to transparency and interpretability in the domain of deep learning and the models employed. These deep learning models are described as "black boxes", producing output based on complex non-linear systems where the decision-making process remains unknown and opaque (Christin, 2020). Although, opaqueness is not introduced purely from the unintelligibility of algorithms. As Seaver (2017) argues, algorithms should not be seen as "in" culture but "as" part of the culture. This makes the technical and non-technical interactions that make up these algorithms not "distinct". Instead, algorithms are a compound of human practices that comprise a broader "algorithmic system". Therefore, it is important to consider the broader context in which algorithms may contribute to ethnopolitical issues, generating the risk of algorithms being biased or making unfair decisions that disadvantage certain groups (Seaver, 2017).

Academics have widely cited that addressing ethnopolitical issues involves incorporating accountability into the systems (Amoore, 2020). In line with Seaver's view, this involves addressing both the human and non-technical interactions that can introduce opacity, including the intelligibility of algorithms, conscious decisions to deceive people or conceal algorithms' trade secrets, usually for monetary gain, and opacity caused by a lack of specialist knowledge to understand the algorithms (Burnell, 2016).

In recognition of these issues, this research adopts an ethical AI approach committed to transparency, accountability, and fairness (Williams, 2014). By considering various ways that biases can influence algorithms during the design process, the research has taken several steps to ensure the mitigation of potential biases. These steps include feature analysis and SHAP explanations. This ensures that the models' employees do not inadvertently disadvantage specific groups and impact case studies based on their institution and other

characteristics. The research will critically analyze the features and their contribution toward understanding significance and reach in REF impact case studies. However, in the case involving intricate patterns and complex interactions between variables, often found in BERT model, we aim to employed SHapley Additive Explanation (SHAP) to improve the interpretability.

Furthermore, to encourage scrutiny and accountability, the underlying code is made publicly available to increase transparency in the research conclusions. This process ensures that the model's decision-making can be reviewed and audited by external reviewers on the conclusion drawn from the model's output. It also conforms to principles of AI solidarity, ensuring the research does not reinforce the existing power structures and hierarchies within institutions by sharing any benefits gained from our analysis of impact case studies with the community (Luengo-Oraz, 2019). By considering these factors and addressing potential biases through the design and evaluation process, this research aims to minimize the risk of model contribution to the ethnopolitical issue and disadvantaging certain groups.

Results

In this section, the research progressively builds from simpler to more advanced models to analyse the classification of impact case studies. Initially, the research provides findings based on the frequency of words and phrases findings using the chi-squared approach, offering insight into the predictive features driving the models employed. Following this topic further drives into the topic modelling, which aids in identifying "hot topics" and investigates whether some topics achieve a higher score with low variance, reflecting potential trends or preferences in particular areas of research. By devoid of the subjectivity associated with predefined categories, this provides a more data-driven understanding of topic distributions. Both baseline Naive Bayes models, which rely on word frequency, and more sophisticated models like XGBoost and BERT are examined with incorporated handcrafted linguistic features to assess whether these features can enhance model performance by capturing the underlying linguistic nuances. By comparing the predictive power of baseline model, advanced models and the model with handcrafted features, the research explores how these models can capture the significance and reach in the impact case studies, revealing the strength and limitation of frequency-based and semantic-based models as well as providing the potential interpretable result from progressive analysis of simple baseline to advanced model.

Words and Phrases

The chi-squared test is a standard statistical test widely used in machine learning as a feature selection method to select the most helpful input features; this is useful as it can reduce the complexity of the model and potentially improve accuracy. This research instead uses the chi-squared statistic to compare the importance of words and phrases in specific class labels under different frequency-based feature extraction techniques, namely Bag of Words (Bow) and Term-Frequency Inverse document frequency (TF-IDF). Figures 8, 9 and 10 illustrate each impact profile's results. The size of the words shown in the cloud corresponds to the chi-squared score, and a higher chi-squared score indicates a bigger word size and greater relevance of the word or phrase to the specific quality profile.

(a)

(b)

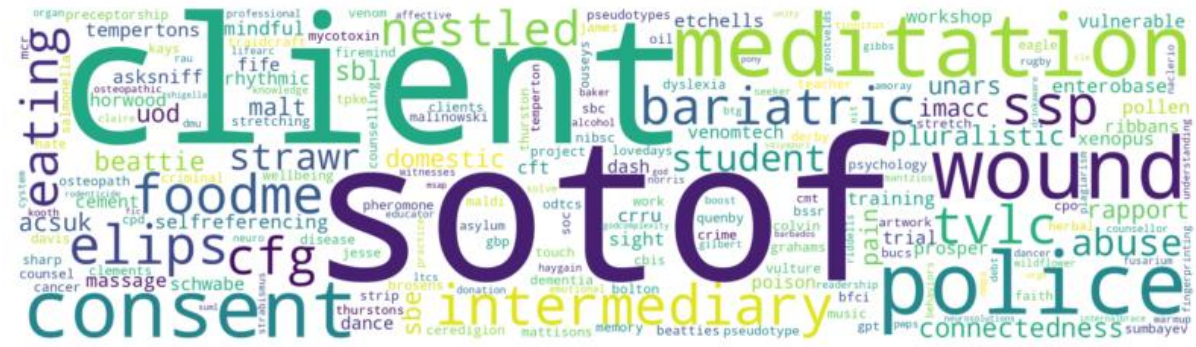
Figure 8: Word Clouds for terms strongly associated with 4* impact case studies: using (a) BoW and (b) TF-IDF for feature extraction.

(c)

(d)

Figure 9: Word Clouds for terms strongly associated with 3* impact case studies: using (a) BoW and (b) TF-IDF for feature extraction.

(e)



(f)

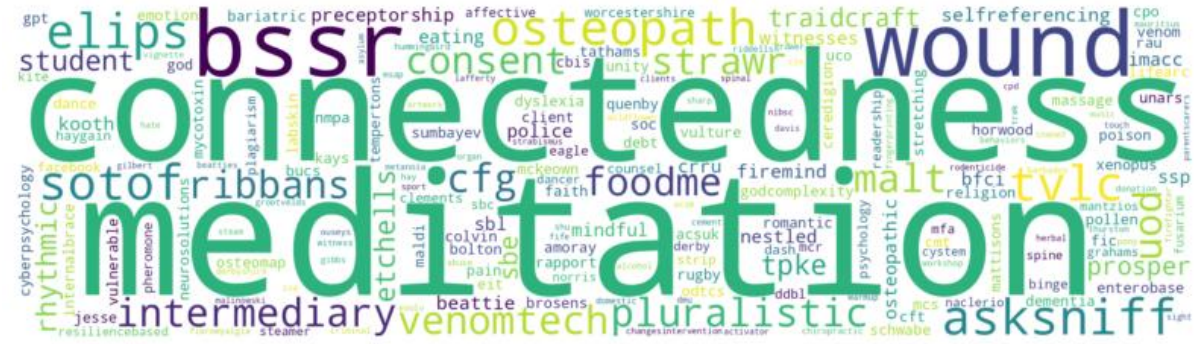


Figure 10: Word Clouds for terms strongly associated with 2* star or less impact case studies: using (a) BoW and (b) TF-IDF for feature extraction.

The illustrations in Figures 8, 9 and 10 suggest that models that use frequency-based methods as input features to predict the quality profiles will likely leverage hot topics across impact case studies. Figure 8

(a) and (b). “stoke”, “patient”, “treatment", “NHS” and “malaria” signifies the relationship of medical topic toward higher class rebel, suggesting the research on the field of medical tends to score higher and potentially indicate additional bias toward research in the medical domain. This results likely presents the “hot topic" in the Uoa from the Medicine, Health and Life Science panel

Biodiversity, animal, restoration and habitat terms are reflected with high chi-squared score within the 3-star quality profile, highlighting the focus on the environment, ecological and genetics studies. On the other hand, the research in class two or below are discovered to associated with the words of “crime”, “domestic”, “abuse”, “children”, “witness”, and “well-being”. It centred on the psychological well-being as well as the diverse set of topics, whether the “asylum seeker”, “dance, “music” and “artwork”. These highlight the niche research area which may lack the prevalent reach or transformational potential toward the significant area of policy and economic.

TF-IDF weights down the importance of words that appear frequently across documents, putting more emphasis on rare terms and overlooking words and phrases that appear commonly across documents as they are considered less meaningful. This is evident in Figures 8, 9 and 10, which show that when TF-IDF is utilised as the input feature, the word cloud produced (i.e. b, d and f) emphasises more specific and unique words. Whilst this may be useful for classification tasks. In specialised documents, such as impact case studies, this may downplay essential terms that appear frequently amongst documents that may be crucial in understanding the core content of impact case studies. For instance, by comparing the word clouds of Figure 8 parts (a) and (b), we find that TF-IDF puts more emphasis on specific words like "rvc", "psa" and "txa" which are rare across documents and disregards terms like "estimating", "guidelines", and "recommend", which are often used to explain the quantifiable significance of the research impact, as well as terms like "women" and "children", which are associated with the level of reach in impact case studies. Consequently, when assessing the significance and reach of impact case studies, classifiers that utilise TF-IDF are likely to overfit specific words and phrases and less likely to identify latent themes across impact case studies, leading to poor generalisation and worse performance.

On the other hand, both input features introduce a high risk of institutional bias regardless of their perspective on the frequency and infrequency of words. This is showcased by institutions with a strong background and reputation in academia like "kings", "oxford", and "lshtm" having relatively high association with 4* impact case studies when compared to other words and phrases in the word cloud for both parts, Figure. 8 (a) and (b). Similarly, although less pronounced, institutions with relatively weaker reputations like "Newcastle", "Aberdeen", "Essex", and "Northumbria" appear to be more closely associated with 3* impact case studies. Should classifiers try to predict the significance and reach of impact case studies, the apparent institutional bias may cause deviations from the excepted results, which may favour impact case studies that mention specific institutions without them necessarily having research that has higher significance and reach deterring from any meaningful patterns in the data.

Furthermore, the research discovered the terms associated with significant criteria measurement in the impact case studies as evident by “guideline” and “estimate”. According to the expert's opinion, the key distinction four-star rating and lower-rating impact case studies attributed to the quantifiable evidence supporting the impact case studies. This means the impact case studies with numerical evidence on its significances tends to score higher. This could potentially explain the appearance of the word “estimation” with its high significant chi-squared value in the four-star classification. The immediate and delay in significant may play a role in this high impact score where the public health tends to be direct and display a broader societal impact compared to others (Oancea, 2013). It is related to the major heath conditions and areas where it indicates the immediate and relevant effect of the research. In addition, these finds from the research can involve in the change in medical guidelines and policies, reflecting the strength in the impact potential

Thematic Analysis: Topic modelling

By employing large language models (LLMs) such as built in BERT topic modelling, this research has mapped the narratives from impact case studies into a semantic latent space. By clustering impact case studies based on their proximity in this latent space, the research has identified 40 clusters of related case studies across Panel A's text data. This approach classifies the topic purely based on the textual content of the impact details, rather than relying on the author's pre-defined categorisations which can be subjective and may lack the nuance needed to capture the interconnections between impact across various sectors. As Pollitt et al. (2023) note, this method aims to reduce the dependency on the subjective "structure" and "defined categorisation", leading to a more objective understanding of the relationships between different impacts, affected sectors, and the diverse beneficiaries that traditional defined categories may oversimplified and not fully capture it.

To assign meaningful labels to the topics, we examined the top 20 most representative words associated with each cluster, as identified by BERTopic. These words reflect the common themes and key terms that are prevalent in the impact case studies within the same cluster. By analyzing the case studies in conjunction with these representative words, we were able to devise meaningful labels that capture the essence of each topic. This process ensures that the topics are based purely on the textual content, providing an objective classification of the case studies.

ID	***Topic name***	***Characteristics (top 20 words)***
1	Mental Health and Psychological Well-being	mental, mental health, psychological, mbct, depression, health, intervention, young, service, disorder, psychosis, young people, iapt, training, itq, wellbeing, mindfulness, ptsd, people, sleep
2	Pharmaceutical Companies and Drug Development	company, product, drug, technology, pharmaceutical, market, ltd, discovery, investment, commercial, development, patent, cell, drug discovery, new, gbp, growth, protein, drug development, camstent
3	Conservation and Biodiversity Protection	conservation, forest, specie, biodiversity, wildlife, climate, iucn, elephant, species, climate change, protect, ecosystem, natural, nature, saiga, endanger, ipcc, peatland, area, trade
4	Dementia and Social Care	dementia, care, social, social prescribing, epicc, care home, home, people, training, carers, arbd, dementia care, prescribing, people dementia, advanced dementia, mhl, spiritual, support, spiritual care, live dementia
5	Eye Health and Optometry	eye, vision, light, visual, optometrist, diabetic, glaucoma, lens, eye health, drex, colour, eyecare, lighting, peek, visual stress, eye care, nystagmus, cataract, neuroeyecoach, arclight
6	Autism and Child Education	autism, child, teacher, comprehension, reading, read, school, phonics, language, asd, parent, qbtest, dyslexia, wait, mdat, teach, elips, literacy, mindmindedness, education
7	Stroke and Rehabilitation	stroke, rehabilitation, cardiac, exercise, heart, esd, aphasia, stroke patient, stroke service, stroke care, cardiac rehabilitation, fnd, edinburgh neuroscience, stroke survivor, edinburgh, thrombolysis, neuroscience, poststroke, reachhf, heart failure
8	Midwifery and Maternity Care	midwifery, baby, maternity, preterm, birth, breastfeeding, stillbirth, pregnancy, breastfeed, maternal, woman, neonatal, dcc, midwifery unit, equist, unit, newborn, baby friendly, miscarriage, midwife
9	Criminal Justice and Policing	police, davis, witness, crime, intermediary, justice, criminal, composite, victim, evofit, force, identification, facial, laughter, ssp, suspect, vulnerable, superrecogniser, officer, superrecognisers
10	Sepsis and Emergency Care	sepsis, pincer, pharmacy, ambulance, prescribe, error, cpr, medication, patient safety, pharmacist, safety, resuscitation, respect,

		community pharmacy, emergency, script, vancomycin, ohca, scotland, medication error
11	Vaccination and Immunisation	vaccine, vaccination, hpv, immunisation, influenza, rabies, cervical, meningitis, flu, rsv, pcv, screening, menacwy, pneumococcal, gavi, influenza vaccination, jcvi, schedule, age, rotavirus
12	Wound Care and Medical Devices	alsmnd, wound, bactiseal, dipals, device, catheter, tvlc, compression, collar, continence, leg, ulcer, venous, wound care, malt, pressure, alarm, infection, hai, patient
13	Radiotherapy and Cancer Treatment	radiotherapy, cancer, prostate cancer, breast, prostate, breast cancer, docetaxel, adjuvant, men, abiraterone, treatment, trastuzumab, boadicea, stampede, woman, metastatic, chemotherapy, fraction, trial, patient
14	Nutrition and Sugar Consumption	sugar, drink, food, breakfast, sodium, intake, tax, salt, dietary, sugary, obesity, levy, school, soft, consumption, sugary drink, soft drink, sugar content, healthy, takeaway
15	Agriculture and Crop Management	seed, wheat, cassava, crop, farmer, potato, variety, pest, fungicide, farm, tomato, resistance, agricultural, barley, plant, agriculture, production, yield, breeding, novapro
16	Alcohol Misuse and Substance Abuse	alcohol, misuse, drug, dry january, dry, thn, antidoping, wada, substance, nps, dope, january, petrczi, naloxone, alcohol change, acmd, safety, abertay, dfsa, fatigue
17	Violence, Bullying, and Abuse	violence, school, abuse, bullying, sexual, domestic, hbsc, glass cliff, cliff, domestic abuse, iris, schools, glass, young, young people, child, ace, behaviour, natsal, cyberbullying
18	Gene Therapy and Genetic Disorders	gene therapy, gene, mirbase, lrrk, frda, disease, eculizumab, microrna, npc, dmd, miglustat, therapy, haemophilia, sma, treatment, patient, trial, gaucher, therapeutic, translarna
19	Dental and Oral Health	dental, oral, oral health, fluoride, toothpaste, caries, dentist, childsmile, dental anxiety, dentistry, tooth, etw, biominf, teeth, biomin, acff, health, decay, toothbrushing, child
20	Malaria and Vector Control	malaria, mosquito, control, vector, schistosomiasis, dengue, vector control, resistance, insecticide, wolbachia, mda, zika, elimination, aegypti, anthrax, itns, malaria control, insecticide resistance, burkina, burkina faso
21	Veterinary and Food Safety	pig, campylobacter, btb, food, farm, mastitis, actiphage, efsa, industry, dairy, cattle, dmcp, animal, sheep, bovine, meat, antibiotic, test, farmer, vras
22	Robotic Surgery and Orthopaedics	surgery, robotic, hip, surgeon, surgical, robotic surgery, orthopaedic, eras, replacement, rolarr, hip replacement, bariatric, covidsurg, knee, csaw, total hip, cement, rectal, patient, urology
23	Arthritis and Rheumatology	arthritis, osteoarthritis, gout, start back, escapepain, rheumatology, back, secukinumab, sycamore, uveitis, adalimumab, eular, start, psa, msk, musculoskeletal, bevacizumab, odyssey, versus arthritis, ivan
24	Suicide Prevention and Mental Health	suicide, suicide prevention, prison, selfharm, prevention, storm, pesticide, prisoner, rail, probation, james place, ncish, suicidal, ban, hmp, staff, suicides, ebm, james, suicide selfharm
25	Cancer Treatment and Clinical Trials	aml, lenvatinib, lymphoma, relapse, alemtuzumab, treatment, cll, trial, chemotherapy, cellbank, flotetuzumab, leukaemia, mylotarg, molecular, myeloma, mcl, patient, obinutuzumab, uhcc, ibrutinib
26	Pollination and Environmental Protection	pollinator, pesticide, bee, pollinators, ppdb, wildflower, pollination, grower, environmental, plant, defra, crop, pheromone, ipbes, urban, pollinator strategy, land, national pollinator, strip, bat
27	Smoking and Air Pollution	tobacco, air, smoking, smoke, packaging, air pollution, pollution, smokefree, ecigarettes, tobacco control, standardised packaging, sps, quit, standardised, air quality, stop smoking, noise, vaping, clean, stoptober

28	Sports and Athlete Performance	sport, football, naclerio, athlete, injury, athletes, company, racerunning, physical activity, sport nutrition, physical, performance, race, stretch, drpe, crown sport, endurance, stirlings, jet lag, jet
29	Lung Cancer and Screening	lung, asl, lung cancer, earlycdtlung, lumpo, oncimmune, predict, fracture, mri, cancer, osteoporosis, frax, breast, predict prostate, prostate, peptest, predict breast, cytospongetff, earlycdtlung test, lvm
30	Cancer Support and Awareness	cancer, macmillan, truenth, pcihn, cancer awareness, cancer support, move pack, followup, cancer survivor, breast, macmillan cancer, secondary breast, education cancer, awareness programme, breast cancer, cancer care, tya, move, telephone, facet
31	Veterinary and Animal Welfare	veterinary, dog, welfare, animal, cat, qba, rvc, ncrs, owner, animal welfare, rabbit, feline, pet, marmoset, vet, mouse, pain, primate, laboratory, xenopus
32	Diabetes and Weight Management	diabetes, type diabetes, glucose, weight, power, insulin, mody, beta, beta score, blood glucose, remission, weight management, monogenic diabetes, type, monogenic, selfmonitoring, calculator, selfmonitoring blood, islet, digem
33	Cardiac Health and Chest Pain	troponin, ctca, chest pain, chest, highsensitivity, cvlprit, heart attack, gaze, attack, statin, cmr, heart, sacubitrilvalsartan, coronary, high sensitivity, cardiac, cemarc, cholesterol, highsensitivity troponin, sensitivity troponin
34	COPD and Respiratory Health	copd, asthma, cough, migraine, mepolizumab, inhaler, pollen, bronchiectasis, respiratory, exacerbation, severe, severe asthma, yohannes, triple, eosinophil, forecast, chronic cough, gold, blood eosinophil, chronic
35	Palliative Care and End of Life	palliative, palliative care, care, breathlessness, homeless, hospice, end life, care home, volunteer, spict, home, life care, end, vivaldi, care people, kings, endoflife, uganda, homelessness, kings research
36	Cancer and PARP Inhibitors	olaparib, inhibitor, brca, parp, cancer, parp inhibitor, ovarian, lynparza, rucaparib, fulvestrant, braf, breast, ovarian cancer, breast cancer, capivasertib, mutation, brca mutation, pancreatic, icr, astrazeneca
37	Genetic Disorders and Genome Sequencing	gene, genetic, rare, hgmd, genomic, genome, mutation, rare disease, genome project, amish, laboratories, diagnosis, deficiency, disease, test, testing, laboratory, leeds, genomics, diagnostic
38	Gambling and Financial Behavior	game, gamble, gambling, debt, fca, voicehearing, credit card, credit, etchells, card, video game, minimum, hellblade, shq, voices, repayment, player, ninja, ninja theory, voicehearers
39	HIV and Sexual Health	hiv, prep, transmission, selftesting, art, hiv selftesting, hiv prevention, hiv transmission, vmmc, circumcision, hiv test, hiv infection, msm, male circumcision, proud, antiretroviral, prevention, ipt, hiv testing, infection
40	Pain Management in Older Adults	pain, pain management, chronic pain, old, old adults, pain old, pincus, chronic, old people, schofield, pain assessment, osteopath, fall old, sensecam, paper also, adults, fall, faculty pain, pain medicine, british pain
41	Genetic study of Infectious disease	Ebola, wbs, sequence, outbreak, sequencing, pathogen, genome, viirus, genomics, coguk, sarscov, enterobase, transmission, infection, salmonella, genomic, isolates, difficile, laboratory, epidemic

Figure 11: Topics identified with BERTopic, with the top words and impact case studies associated with each topic

Topics across Disciplines

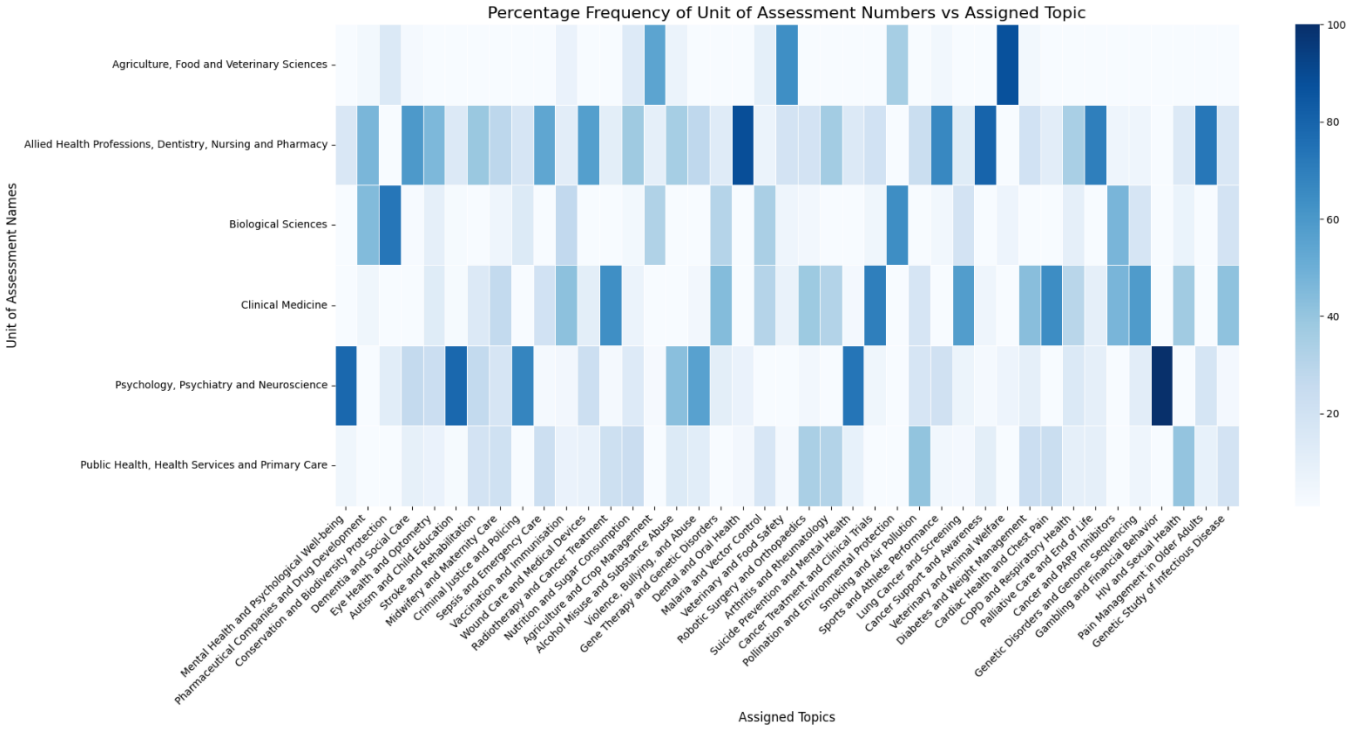


Figure 12: Percentage frequency of unit of assessment numbers across assigned topics

Figure. 13 visualises the percentage frequency of different topics across various Units of Assessment (UoAs) showing how frequently certain topics are associated with specific UoAs. The figures suggest that that some UoA like “Allied Health Professions, Dentistry, Nursing and Pharmacy” have consistent moderate-to-high association across a range of topics, such as “Pharmaceutical Companies and Drug Development”, “Dementia and Social Care” and “Dental and Oral Health” indicating a broad range of engagement with multiple research areas. Similarly, Clinical Medicines are associated with a broad range of interpretations of impact such as “Radiotherapy and Cancer Treatment”, “Cardiac Health and Chest Pain” and “Genetic disorders”.

In comparison, certain UoA such as “Biological Sciences”, are highly concentrated in topics related to genetics, biodiversity and disease like “Pollination and Environmental Protection”, “Conservation and Biodiversity protection” and “Pharmaceutical Companies and Drug Development”. Likewise, “Agriculture, Food and Veterinary Sciences” is highly concentrated in topics like “Veterinary and Animal Welfare” and “Agricultural and crop Management”.

Topics across types of impact

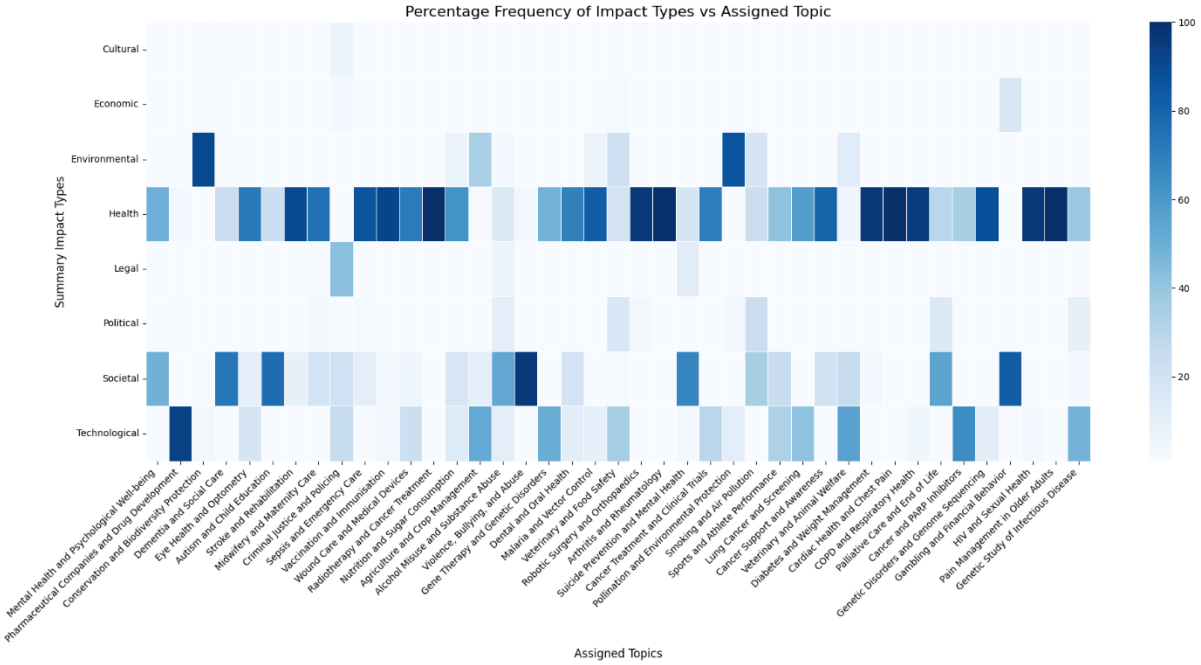


Figure 13: Percentage frequency of impact types vs assigned topic

The Figure. 13 above illustrates the frequency of different impact types across different interpretations of impact. The figure indicates that most of the impact case studies in the Medicine, Health and Life Sciences have “Health”, “Societal” and “Technological” impact, rarely showing “Economic” and “Cultural” impact. Further analysis is needed to determine whether limited exposure to economic and cultural impact factors could affect the model's performance on unseen data in those areas.

Topics and Rating

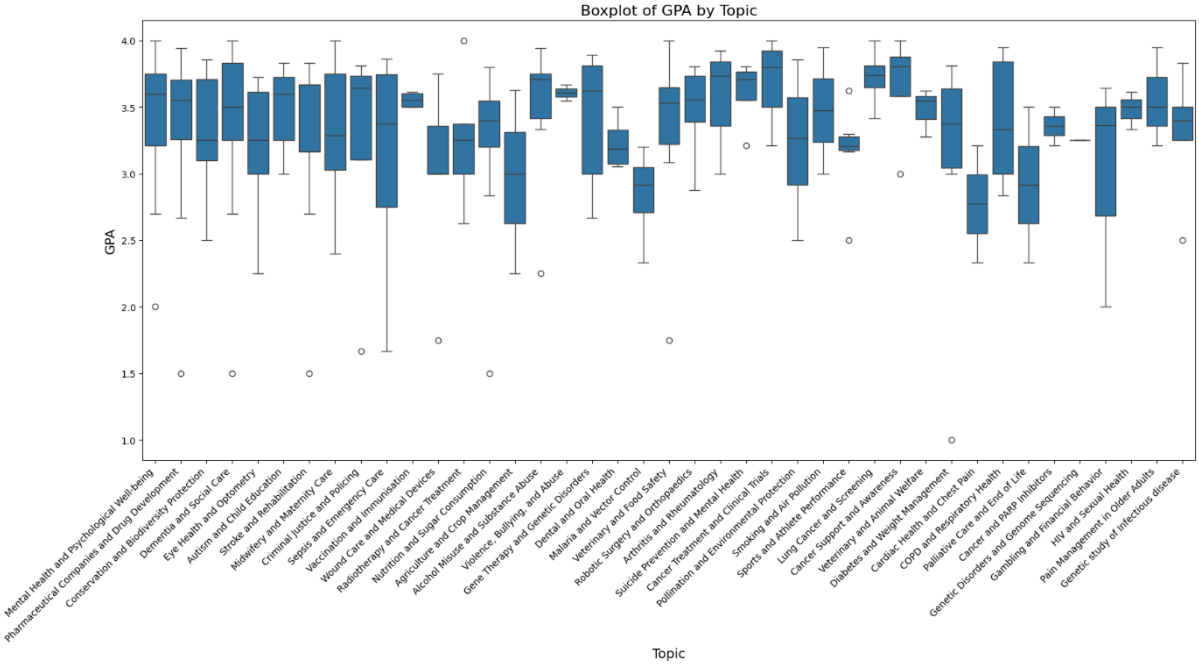


Figure 14: Boxplot of GPA by topic

There has been research that suggested that certain types of impact that are more quantifiable may be favoured (e.g. economic or health) over less tangible impact (e.g. cultural or societal changes) (Weinstein et al., 2019; Manville et al., 2015). This is because these types of impact can be more easily measured and evidenced in impact case studies (Grant et al., 2009). Building on the work of Terama et al (2016), this research attempts to identify if specific interpretations of impact are likely to be favoured by the panellists. T

The results, as presented in Figure 14, indicate that the whiskers of each topic's boxplot do not overlap, implying any statistically significant differences in GPA outcomes across topics. This suggests that no impact interpretation was favoured over another, indicating a consistent impact evaluation across different topics. However, this research found differences in the ranges of GPA scores between impact case studies. Some topics like "Sepsis and Emergency care", "Gene Therapy and Disorders", and "Gambling and Financial Behaviour" had extremely large ranges, indicating greater variability across these topics.

In contrast, specific interpretations of impact related to cancer, like "Lung Cancer and Screening", "Cancer Support and Awareness", and "Cancer treatment and clinical trials", have high means and little variation. This suggests that these topics achieve higher scores more consistently. Similarly, topics like "cardiac health and chest pain" and "Malaria and Vector Control" have low variance and means, suggesting they consistently achieve lower scores. More analysis is required to understand how predictive models would behave with words associated with these topics.

The primary research this paper has obtained suggest that research with impact that reaches multiple impact types is likely to score higher. However, when comparing the GPA of topics and their association with specific impact types, we found that topics associated with multiple impact types like “Sports and Athlete performance”, which is associated with “technological”, “societal” and “health” impact types does not necessarily score higher.

Linguistic features

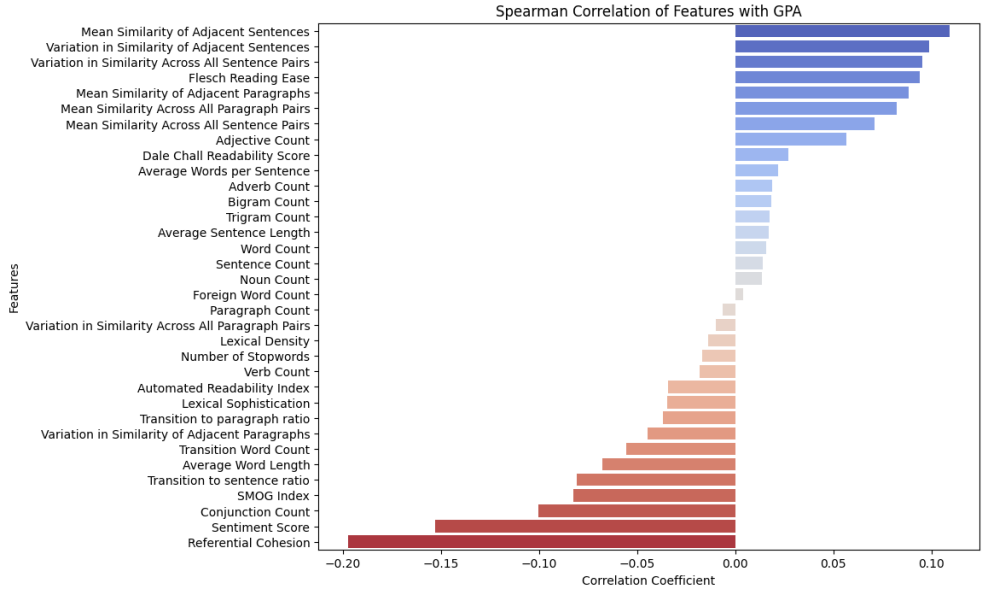


Figure 15: Pearson correlation of handcrafted features with GPA

Before exploring more complex relationships and interactions between the handcrafted features, this research first examines the univariate relationship between each feature and GPA independently. This initial analysis employs the Spearman correlation coefficient to assess the strength and direction of the relationship. This approach is appropriate because we cannot assume a linear relationship between all variables and GPA. The Spearman correlation coefficient, unlike Pearson's, does not assume linearity, making it suitable for evaluating the relationships between the handcrafted features and GPA (McCrum-Gardner, 2008).

The results of the Spearman correlation coefficients between the handcrafted features and GPA are illustrated in Figure 13. The figure shows the strength and direction of their relationship. The top features with positive monotonic relationships are metrics employed by the research to measure the lexical cohesion of impact case studies. Impact case studies with lower lexical cohesion are characterised by more disconnected and loosely correlated ideas. The lack of a consistent and smooth flow of ideas in impact case studies makes it difficult for the reader to discern ideas leading to weaker narratives.

The readability metric with the highest magnitude correlation with GPA is the SMOG index, with a negative correlation of –0.10. In comparison, the Flesch Reading Ease Readability metric had a positive monotonic relationship with a GPA of around 0.08, and the Dale Chall Readability metric had a lower positive correlation of 0.03. Therefore, studies that are easier to read (Flesch) tend to score higher, while those that are more complex (SMOG) may suffer lower scores due to reduced clarity. Vocabulary complexity (Dale-Chall) plays a more minor role in determining GPA. This is consistent with our primary research and findings in the literature review that impact case studies with less jargon and more apparent, concise language indicated by higher Flesch scores, and lower SMOG scores tend to score higher.

There is a weak correlation between length-based features, such as the word count and sentence count. This is expected considering the word limit imposed by the REF criteria. Although, since this limit is not strictly enforced, the results suggest that institutions that do not conform to this word limit do not have an unfair advantage, nor are they at a disadvantaged.

A negative correlation between the sentiment and GPA could suggest that a strongly positive sentiment might be interpreted as overstating the significance and success of the research. Panellists might prefer a more neutral and objective tone. The most substantial negative monotonic relationship with GPA among our handcrafted features was referential integrity, measured as the ratio of pronouns to nouns. While correct use of pronouns can lead to good referential cohesion, overusing specific pronouns without sufficient use of nouns can lead to indirect and unclear language (Stonjic, Stone and Lepore, 2017). Therefore, impact case studies with lower referential cohesion are likely to have less explicit references to specific nouns, such as institutions, beneficiaries or outcomes, which can introduce ambiguity, leading to uncertainty about the impact stated.

In general, the handcrafted features have a very weak to no monotonic correlation with the GPA. This corroborates the findings of Reichard et al. (2024), While there may be some differences between higher and lower-scoring impact case studies, we find that similarities between these impact case studies are more pronounced. This is echoed by the finding from our primary research that linguistic features do not deter from the overall criteria of significance and reach.

Predictive modelling

This section aims to dive into and interpret the behaviour of employed machine learning models to understand the patterns captured by these models and how valuable these patterns are in capturing significance and reach. By comparing and interpreting decisions made by increasingly more complex models under different configurations of input features, this section identifies the strengths and limitations of our chosen models for capturing significance and reach.

This section is split into two parts. The first part identifies the optimal configuration of input features for each of the hyperparameter-tuned models, and the understanding obtained from the previous section attempts to explain the differences in performance between these methods. The second part uses the compares the saliency maps of the models with their optimal configuration to obtain deeper understanding about the model's behaviour.

Quantitative Analysis

Dummy Classifier

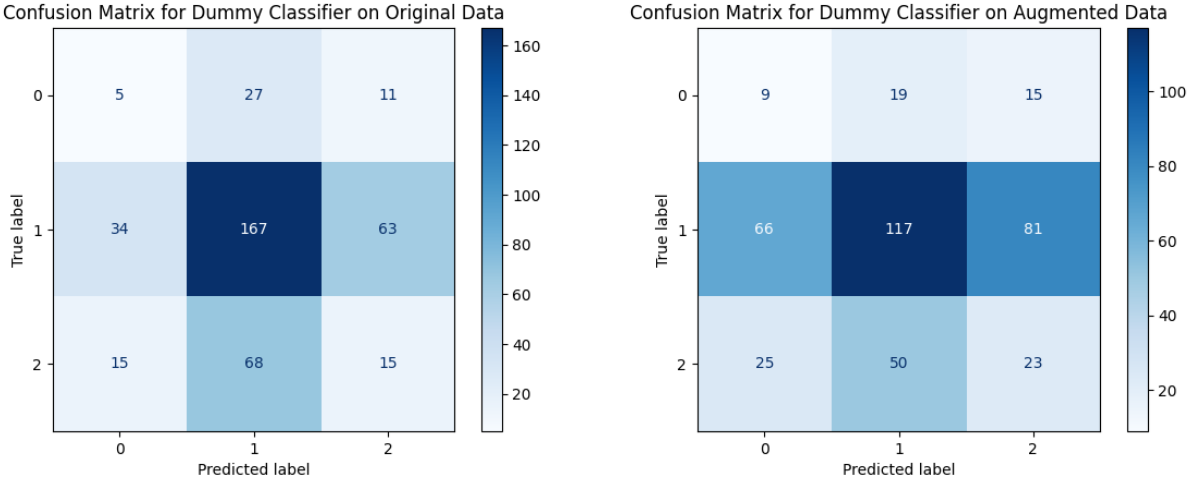


Figure 16: Confusion matrix for dummy classifiers (stratified) on Original and Augmented data

Features	Metrics (macro)	Original Data	Augmented Data
Dummy Features	Precision	0.30	0.30
	Recall	0.30	0.30
	F-1 Score	0.30	0.29
	Accuracy	0.46	0.37

Figure 17: Performance metrics (macro) for dummy features on original and augmented data

The dummy classifier serves as a useful simple baseline model that does not rely on any input features to make predictions. Instead, it makes predictions based solely on the distributions of data. In a typical setup, such as the model of Stratified K-Fold Cross Validation, the dummy classifier will predict the output according to the proportion of class in the training dataset.

This type of classifier is a valuable benchmark for comparison with more complex models. By comparing their performance to the dummy classifier, it provides a better understanding of whether an employed model has learned meaningful patterns from the input features or if its performance is caused by random chance.

Naive Bayes

Features	Metrics (macro)	Original Data	Augmented Data	SMOTE
TF-IDF	Precision	0.22	0.39	0.47
	Recall	0.33	0.37	0.57
	F-1 Score	0.26	0.34	0.45
	Accuracy	0.65	0.65	0.47
BoW	Precision	0.55	0.30	0.52
	Recall	0.49	0.30	0.46
	F-1 Score	0.51	0.29	0.48
	Accuracy	0.64	0.37	0.62
TF-IDF + Handcrafted Features	Precision	0.38	0.34	0.17
	Recall	0.49	0.47	0.47
	F-1 Score	0.29	0.27	0.22
	Accuracy	0.30	0.27	0.20
BoW + Handcrafted Features	Precision	0.44	0.44	0.45
	Recall	0.47	0.47	0.47
	F-1 Score	0.45	0.45	0.45
	Accuracy	0.53	0.53	0.53
Handcrafted Features	Precision	0.41	0.38	0.38
	Recall	0.41	0.40	0.41
	F-1 Score	0.41	0.36	0.36
	Accuracy	0.53	0.40	0.41

Figure 18: Performance metrics (macro) for various feature sets on original, augmented, and SMOTE data

By virtue of their design, both TF-IDF and BoW focus on identifying words and phrases specific to certain classes, helping distinguish between different categories and classes. Whilst this is useful for certain classification tasks, in specialised documents, such as impact case studies, this may downplay important terms that appear frequently amongst documents that may be crucial in understanding the core content of impact case studies. TF-IDF further weights down the importance of words that appear frequently across documents, putting more of an emphasises rare terms, leading to even more words and phrases being overlooked, which could have contributed to a more accurate classification of impact case studies. Leading to the issue being exacerbated when TF-IDF is used as the input feature, which could explain the decrease in performance in models using this method, as seen in Figure 18.

This is demonstrated by the differences in the terms found in the word clouds for each class when using BoW (shown in Figures 8, 9 and 10). These figures suggest that by over-penalising frequent terms across documents, models that utilise TF-IDF could be overfitting to specific words and phrases like “txa”, “mepolizumab” and “wgs”. Disregarding terms like “estimating”, “guidelines”, and “recommend”, which are often used to explain the quantifiable significance of the research impact. This is corroborated by Figure 20 (b) which shows the global feature importance for Naive Bayes, although the words “estimating”, “guidelines”, and “recommend”, are consider meaningless by TF-IDF, they have been shown to be strong predictors in BoW.

Furthermore, both BoW and TF-IDF models exhibit signs of institutional bias, regardless of their different approaches on word frequency. This can be attributed to the TF-IDF model behaviour that balances the frequency of terms in the local document and the global rarity. This suggests that while the words “Oxford”, “Edinburgh”, “them”, and others frequently occur in the local document associated with high-score impact case studies, they occur less frequently in the entire dataset, reflecting the global rarity. Thus, this does not eliminate the potential bias for institutions from both models but rather signifies it in the TF-IDF model. Therefore, to better predict and capture the nuanced aspect of the impact case studies, it necessitates employing a more sophisticated model

that could account for the depth and extent of the topic and impact case studies covered to manage the novel and emerging areas that are equipped with the transformative impact.

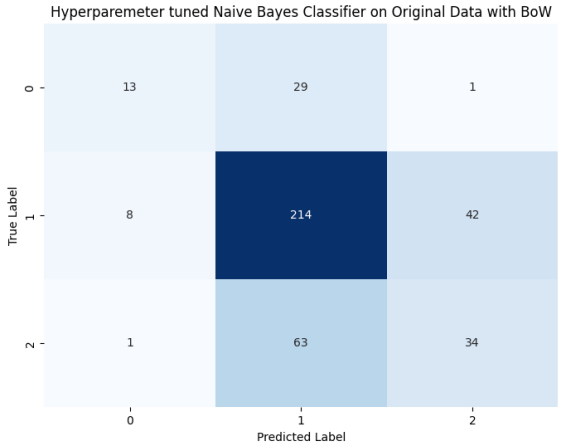
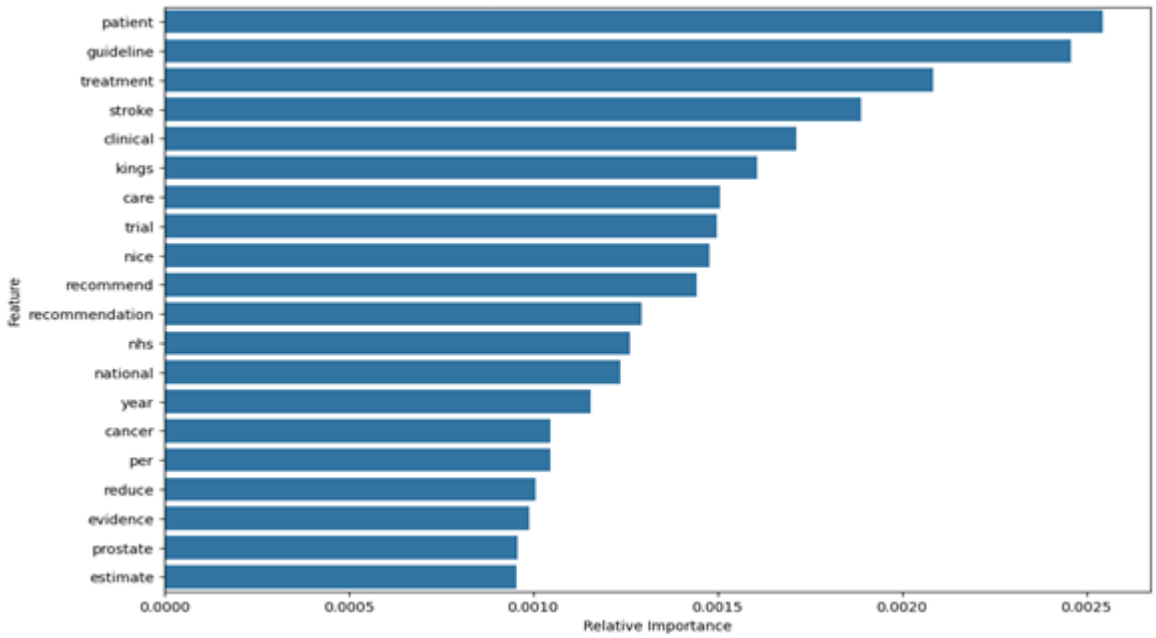
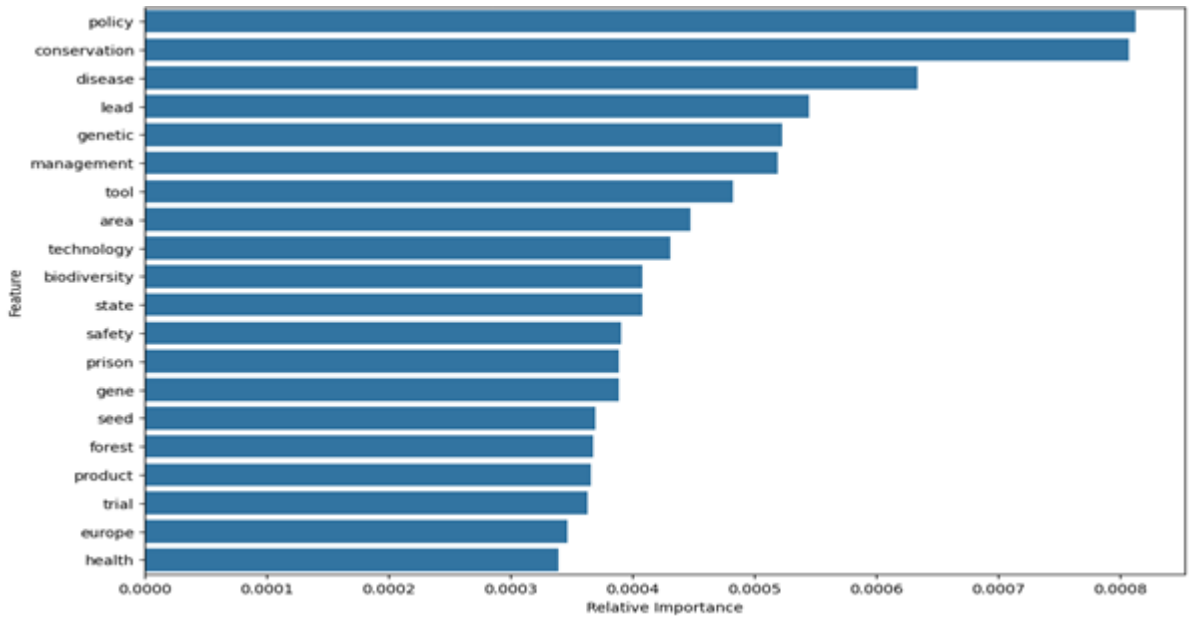


Figure 19: Confusion matrix for Naive Bayes on Original data

(a)



(b)



(c)

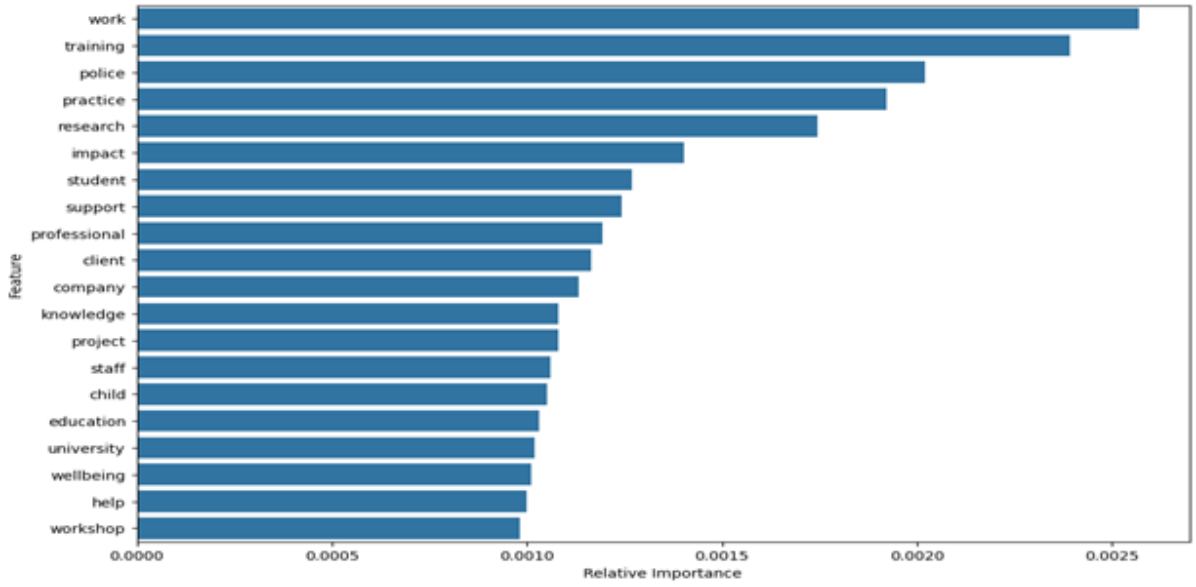


Figure 20: The top 20 discriminative features for each class, class 4*(a), class 3*(b), and class2* or below(c), under the Naive Bayes model with Bag-of-Words (BoW), with features ranked by their relative importance.

XGBoost

Features	Metrics (macro)	Original Data	Augmented Data
TF-IDF	Precision	0.44	0.50
	Recall	0.42	0.44
	F1 Score	0.42	0.46
	Accuracy	0.58	0.63
BoW	Precision	0.46	0.55
	Recall	0.44	0.41
	F1 Score	0.45	0.42
	Accuracy	0.58	0.68
Longformer Embeddings	Precision	0.57	0.55
	Recall	0.44	0.50
	F1 Score	0.46	0.52
	Accuracy	0.66	0.68
Longformer Embeddings + Handcrafted Features	Precision	0.49	0.65
	Recall	0.47	0.52
	F1 Score	0.46	0.55
	Accuracy	0.64	0.70
OpenAI Embeddings	Precision	0.64	0.59
	Recall	0.45	0.46
	F1 Score	0.47	0.49
	Accuracy	0.66	0.67
OpenAI Embeddings + Handcrafted Features	Precision	0.62	0.64
	Recall	0.48	0.48
	F1 Score	0.51	0.51
	Accuracy	0.67	0.68
FastText Embeddings	Precision	0.48	0.52
	Recall	0.44	0.44
	F1 Score	0.45	0.46
	Accuracy	0.59	0.63
FastText + Handcrafted Features	Precision	0.52	0.53
	Recall	0.44	0.44
	F1 Score	0.46	0.46
	Accuracy	0.63	0.64
Handcrafted Features	Precision	0.40	0.38
	Recall	0.38	0.35
	F1 Score	0.38	0.35
	Accuracy	0.40	0.41

d

Figure 21: Performance metrics (macro) for different embedding and handcrafted feature combinations on original and augmented data

The results illustrated in Figure 21 and Figure 18 suggest that Naive Bayes performs better on frequency-based methods than XGBoost. Naive bayes is based on the Bayes Theorem which assumes that the features are conditionally independent, although this assumption is rarely met in practice, it makes Naive Bayes particularly well-suited for frequency-based methods and high-dimensional vector spaces compared to more sophisticated models like XGBoost.

These are Uncontextualized word embeddings, such as FastText, are widely considered an improvement to frequency-based methods because they transform terms into a continuous vector space where more closely related terms are closer together. This means, instead of just relying on the co-occurrence of words, the model can understand words that are closely related in meaning, even if they appear in different contexts. This means that it can capture more nuanced patterns between impact case studies than frequency-based methods justifying the marginal increase in performance.

Fine-tuning a transformer is a computationally extensive task which requires extensive computational resources that may not be easily accessible. In contrast, using BERT embeddings is more lightweight making it suitable for cases where extensive resources are not available (Rogers, Kovaleva and Rumshisky, 2021). These fixed length embeddings from transformers can be seamlessly integrated with handcrafted features in the machine learning pipeline for better interpretability whilst still offering competitive results.

The Longformer embeddings are pretrained on a large corpus and already contain a lot of contextual information without needing to be explicitly fine-tuned on a downstream task. However, using these embeddings without fine-tuning may lead to a compromise in performance, given these embeddings are often extracted from the final layers of the model, abstracting a lot of the details from the intermediate

layers. Including positional information which could be lost or not fully preserved affecting the model's ability to capture long-range dependencies.

Nevertheless, without having to explicitly fine-tune BERT, XGBoost with Longformer Embeddings has received a macro F1-score of 0.52 which is marginally higher than all other configurations. Additionally, by complementing the embeddings with specific structural and linguistic cues from handcrafted features we observed an increase in the macro F1 score to 0.55.

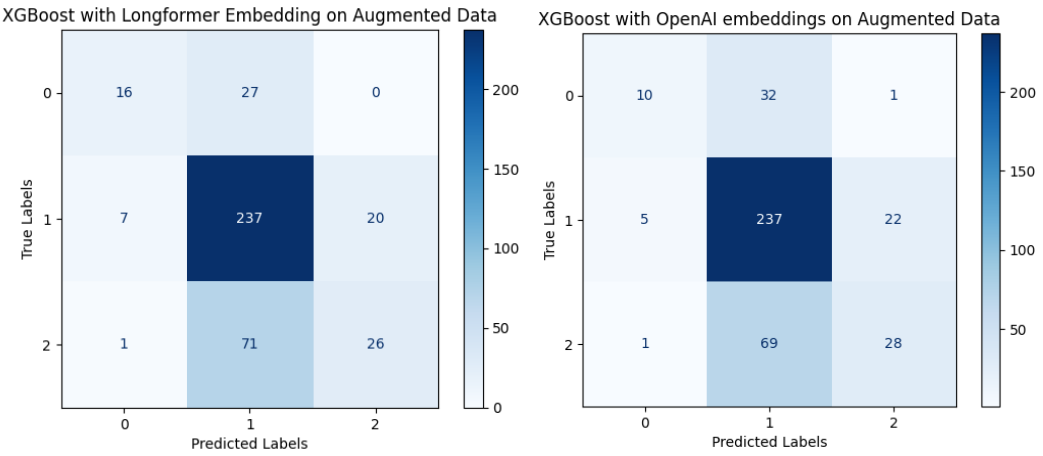


Figure 22: Confusion matrix for XGBoost with longformer and OpenAI embeddings on Augmented Data

Transformer

Features	Metrics (macro)	Original Data	Augmented Data
BERT-base Embeddings	Precision	0.49	0.50
	Recall	0.43	0.56
	F1 Score	0.45	0.52
	Accuracy	0.65	0.60
Longformer Embeddings	Precision	0.48	0.53
	Recall	0.47	0.55
	F1 Score	0.46	0.53
	Accuracy	0.64	0.66

Figure 23: Performance metrics (macro) comparison of BERT-base and Longformer embeddings on original and augmented data

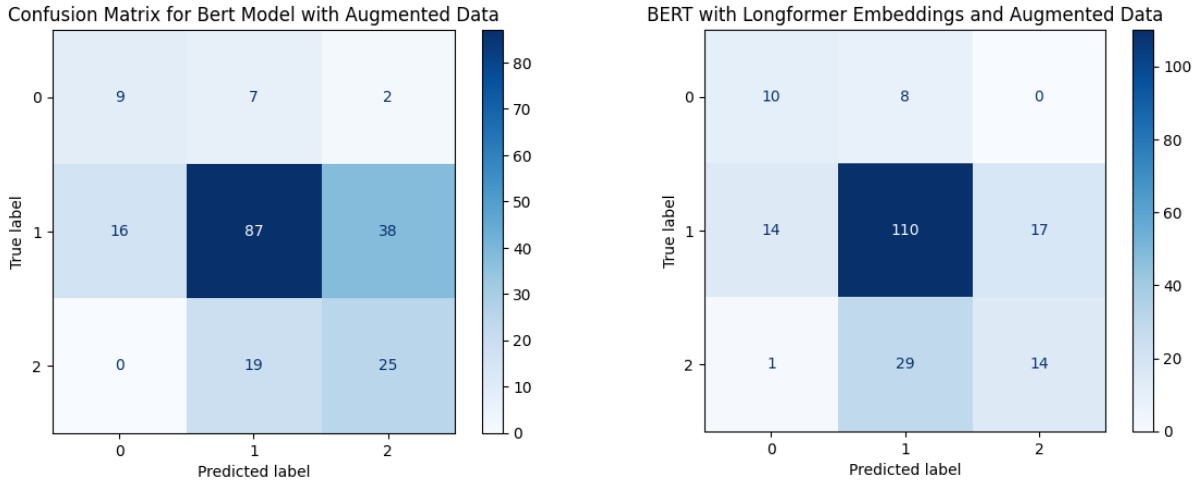


Figure 24: Confusion Matrix for BERT model and BERT with Longformer on Augmented Data.

Both BERT-base and Bert-longformer show improvement across all metrics when augmented data is used, but longformer benefits slightly more from augmentation in terms of precision and F1-score, while BERT-base shows a significant gain in recall and F1-score. Likewise, longformer shows a notable improvement in correctly classifying 3-star impact case studies, whereas BERT-base has a high misclassification in the same label. Although the longformer performs better in correctly predicting label 2 with few misclassifications into other categories.

Although there are improvements in performance between BERT base and longformer, the longformer only performs marginally better despite BERT-base having a much smaller context window. This could suggest most of the significant information is concentrated in the first few tokens, leading to truncation that being that impactful (Liu et al., 2019).

Furthermore, longformer received relatively good performance compared to the other models, with a macro F1-score of 0.53. However, our findings show that XGBoost with BERT Longformer embedding produces comparable and even better results when combined with handcrafted features (F1-score of 0.55) without needing to be explicitly fine-tuned on a downstream task. This suggests that fine-tuning BERT on impact case studies does not lead to much improvement.

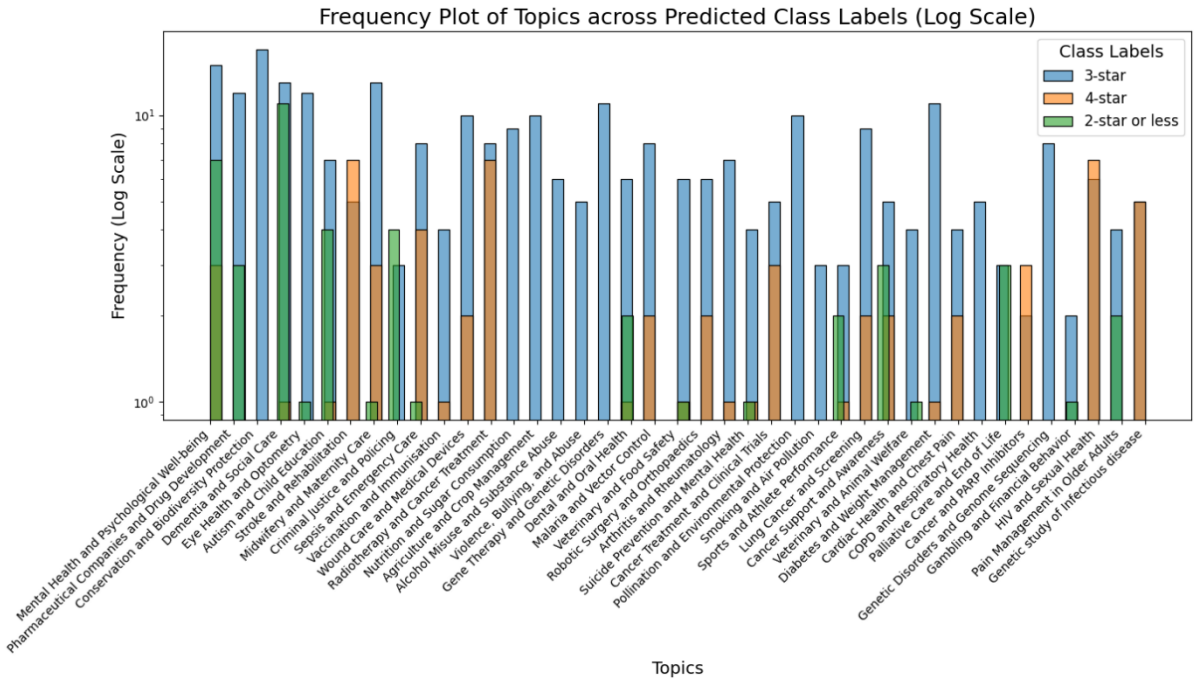


Figure 25: Distribution of topics across the predictions of 2-star or less, 3-star or 4-star impact case studies in the testing set

As illustrated in Figure 24, the 3-star predictions (blue bars) dominate across most topics, which is expected given their overrepresentation in the dataset. Some topics, such as "Mental Health and Psychological Well-Being," "Pharmaceutical Companies and Drug Development," and "Cancer Support and Awareness," are represented equally across all star categories. However, certain topics exhibit an unusually high frequency of 2-star or less impact case studies, including "Dementia and Social Care," "Autism and Child Education," "Criminal Justice and Policing," "Palliative Care and End of Life," and "Pain Management and Older Adults." Conversely, topics such as "HIV and Sexual Health," "Radiotherapy and Cancer Treatment," "Genetic Study of Infectious Diseases," and "Stroke and Rehabilitation" show an unusually high frequency of 4-star predictions. After comparing these patterns with the GPA boxplots in Figure 14, we find no pattern between the distribution of scores in the underlying data and the predicted scores suggesting the model prediction are based on factors beyond interpretations of impact.

Qualitative Analysis

By understanding the models as “black boxes”, this research looks to model-agnostic techniques like LIME (Local Interpretable Model-agnostic Explanations) and SHAP (Shapley Additive exPlanations), which abstract away details of the complex inner workings of these models by focusing entirely on how the input words and phrases influence the output. Both LIME and SHAP highlight the contribution of the individual words and phrases to the final predictions; this means they can show if a particular word or phrase decreases or increases the likelihood of an impact case study being graded into a specific rating category. While these approaches are valuable, there are limitations. Since they only provide individual local explanations of model prediction for each example instead of broad explanations of how the model works. Therefore, we can only understand the model’s predictions in a specific context, losing the detail of the model’s overall behaviour. To help mitigate this, this research compares multiple impact case studies to understand how local patterns generalize across impact case studies.

Another challenge arises from the absence of individual scores in the impact case study, which makes it difficult to evaluate models on specific examples. To alleviate this, the research focuses on impact case studies from the testing set for Higher Education Institutions (HEIs) within Units of Assessment (UoA), where all the impact case studies have been awarded the same rating. This allows precise comparison between different models’ performance on specific impact case studies.

Impact case Study 1 with title - “Development and implementation of UK tobacco control policy”

Quality Profile: 4-star

Assigned Topic: “Smoking and Air Pollution”

Impact type: “Health”

According to Figures shown below, saliency maps offer a valuable tool for understanding the decision-making process of advanced models like BERT with XGBoost beyond their overall accuracy score. These maps highlight the features the model considers important, revealing distinctions between how different models make predictions compared to the baseline model. The Naive Bayes model focuses on the institutional feature, evident by the term "Hertfordshire", and associates it with class three rating despite the features being uncorrelated to the criteria of reach and significance in the impact case studies, reflecting an inability to generalize beyond the surface level features. In addition, the terms "herts", "let", "rsph" may not meaningfully contribute to the prediction of the significant impact case studies, resulting in poor predictive ability in this baseline model. On the other hand, the advanced models, such as Wording embedding with XGBoost, BERT with XGBoost and BERT LongFormer, present the consideration toward the broader semantic understanding of the text. They focus on terms like "habit", "stress," and "exercise" and their context evidence of reach and significance, which contrasts with the baseline model where these terms lead to lower rating classification from the model's lack of semantic understanding. This latter association with a lower rating of these terms in the baseline model may be attributed

to the qualitative characteristics of the terms, which are challenging for the measurement of significance in research toward society. Furthermore, Advanced models like Longformer and Longformer with XGBoost also show predictive behaviour related to user feedback. This is shown by the model often categorizing the text related to the review of users strongly toward class 3 rating, which swayed the predictive outcome toward the medium significant impact case studies.

On the other hand, there are similar feature importances employed across Naive Bayes, Longformer, and Longformer with XGBoost. The saliency maps indicate that the term "client" is often associated with lower ratings or less significant impact case studies. This highlights a potential bias and low generalization in how certain words are processed and interpreted by the models. In addition, this decision may come from the aspect of sectoral impact in the training set whereby the term “client” is often associated with service delivery or business-to-business interactions rather than impacting the significant toward wide-reaching impact of public, society and environment as other research. Therefore, the models trained on these may associated aggressively and negatively with the business sectoral impact, indicating the potential underrepresented sectoral impact in high-rated case studies.

researchbased something different behavioural intervention develop university hertfordshire commercialise spinout company result broad range impacts deliver economic impact commercial benefit within evidence efficacy clinical health intervention find randomise control trial overseas improve peoples health wellbeing productivity internationally within health workplace domestic settings february rilkes room ltd become social enterprise something different ltd follow investment pax holding ltd investment allow company cofounded fletcher pine social entrepreneur ray richards develop dsd platform lowcost scalable accessible behaviour change intervention aim three target beneficiary improve wellbeing mental physical health productivity individuals collaboration charities healthcare organisation public health commissioners businesses dsd ltd generate revenue august december create support six job core staff and one point sustain roles dsd platform underpin award two major horizon grant combined value change use dsd model change lifestyle behavioural risk factor people heart disease hypertension preventomics apply dsd model development digital platform guide consumer healthy food choices grant include total direct funding specifically smes develop new product services one example dutch design firm onmi for fletcher advisor receive onmi offer two behaviour change programme base dsd alius flex create two product via change programme cookit smart spatula help user adopt healthy cook habit public vote award innovation radar prize mysleeve smart device wrap around cup bottle encourage healthy drinking habits efficacy dsd demonstrate rct actrn register feb habitbased intervention carry researcher faculty health science medicine bond university australia dsd find clinically beneficial achieve longterm weight loss improve diet exercise habits approximately participant lose body weight th benchmark successful healthful weight loss keep month intervention ended researcher concluded habitbased intervention potential change think weight management importantly behave change programme result rct nct register oct involve heart disease hypertension patient three hospital spain netherlands taiwan dsdbased intervention show feasible quality life improve intervention group positive costeffectiveness ratio demonstrate spanish hospital dsd ltd create behaviour change programme address various healthrelated issues impact evaluation carry found that user countries report low level anxiety depression number people clinical level anxiety fell clinical level depression participant comment included get motivate take exercise thing differently life rut grandchild love one favourite dos give something somebody unexpected take neighbour never meet flower biscuit end get whole street involve lady meet month basically completely sedentary exercise all exercise result course ive probably add year year life dsd collaborate charity action happiness design intensive sixweek programme help people develop habit proven bring happiness people around them happiness programme sent people regular small positive action dos break exist habit cycles accord evaluation publish action happiness participant country say would recommend dsd others overall average uplift wellbeing comment share zone programme participants dsd collaborate hertfordshire public health run three programme healthy herts focus exercise diet less stress herts explore way combat stress happiness herts develop programme run action happiness three intervention engage residents evaluation show subjective wellbeing people clinical level anxiety depression move healthier levels physical activity weight average six week alcohol consumption fell representative comment participants capture video testimony included definitely help refocus mental energy thinking patterns feel extremely low want anything change positive attitude spite whats cant even send text know look forward text come isnt amazing intervention royal society public health rsph health wellbeing award its innovative creative approach behaviour change individual hertfordshire rsph highlight dsds accessibility scalability underpinned research good practice technology hertfordshires director public health say dsd shows people longterm condition move clinical level depression fraction cost facetoface care hertfordshire public health commission dsd programme healthy herts banner include programme help people type diabetes manage condition improve mental health young people empower victim domestic abuse dsd develop healthy cov coventry city council public health team design engage resident make small change everyday exercise diet make use citys leisure facilities dsd run smoking cessation programme partnership brighton heave clinical commissioning group achieve quit rate result dsd finalist partnership nhs category south east health technology alliance awards medicinfo dutch healthcare innovation company invest threeyear agreement distribute dsd beneficiary netherlands include public health centre educational institution local authority care home people diabetes copd medicinfos director said run eight programmes successful base qualitative feedback participants best result care home complete culture change occur among healthcare professionals dictate organisational structure implement structure base client want need dsd also work well blue collar worker building company like conciseness format large corporate organisation use dsd deliver wellbeing training programme employees focus issue resilience stress inclusivity leadership client impact period include general electric usa aviva american express mondelez remploy tui cisco dsd work ciscos team deliver programme aim build inclusive workplace culture offer employee across ciscos europe middle east africa russia region total employee across country participated publish evaluation find uplift people engage others different cultures one employee commented something different change lot habit introduced real change person work home collaboration tui dsd design bespoke leadership training programme combine tuis management development goal drive improve digital skills make available manager tuis perspective leadership programme one international tax manager said sum programme word would life changing regional finance director commented teach let daytoday become strategic leader work mondelez dsd create agility work programme encourage employee adopt new way working programme make available employee seven site across ireland result say likely look new way things similar programme engaged avivas employee encourage embrace different measures evaluation show likely recommend dsd say thing differently result

Figure 26: Feature importance for predicting Impact Case Study 1 using Naïve Bayes with BoW

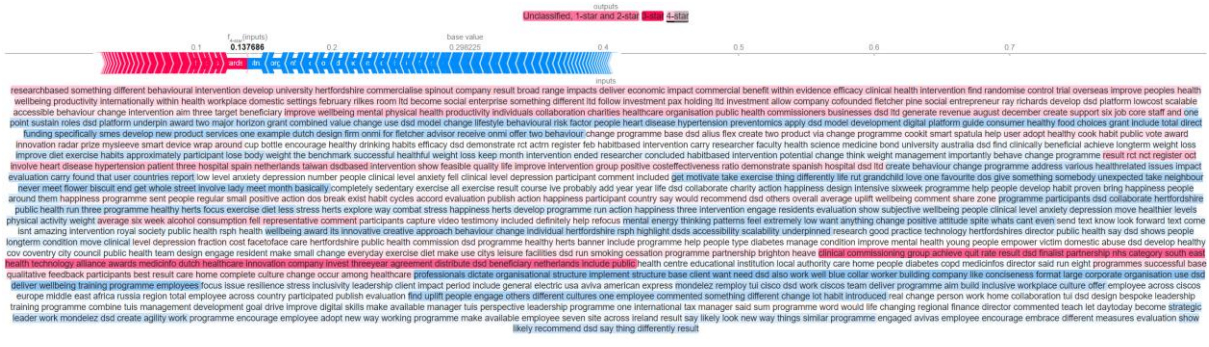


Figure 27: Impact case study predictions using an uncontextualized Embeddings XGBoost.

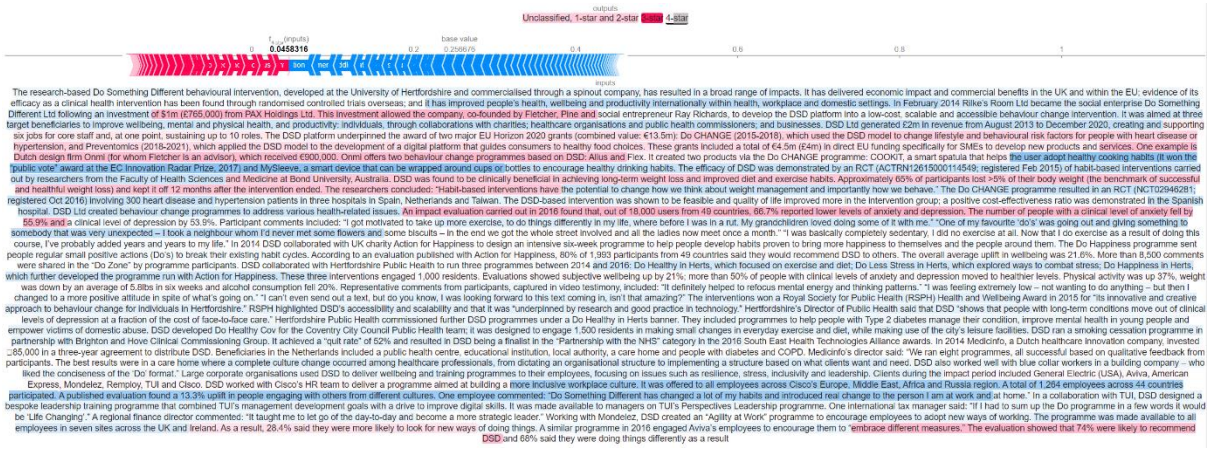


Figure 28: Impact case study predictions using XGBoost with Longformer and saliency analysis.

Figure 29: Impact case study predictions using XGBoost with Longformer and saliency analysis.

Impact case Study 3 with title - "Change the way that we view video games and their effects"

Quality-Profile: 0-Star

Assigned Topic: "Gambling and Financial Behaviour"

Impact type: "Societal"

In this impact case study, the findings across advanced models present a different predictive behaviour than the baseline model, leading to distinct predictive outcomes. The advanced model with Bert indicates the ability to capture semantic meaning and context connotation toward "public perception", and the semantics related to this change in viewpoint contrast to the frequent-based model of Naive Bayes with BoW. This indicates that the models discover the pattern related to the reach and particularly significance, not just relying on the frequency of the words associated with the impact class. This is evident in the red highlights under the class two or less section whereby the sentence relates to significant measurement of "Allowing him to see the screen differently and develop a prospect on video game and screen use" and "allow the opportunities to promote, explain and discuss the underpinning research in a variety of public context". In addition, another element of reach described in impact case studies with "A key stakeholder, in Parent Zone, an organization which provides support and information" has been highlighted as an essential element in predicting the impact case studies with two or lower ratings. This indicates that BERT is factoring in nuanced shifts in meaning and the extent of research significance when making predictions, leading to different classification outcomes.

In contrast, the Naive Bayes model with BoW, which relied on the frequency of the words rather than their semantic meaning, strongly correlates terms like "Journalists," "underpinning," and "viewpoint" with a class three rating. This suggests that the focus of the baseline model may overfit particular words toward specific classes and fail to capture the border context related to the term.

Figure 30: Impact case study predictions using Naive Bayes with BoW.

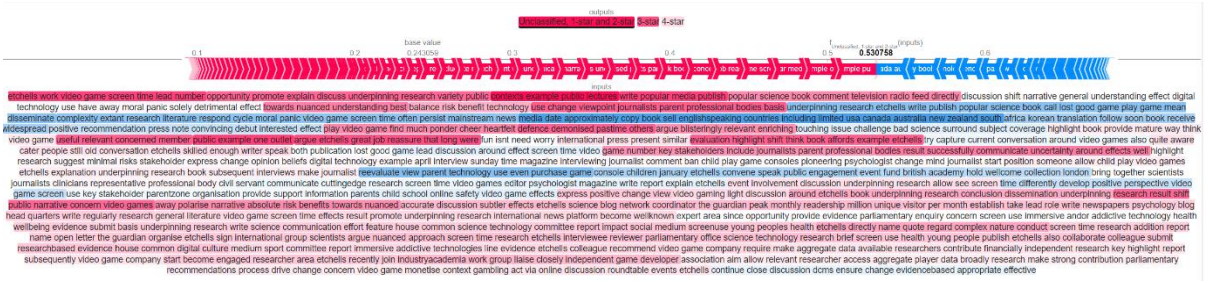


Figure 31: Impact case study predictions using an uncontextualized Embeddings XGBoost.

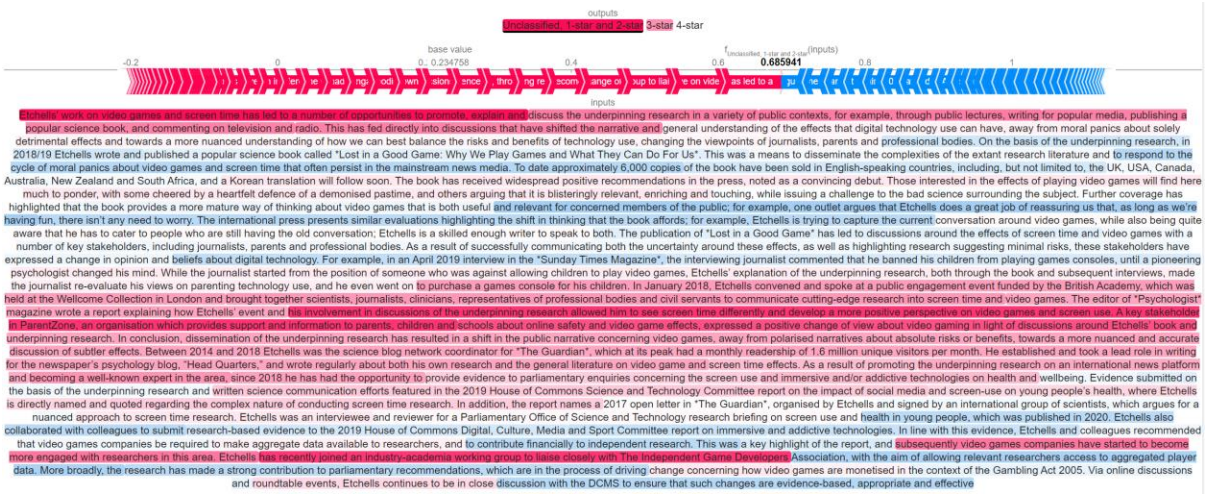


Figure 32: Impact case study predictions using XGBoost with Longformer and saliency analysis.

Figure 33: Impact case study predictions using Bert Longformer and saliency analysis.

Conclusion

This research has mapped the latent space of impact case studies to uncover common themes and topics across impact case studies for a detailed comparison of impact case studies across different topics. As a result, in the thematic model focusing on topic modelling, the research identified 40 topics that tend to focus notably on the healthcare and medicine section, creating an imbalanced dataset. Some topics provided insights into specific beneficiaries and broader interpretations of impact. This assists in identifying certain topics that score consistently, particularly in a cancer-related field. Although these observed topics display higher scores with low variation, they were insignificant enough to produce substantial predictive influence toward the models by building this based on the pioneering literature of Terama et al. (2016) and King et al. (2020), who faced the unidentified board topic of "Asia" and "Cancer" from analysing the whole impact submission. This research, conversely, only focuses on Panel A impact submission, indicating how each identified topic relates to significance, reach, or beneficiary, with many topics directly demonstrating the Unit of Assessment (UoA). However, we found no further evidence of topics associated with high- or low-scoring in impact case studies, providing results similar to those of the previous literature. The linguistic features also show the insignificant predictive influence.

The research observed the presence of institutional bias in a simple baseline model of Naive Bayes with BoW, while the model with IF-IDF further exacerbates this bias that could disadvantage certain researchers from specific universities in classifying the score. It provides some behaviour that discriminates based on the entity as well as company names rather than focusing on the boarder context of research significance and reach criteria aspects. While these simple models do recognise useful predictive features associated with the significance, such as "estimating", "guideline", "bans", "recommendation", and "viewpoint", their understanding of impact remains shallow. They capture surface-level features rather than the full breadth of the impact described.

By comparing the base line model with advanced model, the research found the highest macro F1 score was achieved using the XGBoost with Longformer embeddings and integrating of handcrafted features on augmented data. Interestingly, the model without fine-tuning performs better than the model with fined tuning, indicating the potential overfitting in the small dataset toward the noise and outliers of under and over-representation features in impact case studies. Despite this, the performance differences between the baseline and advanced models were notable. The research founds that transformer-based models, like BERT and Longformer can capture some aspects of significant and reach in impact case studies, but their interpretation of these concepts various, highlighting the complex nature of how impact and its criteria is defied and measured. In addition, no correlation was observed between the high-scoring topic and model prediction outcomes, suggesting that the model may learn from other characteristic features in the impact case studies.

According to the evaluation of model decision-making in saliency maps, we gained deeper insights into the decision-making process of advanced models compared to baseline models. Sophisticated models like Longformer could capture better semantic aspects, leading to more informed classification. On the other hand, simpler models, like Naive Bayes, relied heavily on word frequency without considering the supporting evidence provided in the impact case studies. According to the expert panellists, this distinction between high and low scores is attributed to the ability to quantify research significance and dense evidence, a crucial element in impact evaluation.

While large language models may capture certain aspects of significant and reach, the subjective nature of impact assessment as well as the complexity of real-world effect of research involving the novelty field of research and effect may necessitate better comprehension. Moreover, impact case studies are often carefully crafted to secure funding, reflecting the well-structured narratives and articulation of impact pathways, which may hinder the models' ability to classify based on cohesion, coherence and other linguistic features. It highlights another possible explanation toward poor classification of simpler models with slightly better results in advanced models.

Limitations

The research faces several limitations as it only focuses on the impact of case studies on Panel A, with around 1400 cases from the specific submission period of 2021. This may not provide sufficient representation for the models to fully capture the useful predictive patterns across different fields, leading to overfitting and potential bias in prediction toward a group of beneficiaries and sectoral impact. Additionally, the lack of individual score for impact case studies presents a challenge in evaluation the model performance due to limited number of impacts with known score. Data imbalance with large number of studies found in class 3 further presents difficulties of overrepresentation and skew the result in favour of majority class, particularly baseline model.

Moreover, the black-box nature of the sophisticated models Bert and Longformer employed, despite offering better performance, incurs the issue of poor transparency toward interpretation. This means it necessitates the incorporation of LIME and SHAP to provide local explanation. However, this approach has some constraints that do not offer a complete and robust understanding of the decision-making of the models. Computational limitations, the reliance on specific sections of impact case studies, as well as the lack of individual scores for training and testing samples further limit the research's outcomes. This highlights the broader dataset and details on impact scoring, which could further benefit future research. In future research, it would be good to incorporate more of the logic and process the panellist utilised to assess the significance and reach criteria, along with the evidence evaluation to classify the impact score.

References

1. Alawida, M., Mejri, S., Mehmood, A., Chikhaoui, B., & Isaac Abiodun, O. (2023). A comprehensive study of ChatGPT: advancements, limitations, and ethical considerations in natural language processing and cybersecurity. Information, 14(8), 462.
2. AlZu'bi, S., Mughaid, A., Quiam, F., & Hendawi, S. (2024). Exploring the capabilities and limitations of chatgpt and alternative big language models. In Artificial Intelligence and Applications (Vol. 2, No. 1, pp. 28-37).
3. Amoore, L. (2020). Cloud ethics: Algorithms and the attributes of ourselves and others. Duke University Press.
4. Author(s). (2024). The Role of ChatGPT in Text Preprocessing for Natural Language Processing Tasks. Journal of Artificial Intelligence and Text Processing, 12(3), 45-56.
5. Barbieri, F., Camacho-Collados, J., Neves, L., & Espinosa-Anke, L. (2020). Tweeteval: Unified benchmark and comparative evaluation for tweet classification. arXiv preprint arXiv:2010.12421.
6. Barker, K. (2007). The UK Research Assessment Exercise: the evolution of a national research evaluation system. Research evaluation, 16(1), 3-12.
7. Beltagy, I., Peters, M. E., & Cohan, A. (2020). Longformer: The long-document transformer. arXiv preprint arXiv:2004.05150.
8. Bigo, D., Isin, E., & Ruppert, E. (2019). Data politics: Worlds, subjects, rights (p. 304). Taylor & Francis.
9. Brassington, L. (2022). The future of digital learning resources: Students' expectations versus reality. Higher Education Policy Institute.
10. Brochu, E., Cora, V. M., & De Freitas, N. (2010). A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. arXiv preprint arXiv:1012.2599.
11. Brown, R., & Carasso, H. (2013). Everything for sale? The marketisation of UK higher education. Routledge.
12. Bukar, U. A., Sayeed, M. S., Razak, S. F. A., Yogarayan, S., Amodu, O. A., & Raja Mahmood, R. A. (2024). Text Analysis on Early Reactions to ChatGPT as a Tool for Academic Progress or Exploitation. SN Computer Science, 5(4), 366.
13. Burnell, P. J., Rakner, L., & Randall, V. (Eds.). (2017). Politics in the developing world. Oxford University Press.
14. Choi, H., Kim, J., Joe, S., & Gwon, Y. (2021, January). Evaluation of bert and albert sentence embedding performance on downstream nlp tasks. In 2020 25th International conference on pattern recognition (ICPR) (pp. 5482-5487). IEEE.
15. Christin, A. (2020). The ethnographer and the algorithm: beyond the black box. Theory and Society, 49(5), 897-918.
16. Das, R., Zaheer, M., & Dyer, C. (2015, July). Gaussian LDA for topic models with word embeddings. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers) (pp. 795-804).
17. Devlin, J. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.
18. Dourish, P. (2022). The stuff of bits: An essay on the materialities of information. MIT Press.
19. Duarte, P. O., Alves, H. B., & Raposo, M. B. (2010). Understanding university image: a structural equation model approach. International review on public and nonprofit marketing, 7, 21-36.
20. Fairclough, A. (1995). Martin Luther King, Jr. University of Georgia Press.
21. Farag, Y., Valvoda, J., Yannakoudakis, H., & Briscoe, T. (2020). Analyzing neural discourse coherence models. arXiv preprint arXiv:2011.06306.
22. Feng, S. Y., Gangal, V., Wei, J., Chandar, S., Vosoughi, S., Mitamura, T., & Hovy, E. (2021). A survey of data augmentation approaches for NLP. arXiv preprint arXiv:2105.03075.
23. Gallo, S. A., Schmaling, K. B., Thompson, L. A., & Glisson, S. R. (2021). Grant review feedback: Appropriateness and usefulness. Science and Engineering Ethics, 27, 1-20.
24. Giray, L. (2023). Prompt engineering with ChatGPT: a guide for academic writers. Annals of biomedical engineering, 51(12), 2629-2633.
25. Gow, J., & Redwood, H. (2020). Impact in international affairs: the quest for world-leading research. Routledge.

26. Grant, A. M. (2012). Leading with meaning: Beneficiary contact, prosocial impact, and the performance effects of transformational leadership. Academy of management journal, 55(2), 458-476.
27. Grootendorst, M., & Vanschoren, J. (2020). Beyond bag-of-concepts: Vectors of locally aggregated concepts. In Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2019, Würzburg, Germany, September 16–20, 2019, Proceedings, Part II (pp. 681-696). Springer International Publishing.
28. Gunning, R. (1952). The technique of clear writing.
29. Guthrie, S., Ghiga, I., & Wooding, S. (2017). What do we know about grant peer review in the health sciences?. F1000Research, 6.
30. Halliday, M. A. K., & Hasan, R. (2014). Cohesion in english. Routledge.
31. Heesen, R., & Bright, L. K. (2021). Is peer review a good idea?. The British Journal for the Philosophy of Science.
32. Hochreiter, S., & Schmidhuber, J. (1997). Flat minima. Neural computation, 9(1), 1-42.
33. Horbach, S. P., & Halffman, W. (2020). Innovating editorial practices: academic publishers at work. Research integrity and peer review, 5, 1-15.
34. Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: a review. ACM computing surveys (CSUR), 31(3), 264-323.
35. Jaiswal, A., & Milios, E. (2023). Breaking the Token Barrier: Chunking and Convolution for Efficient Long Text Classification with BERT. arXiv preprint arXiv:2310.20558.
36. Kincaid, J. P. (1975). Derivation of new readability formulas (Automated readability index, fog count, and flesch reading ease formula) for Navy enlisted personnel.
37. King, R. S. (2015). Cluster analysis and data mining: An introduction. Mercury Learning and Information.
38. Kowsari, K., Jafari Meimandi, K., Heidarysafa, M., Mendu, S., Barnes, L., & Brown, D. (2019). Text classification algorithms: A survey. Information, 10(4), 150.
39. Leland, C., Ociepka, A., Kuonen, K., & Bangert, S. (2018). Learning to talk back to texts. Journal of Adolescent & Adult Literacy, 61(6), 643-652.
40. Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... & Stoyanov, V. (2019). RoBERTa: A robustly optimized BERT pretraining approach. arXiv preprint arXiv:1907.11692.
41. Liu, Z., Lin, W., Shi, Y., & Zhao, J. (2021, August). A robustly optimized BERT pre-training approach with post-training. In China National Conference on Chinese Computational Linguistics (pp. 471-484). Cham: Springer International Publishing.
42. Macilwain, C. (2009). The RAE: an assessment too far?. Cell, 139(4), 643-646.
43. Manville, C., Hinrichs, S., Parks, S., Kamenetzky, A., Gunashekar, S., Wilkinson, B., & Grant, J. (2015). Characteristics of high-performing research units. Prepared for the Higher Education Funding Council for England.
44. Manville, C., Jones, M. M., Frearson, M., Castle-Clarke, S., Henham, M. L., Gunashekar, S., & Grant, J. (2015). Preparing impact submissions for REF 2014: An evaluation. RAND Europe: Cambridge, UK.
45. Mc Laughlin, G. H. (1969). SMOG grading-a new readability formula. Journal of reading, 12(8), 639-646.
46. McCann, B., Bradbury, J., Xiong, C., & Socher, R. (2017). Learned in translation: Contextualized word vectors. Advances in neural information processing systems, 30.
47. McCrum-Gardner, E. (2008). Which is the correct statistical test to use?. British Journal of Oral and Maxillofacial Surgery, 46(1), 38-41.
48. McInnes, L., Healy, J., & Melville, J. (2018). Umap: Uniform manifold approximation and projection for dimension reduction. arXiv preprint arXiv:1802.03426.
49. McKenna, H. P., & McKenna, H. P. (2021).
50. McKenna, H. P., & McKenna, H. P. (2021). Research impact: the what, why, when and how (pp. 1-15). Springer International Publishing.
51. Meesad, P., Boonrawd, P., & Nuipian, V. (2011). A chi-square-test for word importance differentiation in text classification. In Proceedings of international conference on information and electronics engineering (pp. 110-114).
52. Merchant, A., Rahimtoroghi, E., Pavlick, E., & Tenney, I. (2020). What happens to BERT embeddings during fine-tuning?. arXiv preprint arXiv:2004.14448.

53. Oancea, A. (2013). Interpretations of research impact in seven disciplines. European Educational Research Journal, 12(2), 242-250.
54. Pidd, M., & Broadbent, J. (2015). Business and management studies in the 2014 Research Excellence Framework. British Journal of Management, 26(4), 569-581.
55. Pinar, M., & Horne, T. J. (2022). Assessing research excellence: evaluating the research excellence framework. Research Evaluation, 31(2), 173-187.
56. Pollitt, A., Sreenan, N., Grant, J., Szomszor, M., Leeworthy, D., & Hughes, D. (2023). The impacts of research from Welsh universities: Final report. The Learned Society of Wales
57. Qin, L., Chen, Q., Feng, X., Wu, Y., Zhang, Y., Li, Y., ... & Yu, P. S. (2024). Large language models meet nlp: A survey. arXiv preprint arXiv:2405.12819.
58. RAND Europe (2012). New challenges for universities. RAND Corporation. Available at: file:///C:/Users/fnatc/Downloads/RAND_CP661.pdf (Accessed: 16 July 2024).
59. Raschka, S. (2014). Naive bayes and text classification i-introduction and theory. arXiv preprint arXiv:1410.5329.
60. REF2021. (2019). Panel criteria and working methods. https://ref.ac.uk/publications-and-reports/panel-criteriaand-working-methods-201902/
61. REF2021. (2022). Results and Submissions. https://results2021.ref.ac.uk/
62. Reichard, B., Reed, M. S., Chubb, J., Hall, G., Jowett, L., Peart, A., & Whittle, A. (2020). Writing impact case studies: a comparative study of high-scoring and low-scoring case studies from REF2014. Palgrave Communications, 6(1), 1-17.
63. Reichard, B., Reed, M., Chubb, J., Hall, G., Jowett, L., & Whittle, A. (2020). The grammar of impact–what can we learn from REF 2014 about writing impact case studies?. Impact of Social Sciences Blog.
64. Reimers, N., Schiller, B., Beck, T., Daxenberger, J., Stab, C., & Gurevych, I. (2019). Classification and clustering of arguments with contextualized word embeddings. arXiv preprint arXiv:1906.09821.
65. Rhodes, R. A. (1994). The hollowing out of the state: The changing nature of the public service in Britain. Political quarterly, 65(2).
66. Robinson, R. (1986). Restructuring the Welfare State: an analysis of public expenditure, 1979/80–1984/85. Journal of Social Policy, 15(1), 1-21.
67. Röder, M., Both, A., & Hinneburg, A. (2015, February). Exploring the space of topic coherence measures. In Proceedings of the eighth ACM international conference on Web search and data mining (pp. 399-408).
68. Rogers, A., Kovaleva, O., & Rumshisky, A. (2021). A primer in BERTology: What we know about how BERT works. Transactions of the Association for Computational Linguistics, 8, 842-866.
69. Shahapure, K. R., & Nicholas, C. (2020, October). Cluster quality analysis using silhouette score. In 2020 IEEE 7th international conference on data science and advanced analytics (DSAA) (pp. 747-748). IEEE.
70. Sivertsen, G. (2017). Unique, but still best practice? The Research Excellence Framework (REF) from an international perspective. Palgrave Communications, 3(1), 1-6.
71. Terama, E., Milligan, B., Jiménez-Aybar, R., Mace, G. M., & Ekins, P. (2016). Accounting for the environment as an economic asset: global progress and realizing the 2030 Agenda for Sustainable Development. Sustainability science, 11, 945-950.
72. Terämä, E., Smallman, M., Lock, S. J., Johnson, C., & Austwick, M. Z. (2016). Beyond academia–Interrogating research impact in the research excellence framework. PLoS one, 11(12), e0168533.
73. Thelwall, M., Kousha, K., Wilson, P., Makita, M., Abdoli, M., Stuart, E., ... & Cancellieri, M. (2023). Predicting article quality scores with machine learning: The UK Research Excellence Framework. Quantitative Science Studies, 4(2), 547-573.
74. Thorpe, A., Craig, R., Tourish, D., Hadikin, G., & Batistic, S. (2018). 'Environment'Submissions in the UK's Research Excellence Framework 2014. British Journal of Management, 29(3), 571-587.
75. Thudumu, Srikanth, et al. "A comprehensive survey of anomaly detection techniques for high dimensional big data." Journal of Big Data 7 (2020): 1-30.
76. Turney, P. D., & Pantel, P. (2010). From frequency to meaning: Vector space models of semantics. Journal of artificial intelligence research, 37, 141-188.
77. Uto, M., Xie, Y., & Ueno, M. (2020, December). Neural automated essay scoring incorporating handcrafted features. In Proceedings of the 28th international conference on computational linguistics (pp. 6077-6088).
78. Vaswani, A. (2017). Attention is all you need. Advances in Neural Information Processing Systems.

79. Watermeyer, R. (2019). *Competitive accountability in academic life: The struggle for social impact and public legitimacy*. Edward Elgar Publishing.
80. *Watermeyer, R. (2019). Mediating the Duality of Universities and Society: Arts and Humanities Confronting the Obstacles of 'Authentic Engagement'. Higher Education in the World 7, 216.*
81. Watermeyer, R., & Chubb, J. (2019). Evaluating 'impact'in the UK's Research Excellence Framework (REF): liminality, looseness and new modalities of scholarly distinction. *Studies in Higher Education*, *44*(9), 1554-1566.
82. Watermeyer, R., Derrick, G. E., & Borras Batalla, M. (2022). Affective auditing: The emotional weight of the research excellence framework. Research Evaluation, 31(4), 498-506.
83. Weinstein, N., Wilsdon, J., Chubb, J., & Haddock, G. (2019). The real-time REF review: A pilot study to examine the feasibility of a longitudinal evaluation of perceptions and attitudes towards REF 2021.
84. Williams, A. (2014). A global index of information and political transparency.
85. Young, T., Hazarika, D., Poria, S., & Cambria, E. (2018). Recent trends in deep learning based natural language processing. IEEE Computational Intelligence Magazine, 13(3), 55-75.
86. Zhao, H., Wu, J., & Li, C. (2019). A comparative study of machine learning algorithms for classification. Journal of Machine Learning Research, 20(1), 1-9.

Appendix

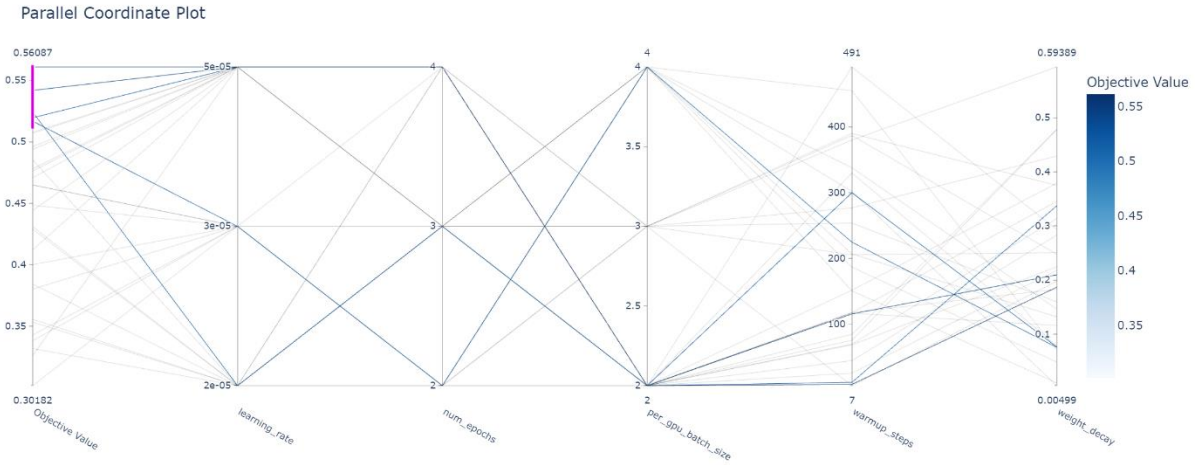
Appendix 1: Word clouds of Naive Bayes with Bag of Words and TF-IDF.

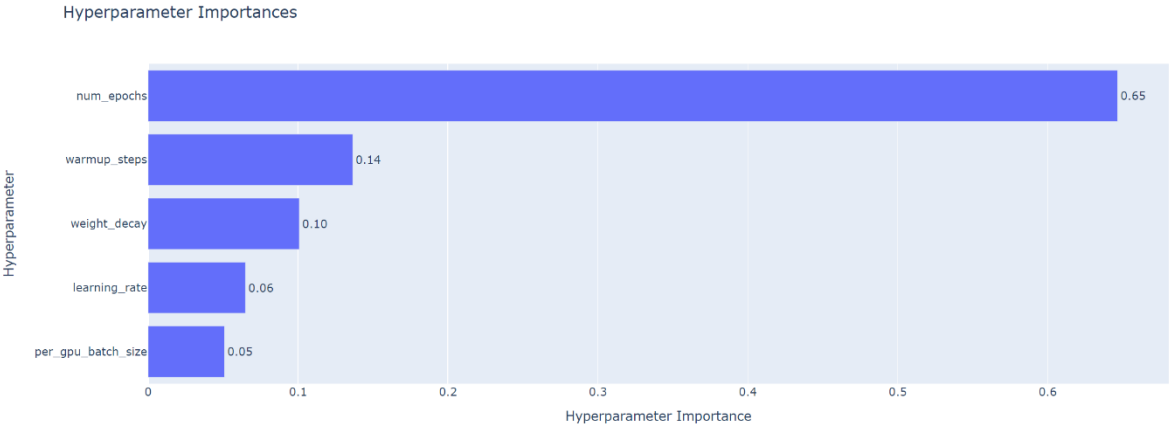
BoW N-gram

TF-IDF

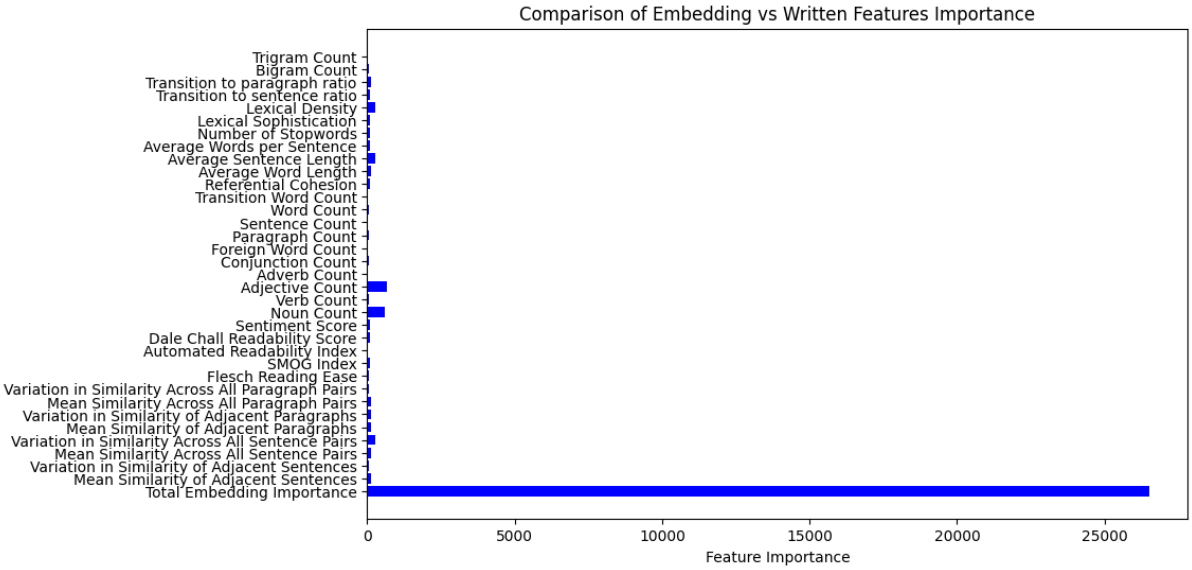
Appendix 2: Visualisation of hyperparameter tuning BERT topic

Appendix 3: Visualisation of hyperparameter tuning for BERT-longformer





Appendix 4: XGBoost with Longformer feature importance

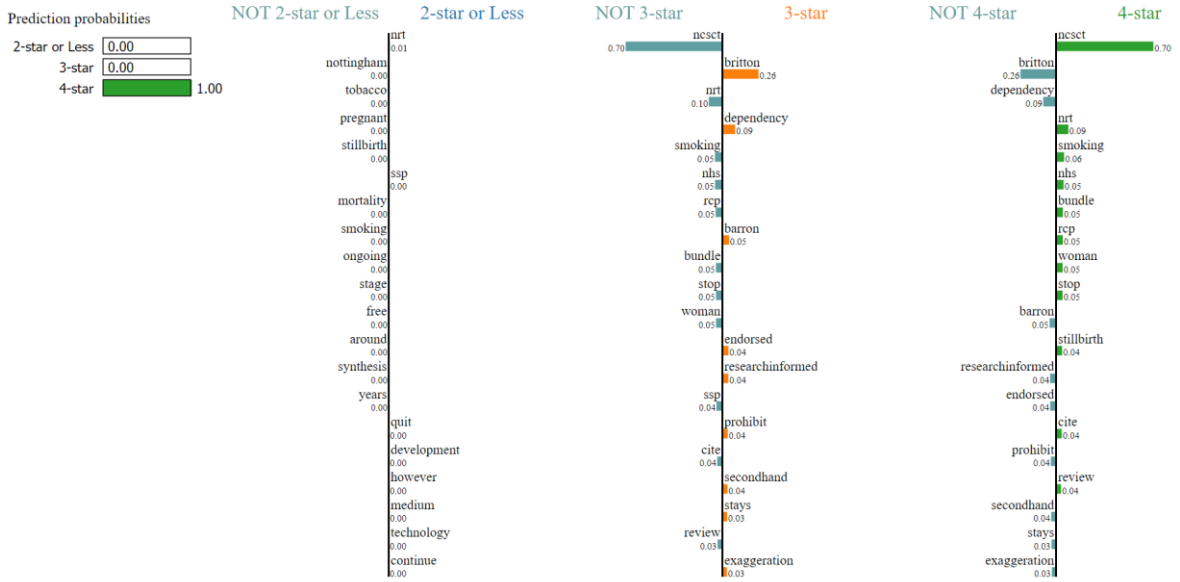


Appendix 5: Bag of Words Naïve Bayes on Impact Case Studies on Index 21, 32, 35, and 136.

Index 21: actual: 4-star predicted: 3-star

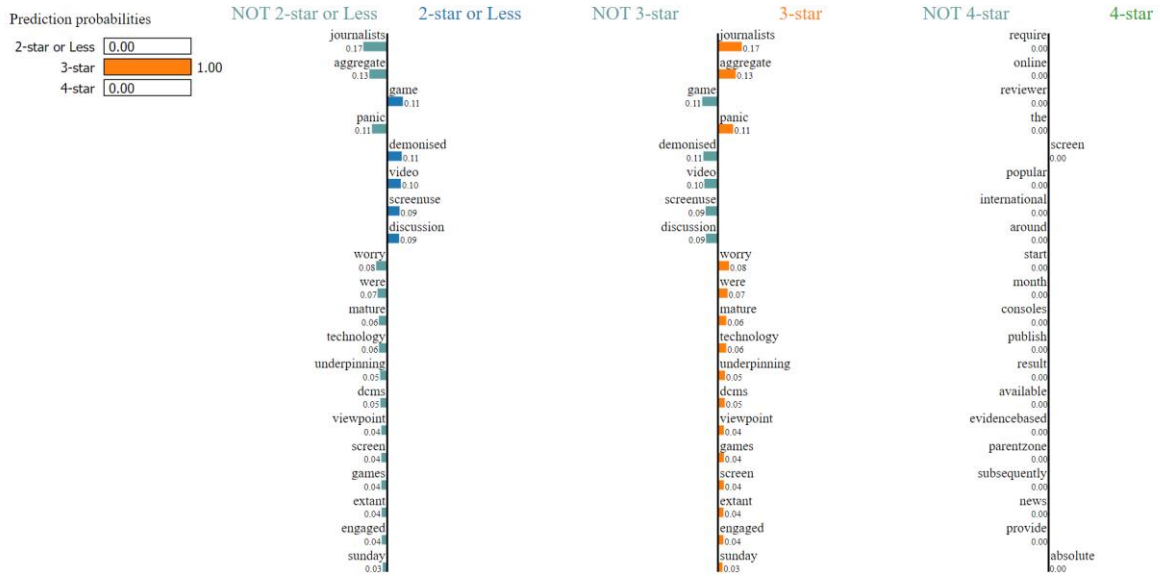
researchbased something different behavioural intervention develop university hertfordshire commercialise spinout company result broad range impacts deliver economic impact commercial benefit within evidence efficacy clinical health intervention find randomise control trial overseas improve peoples health wellbeing productivity internationally within health workplace domestic settings february rilkes room ltd become social enterprise something different ltd follow investment pax holding ltd investment allow company cofounded fletcher pine social entrepreneur ray richards develop dsd platform lowcost scalable accessible behaviour change intervention aim three target beneficiary improve wellbeing mental physical health productivity individuals collaboration charities healthcare organisation public health commissioners businesses dsd ltd generate revenue august december create support six job core staff and one point sustain roles dsd platform underpin award two major horizon grant combined value change use dsd model change lifestyle behavioural risk factor people heart disease hypertension preventomics apply dsd model development digital platform guide consumer healthy food choices grant include total direct funding specifically smes develop new product services one example dutch design firm onmi for fletcher advisor receive onmi offer two behaviour change programme base dsd alius flex create two product via change programme cookit smart spatula help user adopt healthy cook habit public vote award innovation radar prize mysleeve smart device wrap around cup bottle encourage healthy drinking habits efficacy dsd demonstrate rct actrn register feb habitbased intervention carry researcher faculty health science medicine bond university australia dsd find clinically beneficial achieve longterm weight loss improve diet exercise habits approximately participant lose body weight th benchmark successful healthful weight loss keep month intervention ended researcher concluded habitbased intervention potential change think weight management importantly behave change programme result rct nct register oct involve heart disease hypertension patient three hospital spain netherlands taiwan dsdbased intervention show feasible quality life improve intervention group positive costeffectiveness ratio demonstrate spanish hospital dsd ltd create behaviour change programme address various healthrelated issues impact evaluation carry found that user countries report low level anxiety depression number people clinical level anxiety fell clinical level depression participant comment included get motivate take exercise thing differently life rut grandchild love one favourite dos give something somebody unexpected take neighbour never meet flower biscuit end get whole street involve lady meet month basically completely sedentary exercise all exercise result course ive probably add year year life dsd collaborate charity action happiness design intensive sixweek programme help people develop habit proven bring happiness people around them happiness programme sent people regular small positive action dos break exist habit cycles accord evaluation publish action happiness participant country say would recommend dsd others overall average uplift wellbeing comment share zone programme participants dsd collaborate hertfordshire public health run three programme healthy herts focus exercise diet less stress herts explore way combat stress happiness herts develop programme run action happiness three intervention engage residents evaluation show subjective wellbeing people clinical level anxiety depression move healthier levels physical activity weight average six week alcohol consumption fell representative comment participants capture video testimony included definitely help refocus mental energy thinking patterns feel extremely low want anything change positive attitude spite whats cant even send text know look forward text come isnt amazing intervention royal society public health rsph health wellbeing award its innovative creative approach behaviour change individual hertfordshire rsph highlight dsds accessibility scalability underpinned research good practice technology hertfordshires director public health say dsd shows people longterm condition move clinical level depression fraction cost facetoface care hertfordshire public health commission dsd programme healthy herts banner include programme help people type diabetes manage condition improve mental health young people empower victim domestic abuse dsd develop healthy cov coventry city council public health team design engage resident make small change everyday exercise diet make use citys leisure facilities dsd run smoking cessation programme partnership brighton heave clinical commissioning group achieve quit rate result dsd finalist partnership nhs category south east health technology alliance awards medicinfo dutch healthcare innovation company invest threeyear agreement distribute dsd beneficiary netherlands include public health centre educational institution local authority care home people diabetes copd medicinfos director said run eight programmes successful base qualitative feedback participants best result care home complete culture change occur among healthcare professionals dictate organisational structure implement structure base client want need dsd also work well blue collar worker building company like conciseness format large corporate organisation use dsd deliver wellbeing training programme employees focus issue resilience stress inclusivity leadership client impact period include general electric usa aviva american express mondelez remploy tui cisco dsd work ciscos team deliver programme aim build inclusive workplace culture offer employee across ciscos europe middle east africa russia region total employee across country participated publish evaluation find uplift people engage others different cultures one employee commented something different change lot habit introduced real change person work home collaboration tui dsd design bespoke leadership training programme combine tuis management development goal drive improve digital skills make available manager tuis perspective leadership programme one international tax manager said sum programme word would life changing regional finance director commented teach let daytoday become strategic leader work mondelez dsd create agility work programme encourage employee adopt new way working programme make available employee seven site across ireland result say likely look new way things similar programme engaged avivas employee encourage embrace different measures evaluation show likely recommend dsd say thing differently result

Index 32: actual: 4-star predicted: 4-star



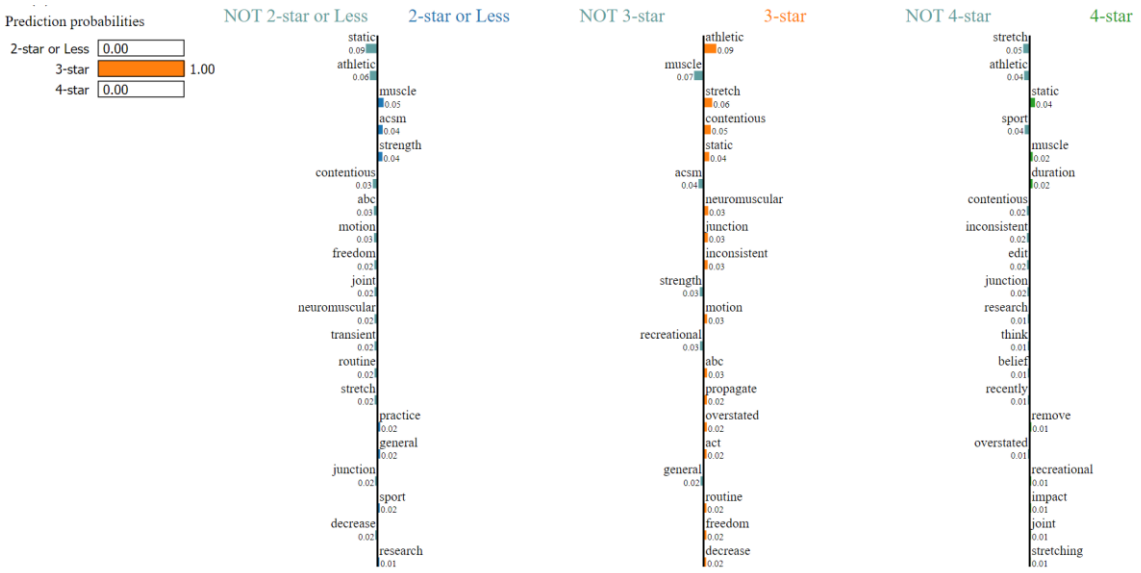
smoking large avoidable cause death disability typically reduce life expectancy ten years work contribute reduction smoke prevalence equivalent around few smoker large usa rich countries already help prevent thousand premature deaths case morbidity hundred stillbirth substantial costs quit smoke improves life expectancy wellbeing protects child unborn child reduces poverty improve productivity work closely evidence translation tobacco advisory group rcp national institute health care excellence britton chair public health guideline development group public health advisory committee member guideline group quality standard committee medicine healthcare product regulatory agency ecigarette safety britton member expert work group policy advocacy action smoking health britton board member since national centre smoke cessation training coproducing researchinformed clinician guidance training dissemination also occur public health england via tobacco control implementation board britton chair since describe detail below acknowledge public health england research range policy area extremely valuable demonstrate policy effectiveness hence ensure adoption maintenance new policies tobacco control programme lead public health england state exaggeration say research university nottingham central role shape tobacco control policy practice change promote public health england impact include tobacco control plan england define policy reduce smoke prevalence england cite two rcp report evidence harm children need help smoker mental health problem quit recommend screen smoke woman attend antenatal care routine use carbon monoxide monitor part new care bundle reduce stillbirth work underpinning evidence recommend automatic provision cessation support woman identify smokers require woman optout cite work justification long term plan set commitment future service cite rcp report demonstrate impact implement tobacco dependency treatment nhs services statement recommend model adopt thus influence future tobacco dependency treatment tobacco control programme lead public health england state work john britton nottingham colleague lead publication series systematic review economic model support approach royal college physician report lead directly adoption approach long term plan britton advise senior nhs england staff design cost service provision use electronic cigarette reduce tobaccorelated harm partial complete substitution smoking controversial meet strong opposition usa australia countries however adopt internationally groundbreaking policy promote electronic cigarette tobacco harm reduction strategy policy substantially underpin research tobacco control programme lead public health england state research evidence synthesis carry nottingham group fundamental adoption endorsement public health england vaping reduced harm substitute smoking rcp report advocate promotion electronic cigarette replace smoking quote verbatim introduction section house common science technology committee report ecigarettes ukctas submission committee lead britton cite research commission public health england rcp also quote verbatim cite extensively throughout report concluded report submission advocated ecigarette use promote harm reduction strategy world leader encourage vaping instead smoking see great annual reduction smoke prevalence usa vaping medically endorsed australia vaping nicotine prohibited report indoor air quality national offender management service subsequently publish peerreviewed article demonstrate high level tobacco smoke exposure prisons lead noms commission independent organisation repeat corroborate finding cite source evidence justify english prison become smoke free since march legislation prohibit smoking car carry child cite department health consultation legislation enact review smoking pregnancy stillbirth risk underpinning evidence introduction optout smoking cessation provision routine within nhs antenatal care core element saving babies life care bundle reduce stillbirth independent care bundle evaluation find implementation across england result decrease stillbirths per year approximately fewer cite review review cite core rationale reduce smoking pregnancy version two care bundle design reduce perinatal mortality world health organisation who ask share finding unpublished cochrane systematic review nrt pregnancy feature prominently recommendation prevention management tobacco use secondhand smoke exposure pregnancy concluding insufficient evidence recommend nrt pregnancy call research update review new trial show nrt effective publish impact work continues updated review justify pregnant womens nrt use european guideline well united kingdom australia new zealand canada updated review finalized influence efficacy data major nrt trial nice propose update guidance nrt pregnancy update ongoing due service design costing proposal cessation services summarise rcp report draw trial similar model develop ottawa cite long term plan work provision cessation service mental health setting march rcp report smoking mental health heavily cite phe implementation guidance commissioner provider mental health services cochrane review nrt pregnancy nrt trial feature prominently two evidence review november nice guidance justify recommendation offer nrt pregnant woman unable stop smoking aim reduce craving cigarette hospital stays impact nhs training detail evidence letter chief executive national centre smoke cessation training consensus support pregnant woman stop smoking produce support smokefree pregnancy project ncsct use formulate first ever nhs standard treatment programme smoke pregnancy update nhs staff training treat patient accord this ncsct use emergent project finding underpin stp change three nhs online training package facetoface training course incorporate key consensus recommendations permit support pregnant smokers health professional must successfully complete ncsct training module course follow stp hence work substantially underpin current nhs clinical practice chief executive ncsct state ssp project finding substantially help ncsct improve quality stop smoking support pregnant woman ncsct guidance standard treatment intervention deliver pregnant woman significantly improve base evidence base develop ssp project ssp project finding continue impact pregnant woman help stop smoking year come training module revise november then nhs professional complete these october complete new versions similarly prior course curriculum change around health professional complete interpersonal training course november october work demonstrate presence substantial tobacco imagery television music video see young people quote kevin barron justification standardised packaging committee stage debate draft standardise packaging tobacco product regulation subsequently enact may study smoke frequency itv love island series receive widespread medium attention present ofcom subsequently depiction smoking remove series britton list health service journal top clinical leader innovation coleman receive year senior investigator award national institute health research

Index 35: actual: 2-star or less predicted: 3-star



etchells work video game screen time lead number opportunity promote explain discuss underpinning research variety public contexts example public lectures write popular media publish popular science book comment television radio feed directly discussion shift narrative general understanding effect digital technology use have away moral panic solely detrimental effect towards nuanced understanding best balance risk benefit technology use change viewpoint journalists parent professional bodies basis underpinning research etchells write publish popular science book call lost good game play game mean disseminate complexity extant research literature respond cycle moral panic video game screen time often persist mainstream news media date approximately copy book sell englishspeaking countries including limited usa canada australia new zealand south africa korean translation follow soon book receive widespread positive recommendation press note convincing debut interested effect play video game find much ponder cheer heartfelt defence demonised pastime others argue blisteringly relevant enriching touching issue challenge bad science surround subject coverage highlight book provide mature way think video game useful relevant concerned member public example one outlet argue etchells great job reassure that long were fun isnt need worry international press present similar evaluation highlight shift think book affords example etchells try capture current conversation around video games also quite aware cater people still old conversation etchells skilled enough writer speak both publication lost good game lead discussion around effect screen time video game number key stakeholders include journalists parent professional bodies result successfully communicate uncertainty around effects well highlight research suggest minimal risks stakeholder express change opinion beliefs digital technology example april interview sunday time magazine interviewing journalist comment ban child play game consoles pioneering psychologist change mind journalist start position someone allow child play video games etchells explanation underpinning research book subsequent interviews make journalist reevaluate view parent technology use even purchase game console children january etchells convene speak public engagement event fund british academy hold wellcome collection london bring together scientists journalists clinicians representative professional body civil servant communicate cuttingedge research screen time video games editor psychologist magazine write report explain etchells event involvement discussion underpinning research allow see screen time differently develop positive perspective video game screen use key stakeholder parentzone organisation provide support information parents child school online safety video game effects express positive change view video gaming light discussion around etchells book underpinning research conclusion dissemination underpinning research result shift public narrative concern video games away polarise narrative absolute risk benefits towards nuanced accurate discussion subtler effects etchells science blog network coordinator the guardian peak monthly readership million unique visitor per month establish take lead role write newspapers psychology blog head quarters write regularly research general literature video game screen time effects result promote underpinning research international news platform become wellknown expert area since opportunity provide evidence parliamentary enquiry concern screen use immersive andor addictive technology health wellbeing evidence submit basis underpinning research write science communication effort feature house common science technology committee report impact social medium screenuse young peoples health etchells directly name quote regard complex nature conduct screen time research addition report name open letter the guardian organise etchells sign international group scientists argue nuanced approach screen time research etchells interviewee reviewer parliamentary office science technology research brief screen use health young people publish etchells also collaborate colleague submit researchbased evidence house common digital culture medium sport committee report immersive addictive technologies line evidence etchells colleague recommend video game company require make aggregate data available researchers contribute financially independent research key highlight report subsequently video game company start become engaged researcher area etchells recently join industryacademia work group liaise closely independent game developer association aim allow relevant researcher access aggregate player data broadly research make strong contribution parliamentary recommendations process drive change concern video game monetise context gambling act via online discussion roundtable events etchells continue close discussion dcms ensure change evidencebased appropriate effective

Index 136: actual: 2-star or less predicted: 3-star

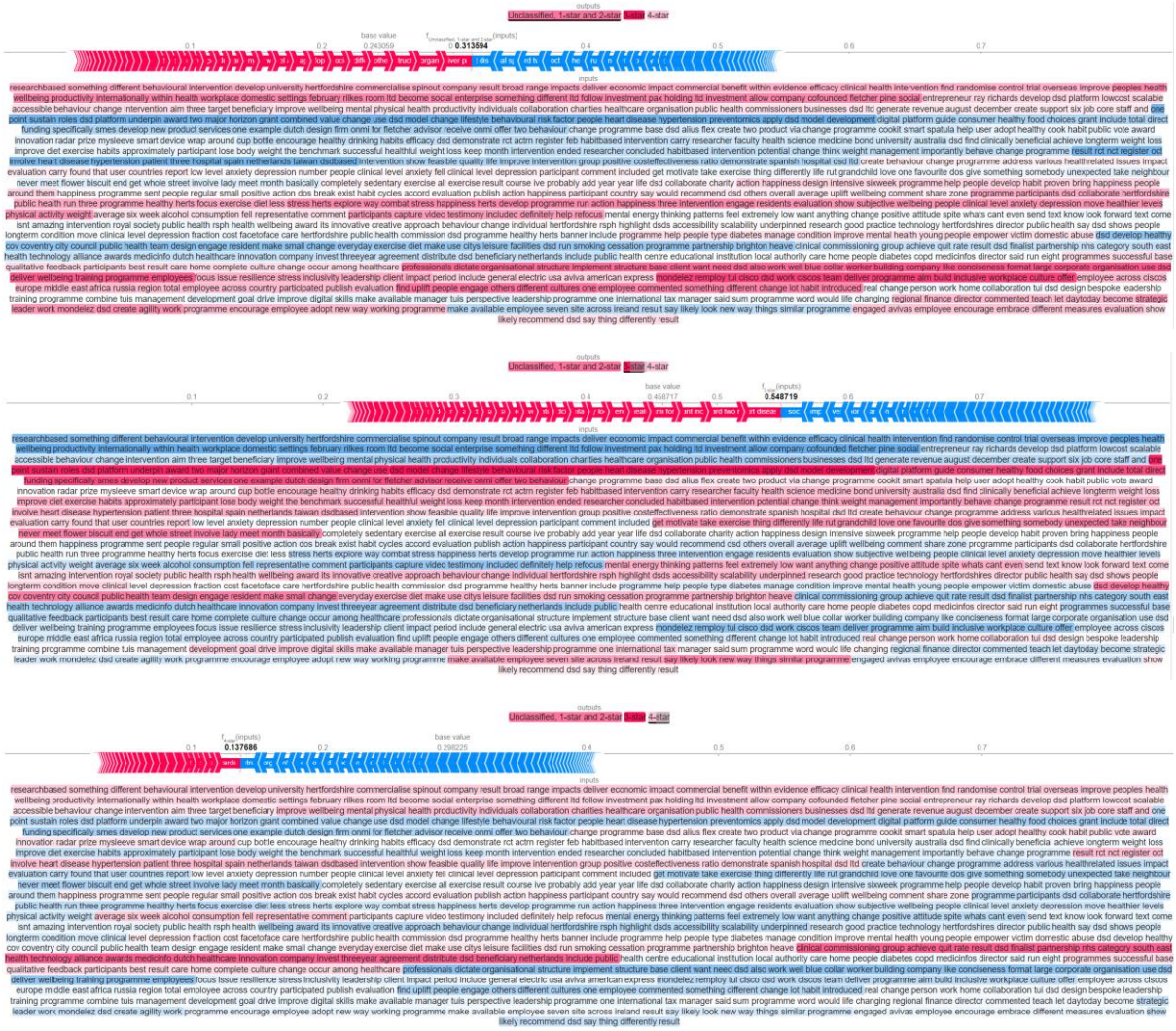


kays research help challenge notion static stretching duration use athletic environment detrimental effect remove warmup practices clarify doseresponse effect stretch maximum safe duration stretch warmup research also put rest one contentious issue confirm static muscle stretch exercise reduce risk muscle strain injury common injury sport research finding disseminate among key stakeholder lead change guidance practice effect stretch finding also propagate international medium increase public awareness benefit research dominant medium narrative recently focus study report potentially detrimental effect static stretch exercise include negative effect performance injury preventative benefit four kays study position stand challenge narrative widely cite international media change medium coverage stretch practice exercise increase public understanding benefit stretch different type stretching kays systematic review acute effect several type muscle stretch technique physical performance flexibility injury prevention cite news story international news outlet include new york time guardian kays research pnf stretching highlight shape magazine bring attention benefit type stretching something widely know general public subsequent research clarify negative effect several type muscle stretching use within full dynamic warmup athletic performance also widely cite international news source include new york time abc life comprehensive list medium coverage see kay clarify misconception report medium negative effect muscle stretch muscle strength increase public awareness way improve muscle strength study conduct bbc trust i am doctor programme base research view figure study look possibility increase muscle strength simply think exercise something call motor imagery study directly impact participants improve muscle strength average with one participant increase strength introduce viewer new scientific research stretch muscle strength accessible way kay produce two think stronger audioguides complement programme focus low upper body strength available download free increase publics access techniques kays finding limited detrimental effect static stretching athletic general population help clarify common misconception report general public wider media prior kays research govern bodies american college sport medicine recommend removal static stretching warmup routine due belief would cause performance impairment strength speed powerrelated activities injury preventative benefit kays research instrumental identify limitation within research systematic metaanalytical review clarify actual effect stretch within warmup research correct previous misconceptions provide evidencebased guidance health body clarify govern body athletes mislead information previously inconsistent inform new position stand public health guidance previously contentious subjects englands national health service nhs approach kay help develop write edit nhs choice health advice section stretch flexibility guidance quote kay likely duration stretch use warmup routine recreational exerciser produce negligible transient

reduction strength cite kays research report the reduction performance preexercise stretching overstated first guidance release stretch flexibility kays research key development recommendations accord freedom information act request page view time since november canadian society exercise physiology csep canadas lead authority exercise science focus promote research evidencebased fitness performance health outcome canadians approve directly cite kays research first position stand topic coauthor kay csep chair states the recommendation csep position stand component warmup include appropriate duration stretching inclusion static proprioceptive neuromuscular facilitation pnf stretching recommend potential positively influence standard warmup routine large number athletes position stand use kays extensive research enable change position stand csep inform national academy sport medicine nasm provide athletic guidance stretch acsm american college sport medicine national academy sport medicine nasm united states leading authority personal training certification worldwide directly cite kays research regard best practice stretching kays research use inform narrative include static stretch second use preexercise protocol without significant decrease strength power speeddependent task performances static stretch second consider effective method increase joint range motion rom often think improve performance reduce incidence activityrelated injuries nasm conclude static stretching beneficial many ways correct muscle imbalances decrease muscle hypertonicity increase joint rom relieve joint stress improve extensibility musculotendinous junction maintain normal functional length muscle lengthtension relationships decrease chance injury turn enhance power strength recommendation use kays research provide evidencebased exercise prescription guide use stretch within warmup

Appendix 3: Uncontextualized Embeddings XGBoost analysis on impact case studies with index 21, 32, 35, and 136.

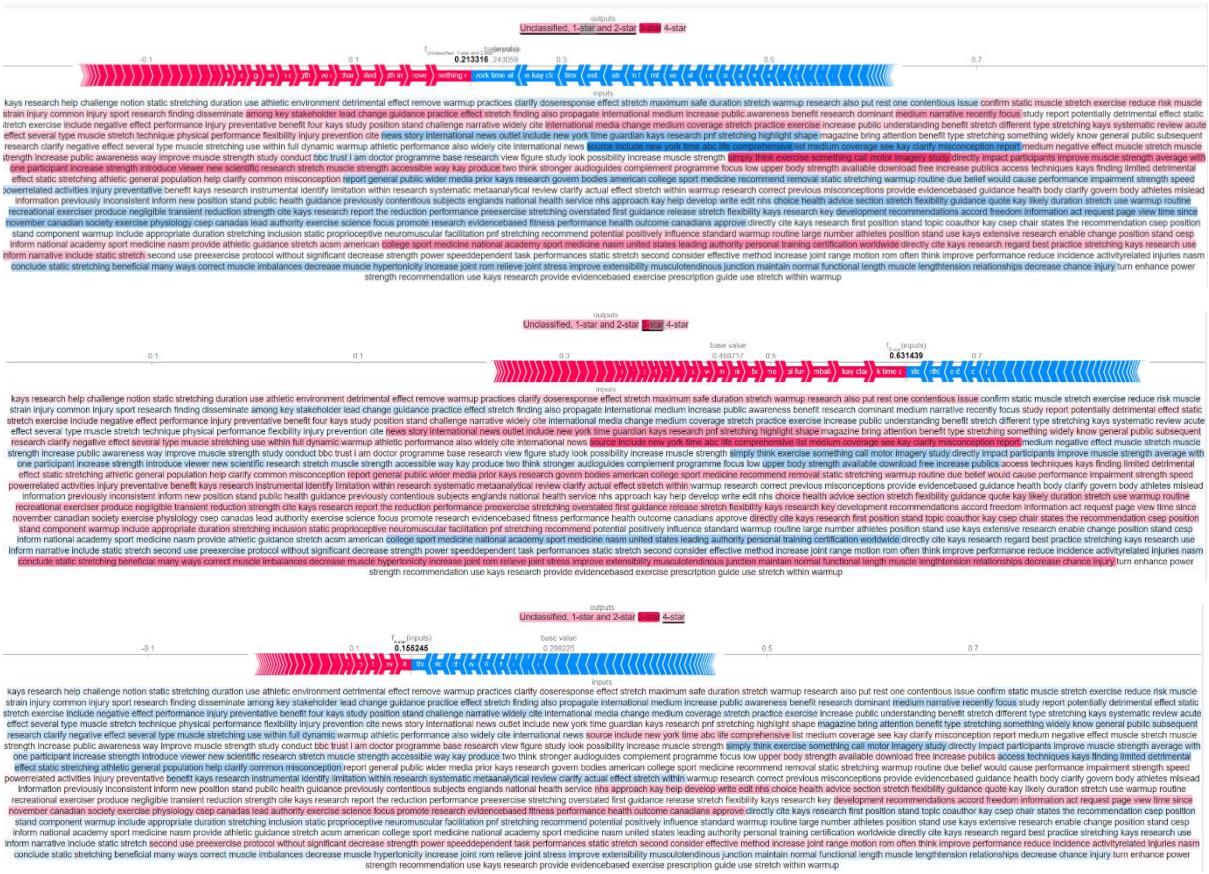
Index 21: actual: 4-star predicted: 4-star



Index 32: actual: 4-star predicted: 4-star

Index 35: actual: 2-star or less predicted: 2-star or less

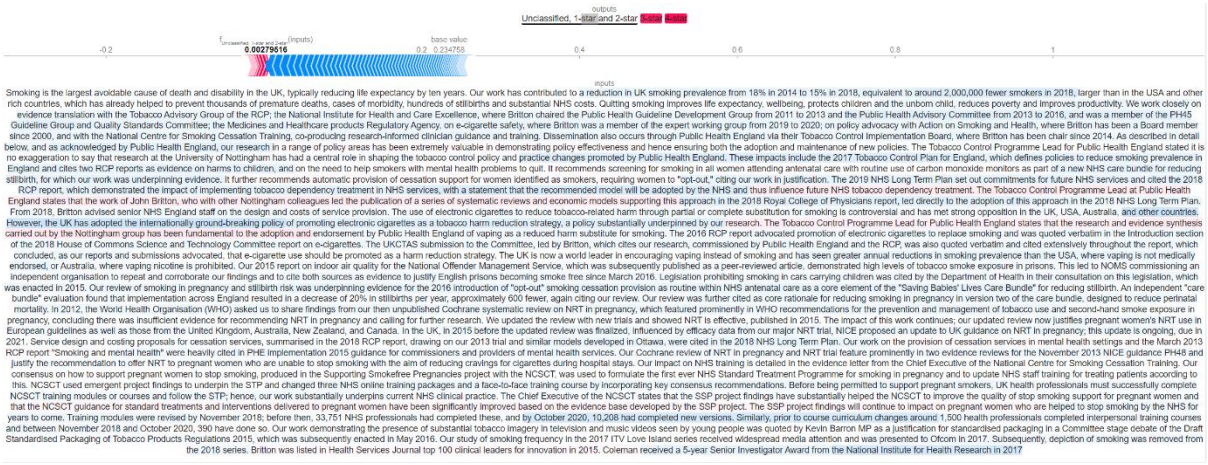
Index 136: actual: 2-star or less predicted: 3-star



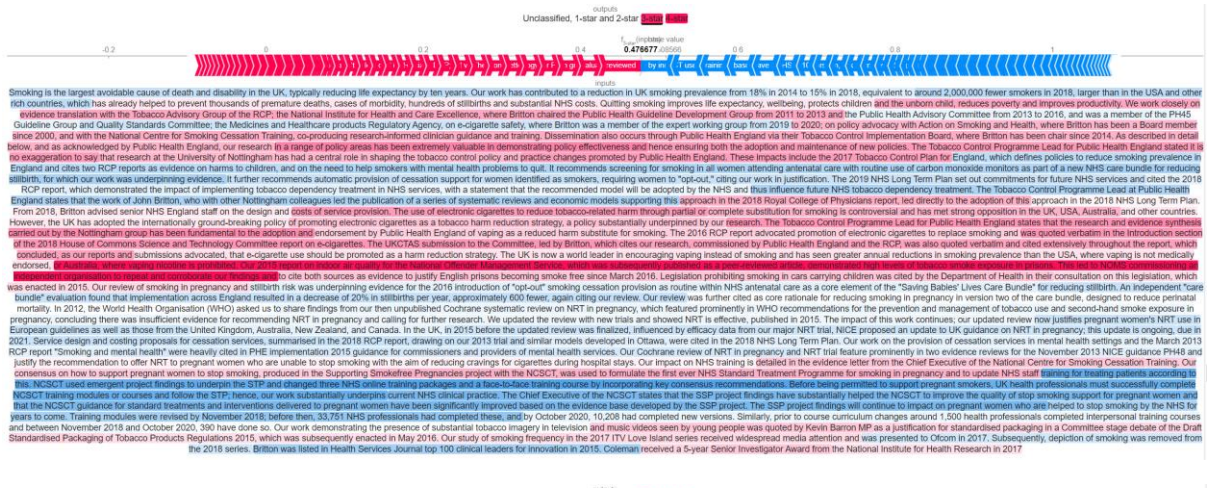
Appendix 4: Longformer XGBoost analysis on impact case studies with Index 21, 31, 35, and 136

Index 21: actual: 4-star predicted: 3-star

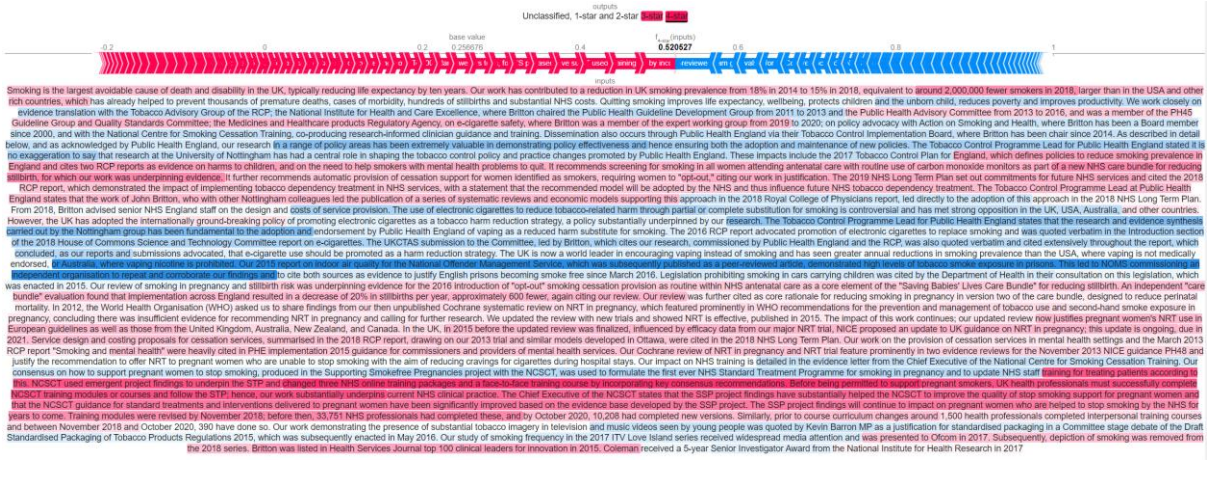
Index 31: actual: 4-star predicted: 4-star



Smoking is the largest avoidable cause of death and disability in the UK, typically reducing life expectancy by ten years. Our work has contributed to a reduction in UK smoking prevalence from 18% in 2014 to 15% in 2018, equivalent to around 2,000,000 fewer smokers in 2018, larger than in the USA and other rich countries, which has already helped to prevent thousands of premature deaths, cases of morbidity, hundreds of stillbirths and substantial NHS costs. Quitting smoking improves life expectancy, wellbeing, protects children and the unborn child, reduces poverty and improves productivity. We work closely on evidence translation with the Tobacco Advisory Group of the RCP, the National Institute for Health and Care Excellence, where Britton chaired the Public Health Guideline Development Group from 2011 to 2013 and the Public Health Advisory Committee from 2013 to 2016, and was a member of the PH45 Guideline Group and Quality Standards Committee; the Medicines and Healthcare products Regulatory Agency, on e-cigarette safety, where Britton was a member of the expert working group from 2019 to 2020; on policy advocacy with Action on Smoking and Health, where Britton has been a Board member since 2000, and with the National Centre for Smoking Cessation Training, co-producing research-informed clinician guidance and training. Dissemination also occurs through Public Health England via their Tobacco Control Implementation Board, where Britton has been chair since 2014. As described in detail below, and as acknowledged by Public Health England, our research in a range of policy areas has been extremely valuable in demonstrating policy effectiveness and hence ensuring both the adoption and maintenance of new policies. The Tobacco Control Programme Lead for Public Health England stated it is no exaggeration to say that research at the University of Nottingham has had a central role in shaping the tobacco control policy and practice changes promoted by Public Health England. These impacts include the 2017 Tobacco Control Plan for England, which defines policies to reduce smoking prevalence in England and cites two RCP reports as evidence on harms to children, and on the need to help smokers with mental health problems to quit. It recommends screening for smoking in all women attending antenatal care with routine use of carbon monoxide monitors as part of a new NHS care bundle for reducing stillbirth, for which our work was underpinning evidence. It further recommends automatic provision of cessation support for women identified as smokers, requiring women to "opt-out," citing our work in justification. The 2019 NHS Long Term Plan set out commitments for future NHS services and cited the 2018 RCP report, which demonstrated the impact of implementing tobacco dependency treatment in NHS services, with a statement that the recommended model will be adopted by the NHS and thus influence future NHS tobacco dependency treatment. The Tobacco Control Programme Lead at Public Health England states that the work of John Britton, who with other Nottingham colleagues led the publication of a series of systematic reviews and economic models supporting this approach in the 2018 Royal College of Physicians report, led directly to the adoption of this approach in the 2018 NHS Long Term Plan. From 2018, Britton advised senior NHS England staff on the design and costs of service provision. The use of electronic cigarettes to reduce tobacco-related harm through partial or complete substitution for smoking is controversial and has met strong opposition in the UK, USA, Australia, and other countries. However, the UK has adopted the internationally ground-breaking policy of promoting electronic cigarettes as a tobacco harm reduction strategy, a policy substantially underpinned by our research. The Tobacco Control Programme Lead for Public Health England states that the research and evidence synthesis carried out by the Nottingham group has been fundamental to the adoption and endorsement by Public Health England of vaping as a reduced harm substitute for smoking. The 2016 RCP report advocated promotion of electronic cigarettes to replace smoking and was quoted verbatim in the Introduction section of the 2018 House of Commons Science and Technology Committee report on e-cigarettes. The UKCTAS submission to the Committee, led by Britton, which cites our research, commissioned by Public Health England and the RCP, was also quoted verbatim and cited extensively throughout the report, which concluded, as our reports and submissions advocated, that e-cigarette use should be promoted as a harm reduction strategy. The UK is now a world leader in encouraging vaping instead of smoking and has seen greater annual reductions in smoking prevalence than the USA, where vaping is not medically endorsed, or Australia, where vaping nicotine is prohibited. Our 2015 report on indoor air quality for the National Offender Management Service, which was subsequently published as a peer-reviewed article, demonstrated high levels of tobacco smoke exposure in prisons. This led to NOMS commissioning an independent organisation to repeat and corroborate our findings and to cite both sources as evidence to justify English prisons becoming smoke free since March 2016. Legislation prohibiting smoking in cars carrying children was cited by the Department of Health in their consultation on this legislation, which was enacted in 2015. Our review of smoking in pregnancy and stillbirth risk was underpinning evidence for the 2016 introduction of "opt-out" smoking cessation provision as routine within NHS antenatal care as a core element of the "Saving Babies' Lives Care Bundle" for reducing stillbirth. An independent "care bundle" evaluation found that implementation across England resulted in a decrease of 20% in stillbirths per year, approximately 600 fewer, again citing our review. Our review was further cited as core rationale for reducing smoking in pregnancy in version two of the care bundle, designed to reduce perinatal mortality. In 2012, the World Health Organisation (WHO) asked us to share findings from our then unpublished Cochrane systematic review on NRT in pregnancy, which featured prominently in WHO recommendations for the prevention and management of tobacco use and second-hand smoke exposure in pregnancy, concluding there was insufficient evidence for recommending NRT in pregnancy and calling for further research. We updated the review with new trials and showed NRT is effective, published in 2015. The impact of this work continues; our updated review now justifies pregnant women's NRT use in European guidelines as well as those from the United Kingdom, Australia, New Zealand, and Canada. In the UK, in 2015 before the updated review was finalized, influenced by efficacy data from our major NRT trial, NICE proposed an update to UK guidance on NRT in pregnancy; this update is ongoing, due in 2021. Service design and costing proposals for cessation services, summarised in the 2018 RCP report, drawing on our 2013 trial and similar models developed in Ottawa, were cited in the 2018 NHS Long Term Plan. Our work on the provision of cessation services in mental health settings and the March 2013 RCP report "Smoking and mental health" were heavily cited in PHE implementation 2015 guidance for commissioners and providers of mental health services. Our Cochrane review of NRT in pregnancy and NRT trial feature prominently in two evidence reviews for the November 2013 NICE guidance PH48 and justify the recommendation to offer NRT to pregnant women who are unable to stop smoking with the aim of reducing cravings for cigarettes during hospital stays. Our impact on NHS training is detailed in the evidence letter from the Chief Executive of the National Centre for Smoking Cessation Training. Our consensus on how to support pregnant women to stop smoking, produced in the Supporting Smokefree Pregnancies project with the NCSCT, was used to formulate the first ever NHS Standard Treatment Programme for smoking in pregnancy and to update NHS staff training for treating patients according to this. NCSCT used emergent project findings to underpin the STP and changed three NHS online training packages and a face-to-face training course by incorporating key consensus recommendations. Before being permitted to support pregnant smokers, UK health professionals must successfully complete NCSCT training modules or courses and follow the STP; hence, our work substantially underpins current NHS clinical practice. The Chief Executive of the NCSCT states that the SSP project findings have substantially helped the NCSCT to improve the quality of stop smoking support for pregnant women and that the NCSCT guidance for standard treatments and interventions delivered to pregnant women have been significantly improved based on the evidence base developed by the SSP project. The SSP project findings will continue to impact on pregnant women who are helped to stop smoking by the NHS for years to come. Training modules were revised by November 2018; before then, 33,751 NHS professionals had completed these, and by October 2020, 10,208 had completed new versions. Similarly, prior to course curriculum changes around 1,500 health professionals completed interpersonal training courses and between November 2018 and October 2020, 390 have done so. Our work demonstrating the presence of substantial tobacco imagery in television and music videos seen by young people was quoted by Kevin Barron MP as a justification for standardised packaging in a Committee stage debate of the Draft Standardised Packaging of Tobacco Products Regulations 2015, which was subsequently enacted in May 2016. Our study of smoking frequency in the 2017 ITV Love Island series received widespread media attention and was presented to Ofcom in 2017. Subsequently, depiction of smoking was removed from the 2018 series. Britton was listed in Health Services Journal top 100 clinical leaders for innovation in 2015. Coleman received a 5-year Senior Investigator Award from the National Institute for Health Research in 2017



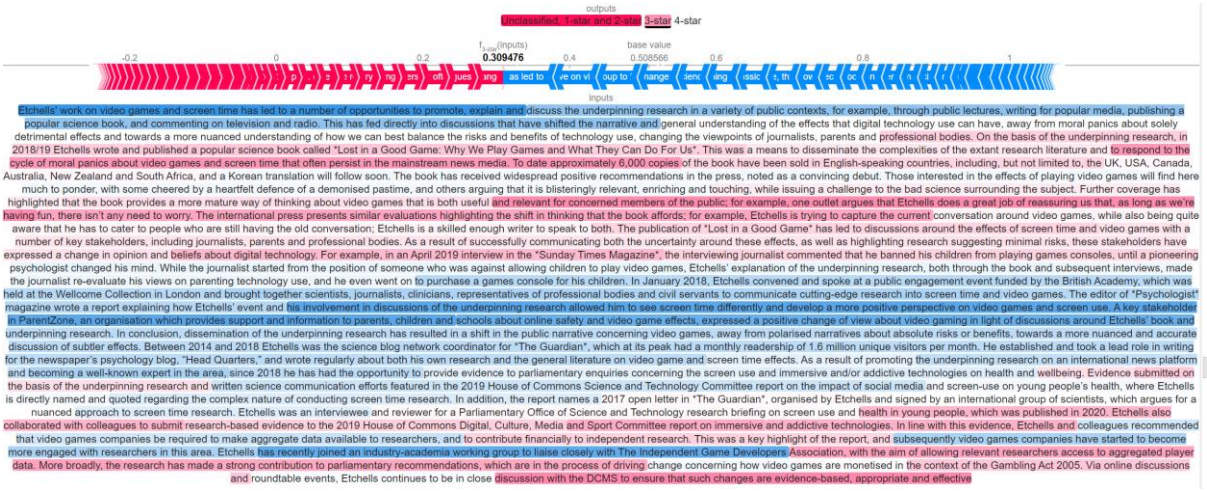
Smoking is the largest avoidable cause of death and disability in the UK, typically reducing life expectancy by ten years. Our work has contributed to a reduction in UK smoking prevalence from 18% in 2014 to 15% in 2018, equivalent to around 2,000,000 fewer smokers in 2018, larger than in the USA and other rich countries, which has already helped to prevent thousands of premature deaths, cases of morbidity, hundreds of stillbirths and substantial NHS costs. Quitting smoking improves life expectancy, wellbeing, protects children and the unborn child, reduces poverty and improves productivity. We work closely on evidence translation with the Tobacco Advisory Group of the RCP, the National Institute for Health and Care Excellence, where Britton chaired the Public Health Guideline Development Group from 2011 to 2013 and the Public Health Advisory Committee from 2013 to 2016, and was a member of the PH45 Guideline Group and Quality Standards Committee; the Medicines and Healthcare products Regulatory Agency, on e-cigarette safety, where Britton was a member of the expert working group from 2019 to 2020; on policy advocacy with Action on Smoking and Health, where Britton has been a Board member since 2000, and with the National Centre for Smoking Cessation Training, co-producing research-informed clinician guidance and training. Dissemination also occurs through Public Health England via their Tobacco Control Implementation Board, where Britton has been chair since 2014. As described in detail below, and as acknowledged by Public Health England, our research in a range of policy areas has been extremely valuable in demonstrating policy effectiveness and hence ensuring both the adoption and maintenance of new policies. The Tobacco Control Programme Lead for Public Health England stated it is no exaggeration to say that research at the University of Nottingham has had a central role in shaping the tobacco control policy and practice changes promoted by Public Health England. These impacts include the 2017 Tobacco Control Plan for England, which defines policies to reduce smoking prevalence in England and cites two RCP reports as evidence on harms to children, and on the need to help smokers with mental health problems to quit. It recommends screening for smoking in all women attending antenatal care with routine use of carbon monoxide monitors as part of a new NHS care bundle for reducing stillbirth, for which our work was underpinning evidence. It further recommends automatic provision of cessation support for women identified as smokers, requiring women to "opt-out," citing our work in justification. The 2019 NHS Long Term Plan set out commitments for future NHS services and cited the 2018 RCP report, which demonstrated the impact of implementing tobacco dependency treatment in NHS services, with a statement that the recommended model will be adopted by the NHS and thus influence future NHS tobacco dependency treatment. The Tobacco Control Programme Lead at Public Health England states that the work of John Britton, who with other Nottingham colleagues led the publication of a series of systematic reviews and economic models supporting this approach in the 2018 Royal College of Physicians report, led directly to the adoption of this approach in the 2018 NHS Long Term Plan. From 2018, Britton advised senior NHS England staff on the design and costs of service provision. The use of electronic cigarettes to reduce tobacco-related harm through partial or complete substitution for smoking is controversial and has met strong opposition in the UK, USA, Australia, and other countries. However, the UK has adopted the internationally ground-breaking policy of promoting electronic cigarettes as a tobacco harm reduction strategy, a policy substantially underpinned by our research. The Tobacco Control Programme Lead for Public Health England states that the research and evidence synthesis carried out by the Nottingham group has been fundamental to the adoption and endorsement by Public Health England of vaping as a reduced harm substitute for smoking. The 2016 RCP report advocated promotion of electronic cigarettes to replace smoking and was quoted verbatim in the Introduction section of the 2018 House of Commons Science and Technology Committee report on e-cigarettes. The UKCTAS submission to the Committee, led by Britton, which cites our research, commissioned by Public Health England and the RCP, was also quoted verbatim and cited extensively throughout the report, which concluded, as our reports and submissions advocated, that e-cigarette use should be promoted as a harm reduction strategy. The UK is now a world leader in encouraging vaping instead of smoking and has seen greater annual reductions in smoking prevalence than the USA, where vaping is not medically endorsed, or Australia, where vaping nicotine is prohibited. Our 2015 report on indoor air quality for the National Offender Management Service, which was subsequently published as a peer-reviewed article, demonstrated high levels of tobacco smoke exposure in prisons. This led to NOMS commissioning an independent organisation to repeat and corroborate our findings and to cite both sources as evidence to justify English prisons becoming smoke free since March 2016. Legislation prohibiting smoking in cars carrying children was cited by the Department of Health in their consultation on this legislation, which was enacted in 2015. Our review of smoking in pregnancy and stillbirth risk was underpinning evidence for the 2016 introduction of "opt-out" smoking cessation provision as routine within NHS antenatal care as a core element of the "Saving Babies' Lives Care Bundle" for reducing stillbirth. An independent "care bundle" evaluation found that implementation across England resulted in a decrease of 20% in stillbirths per year, approximately 600 fewer, again citing our review. Our review was further cited as core rationale for reducing smoking in pregnancy in version two of the care bundle, designed to reduce perinatal mortality. In 2012, the World Health Organisation (WHO) asked us to share findings from our then unpublished Cochrane systematic review on NRT in pregnancy, which featured prominently in WHO recommendations for the prevention and management of tobacco use and second-hand smoke exposure in pregnancy, concluding there was insufficient evidence for recommending NRT in pregnancy and calling for further research. We updated the review with new trials and showed NRT is effective, published in 2015. The impact of this work continues; our updated review now justifies pregnant women's NRT use in European guidelines as well as those from the United Kingdom, Australia, New Zealand, and Canada. In the UK, in 2015 before the updated review was finalized, influenced by efficacy data from our major NRT trial, NICE proposed an update to UK guidance on NRT in pregnancy; this update is ongoing, due in 2021. Service design and costing proposals for cessation services, summarised in the 2018 RCP report, drawing on our 2013 trial and similar models developed in Ottawa, were cited in the 2018 NHS Long Term Plan. Our work on the provision of cessation services in mental health settings and the March 2013 RCP report "Smoking and mental health" were heavily cited in PHE implementation 2015 guidance for commissioners and providers of mental health services. Our Cochrane review of NRT in pregnancy and NRT trial feature prominently in two evidence reviews for the November 2013 NICE guidance PH48 and justify the recommendation to offer NRT to pregnant women who are unable to stop smoking with the aim of reducing cravings for cigarettes during hospital stays. Our impact on NHS training is detailed in the evidence letter from the Chief Executive of the National Centre for Smoking Cessation Training. Our consensus on how to support pregnant women to stop smoking, produced in the Supporting Smokefree Pregnancies project with the NCSCT, was used to formulate the first ever NHS Standard Treatment Programme for smoking in pregnancy and to update NHS staff training for treating patients according to this. NCSCT used emergent project findings to underpin the STP and changed three NHS online training packages and a face-to-face training course by incorporating key consensus recommendations. Before being permitted to support pregnant smokers, UK health professionals must successfully complete NCSCT training modules or courses and follow the STP; hence, our work substantially underpins current NHS clinical practice. The Chief Executive of the NCSCT states that the SSP project findings have substantially helped the NCSCT to improve the quality of stop smoking support for pregnant women and that the NCSCT guidance for standard treatments and interventions delivered to pregnant women have been significantly improved based on the evidence base developed by the SSP project. The SSP project findings will continue to impact on pregnant women who are helped to stop smoking by the NHS for years to come. Training modules were revised by November 2018; before then, 33,751 NHS professionals had completed these, and by October 2020, 10,208 had completed new versions. Similarly, prior to course curriculum changes around 1,500 health professionals completed interpersonal training courses and between November 2018 and October 2020, 390 have done so. Our work demonstrating the presence of substantial tobacco imagery in television and music videos seen by young people was quoted by Kevin Barron MP as a justification for standardised packaging in a Committee stage debate of the Draft Standardised Packaging of Tobacco Products Regulations 2015, which was subsequently enacted in May 2016. Our study of smoking frequency in the 2017 ITV Love Island series received widespread media attention and was presented to Ofcom in 2017. Subsequently, depiction of smoking was removed from the 2018 series. Britton was listed in Health Services Journal top 100 clinical leaders for innovation in 2015. Coleman received a 5-year Senior Investigator Award from the National Institute for Health Research in 2017



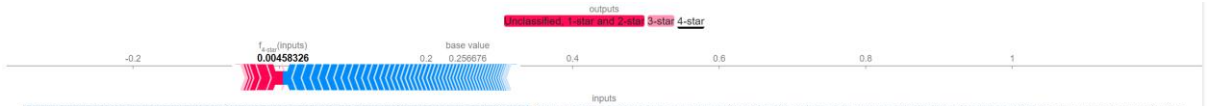
Smoking is the largest avoidable cause of death and disability in the UK, typically reducing life expectancy by ten years. Our work has contributed to a reduction in UK smoking prevalence from 18% in 2014 to 15% in 2018, equivalent to around 2,000,000 fewer smokers in 2018, larger than in the USA and other rich countries, which has already helped to prevent thousands of premature deaths, cases of morbidity, hundreds of stillbirths and substantial NHS costs. Quitting smoking improves life expectancy, wellbeing, protects children and the unborn child, reduces poverty and improves productivity. We work closely on evidence translation with the Tobacco Advisory Group of the RCP, the National Institute for Health and Care Excellence, where Britton chaired the Public Health Guideline Development Group from 2011 to 2013 and the Public Health Advisory Committee from 2013 to 2016, and was a member of the PH45 Guideline Group and Quality Standards Committee; the Medicines and Healthcare products Regulatory Agency, on e-cigarette safety, where Britton was a member of the expert working group from 2019 to 2020; on policy advocacy with Action on Smoking and Health, where Britton has been a Board member since 2000, and with the National Centre for Smoking Cessation Training, co-producing research-informed clinician guidance and training. Dissemination also occurs through Public Health England via their Tobacco Control Implementation Board, where Britton has been chair since 2014. As described in detail below, and as acknowledged by Public Health England, our research in a range of policy areas has been extremely valuable in demonstrating policy effectiveness and hence ensuring both the adoption and maintenance of new policies. The Tobacco Control Programme Lead for Public Health England stated it is no exaggeration to say that research at the University of Nottingham has had a central role in shaping the tobacco control policy and practice changes promoted by Public Health England. These impacts include the 2017 Tobacco Control Plan for England, which defines policies to reduce smoking prevalence in England and cites two RCP reports as evidence on harms to children, and on the need to help smokers with mental health problems to quit. It recommends screening for smoking in all women attending antenatal care with routine use of carbon monoxide monitors as part of a new NHS care bundle for reducing stillbirth, for which our work was underpinning evidence. It further recommends automatic provision of cessation support for women identified as smokers, requiring women to "opt-out," citing our work in justification. The 2019 NHS Long Term Plan set out commitments for future NHS services and cited the 2018 RCP report, which demonstrated the impact of implementing tobacco dependency treatment in NHS services, with a statement that the recommended model will be adopted by the NHS and thus influence future NHS tobacco dependency treatment. The Tobacco Control Programme Lead at Public Health England states that the work of John Britton, who with other Nottingham colleagues led the publication of a series of systematic reviews and economic models supporting this approach in the 2018 Royal College of Physicians report, led directly to the adoption of this approach in the 2018 NHS Long Term Plan. From 2018, Britton advised senior NHS England staff on the design and costs of service provision. The use of electronic cigarettes to reduce tobacco-related harm through partial or complete substitution for smoking is controversial and has met strong opposition in the UK, USA, Australia, and other countries. However, the UK has adopted the internationally ground-breaking policy of promoting electronic cigarettes as a tobacco harm reduction strategy, a policy substantially underpinned by our research. The Tobacco Control Programme Lead for Public Health England states that the research and evidence synthesis carried out by the Nottingham group has been fundamental to the adoption and endorsement by Public Health England of vaping as a reduced harm substitute for smoking. The 2016 RCP report advocated promotion of electronic cigarettes to replace smoking and was quoted verbatim in the Introduction section of the 2018 House of Commons Science and Technology Committee report on e-cigarettes. The UKCTAS submission to the Committee, led by Britton, which cites our research, commissioned by Public Health England and the RCP, was also quoted verbatim and cited extensively throughout the report, which concluded, as our reports and submissions advocated, that e-cigarette use should be promoted as a harm reduction strategy. The UK is now a world leader in encouraging vaping instead of smoking and has seen greater annual reductions in smoking prevalence than the USA, where vaping is not medically endorsed, or Australia, where vaping nicotine is prohibited. Our 2015 report on indoor air quality for the National Offender Management Service, which was subsequently published as a peer-reviewed article, demonstrated high levels of tobacco smoke exposure in prisons. This led to NOMS commissioning an independent organisation to repeat and corroborate our findings and to cite both sources as evidence to justify English prisons becoming smoke free since March 2016. Legislation prohibiting smoking in cars carrying children was cited by the Department of Health in their consultation on this legislation, which was enacted in 2015. Our review of smoking in pregnancy and stillbirth risk was underpinning evidence for the 2016 introduction of "opt-out" smoking cessation provision as routine within NHS antenatal care as a core element of the "Saving Babies' Lives Care Bundle" for reducing stillbirth. An independent "care bundle" evaluation found that implementation across England resulted in a decrease of 20% in stillbirths per year, approximately 600 fewer, again citing our review. Our review was further cited as core rationale for reducing smoking in pregnancy in version two of the care bundle, designed to reduce perinatal mortality. In 2012, the World Health Organisation (WHO) asked us to share findings from our then unpublished Cochrane systematic review on NRT in pregnancy, which featured prominently in WHO recommendations for the prevention and management of tobacco use and second-hand smoke exposure in pregnancy, concluding there was insufficient evidence for recommending NRT in pregnancy and calling for further research. We updated the review with new trials and showed NRT is effective, published in 2015. The impact of this work continues; our updated review now justifies pregnant women's NRT use in European guidelines as well as those from the United Kingdom, Australia, New Zealand, and Canada. In the UK, in 2015 before the updated review was finalized, influenced by efficacy data from our major NRT trial, NICE proposed an update to UK guidance on NRT in pregnancy; this update is ongoing, due in 2021. Service design and costing proposals for cessation services, summarised in the 2018 RCP report, drawing on our 2013 trial and similar models developed in Ottawa, were cited in the 2018 NHS Long Term Plan. Our work on the provision of cessation services in mental health settings and the March 2013 RCP report "Smoking and mental health" were heavily cited in PHE implementation 2015 guidance for commissioners and providers of mental health services. Our Cochrane review of NRT in pregnancy and NRT trial feature prominently in two evidence reviews for the November 2013 NICE guidance PH48 and justify the recommendation to offer NRT to pregnant women who are unable to stop smoking with the aim of reducing cravings for cigarettes during hospital stays. Our impact on NHS training is detailed in the evidence letter from the Chief Executive of the National Centre for Smoking Cessation Training. Our consensus on how to support pregnant women to stop smoking, produced in the Supporting Smokefree Pregnancies project with the NCSCT, was used to formulate the first ever NHS Standard Treatment Programme for smoking in pregnancy and to update NHS staff training for treating patients according to this. NCSCT used emergent project findings to underpin the STP and changed three NHS online training packages and a face-to-face training course by incorporating key consensus recommendations. Before being permitted to support pregnant smokers, UK health professionals must successfully complete NCSCT training modules or courses and follow the STP; hence, our work substantially underpins current NHS clinical practice. The Chief Executive of the NCSCT states that the SSP project findings have substantially helped the NCSCT to improve the quality of stop smoking support for pregnant women and that the NCSCT guidance for standard treatments and interventions delivered to pregnant women have been significantly improved based on the evidence base developed by the SSP project. The SSP project findings will continue to impact on pregnant women who are helped to stop smoking by the NHS for years to come. Training modules were revised by November 2018; before then, 33,751 NHS professionals had completed these, and by October 2020, 10,208 had completed new versions. Similarly, prior to course curriculum changes around 1,500 health professionals completed interpersonal training courses and between November 2018 and October 2020, 390 have done so. Our work demonstrating the presence of substantial tobacco imagery in television and music videos seen by young people was quoted by Kevin Barron MP as a justification for standardised packaging in a Committee stage debate of the Draft Standardised Packaging of Tobacco Products Regulations 2015, which was subsequently enacted in May 2016. Our study of smoking frequency in the 2017 ITV Love Island series received widespread media attention and was presented to Ofcom in 2017. Subsequently, depiction of smoking was removed from the 2018 series. Britton was listed in Health Services Journal top 100 clinical leaders for innovation in 2015. Coleman received a 5-year Senior Investigator Award from the National Institute for Health Research in 2017

Index 35: actual: 2-star or less predicted:

Etchells' work on video games and screen time has led to a number of opportunities to promote, explain and discuss the underpinning research in a variety of public contexts, for example, through public lectures, writing for popular media, publishing a popular science book, and commenting on television and radio. This has fed directly into discussions that have shifted the narrative and general understanding of the effects that digital technology use can have, away from moral panics about solely detrimental effects and towards a more nuanced understanding of how we can best balance the risks and benefits of technology use, changing the viewpoints of journalists, parents and professional bodies. On the basis of the underpinning research, in 2018/19 Etchells wrote and published a popular science book called "Lost in a Good Game: Why We Play Games and What They Can Do For Us". This was a means to disseminate the complexities of the extant research literature and to respond to the cycle of moral panics about video games and screen time that often persist in the mainstream news media. To date approximately 6,000 copies of the book have been sold in English-speaking countries, including, but not limited to, the UK, USA, Canada, Australia, New Zealand and South Africa, and a Korean translation will follow soon. The book has received widespread positive recommendations in the press, noted as a convincing debut. Those interested in the effects of playing video games will find here much to ponder, with some cheered by a heartfelt defence of a demonised pastime, and others arguing that it is blisteringly relevant, enriching and touching, while issuing a challenge to the bad science surrounding the subject. Further coverage has highlighted that the book provides a more mature way of thinking about video games that is both useful and relevant for concerned members of the public; for example, one outlet argues that Etchells does a great job of reassuring us that, as long as we're having fun, there isn't any need to worry. The international press presents similar evaluations highlighting the shift in thinking that the book affords; for example, Etchells is trying to capture the current conversation around video games, while also being quite aware that he has to cater to people who are still having the old conversation; Etchells is a skilled enough writer to speak to both. The publication of "Lost in a Good Game" has led to discussions around the effects of screen time and video games with a number of key stakeholders, including journalists, parents and professional bodies. As a result of successfully communicating both the uncertainty around these effects, as well as highlighting research suggesting minimal risks, these stakeholders have expressed a change in opinion and beliefs about digital technology. For example, in an April 2019 interview in the "Sunday Times Magazine", the interviewing journalist commented that he banned his children from playing games consoles, until a pioneering psychologist changed his mind. While the journalist started from the position of someone who was against allowing children to play video games, Etchells' explanation of the underpinning research, both through the book and subsequent interviews, made the journalist re-evaluate his views on parenting technology use, and he even went on to purchase a games console for his children. In January 2018, Etchells convened and spoke at a public engagement event funded by the British Academy, which was held at the Wellcome Collection in London and brought together scientists, journalists, clinicians, representatives of professional bodies and civil servants to communicate cutting-edge research into screen time and video games. The editor of "Psychologist" magazine wrote a report explaining how Etchells' event and his involvement in discussions of the underpinning research allowed him to see screen time differently and develop a more positive perspective on video games and screen use. A key stakeholder in ParentZone, an organisation which provides support and information to parents, children and schools about online safety and video game effects, expressed a positive change of view about video gaming in light of discussions around Etchells' book and underpinning research. In conclusion, dissemination of the underpinning research has resulted in a shift in the public narrative concerning video games, away from polarised narratives about absolute risks or benefits, towards a more nuanced and accurate discussion of subtler effects. Between 2014 and 2018 Etchells was the science blog network coordinator for "The Guardian", which at its peak had a monthly readership of 1.6 million unique visitors per month. He established and took a lead role in writing for the newspaper's psychology blog, "Head Quarters," and wrote regularly about both his own research and the general literature on video game and screen time effects. As a result of promoting the underpinning research on an international news platform and becoming a well-known expert in the area, since 2018 he has had the opportunity to provide evidence to parliamentary enquiries concerning the screen use and immersive and/or addictive technologies on health and wellbeing. Evidence submitted on the basis of the underpinning research and written science communication efforts featured in the 2019 House of Commons Science and Technology Committee report on the impact of social media and screen-use on young people's health, where Etchells is directly named and quoted regarding the complex nature of conducting screen time research. In addition, the report names a 2017 open letter in "The Guardian", organised by Etchells and signed by an international group of scientists, which argues for a nuanced approach to screen time research. Etchells was an interviewee and reviewer for a Parliamentary Office of Science and Technology research briefing on screen use and health in young people, which was published in 2020. Etchells also collaborated with colleagues to submit research-based evidence to the 2019 House of Commons Digital, Culture, Media and Sport Committee report on immersive and addictive technologies. In line with this evidence, Etchells and colleagues recommended that video games companies be required to make aggregate data available to researchers, and to contribute financially to independent research. This was a key highlight of the report, and subsequently video games companies have started to become more engaged with researchers in this area. Etchells has recently joined an industry-academia working group to liaise closely with The Independent Game Developers Association, with the aim of allowing relevant researchers access to aggregated player data. More broadly, the research has made a strong contribution to parliamentary recommendations, which are in the process of driving change concerning how video games are monetised in the context of the Gambling Act 2005. Via online discussions and roundtable events, Etchells continues to be in close discussion with the DCMS to ensure that such changes are evidence-based, appropriate and effective



Etchells' work on video games and screen time has led to a number of opportunities to promote, explain and discuss the underpinning research in a variety of public contexts, for example, through public lectures, writing for popular media, publishing a popular science book, and commenting on television and radio. This has fed directly into discussions that have shifted the narrative and general understanding of the effects that digital technology use can have, away from moral panics about solely detrimental effects and towards a more nuanced understanding of how we can best balance the risks and benefits of technology use, changing the viewpoints of journalists, parents and professional bodies. On the basis of the underpinning research, in 2018/19 Etchells wrote and published a popular science book called "Lost in a Good Game: Why We Play Games and What They Can Do For Us". This was a means to disseminate the complexities of the extant research literature and to respond to the cycle of moral panics about video games and screen time that often persist in the mainstream news media. To date approximately 6,000 copies of the book have been sold in English-speaking countries, including, but not limited to, the UK, USA, Canada, Australia, New Zealand and South Africa, and a Korean translation will follow soon. The book has received widespread positive recommendations in the press, noted as a convincing debut. Those interested in the effects of playing video games will find here much to ponder, with some cheered by a heartfelt defence of a demonised pastime, and others arguing that it is blisteringly relevant, enriching and touching, while issuing a challenge to the bad science surrounding the subject. Further coverage has highlighted that the book provides a more mature way of thinking about video games that is both useful and relevant for concerned members of the public; for example, one outlet argues that Etchells does a great job of reassuring us that, as long as we're having fun, there isn't any need to worry. The international press presents similar evaluations highlighting the shift in thinking that the book affords; for example, Etchells is trying to capture the current conversation around video games, while also being quite aware that he has to cater to people who are still having the old conversation; Etchells is a skilled enough writer to speak to both. The publication of "Lost in a Good Game" has led to discussions around the effects of screen time and video games with a number of key stakeholders, including journalists, parents and professional bodies. As a result of successfully communicating both the uncertainty around these effects, as well as highlighting research suggesting minimal risks, these stakeholders have expressed a change in opinion and beliefs about digital technology. For example, in an April 2019 interview in the "Sunday Times Magazine", the interviewing journalist commented that he banned his children from playing games consoles, until a pioneering psychologist changed his mind. While the journalist started from the position of someone who was against allowing children to play video games, Etchells' explanation of the underpinning research, both through the book and subsequent interviews, made the journalist re-evaluate his views on parenting technology use, and he even went on to purchase a games console for his children. In January 2018, Etchells convened and spoke at a public engagement event funded by the British Academy, which was held at the Wellcome Collection in London and brought together scientists, journalists, clinicians, representatives of professional bodies and civil servants to communicate cutting-edge research into screen time and video games. The editor of "Psychologist" magazine wrote a report explaining how Etchells' event and his involvement in discussions of the underpinning research allowed him to see screen time differently and develop a more positive perspective on video games and screen use. A key stakeholder in ParentZone, an organisation which provides support and information to parents, children and schools about online safety and video game effects, expressed a positive change of view about video gaming in light of discussions around Etchells' book and underpinning research. In conclusion, dissemination of the underpinning research has resulted in a shift in the public narrative concerning video games, away from polarised narratives about absolute risks or benefits, towards a more nuanced and accurate discussion of subtler effects. Between 2014 and 2018 Etchells was the science blog network coordinator for "The Guardian", which at its peak had a monthly readership of 1.6 million unique visitors per month. He established and took a lead role in writing for the newspaper's psychology blog, "Head Quarters," and wrote regularly about both his own research and the general literature on video game and screen time effects. As a result of promoting the underpinning research on an international news platform and becoming a well-known expert in the area, since 2018 he has had the opportunity to provide evidence to parliamentary enquiries concerning the screen use and immersive and/or addictive technologies on health and wellbeing. Evidence submitted on the basis of the underpinning research and written science communication efforts featured in the 2019 House of Commons Science and Technology Committee report on the impact of social media and screen-use on young people's health, where Etchells is directly named and quoted regarding the complex nature of conducting screen time research. In addition, the report names a 2017 open letter in "The Guardian", organised by Etchells and signed by an international group of scientists, which argues for a nuanced approach to screen time research. Etchells was an interviewee and reviewer for a Parliamentary Office of Science and Technology research briefing on screen use and health in young people, which was published in 2020. Etchells also collaborated with colleagues to submit research-based evidence to the 2019 House of Commons Digital, Culture, Media and Sport Committee report on immersive and addictive technologies. In line with this evidence, Etchells and colleagues recommended that video games companies be required to make aggregate data available to researchers, and to contribute financially to independent research. This was a key highlight of the report, and subsequently video games companies have started to become more engaged with researchers in this area. Etchells has recently joined an industry-academia working group to liaise closely with The Independent Game Developers Association, with the aim of allowing relevant researchers access to aggregated player data. More broadly, the research has made a strong contribution to parliamentary recommendations, which are in the process of driving change concerning how video games are monetised in the context of the Gambling Act 2005. Via online discussions and roundtable events, Etchells continues to be in close discussion with the DCMS to ensure that such changes are evidence-based, appropriate and effective

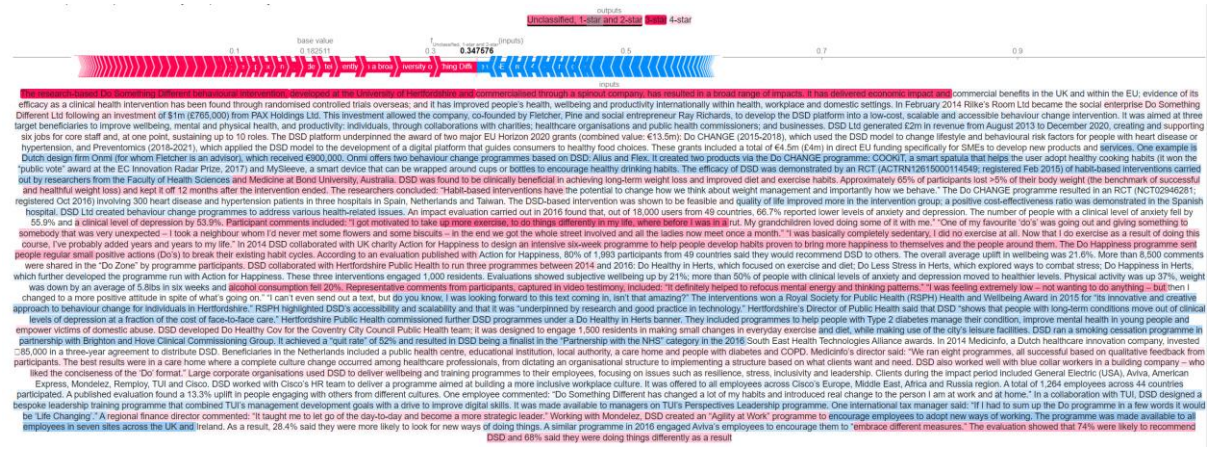


Etchells' work on video games and screen time has led to a number of opportunities to promote, explain and discuss the underpinning research in a variety of public contexts, for example, through public lectures, writing for popular media, publishing a popular science book, and commenting on television and radio. This has fed directly into discussions that have shifted the narrative and general understanding of the effects that digital technology use can have, away from moral panics about solely detrimental effects and towards a more nuanced understanding of how we can best balance the risks and benefits of technology use, changing the viewpoints of journalists, parents and professional bodies. On the basis of the underpinning research, in 2018/19 Etchells wrote and published a popular science book called "Lost in a Good Game: Why We Play Games and What They Can Do For Us". This was a means to disseminate the complexities of the extant research literature and to respond to the cycle of moral panics about video games and screen time that often persist in the mainstream news media. To date approximately 6,000 copies of the book have been sold in English-speaking countries, including, but not limited to, the UK, USA, Canada, Australia, New Zealand and South Africa, and a Korean translation will follow soon. The book has received widespread positive recommendations in the press, noted as a convincing debut. Those interested in the effects of playing video games will find here much to ponder, with some cheered by a heartfelt defence of a demonised pastime, and others arguing that it is blisteringly relevant, enriching and touching, while issuing a challenge to the bad science surrounding the subject. Further coverage has highlighted that the book provides a more mature way of thinking about video games that is both useful and relevant for concerned members of the public; for example, one outlet argues that Etchells does a great job of reassuring us that, as long as we're having fun, there isn't any need to worry. The international press presents similar evaluations highlighting the shift in thinking that the book affords; for example, Etchells is trying to capture the current conversation around video games, while also being quite aware that he has to cater to people who are still having the old conversation; Etchells is a skilled enough writer to speak to both. The publication of "Lost in a Good Game" has led to discussions around the effects of screen time and video games with a number of key stakeholders, including journalists, parents and professional bodies. As a result of successfully communicating both the uncertainty around these effects, as well as highlighting research suggesting minimal risks, these stakeholders have expressed a change in opinion and beliefs about digital technology. For example, in an April 2019 interview in the "Sunday Times Magazine", the interviewing journalist commented that he banned his children from playing games consoles, until a pioneering psychologist changed his mind. While the journalist started from the position of someone who was against allowing children to play video games, Etchells' explanation of the underpinning research, both through the book and subsequent interviews, made the journalist re-evaluate his views on parenting technology use, and he even went on to purchase a games console for his children. In January 2018, Etchells convened and spoke at a public engagement event funded by the British Academy, which was held at the Wellcome Collection in London and brought together scientists, journalists, clinicians, representatives of professional bodies and civil servants to communicate cutting-edge research into screen time and video games. The editor of "Psychologist" magazine wrote a report explaining how Etchells' event and his involvement in discussions of the underpinning research allowed him to see screen time differently and develop a more positive perspective on video games and screen use. A key stakeholder in ParentZone, an organisation which provides support and information to parents, children and schools about online safety and video game effects, expressed a positive change of view about video gaming in light of discussions around Etchells' book and underpinning research. In conclusion, dissemination of the underpinning research has resulted in a shift in the public narrative concerning video games, away from polarised narratives about absolute risks or benefits, towards a more nuanced and accurate discussion of subtler effects. Between 2014 and 2018 Etchells was the science blog network coordinator for "The Guardian", which at its peak had a monthly readership of 1.6 million unique visitors per month. He established and took a lead role in writing for the newspaper's psychology blog, "Head Quarters," and wrote regularly about both his own research and the general literature on video game and screen time effects. As a result of promoting the underpinning research on an international news platform and becoming a well-known expert in the area, since 2018 he has had the opportunity to provide evidence to parliamentary enquiries concerning the screen use and immersive and/or addictive technologies on health and wellbeing. Evidence submitted on the basis of the underpinning research and written science communication efforts featured in the 2019 House of Commons Science and Technology Committee report on the impact of social media and screen-use on young people's health, where Etchells is directly named and quoted regarding the complex nature of conducting screen time research. In addition, the report names a 2017 open letter in "The Guardian", organised by Etchells and signed by an international group of scientists, which argues for a nuanced approach to screen time research. Etchells was an interviewee and reviewer for a Parliamentary Office of Science and Technology research briefing on screen use and health in young people, which was published in 2020. Etchells also collaborated with colleagues to submit research-based evidence to the 2019 House of Commons Digital, Culture, Media and Sport Committee report on immersive and addictive technologies. In line with this evidence, Etchells and colleagues recommended that video games companies be required to make aggregate data available to researchers, and to contribute financially to independent research. This was a key highlight of the report, and subsequently video games companies have started to become more engaged with researchers in this area. Etchells has recently joined an industry-academia working group to liaise closely with The Independent Game Developers Association, with the aim of allowing relevant researchers access to aggregated player data. More broadly, the research has made a strong contribution to parliamentary recommendations, which are in the process of driving change concerning how video games are monetised in the context of the Gambling Act 2005. Via online discussions and roundtable events, Etchells continues to be in close discussion with the DCMS to ensure that such changes are evidence-based, appropriate and effective

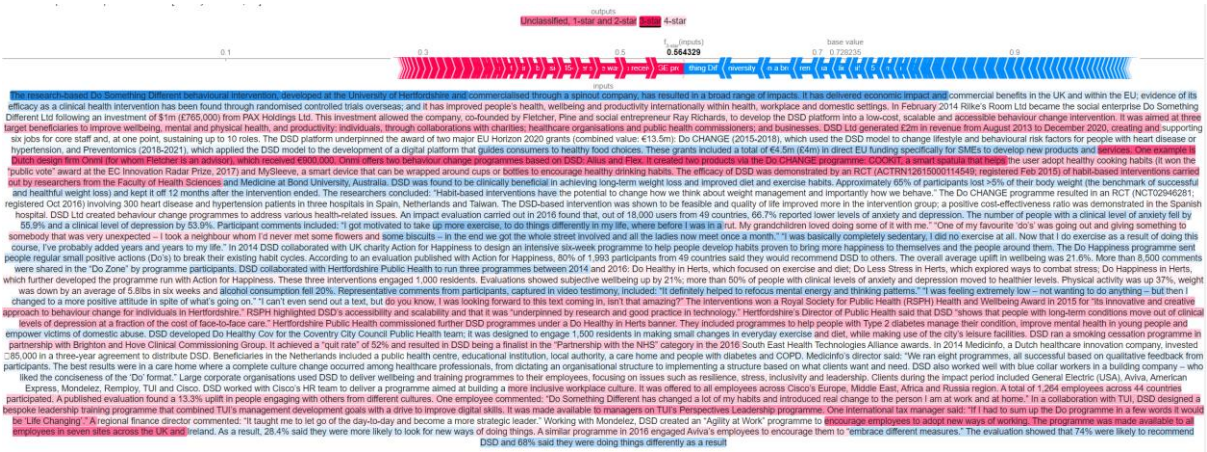
Index 136: actual: 2-star or less predicted:

Appendix 5: XGBoost with OpenAI Embeddings of impact case studies in Index 21, 31,and 32.

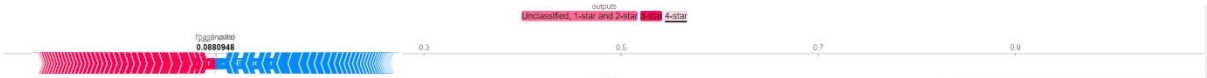
Index 21: actual: 4-star predicted: 3-star



The research-based Do Something Different behavioural intervention, developed at the University of Hertfordshire and commercialised through a spinout company, has resulted in a broad range of impacts. It has delivered economic impact and commercial benefits in the UK and within the EU; evidence of its efficacy as a clinical health intervention has been found through randomised controlled trials overseas; and it has improved people's health, wellbeing and productivity internationally within health, workplace and domestic settings. In February 2014 Rilke's Room Ltd became the social enterprise Do Something Different Ltd following an investment of $1m (£765,000) from PAX Holdings Ltd. This investment allowed the company, co-founded by Fletcher, Pine and social entrepreneur Ray Richards, to develop the DSD platform into a low-cost, scalable and accessible behaviour change intervention. It was aimed at three target beneficiaries to improve wellbeing, mental and physical health, and productivity: individuals, through collaborations with charities; healthcare organisations and public health commissioners; and businesses. DSD Ltd generated £2m in revenue from August 2013 to December 2020, creating and supporting six jobs for core staff and, at one point, sustaining up to 10 roles. The DSD platform underpinned the award of two major EU Horizon 2020 grants (combined value: €13.5m): Do CHANGE (2015-2018), which used the DSD model to change lifestyle and behavioural risk factors for people with heart disease or hypertension, and Preventomics (2018-2021), which applied the DSD model to the development of a digital platform that guides consumers to healthy food choices. These grants included a total of €4.5m (£4m) in direct EU funding specifically for SMEs to develop new products and services. One example is Dutch design firm Onmi (for whom Fletcher is an advisor), which received €900,000. Onmi offers two behaviour change programmes based on DSD: Alius and Flex. It created two products via the Do CHANGE programme: COOKIT, a smart spatula that helps the user adopt healthy cooking habits (it won the "public vote" award at the EC Innovation Radar Prize, 2017) and MySleeve, a smart device that can be wrapped around cups or bottles to encourage healthy drinking habits. The efficacy of DSD was demonstrated by an RCT (ACTRN12615000114549; registered Feb 2015) of habit-based interventions carried out by researchers from the Faculty of Health Sciences and Medicine at Bond University, Australia. DSD was found to be clinically beneficial in achieving long-term weight loss and improved diet and exercise habits. Approximately 65% of participants lost >5% of their body weight (the benchmark of successful and healthful weight loss) and kept it off 12 months after the intervention ended. The researchers concluded: "Habit-based interventions have the potential to change how we think about weight management and importantly how we behave." The Do CHANGE programme resulted in an RCT (NCT02946281; registered Oct 2016) involving 300 heart disease and hypertension patients in three hospitals in Spain, Netherlands and Taiwan. The DSD-based intervention was shown to be feasible and quality of life improved more in the intervention group; a positive cost-effectiveness ratio was demonstrated in the Spanish hospital. DSD Ltd created behaviour change programmes to address various health-related issues. An impact evaluation carried out in 2016 found that, out of 18,000 users from 49 countries, 66.7% reported lower levels of anxiety and depression. The number of people with a clinical level of anxiety fell by 55.9% and a clinical level of depression by 53.9%. Participant comments included: "I got motivated to take up more exercise, to do things differently in my life, where before I was in a rut. My grandchildren loved doing some of it with me," "One of my favourite 'do's' was going out and giving something to somebody that was very unexpected – I took a neighbour whom I'd never met some flowers and some biscuits – in the end we got the whole street involved and all the ladies now meet once a month," "I was basically completely sedentary, I did no exercise at all. Now that I do exercise as a result of doing this course, I've probably added years and years to my life." In 2014 DSD collaborated with UK charity Action for Happiness to design an intensive six-week programme to help people develop habits proven to bring more happiness to themselves and the people around them. The Do Happiness programme sent people regular small positive actions (Do's) to break their existing habit cycles. According to an evaluation published with Action for Happiness, 80% of 1,993 participants from 49 countries said they would recommend DSD to others. The overall average uplift in wellbeing was 21.6%. More than 8,500 comments were shared in the "Do Zone" by programme participants. DSD collaborated with Hertfordshire Public Health to run three programmes between 2014 and 2016: Do Healthy in Herts, which focused on exercise and diet; Do Less Stress in Herts, which explored ways to combat stress; Do Happiness in Herts, which further developed the programme run with Action for Happiness. These three interventions engaged 1,000 residents. Evaluations showed subjective wellbeing up by 21%; more than 50% of people with clinical levels of anxiety and depression moved to healthier levels. Physical activity was up 37%, weight was down by an average of 5.8lbs in six weeks and alcohol consumption fell 20%. Representative comments from participants, captured in video testimony, included: "It definitely helped to refocus mental energy and thinking patterns," "I was feeling extremely low – not wanting to do anything – but then I changed to a more positive attitude in spite of what's going on," "I can't even send out a text, but do you know, I was looking forward to this text coming in, isn't that amazing?" The interventions won a Royal Society for Public Health (RSPH) Health and Wellbeing Award in 2015 for "its innovative and creative approach to behaviour change for individuals in Hertfordshire." RSPH highlighted DSD's accessibility and scalability and that it was "underpinned by research and good practice in technology." Hertfordshire's Director of Public Health said that DSD "shows that people with long-term conditions move out of clinical levels of depression at a fraction of the cost of face-to-face care." Hertfordshire Public Health commissioned further DSD programmes under a Do Healthy in Herts banner. They included programmes to help people with Type 2 diabetes manage their condition, improve mental health in young people and empower victims of domestic abuse. DSD developed Do Healthy Cov for the Coventry City Council Public Health team; it was designed to engage 1,500 residents in making small changes in everyday exercise and diet, while making use of the city's leisure facilities. DSD ran a smoking cessation programme in partnership with Brighton and Hove Clinical Commissioning Group. It achieved a "quit rate" of 52% and resulted in DSD being a finalist in the "Partnership with the NHS" category in the 2016 South East Health Technologies Alliance awards. In 2014 Medicinfo, a Dutch healthcare innovation company, invested □85,000 in a three-year agreement to distribute DSD. Beneficiaries in the Netherlands included a public health centre, educational institution, local authority, a care home and people with diabetes and COPD. Medicinfo's director said: "We ran eight programmes, all successful based on qualitative feedback from participants. The best results were in a care home where a complete culture change occurred among healthcare professionals, from dictating an organisational structure to implementing a structure based on what clients want and need. DSD also worked well with blue collar workers in a building company – who liked the conciseness of the 'Do' format." Large corporate organisations used DSD to deliver wellbeing and training programmes to their employees, focusing on issues such as resilience, stress, inclusivity and leadership. Clients during the impact period included General Electric (USA), Aviva, American Express, Mondelez, Remploy, TUI and Cisco. DSD worked with Cisco's HR team to deliver a programme aimed at building a more inclusive workplace culture. It was offered to all employees across Cisco's Europe, Middle East, Africa and Russia region. A total of 1,264 employees across 44 countries participated. A published evaluation found a 13.3% uplift in people engaging with others from different cultures. One employee commented: "Do Something Different has changed a lot of my habits and introduced real change to the person I am at work and at home." In a collaboration with TUI, DSD designed a bespoke leadership training programme that combined TUI's management development goals with a drive to improve digital skills. It was made available to managers on TUI's Perspectives Leadership programme. One international tax manager said: "If I had to sum up the Do programme in a few words it would be 'Life Changing'." A regional finance director commented: "It taught me to let go of the day-to-day and become a more strategic leader." Working with Mondelez, DSD created an "Agility at Work" programme to encourage employees to adopt new ways of working. The programme was made available to all employees in seven sites across the UK and Ireland. As a result, 28.4% said they were more likely to look for new ways of doing things. A similar programme in 2016 engaged Aviva's employees to encourage them to "embrace different measures." The evaluation showed that 74% were likely to recommend DSD and 68% said they were doing things differently as a result

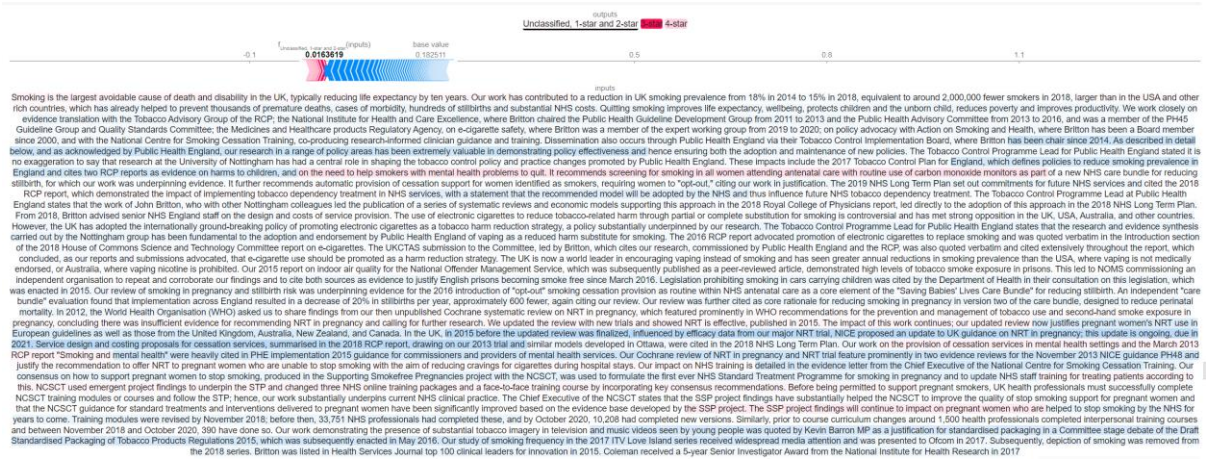


The research-based Do Something Different behavioural intervention, developed at the University of Hertfordshire and commercialised through a spinout company, has resulted in a broad range of impacts. It has delivered economic impact and commercial benefits in the UK and within the EU; evidence of its efficacy as a clinical health intervention has been found through randomised controlled trials overseas; and it has improved people's health, wellbeing and productivity internationally within health, workplace and domestic settings. In February 2014 Rilke's Room Ltd became the social enterprise Do Something Different Ltd following an investment of $1m (£765,000) from PAX Holdings Ltd. This investment allowed the company, co-founded by Fletcher, Pine and social entrepreneur Ray Richards, to develop the DSD platform into a low-cost, scalable and accessible behaviour change intervention. It was aimed at three target beneficiaries to improve wellbeing, mental and physical health, and productivity: individuals, through collaborations with charities; healthcare organisations and public health commissioners; and businesses. DSD Ltd generated £2m in revenue from August 2013 to December 2020, creating and supporting six jobs for core staff and, at one point, sustaining up to 10 roles. The DSD platform underpinned the award of two major EU Horizon 2020 grants (combined value: €13.5m): Do CHANGE (2015-2018), which used the DSD model to change lifestyle and behavioural risk factors for people with heart disease or hypertension, and Preventomics (2018-2021), which applied the DSD model to the development of a digital platform that guides consumers to healthy food choices. These grants included a total of €4.5m (£4m) in direct EU funding specifically for SMEs to develop new products and services. One example is Dutch design firm Onmi (for whom Fletcher is an advisor), which received €900,000. Onmi offers two behaviour change programmes based on DSD: Alius and Flex. It created two products via the Do CHANGE programme: COOKIT, a smart spatula that helps the user adopt healthy cooking habits (it won the "public vote" award at the EC Innovation Radar Prize, 2017) and MySleeve, a smart device that can be wrapped around cups or bottles to encourage healthy drinking habits. The efficacy of DSD was demonstrated by an RCT (ACTRN12615000114549; registered Feb 2015) of habit-based interventions carried out by researchers from the Faculty of Health Sciences and Medicine at Bond University, Australia. DSD was found to be clinically beneficial in achieving long-term weight loss and improved diet and exercise habits. Approximately 65% of participants lost >5% of their body weight (the benchmark of successful and healthful weight loss) and kept it off 12 months after the intervention ended. The researchers concluded: "Habit-based interventions have the potential to change how we think about weight management and importantly how we behave." The Do CHANGE programme resulted in an RCT (NCT02946281; registered Oct 2016) involving 300 heart disease and hypertension patients in three hospitals in Spain, Netherlands and Taiwan. The DSD-based intervention was shown to be feasible and quality of life improved more in the intervention group; a positive cost-effectiveness ratio was demonstrated in the Spanish hospital. DSD Ltd created behaviour change programmes to address various health-related issues. An impact evaluation carried out in 2016 found that, out of 18,000 users from 49 countries, 66.7% reported lower levels of anxiety and depression. The number of people with a clinical level of anxiety fell by 55.9% and a clinical level of depression by 53.9%. Participant comments included: "I got motivated to take up more exercise, to do things differently in my life, where before I was in a rut. My grandchildren loved doing some of it with me," "One of my favourite 'do's' was going out and giving something to somebody that was very unexpected – I took a neighbour whom I'd never met some flowers and some biscuits – in the end we got the whole street involved and all the ladies now meet once a month," "I was basically completely sedentary, I did no exercise at all. Now that I do exercise as a result of doing this course, I've probably added years and years to my life." In 2014 DSD collaborated with UK charity Action for Happiness to design an intensive six-week programme to help people develop habits proven to bring more happiness to themselves and the people around them. The Do Happiness programme sent people regular small positive actions (Do's) to break their existing habit cycles. According to an evaluation published with Action for Happiness, 80% of 1,993 participants from 49 countries said they would recommend DSD to others. The overall average uplift in wellbeing was 21.6%. More than 8,500 comments were shared in the "Do Zone" by programme participants. DSD collaborated with Hertfordshire Public Health to run three programmes between 2014 and 2016: Do Healthy in Herts, which focused on exercise and diet; Do Less Stress in Herts, which explored ways to combat stress; Do Happiness in Herts, which further developed the programme run with Action for Happiness. These three interventions engaged 1,000 residents. Evaluations showed subjective wellbeing up by 21%; more than 50% of people with clinical levels of anxiety and depression moved to healthier levels. Physical activity was up 37%, weight was down by an average of 5.8lbs in six weeks and alcohol consumption fell 20%. Representative comments from participants, captured in video testimony, included: "It definitely helped to refocus mental energy and thinking patterns," "I was feeling extremely low – not wanting to do anything – but then I changed to a more positive attitude in spite of what's going on," "I can't even send out a text, but do you know, I was looking forward to this text coming in, isn't that amazing?" The interventions won a Royal Society for Public Health (RSPH) Health and Wellbeing Award in 2015 for "its innovative and creative approach to behaviour change for individuals in Hertfordshire." RSPH highlighted DSD's accessibility and scalability and that it was "underpinned by research and good practice in technology." Hertfordshire's Director of Public Health said that DSD "shows that people with long-term conditions move out of clinical levels of depression at a fraction of the cost of face-to-face care." Hertfordshire Public Health commissioned further DSD programmes under a Do Healthy in Herts banner. They included programmes to help people with Type 2 diabetes manage their condition, improve mental health in young people and empower victims of domestic abuse. DSD developed Do Healthy Cov for the Coventry City Council Public Health team; it was designed to engage 1,500 residents in making small changes in everyday exercise and diet, while making use of the city's leisure facilities. DSD ran a smoking cessation programme in partnership with Brighton and Hove Clinical Commissioning Group. It achieved a "quit rate" of 52% and resulted in DSD being a finalist in the "Partnership with the NHS" category in the 2016 South East Health Technologies Alliance awards. In 2014 Medicinfo, a Dutch healthcare innovation company, invested □85,000 in a three-year agreement to distribute DSD. Beneficiaries in the Netherlands included a public health centre, educational institution, local authority, a care home and people with diabetes and COPD. Medicinfo's director said: "We ran eight programmes, all successful based on qualitative feedback from participants. The best results were in a care home where a complete culture change occurred among healthcare professionals, from dictating an organisational structure to implementing a structure based on what clients want and need. DSD also worked well with blue collar workers in a building company – who liked the conciseness of the 'Do' format." Large corporate organisations used DSD to deliver wellbeing and training programmes to their employees, focusing on issues such as resilience, stress, inclusivity and leadership. Clients during the impact period included General Electric (USA), Aviva, American Express, Mondelez, Remploy, TUI and Cisco. DSD worked with Cisco's HR team to deliver a programme aimed at building a more inclusive workplace culture. It was offered to all employees across Cisco's Europe, Middle East, Africa and Russia region. A total of 1,264 employees across 44 countries participated. A published evaluation found a 13.3% uplift in people engaging with others from different cultures. One employee commented: "Do Something Different has changed a lot of my habits and introduced real change to the person I am at work and at home." In a collaboration with TUI, DSD designed a bespoke leadership training programme that combined TUI's management development goals with a drive to improve digital skills. It was made available to managers on TUI's Perspectives Leadership programme. One international tax manager said: "If I had to sum up the Do programme in a few words it would be 'Life Changing'." A regional finance director commented: "It taught me to let go of the day-to-day and become a more strategic leader." Working with Mondelez, DSD created an "Agility at Work" programme to encourage employees to adopt new ways of working. The programme was made available to all employees in seven sites across the UK and Ireland. As a result, 28.4% said they were more likely to look for new ways of doing things. A similar programme in 2016 engaged Aviva's employees to encourage them to "embrace different measures." The evaluation showed that 74% were likely to recommend DSD and 68% said they were doing things differently as a result

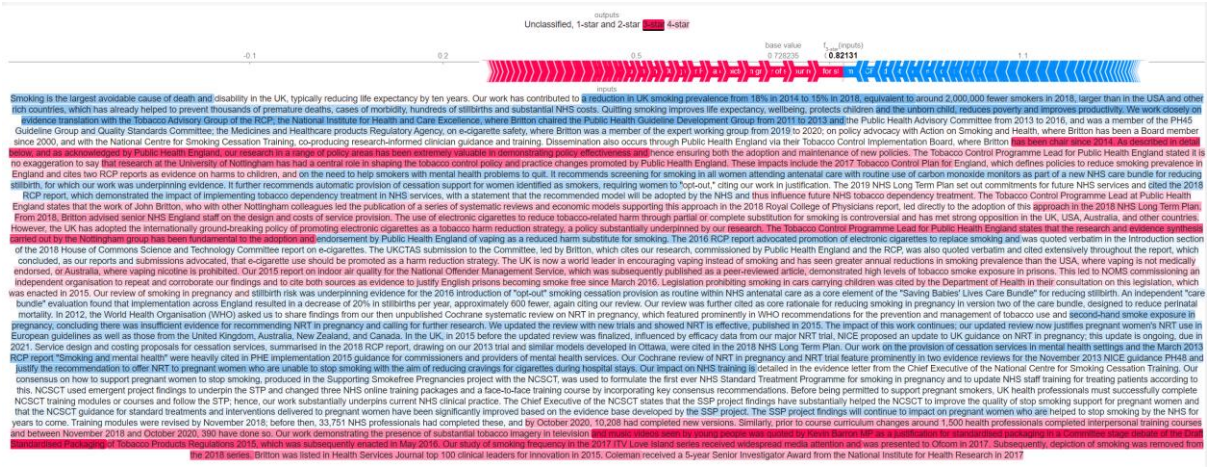


The research-based Do Something Different behavioural intervention, developed at the University of Hertfordshire and commercialised through a spinout company, has resulted in a broad range of impacts. It has delivered economic impact and commercial benefits in the UK and within the EU; evidence of its efficacy as a clinical health intervention has been found through randomised controlled trials overseas; and it has improved people's health, wellbeing and productivity internationally within health, workplace and domestic settings. In February 2014 Rilke's Room Ltd became the social enterprise Do Something Different Ltd following an investment of $1m (£765,000) from PAX Holdings Ltd. This investment allowed the company, co-founded by Fletcher, Pine and social entrepreneur Ray Richards, to develop the DSD platform into a low-cost, scalable and accessible behaviour change intervention. It was aimed at three target beneficiaries to improve wellbeing, mental and physical health, and productivity: individuals, through collaborations with charities; healthcare organisations and public health commissioners; and businesses. DSD Ltd generated £2m in revenue from August 2013 to December 2020, creating and supporting six jobs for core staff and, at one point, sustaining up to 10 roles. The DSD platform underpinned the award of two major EU Horizon 2020 grants (combined value: €13.5m): Do CHANGE (2015-2018), which used the DSD model to change lifestyle and behavioural risk factors for people with heart disease or hypertension, and Preventomics (2018-2021), which applied the DSD model to the development of a digital platform that guides consumers to healthy food choices. These grants included a total of €4.5m (£4m) in direct EU funding specifically for SMEs to develop new products and services. One example is Dutch design firm Onmi (for whom Fletcher is an advisor), which received €900,000. Onmi offers two behaviour change programmes based on DSD: Alius and Flex. It created two products via the Do CHANGE programme: COOKIT, a smart spatula that helps the user adopt healthy cooking habits (it won the "public vote" award at the EC Innovation Radar Prize, 2017) and MySleeve, a smart device that can be wrapped around cups or bottles to encourage healthy drinking habits. The efficacy of DSD was demonstrated by an RCT (ACTRN12615000114549; registered Feb 2015) of habit-based interventions carried out by researchers from the Faculty of Health Sciences and Medicine at Bond University, Australia. DSD was found to be clinically beneficial in achieving long-term weight loss and improved diet and exercise habits. Approximately 65% of participants lost >5% of their body weight (the benchmark of successful and healthful weight loss) and kept it off 12 months after the intervention ended. The researchers concluded: "Habit-based interventions have the potential to change how we think about weight management and importantly how we behave." The Do CHANGE programme resulted in an RCT (NCT02946281; registered Oct 2016) involving 300 heart disease and hypertension patients in three hospitals in Spain, Netherlands and Taiwan. The DSD-based intervention was shown to be feasible and quality of life improved more in the intervention group; a positive cost-effectiveness ratio was demonstrated in the Spanish hospital. DSD Ltd created behaviour change programmes to address various health-related issues. An impact evaluation carried out in 2016 found that, out of 18,000 users from 49 countries, 66.7% reported lower levels of anxiety and depression. The number of people with a clinical level of anxiety fell by 55.9% and a clinical level of depression by 53.9%. Participant comments included: "I got motivated to take up more exercise, to do things differently in my life, where before I was in a rut. My grandchildren loved doing some of it with me," "One of my favourite 'do's' was going out and giving something to somebody that was very unexpected – I took a neighbour whom I'd never met some flowers and some biscuits – in the end we got the whole street involved and all the ladies now meet once a month," "I was basically completely sedentary, I did no exercise at all. Now that I do exercise as a result of doing this course, I've probably added years and years to my life." In 2014 DSD collaborated with UK charity Action for Happiness to design an intensive six-week programme to help people develop habits proven to bring more happiness to themselves and the people around them. The Do Happiness programme sent people regular small positive actions (Do's) to break their existing habit cycles. According to an evaluation published with Action for Happiness, 80% of 1,993 participants from 49 countries said they would recommend DSD to others. The overall average uplift in wellbeing was 21.6%. More than 8,500 comments were shared in the "Do Zone" by programme participants. DSD collaborated with Hertfordshire Public Health to run three programmes between 2014 and 2016: Do Healthy in Herts, which focused on exercise and diet; Do Less Stress in Herts, which explored ways to combat stress; Do Happiness in Herts, which further developed the programme run with Action for Happiness. These three interventions engaged 1,000 residents. Evaluations showed subjective wellbeing up by 21%; more than 50% of people with clinical levels of anxiety and depression moved to healthier levels. Physical activity was up 37%, weight was down by an average of 5.8lbs in six weeks and alcohol consumption fell 20%. Representative comments from participants, captured in video testimony, included: "It definitely helped to refocus mental energy and thinking patterns," "I was feeling extremely low – not wanting to do anything – but then I changed to a more positive attitude in spite of what's going on," "I can't even send out a text, but do you know, I was looking forward to this text coming in, isn't that amazing?" The interventions won a Royal Society for Public Health (RSPH) Health and Wellbeing Award in 2015 for "its innovative and creative approach to behaviour change for individuals in Hertfordshire." RSPH highlighted DSD's accessibility and scalability and that it was "underpinned by research and good practice in technology." Hertfordshire's Director of Public Health said that DSD "shows that people with long-term conditions move out of clinical levels of depression at a fraction of the cost of face-to-face care." Hertfordshire Public Health commissioned further DSD programmes under a Do Healthy in Herts banner. They included programmes to help people with Type 2 diabetes manage their condition, improve mental health in young people and empower victims of domestic abuse. DSD developed Do Healthy Cov for the Coventry City Council Public Health team; it was designed to engage 1,500 residents in making small changes in everyday exercise and diet, while making use of the city's leisure facilities. DSD ran a smoking cessation programme in partnership with Brighton and Hove Clinical Commissioning Group. It achieved a "quit rate" of 52% and resulted in DSD being a finalist in the "Partnership with the NHS" category in the 2016 South East Health Technologies Alliance awards. In 2014 Medicinfo, a Dutch healthcare innovation company, invested □85,000 in a three-year agreement to distribute DSD. Beneficiaries in the Netherlands included a public health centre, educational institution, local authority, a care home and people with diabetes and COPD. Medicinfo's director said: "We ran eight programmes, all successful based on qualitative feedback from participants. The best results were in a care home where a complete culture change occurred among healthcare professionals, from dictating an organisational structure to implementing a structure based on what clients want and need. DSD also worked well with blue collar workers in a building company – who liked the conciseness of the 'Do' format." Large corporate organisations used DSD to deliver wellbeing and training programmes to their employees, focusing on issues such as resilience, stress, inclusivity and leadership. Clients during the impact period included General Electric (USA), Aviva, American Express, Mondelez, Remploy, TUI and Cisco. DSD worked with Cisco's HR team to deliver a programme aimed at building a more inclusive workplace culture. It was offered to all employees across Cisco's Europe, Middle East, Africa and Russia region. A total of 1,264 employees across 44 countries participated. A published evaluation found a 13.3% uplift in people engaging with others from different cultures. One employee commented: "Do Something Different has changed a lot of my habits and introduced real change to the person I am at work and at home." In a collaboration with TUI, DSD designed a bespoke leadership training programme that combined TUI's management development goals with a drive to improve digital skills. It was made available to managers on TUI's Perspectives Leadership programme. One international tax manager said: "If I had to sum up the Do programme in a few words it would be 'Life Changing'." A regional finance director commented: "It taught me to let go of the day-to-day and become a more strategic leader." Working with Mondelez, DSD created an "Agility at Work" programme to encourage employees to adopt new ways of working. The programme was made available to all employees in seven sites across the UK and Ireland. As a result, 28.4% said they were more likely to look for new ways of doing things. A similar programme in 2016 engaged Aviva's employees to encourage them to "embrace different measures." The evaluation showed that 74% were likely to recommend DSD and 68% said they were doing things differently as a result

Index 31: actual: 4-star predicted: 3-star



Smoking is the largest avoidable cause of death and disability in the UK, typically reducing life expectancy by ten years. Our work has contributed to a reduction in UK smoking prevalence from 18% in 2014 to 15% in 2018, equivalent to around 2,000,000 fewer smokers in 2018, larger than in the USA and other rich countries, which has already helped to prevent thousands of premature deaths, cases of morbidity, hundreds of stillbirths and substantial NHS costs. Quitting smoking improves life expectancy, wellbeing, protects children and the unborn child, reduces poverty and improves productivity. We work closely on evidence translation with the Tobacco Advisory Group of the RCP; the National Institute for Health and Care Excellence, where Britton chaired the Public Health Guideline Development Group from 2011 to 2013 and the Public Health Advisory Committee from 2013 to 2016, and was a member of the PH45 Guideline Group and Quality Standards Committee; the Medicines and Healthcare products Regulatory Agency, on e-cigarette safety, where Britton was a member of the expert working group from 2019 to 2020; on policy advocacy with Action on Smoking and Health, where Britton has been a Board member since 2000, and with the National Centre for Smoking Cessation Training, co-producing research-informed clinician guidance and training. Dissemination also occurs through Public Health England via their Tobacco Control Implementation Board, where Britton has been chair since 2014. As described in detail below, and as acknowledged by Public Health England, our research in a range of policy areas has been extremely valuable in demonstrating policy effectiveness and hence ensuring both the adoption and maintenance of new policies. The Tobacco Control Programme Lead for Public Health England stated it is no exaggeration to say that research at the University of Nottingham has had a central role in shaping the tobacco control policy and practice changes promoted by Public Health England. These impacts include the 2017 Tobacco Control Plan for England, which defines policies to reduce smoking prevalence in England and cites two RCP reports as evidence on harms to children, and on the need to help smokers with mental health problems to quit. It recommends screening for smoking in all women attending antenatal care with routine use of carbon monoxide monitors as part of a new NHS care bundle for reducing stillbirth, for which our work was underpinning evidence. It further recommends automatic provision of cessation support for women identified as smokers, requiring women to "opt-out," citing our work in justification. The 2019 NHS Long Term Plan set out commitments for future NHS services and cited the 2018 RCP report, which demonstrated the impact of implementing tobacco dependency treatment in NHS services, with a statement that the recommended model will be adopted by the NHS and thus influence future NHS tobacco dependency treatment. The Tobacco Control Programme Lead at Public Health England states that the work of John Britton, who with other Nottingham colleagues led the publication of a series of systematic reviews and economic models supporting this approach in the 2018 Royal College of Physicians report, led directly to the adoption of this approach in the 2018 NHS Long Term Plan. From 2018, Britton advised senior NHS England staff on the design and costs of service provision. The use of electronic cigarettes to reduce tobacco-related harm through partial or complete substitution for smoking is controversial and has met strong opposition in the UK, USA, Australia, and other countries. However, the UK has adopted the internationally ground-breaking policy of promoting electronic cigarettes as a tobacco harm reduction strategy, a policy substantially underpinned by our research. The Tobacco Control Programme Lead for Public Health England states that the research and evidence synthesis carried out by the Nottingham group has been fundamental to the adoption and endorsement by Public Health England of vaping as a reduced harm substitute for smoking. The 2016 RCP report advocated promotion of electronic cigarettes to replace smoking and was quoted verbatim in the Introduction section of the 2018 House of Commons Science and Technology Committee report on e-cigarettes. The UKCTAS submission to the Committee, led by Britton, which cites our research, commissioned by Public Health England and the RCP, was also quoted verbatim and cited extensively throughout the report, which concluded, as our reports and submissions advocated, that e-cigarette use should be promoted as a harm reduction strategy. The UK is now a world leader in encouraging vaping instead of smoking and has seen greater annual reductions in smoking prevalence than the USA, where vaping is not medically endorsed, or Australia, where vaping nicotine is prohibited. Our 2015 report on indoor air quality for the National Offender Management Service, which was subsequently published as a peer-reviewed article, demonstrated high levels of tobacco smoke exposure in prisons. This led to NOMS commissioning an independent organisation to repeat and corroborate our findings and to cite both sources as evidence to justify English prisons becoming smoke free since March 2016. Legislation prohibiting smoking in cars carrying children was cited by the Department of Health in their consultation on this legislation, which was enacted in 2015. Our review of smoking in pregnancy and stillbirth risk was underpinning evidence for the 2016 introduction of "opt-out" smoking cessation provision as routine within NHS antenatal care as a core element of the "Saving Babies' Lives Care Bundle" for reducing stillbirth. An independent "care bundle" evaluation found that implementation across England resulted in a decrease of 20% in stillbirths per year, approximately 600 fewer, again citing our review. Our review was further cited as core rationale for reducing smoking in pregnancy in version two of the care bundle, designed to reduce perinatal mortality. In 2012, the World Health Organisation (WHO) asked us to share findings from our then unpublished Cochrane systematic review on NRT in pregnancy, which featured prominently in WHO recommendations for the prevention and management of tobacco use and second-hand smoke exposure in pregnancy, concluding there was insufficient evidence for recommending NRT in pregnancy and calling for further research. We updated the review with new trials and showed NRT is effective, published in 2015. The impact of this work continues; our updated review now justifies pregnant women's NRT use in European guidelines as well as those from the United Kingdom, Australia, New Zealand, and Canada. In the UK, in 2015 before the updated review was finalized, influenced by efficacy data from our major NRT trial, NICE proposed an update to UK guidance on NRT in pregnancy; this update is ongoing, due in 2021. Service design and costing proposals for cessation services, summarised in the 2018 RCP report, drawing on our 2013 trial and similar models developed in Ottawa, were cited in the 2018 NHS Long Term Plan. Our work on the provision of cessation services in mental health settings and the March 2013 RCP report "Smoking and mental health" were heavily cited in PHE implementation 2015 guidance for commissioners and providers of mental health services. Our Cochrane review of NRT in pregnancy and NRT trial feature prominently in two evidence reviews for the November 2013 NICE guidance PH48 and justify the recommendation to offer NRT to pregnant women who are unable to stop smoking with the aim of reducing cravings for cigarettes during hospital stays. Our impact on NHS training is detailed in the evidence letter from the Chief Executive of the National Centre for Smoking Cessation Training. Our consensus on how to support pregnant women to stop smoking, produced in the Supporting Smokefree Pregnancies project with the NCSCT, was used to formulate the first ever NHS Standard Treatment Programme for smoking in pregnancy and to update NHS staff training for treating patients according to this. NCSCT used emergent project findings to underpin the STP and changed three NHS online training packages and a face-to-face training course by incorporating key consensus recommendations. Before being permitted to support pregnant smokers, UK health professionals must successfully complete NCSCT training modules or courses and follow the STP; hence, our work substantially underpins current NHS clinical practice. The Chief Executive of the NCSCT states that the SSP project findings have substantially helped the NCSCT to improve the quality of stop smoking support for pregnant women and that the NCSCT guidance for standard treatments and interventions delivered to pregnant women have been significantly improved based on the evidence base developed by the SSP project. The SSP project findings will continue to impact on pregnant women who are helped to stop smoking by the NHS for years to come. Training modules were revised by November 2018; before then, 33,751 NHS professionals had completed these, and by October 2020, 10,208 had completed new versions. Similarly, prior to course curriculum changes around 1,500 health professionals completed interpersonal training courses and between November 2018 and October 2020, 390 have done so. Our work demonstrating the presence of substantial tobacco imagery in television and music videos seen by young people was quoted by Kevin Barron MP as a justification for standardised packaging in a Committee stage debate of the Draft Standardised Packaging of Tobacco Products Regulations 2015, which was subsequently enacted in May 2016. Our study of smoking frequency in the 2017 ITV Love Island series received widespread media attention and was presented to Ofcom in 2017. Subsequently, depiction of smoking was removed from the 2018 series. Britton was listed in Health Services Journal top 100 clinical leaders for innovation in 2015. Coleman received a 5-year Senior Investigator Award from the National Institute for Health Research in 2017



Smoking is the largest avoidable cause of death and disability in the UK, typically reducing life expectancy by ten years. Our work has contributed to a reduction in UK smoking prevalence from 18% in 2014 to 15% in 2018, equivalent to around 2,000,000 fewer smokers in 2018, larger than in the USA and other rich countries, which has already helped to prevent thousands of premature deaths, cases of morbidity, hundreds of stillbirths and substantial NHS costs. Quitting smoking improves life expectancy, wellbeing, protects children and the unborn child, reduces poverty and improves productivity. We work closely on evidence translation with the Tobacco Advisory Group of the RCP; the National Institute for Health and Care Excellence, where Britton chaired the Public Health Guideline Development Group from 2011 to 2013 and the Public Health Advisory Committee from 2013 to 2016, and was a member of the PH45 Guideline Group and Quality Standards Committee; the Medicines and Healthcare products Regulatory Agency, on e-cigarette safety, where Britton was a member of the expert working group from 2019 to 2020; on policy advocacy with Action on Smoking and Health, where Britton has been a Board member since 2000, and with the National Centre for Smoking Cessation Training, co-producing research-informed clinician guidance and training. Dissemination also occurs through Public Health England via their Tobacco Control Implementation Board, where Britton has been chair since 2014. As described in detail below, and as acknowledged by Public Health England, our research in a range of policy areas has been extremely valuable in demonstrating policy effectiveness and hence ensuring both the adoption and maintenance of new policies. The Tobacco Control Programme Lead for Public Health England stated it is no exaggeration to say that research at the University of Nottingham has had a central role in shaping the tobacco control policy and practice changes promoted by Public Health England. These impacts include the 2017 Tobacco Control Plan for England, which defines policies to reduce smoking prevalence in England and cites two RCP reports as evidence on harms to children, and on the need to help smokers with mental health problems to quit. It recommends screening for smoking in all women attending antenatal care with routine use of carbon monoxide monitors as part of a new NHS care bundle for reducing stillbirth, for which our work was underpinning evidence. It further recommends automatic provision of cessation support for women identified as smokers, requiring women to "opt-out," citing our work in justification. The 2019 NHS Long Term Plan set out commitments for future NHS services and cited the 2018 RCP report, which demonstrated the impact of implementing tobacco dependency treatment in NHS services, with a statement that the recommended model will be adopted by the NHS and thus influence future NHS tobacco dependency treatment. The Tobacco Control Programme Lead at Public Health England states that the work of John Britton, who with other Nottingham colleagues led the publication of a series of systematic reviews and economic models supporting this approach in the 2018 Royal College of Physicians report, led directly to the adoption of this approach in the 2018 NHS Long Term Plan. From 2018, Britton advised senior NHS England staff on the design and costs of service provision. The use of electronic cigarettes to reduce tobacco-related harm through partial or complete substitution for smoking is controversial and has met strong opposition in the UK, USA, Australia, and other countries. However, the UK has adopted the internationally ground-breaking policy of promoting electronic cigarettes as a tobacco harm reduction strategy, a policy substantially underpinned by our research. The Tobacco Control Programme Lead for Public Health England states that the research and evidence synthesis carried out by the Nottingham group has been fundamental to the adoption and endorsement by Public Health England of vaping as a reduced harm substitute for smoking. The 2016 RCP report advocated promotion of electronic cigarettes to replace smoking and was quoted verbatim in the Introduction section of the 2018 House of Commons Science and Technology Committee report on e-cigarettes. The UKCTAS submission to the Committee, led by Britton, which cites our research, commissioned by Public Health England and the RCP, was also quoted verbatim and cited extensively throughout the report, which concluded, as our reports and submissions advocated, that e-cigarette use should be promoted as a harm reduction strategy. The UK is now a world leader in encouraging vaping instead of smoking and has seen greater annual reductions in smoking prevalence than the USA, where vaping is not medically endorsed, or Australia, where vaping nicotine is prohibited. Our 2015 report on indoor air quality for the National Offender Management Service, which was subsequently published as a peer-reviewed article, demonstrated high levels of tobacco smoke exposure in prisons. This led to NOMS commissioning an independent organisation to repeat and corroborate our findings and to cite both sources as evidence to justify English prisons becoming smoke free since March 2016. Legislation prohibiting smoking in cars carrying children was cited by the Department of Health in their consultation on this legislation, which was enacted in 2015. Our review of smoking in pregnancy and stillbirth risk was underpinning evidence for the 2016 introduction of "opt-out" smoking cessation provision as routine within NHS antenatal care as a core element of the "Saving Babies' Lives Care Bundle" for reducing stillbirth. An independent "care bundle" evaluation found that implementation across England resulted in a decrease of 20% in stillbirths per year, approximately 600 fewer, again citing our review. Our review was further cited as core rationale for reducing smoking in pregnancy in version two of the care bundle, designed to reduce perinatal mortality. In 2012, the World Health Organisation (WHO) asked us to share findings from our then unpublished Cochrane systematic review on NRT in pregnancy, which featured prominently in WHO recommendations for the prevention and management of tobacco use and second-hand smoke exposure in pregnancy, concluding there was insufficient evidence for recommending NRT in pregnancy and calling for further research. We updated the review with new trials and showed NRT is effective, published in 2015. The impact of this work continues; our updated review now justifies pregnant women's NRT use in European guidelines as well as those from the United Kingdom, Australia, New Zealand, and Canada. In the UK, in 2015 before the updated review was finalized, influenced by efficacy data from our major NRT trial, NICE proposed an update to UK guidance on NRT in pregnancy; this update is ongoing, due in 2021. Service design and costing proposals for cessation services, summarised in the 2018 RCP report, drawing on our 2013 trial and similar models developed in Ottawa, were cited in the 2018 NHS Long Term Plan. Our work on the provision of cessation services in mental health settings and the March 2013 RCP report "Smoking and mental health" were heavily cited in PHE implementation 2015 guidance for commissioners and providers of mental health services. Our Cochrane review of NRT in pregnancy and NRT trial feature prominently in two evidence reviews for the November 2013 NICE guidance PH48 and justify the recommendation to offer NRT to pregnant women who are unable to stop smoking with the aim of reducing cravings for cigarettes during hospital stays. Our impact on NHS training is detailed in the evidence letter from the Chief Executive of the National Centre for Smoking Cessation Training. Our consensus on how to support pregnant women to stop smoking, produced in the Supporting Smokefree Pregnancies project with the NCSCT, was used to formulate the first ever NHS Standard Treatment Programme for smoking in pregnancy and to update NHS staff training for treating patients according to this. NCSCT used emergent project findings to underpin the STP and changed three NHS online training packages and a face-to-face training course by incorporating key consensus recommendations. Before being permitted to support pregnant smokers, UK health professionals must successfully complete NCSCT training modules or courses and follow the STP; hence, our work substantially underpins current NHS clinical practice. The Chief Executive of the NCSCT states that the SSP project findings have substantially helped the NCSCT to improve the quality of stop smoking support for pregnant women and that the NCSCT guidance for standard treatments and interventions delivered to pregnant women have been significantly improved based on the evidence base developed by the SSP project. The SSP project findings will continue to impact on pregnant women who are helped to stop smoking by the NHS for years to come. Training modules were revised by November 2018; before then, 33,751 NHS professionals had completed these, and by October 2020, 10,208 had completed new versions. Similarly, prior to course curriculum changes around 1,500 health professionals completed interpersonal training courses and between November 2018 and October 2020, 390 have done so. Our work demonstrating the presence of substantial tobacco imagery in television and music videos seen by young people was quoted by Kevin Barron MP as a justification for standardised packaging in a Committee stage debate of the Draft Standardised Packaging of Tobacco Products Regulations 2015, which was subsequently enacted in May 2016. Our study of smoking frequency in the 2017 ITV Love Island series received widespread media attention and was presented to Ofcom in 2017. Subsequently, depiction of smoking was removed from the 2018 series. Britton was listed in Health Services Journal top 100 clinical leaders for innovation in 2015. Coleman received a 5-year Senior Investigator Award from the National Institute for Health Research in 2017

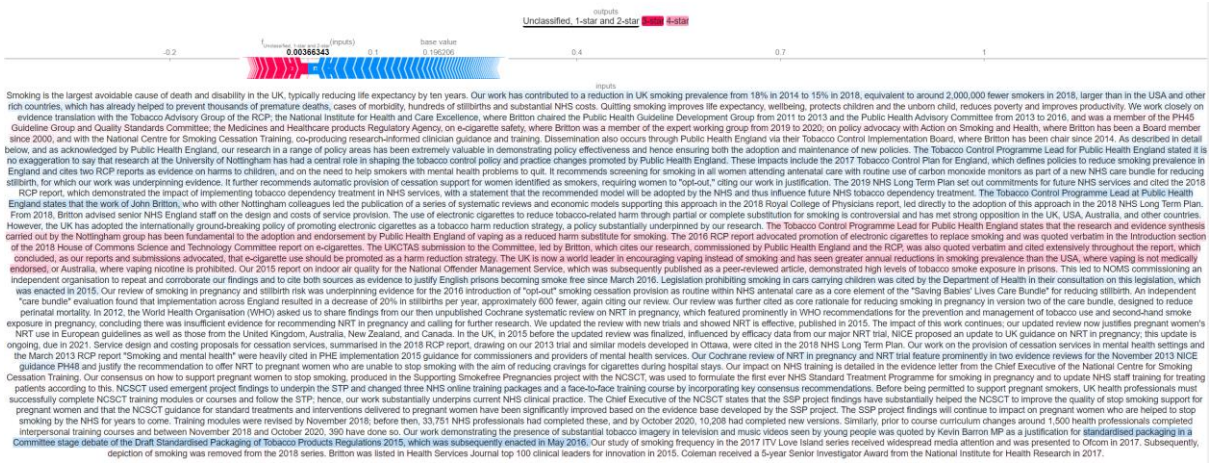
Smoking is the largest avoidable cause of death and disability in the UK, typically reducing life expectancy by ten years. Our work has contributed to a reduction in UK smoking prevalence from 18% in 2014 to 15% in 2018, equivalent to around 2,000,000 fewer smokers in 2018, larger than in the USA and other rich countries, which has already helped to prevent thousands of premature deaths, cases of morbidity, hundreds of stillbirths and substantial NHS costs. Quitting smoking improves life expectancy, wellbeing, protects children and the unborn child, reduces poverty and improves productivity. We work closely on evidence translation with the Tobacco Advisory Group of the RCP; the National Institute for Health and Care Excellence, where Britton chaired the Public Health Guideline Development Group from 2011 to 2013 and the Public Health Advisory Committee from 2013 to 2016, and was a member of the PH45 Guideline Group and Quality Standards Committee; the Medicines and Healthcare products Regulatory Agency, on e-cigarette safety, where Britton was a member of the expert working group from 2019 to 2020; on policy advocacy with Action on Smoking and Health, where Britton has been a Board member since 2000, and with the National Centre for Smoking Cessation Training, co-producing research-informed clinician guidance and training. Dissemination also occurs through Public Health England via their Tobacco Control Implementation Board, where Britton has been chair since 2014. As described in detail below, and as acknowledged by Public Health England, our research in a range of policy areas has been extremely valuable in demonstrating policy effectiveness and hence ensuring both the adoption and maintenance of new policies. The Tobacco Control Programme Lead for Public Health England stated it is no exaggeration to say that research at the University of Nottingham has had a central role in shaping the tobacco control policy and practice changes promoted by Public Health England. These impacts include the 2017 Tobacco Control Plan for England, which defines policies to reduce smoking prevalence in England and cites two RCP reports as evidence on harms to children, and on the need to help smokers with mental health problems to quit. It recommends screening for smoking in all women attending antenatal care with routine use of carbon monoxide monitors as part of a new NHS care bundle for reducing stillbirth, for which our work was underpinning evidence. It further recommends automatic provision of cessation support for women identified as smokers, requiring women to "opt-out," citing our work in justification. The 2019 NHS Long Term Plan set out commitments for future NHS services and cited the 2018 RCP report, which demonstrated the impact of implementing tobacco dependency treatment in NHS services, with a statement that the recommended model will be adopted by the NHS and thus influence future NHS tobacco dependency treatment. The Tobacco Control Programme Lead at Public Health England states that the work of John Britton, who with other Nottingham colleagues led the publication of a series of systematic reviews and economic models supporting this approach in the 2018 Royal College of Physicians report, led directly to the adoption of this approach in the 2018 NHS Long Term Plan. From 2018, Britton advised senior NHS England staff on the design and costs of service provision. The use of electronic cigarettes to reduce tobacco-related harm through partial or complete substitution for smoking is controversial and has met strong opposition in the UK, USA, Australia, and other countries. However, the UK has adopted the internationally ground-breaking policy of promoting electronic cigarettes as a tobacco harm reduction strategy, a policy substantially underpinned by our research. The Tobacco Control Programme Lead for Public Health England states that the research and evidence synthesis carried out by the Nottingham group has been fundamental to the adoption and endorsement by Public Health England of vaping as a reduced harm substitute for smoking. The 2016 RCP report advocated promotion of electronic cigarettes to replace smoking and was quoted verbatim in the Introduction section of the 2018 House of Commons Science and Technology Committee report on e-cigarettes. The UKCTAS submission to the Committee, led by Britton, which cites our research, commissioned by Public Health England and the RCP, was also quoted verbatim and cited extensively throughout the report, which concluded, as our reports and submissions advocated, that e-cigarette use should be promoted as a harm reduction strategy. The UK is now a world leader in encouraging vaping instead of smoking and has seen greater annual reductions in smoking prevalence than the USA, where vaping is not medically endorsed, or Australia, where vaping nicotine is prohibited. Our 2015 report on indoor air quality for the National Offender Management Service, which was subsequently published as a peer-reviewed article, demonstrated high levels of tobacco smoke exposure in prisons. This led to NOMS commissioning an independent organisation to repeat and corroborate our findings and to cite both sources as evidence to justify English prisons becoming smoke free since March 2016. Legislation prohibiting smoking in cars carrying children was cited by the Department of Health in their consultation on this legislation, which was enacted in 2015. Our review of smoking in pregnancy and stillbirth risk was underpinning evidence for the 2016 introduction of "opt-out" smoking cessation provision as routine within NHS antenatal care as a core element of the "Saving Babies' Lives Care Bundle" for reducing stillbirth. An independent "care bundle" evaluation found that implementation across England resulted in a decrease of 20% in stillbirths per year, approximately 600 fewer, again citing our review. Our review was further cited as core rationale for reducing smoking in pregnancy in version two of the care bundle, designed to reduce perinatal mortality. In 2012, the World Health Organisation (WHO) asked us to share findings from our then unpublished Cochrane systematic review on NRT in pregnancy, which featured prominently in WHO recommendations for the prevention and management of tobacco use and second-hand smoke exposure in pregnancy, concluding there was insufficient evidence for recommending NRT in pregnancy and calling for further research. We updated the review with new trials and showed NRT is effective, published in 2015. The impact of this work continues; our updated review now justifies pregnant women's NRT use in European guidelines as well as those from the United Kingdom, Australia, New Zealand, and Canada. In the UK, in 2015 before the updated review was finalized, influenced by efficacy data from our major NRT trial, NICE proposed an update to UK guidance on NRT in pregnancy; this update is ongoing, due in 2021. Service design and costing proposals for cessation services, summarised in the 2018 RCP report, drawing on our 2013 trial and similar models developed in Ottawa, were cited in the 2018 NHS Long Term Plan. Our work on the provision of cessation services in mental health settings and the March 2013 RCP report "Smoking and mental health" were heavily cited in PHE implementation 2015 guidance for commissioners and providers of mental health services. Our Cochrane review of NRT in pregnancy and NRT trial feature prominently in two evidence reviews for the November 2013 NICE guidance PH48 and justify the recommendation to offer NRT to pregnant women who are unable to stop smoking with the aim of reducing cravings for cigarettes during hospital stays. Our impact on NHS training is detailed in the evidence letter from the Chief Executive of the National Centre for Smoking Cessation Training. Our consensus on how to support pregnant women to stop smoking, produced in the Supporting Smokefree Pregnancies project with the NCSCT, was used to formulate the first ever NHS Standard Treatment Programme for smoking in pregnancy and to update NHS staff training for treating patients according to this. NCSCT used emergent project findings to underpin the STP and changed three NHS online training packages and a face-to-face training course by incorporating key consensus recommendations. Before being permitted to support pregnant smokers, UK health professionals must successfully complete NCSCT training modules or courses and follow the STP; hence, our work substantially underpins current NHS clinical practice. The Chief Executive of the NCSCT states that the SSP project findings have substantially helped the NCSCT to improve the quality of stop smoking support for pregnant women and that the NCSCT guidance for standard treatments and interventions delivered to pregnant women have been significantly improved based on the evidence base developed by the SSP project. The SSP project findings will continue to impact on pregnant women who are helped to stop smoking by the NHS for years to come. Training modules were revised by November 2018; before then, 33,751 NHS professionals had completed these, and by October 2020, 10,208 had completed new versions. Similarly, prior to course curriculum changes around 1,500 health professionals completed interpersonal training courses and between November 2018 and October 2020, 390 have done so. Our work demonstrating the presence of substantial tobacco imagery in television and music videos seen by young people was quoted by Kevin Barron MP as a justification for standardised packaging in a Committee stage debate of the Draft Standardised Packaging of Tobacco Products Regulations 2015, which was subsequently enacted in May 2016. Our study of smoking frequency in the 2017 ITV Love Island series received widespread media attention and was presented to Ofcom in 2017. Subsequently, depiction of smoking was removed from the 2018 series. Britton was listed in Health Services Journal top 100 clinical leaders for innovation in 2015. Coleman received a 5-year Senior Investigator Award from the National Institute for Health Research in 2017

Index 35: actual: 2-star or less : 2-star or less

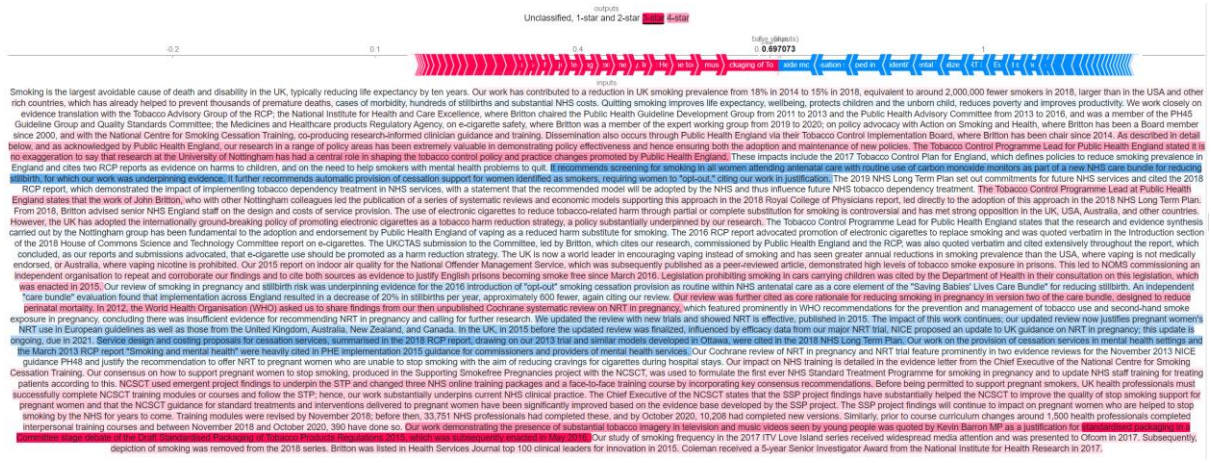
Appendix 6: BERT Longformer Variant local analysis with the impact case studies of Index 21, 31, 32, and 136.

Index 21: actual: 4-star predicted: 4-star

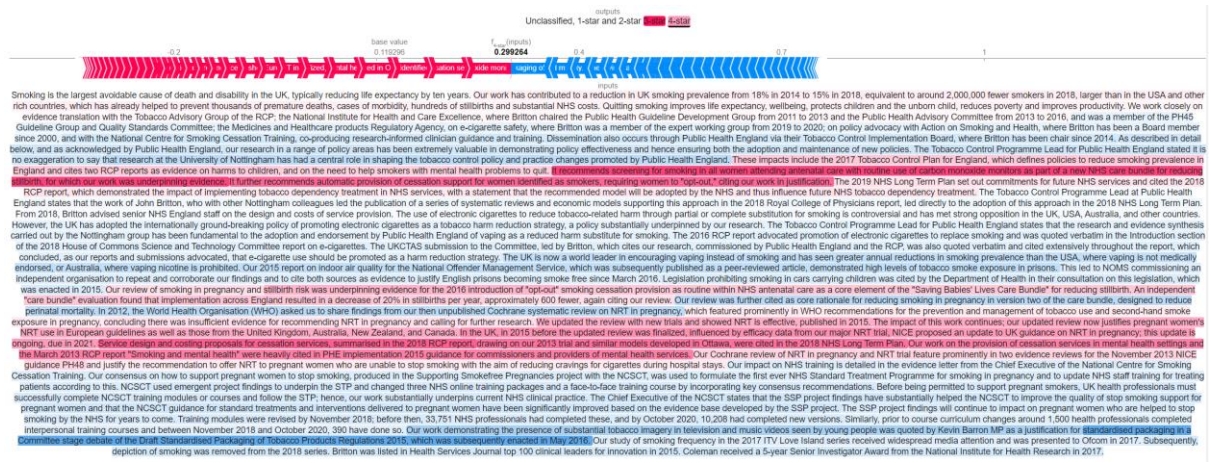
Index 31: actual: 4-star predicted: 3-star



Smoking is the largest avoidable cause of death and disability in the UK, typically reducing life expectancy by ten years. Our work has contributed to a reduction in UK smoking prevalence from 18% in 2014 to 15% in 2018, equivalent to around 2,000,000 fewer smokers in 2018, larger than in the USA and other rich countries, which has already helped to prevent thousands of premature deaths, cases of morbidity, hundreds of stillbirths and substantial NHS costs. Quitting smoking improves life expectancy, wellbeing, protects children and the unborn child, reduces poverty and improves productivity. We work closely on evidence translation with the Tobacco Advisory Group of the RCP, the National Institute for Health and Care Excellence, where Britton chaired the Public Health Guideline Development Group from 2011 to 2013 and the Public Health Advisory Committee from 2013 to 2016, and was a member of the PH45 Guideline Group and Quality Standards Committee; the Medicines and Healthcare products Regulatory Agency, on e-cigarette safety, where Britton was a member of the expert working group from 2019 to 2020; on policy advocacy with Action on Smoking and Health, where Britton has been a Board member since 2000, and with the National Centre for Smoking Cessation Training, co-producing research-informed clinician guidance and training. Dissemination also occurs through Public Health England via their Tobacco Control Implementation Board, where Britton has been chair since 2014. As described in detail below, and as acknowledged by Public Health England, our research in a range of policy areas has been extremely valuable in demonstrating policy effectiveness and hence ensuring both the adoption and maintenance of new policies. The Tobacco Control Programme Lead for Public Health England stated it is no exaggeration to say that research at the University of Nottingham has had a central role in shaping the tobacco control policy and practice changes promoted by Public Health England. These impacts include the 2017 Tobacco Control Plan for England, which defines policies to reduce smoking prevalence in England and cites two RCP reports as evidence on harms to children, and on the need to help smokers with mental health problems to quit. It recommends screening for smoking in all women attending antenatal care with routine use of carbon monoxide monitors as part of a new NHS care bundle for reducing stillbirth, for which our work was underpinning evidence. It further recommends automatic provision of cessation support for women identified as smokers, requiring women to "opt-out," citing our work in justification. The 2019 NHS Long Term Plan set out commitments for future NHS services and cited the 2018 RCP report, which demonstrated the impact of implementing tobacco dependency treatment in NHS services, with a statement that the recommended model will be adopted by the NHS and thus influence future NHS tobacco dependency treatment. The Tobacco Control Programme Lead at Public Health England states that the work of John Britton, who with other Nottingham colleagues led the publication of a series of systematic reviews and economic models supporting this approach in the 2018 Royal College of Physicians report, led directly to the adoption of this approach in the 2018 NHS Long Term Plan. From 2018, Britton advised senior NHS England staff on the design and costs of service provision. The use of electronic cigarettes to reduce tobacco-related harm through partial or complete substitution for smoking is controversial and has met strong opposition in the UK, USA, Australia, and other countries. However, the UK has adopted the internationally ground-breaking policy of promoting electronic cigarettes as a tobacco harm reduction strategy, a policy substantially underpinned by our research. The Tobacco Control Programme Lead for Public Health England states that the research and evidence synthesis carried out by the Nottingham group has been fundamental to the adoption and endorsement by Public Health England of vaping as a reduced harm substitute for smoking. The 2016 RCP report advocated promotion of electronic cigarettes to replace smoking and was quoted verbatim in the Introduction section of the 2018 House of Commons Science and Technology Committee report on e-cigarettes. The UKCTAS submission to the Committee, led by Britton, which cites our research, commissioned by Public Health England and the RCP, was also quoted verbatim and cited extensively throughout the report, which concluded, as our reports and submissions advocated, that e-cigarette use should be promoted as a harm reduction strategy. The UK is now a world leader in encouraging vaping instead of smoking and has seen greater annual reductions in smoking prevalence than the USA, where vaping is not medically endorsed, or Australia, where vaping nicotine is prohibited. Our 2015 report on indoor air quality for the National Offender Management Service, which was subsequently published as a peer-reviewed article, demonstrated high levels of tobacco smoke exposure in prisons. This led to NOMS commissioning an independent organisation to repeat and corroborate our findings and to cite both sources as evidence to justify English prisons becoming smoke free since March 2016. Legislation prohibiting smoking in cars carrying children was cited by the Department of Health in their consultation on this legislation, which was enacted in 2015. Our review of smoking in pregnancy and stillbirth risk was underpinning evidence for the 2016 introduction of "opt-out" smoking cessation provision as routine within NHS antenatal care as a core element of the "Saving Babies' Lives Care Bundle" for reducing stillbirth. An independent "care bundle" evaluation found that implementation across England resulted in a decrease of 20% in stillbirths per year, approximately 600 fewer, again citing our review. Our review was further cited as core rationale for reducing smoking in pregnancy in version two of the care bundle, designed to reduce perinatal mortality. In 2012, the World Health Organisation (WHO) asked us to share findings from our then unpublished Cochrane systematic review on NRT in pregnancy, which featured prominently in WHO recommendations for the prevention and management of tobacco use and second-hand smoke exposure in pregnancy, concluding there was insufficient evidence for recommending NRT in pregnancy and calling for further research. We updated the review with new trials and showed NRT is effective, published in 2015. The impact of this work continues; our updated review now justifies pregnant women's NRT use in European guidelines as well as those from the United Kingdom, Australia, New Zealand, and Canada. In the UK, in 2015 before the updated review was finalized, influenced by efficacy data from our major NRT trial, NICE proposed an update to UK guidance on NRT in pregnancy; this update is ongoing, due in 2021. Service design and costing proposals for cessation services, summarised in the 2018 RCP report, drawing on our 2013 trial and similar models developed in Ottawa, were cited in the 2018 NHS Long Term Plan. Our work on the provision of cessation services in mental health settings and the March 2013 RCP report "Smoking and mental health" were heavily cited in PHE implementation 2015 guidance for commissioners and providers of mental health services. Our Cochrane review of NRT in pregnancy and NRT trial feature prominently in two evidence reviews for the November 2013 NICE guidance PH48 and justify the recommendation to offer NRT to pregnant women who are unable to stop smoking with the aim of reducing cravings for cigarettes during hospital stays. Our impact on NHS training is detailed in the evidence letter from the Chief Executive of the National Centre for Smoking Cessation Training. Our consensus on how to support pregnant women to stop smoking, produced in the Supporting Smokefree Pregnancies project with the NCSCT, was used to formulate the first ever NHS Standard Treatment Programme for smoking in pregnancy and to update NHS staff training for treating patients according to this. NCSCT used emergent project findings to underpin the STP and changed three NHS online training packages and a face-to-face training course by incorporating key consensus recommendations. Before being permitted to support pregnant smokers, UK health professionals must successfully complete NCSCT training modules or courses and follow the STP; hence, our work substantially underpins current NHS clinical practice. The Chief Executive of the NCSCT states that the SSP project findings have substantially helped the NCSCT to improve the quality of stop smoking support for pregnant women and that the NCSCT guidance for standard treatments and interventions delivered to pregnant women have been significantly improved based on the evidence base developed by the SSP project. The SSP project findings will continue to impact on pregnant women who are helped to stop smoking by the NHS for years to come. Training modules were revised by November 2018; before then, 33,751 NHS professionals had completed these, and by October 2020, 10,208 had completed new versions. Similarly, prior to course curriculum changes around 1,500 health professionals completed interpersonal training courses and between November 2018 and October 2020, 390 have done so. Our work demonstrating the presence of substantial tobacco imagery in television and music videos seen by young people was quoted by Kevin Barron MP as a justification for standardised packaging in a Committee stage debate of the Draft Standardised Packaging of Tobacco Products Regulations 2015, which was subsequently enacted in May 2016. Our study of smoking frequency in the 2017 ITV Love Island series received widespread media attention and was presented to Ofcom in 2017. Subsequently, depiction of smoking was removed from the 2018 series. Britton was listed in Health Services Journal top 100 clinical leaders for innovation in 2015. Coleman received a 5-year Senior Investigator Award from the National Institute for Health Research in 2017.



Smoking is the largest avoidable cause of death and disability in the UK, typically reducing life expectancy by ten years. Our work has contributed to a reduction in UK smoking prevalence from 18% in 2014 to 15% in 2018, equivalent to around 2,000,000 fewer smokers in 2018, larger than in the USA and other rich countries, which has already helped to prevent thousands of premature deaths, cases of morbidity, hundreds of stillbirths and substantial NHS costs. Quitting smoking improves life expectancy, wellbeing, protects children and the unborn child, reduces poverty and improves productivity. We work closely on evidence translation with the Tobacco Advisory Group of the RCP, the National Institute for Health and Care Excellence, where Britton chaired the Public Health Guideline Development Group from 2011 to 2013 and the Public Health Advisory Committee from 2013 to 2016, and was a member of the PH45 Guideline Group and Quality Standards Committee; the Medicines and Healthcare products Regulatory Agency, on e-cigarette safety, where Britton was a member of the expert working group from 2019 to 2020; on policy advocacy with Action on Smoking and Health, where Britton has been a Board member since 2000, and with the National Centre for Smoking Cessation Training, co-producing research-informed clinician guidance and training. Dissemination also occurs through Public Health England via their Tobacco Control Implementation Board, where Britton has been chair since 2014. As described in detail below, and as acknowledged by Public Health England, our research in a range of policy areas has been extremely valuable in demonstrating policy effectiveness and hence ensuring both the adoption and maintenance of new policies. The Tobacco Control Programme Lead for Public Health England stated it is no exaggeration to say that research at the University of Nottingham has had a central role in shaping the tobacco control policy and practice changes promoted by Public Health England. These impacts include the 2017 Tobacco Control Plan for England, which defines policies to reduce smoking prevalence in England and cites two RCP reports as evidence on harms to children, and on the need to help smokers with mental health problems to quit. It recommends screening for smoking in all women attending antenatal care with routine use of carbon monoxide monitors as part of a new NHS care bundle for reducing stillbirth, for which our work was underpinning evidence. It further recommends automatic provision of cessation support for women identified as smokers, requiring women to "opt-out," citing our work in justification. The 2019 NHS Long Term Plan set out commitments for future NHS services and cited the 2018 RCP report, which demonstrated the impact of implementing tobacco dependency treatment in NHS services, with a statement that the recommended model will be adopted by the NHS and thus influence future NHS tobacco dependency treatment. The Tobacco Control Programme Lead at Public Health England states that the work of John Britton, who with other Nottingham colleagues led the publication of a series of systematic reviews and economic models supporting this approach in the 2018 Royal College of Physicians report, led directly to the adoption of this approach in the 2018 NHS Long Term Plan. From 2018, Britton advised senior NHS England staff on the design and costs of service provision. The use of electronic cigarettes to reduce tobacco-related harm through partial or complete substitution for smoking is controversial and has met strong opposition in the UK, USA, Australia, and other countries. However, the UK has adopted the internationally ground-breaking policy of promoting electronic cigarettes as a tobacco harm reduction strategy, a policy substantially underpinned by our research. The Tobacco Control Programme Lead for Public Health England states that the research and evidence synthesis carried out by the Nottingham group has been fundamental to the adoption and endorsement by Public Health England of vaping as a reduced harm substitute for smoking. The 2016 RCP report advocated promotion of electronic cigarettes to replace smoking and was quoted verbatim in the Introduction section of the 2018 House of Commons Science and Technology Committee report on e-cigarettes. The UKCTAS submission to the Committee, led by Britton, which cites our research, commissioned by Public Health England and the RCP, was also quoted verbatim and cited extensively throughout the report, which concluded, as our reports and submissions advocated, that e-cigarette use should be promoted as a harm reduction strategy. The UK is now a world leader in encouraging vaping instead of smoking and has seen greater annual reductions in smoking prevalence than the USA, where vaping is not medically endorsed, or Australia, where vaping nicotine is prohibited. Our 2015 report on indoor air quality for the National Offender Management Service, which was subsequently published as a peer-reviewed article, demonstrated high levels of tobacco smoke exposure in prisons. This led to NOMS commissioning an independent organisation to repeat and corroborate our findings and to cite both sources as evidence to justify English prisons becoming smoke free since March 2016. Legislation prohibiting smoking in cars carrying children was cited by the Department of Health in their consultation on this legislation, which was enacted in 2015. Our review of smoking in pregnancy and stillbirth risk was underpinning evidence for the 2016 introduction of "opt-out" smoking cessation provision as routine within NHS antenatal care as a core element of the "Saving Babies' Lives Care Bundle" for reducing stillbirth. An independent "care bundle" evaluation found that implementation across England resulted in a decrease of 20% in stillbirths per year, approximately 600 fewer, again citing our review. Our review was further cited as core rationale for reducing smoking in pregnancy in version two of the care bundle, designed to reduce perinatal mortality. In 2012, the World Health Organisation (WHO) asked us to share findings from our then unpublished Cochrane systematic review on NRT in pregnancy, which featured prominently in WHO recommendations for the prevention and management of tobacco use and second-hand smoke exposure in pregnancy, concluding there was insufficient evidence for recommending NRT in pregnancy and calling for further research. We updated the review with new trials and showed NRT is effective, published in 2015. The impact of this work continues; our updated review now justifies pregnant women's NRT use in European guidelines as well as those from the United Kingdom, Australia, New Zealand, and Canada. In the UK, in 2015 before the updated review was finalized, influenced by efficacy data from our major NRT trial, NICE proposed an update to UK guidance on NRT in pregnancy; this update is ongoing, due in 2021. Service design and costing proposals for cessation services, summarised in the 2018 RCP report, drawing on our 2013 trial and similar models developed in Ottawa, were cited in the 2018 NHS Long Term Plan. Our work on the provision of cessation services in mental health settings and the March 2013 RCP report "Smoking and mental health" were heavily cited in PHE implementation 2015 guidance for commissioners and providers of mental health services. Our Cochrane review of NRT in pregnancy and NRT trial feature prominently in two evidence reviews for the November 2013 NICE guidance PH48 and justify the recommendation to offer NRT to pregnant women who are unable to stop smoking with the aim of reducing cravings for cigarettes during hospital stays. Our impact on NHS training is detailed in the evidence letter from the Chief Executive of the National Centre for Smoking Cessation Training. Our consensus on how to support pregnant women to stop smoking, produced in the Supporting Smokefree Pregnancies project with the NCSCT, was used to formulate the first ever NHS Standard Treatment Programme for smoking in pregnancy and to update NHS staff training for treating patients according to this. NCSCT used emergent project findings to underpin the STP and changed three NHS online training packages and a face-to-face training course by incorporating key consensus recommendations. Before being permitted to support pregnant smokers, UK health professionals must successfully complete NCSCT training modules or courses and follow the STP; hence, our work substantially underpins current NHS clinical practice. The Chief Executive of the NCSCT states that the SSP project findings have substantially helped the NCSCT to improve the quality of stop smoking support for pregnant women and that the NCSCT guidance for standard treatments and interventions delivered to pregnant women have been significantly improved based on the evidence base developed by the SSP project. The SSP project findings will continue to impact on pregnant women who are helped to stop smoking by the NHS for years to come. Training modules were revised by November 2018; before then, 33,751 NHS professionals had completed these, and by October 2020, 10,208 had completed new versions. Similarly, prior to course curriculum changes around 1,500 health professionals completed interpersonal training courses and between November 2018 and October 2020, 390 have done so. Our work demonstrating the presence of substantial tobacco imagery in television and music videos seen by young people was quoted by Kevin Barron MP as a justification for standardised packaging in a Committee stage debate of the Draft Standardised Packaging of Tobacco Products Regulations 2015, which was subsequently enacted in May 2016. Our study of smoking frequency in the 2017 ITV Love Island series received widespread media attention and was presented to Ofcom in 2017. Subsequently, depiction of smoking was removed from the 2018 series. Britton was listed in Health Services Journal top 100 clinical leaders for innovation in 2015. Coleman received a 5-year Senior Investigator Award from the National Institute for Health Research in 2017.



Smoking is the largest avoidable cause of death and disability in the UK, typically reducing life expectancy by ten years. Our work has contributed to a reduction in UK smoking prevalence from 18% in 2014 to 15% in 2018, equivalent to around 2,000,000 fewer smokers in 2018, larger than in the USA and other rich countries, which has already helped to prevent thousands of premature deaths, cases of morbidity, hundreds of stillbirths and substantial NHS costs. Quitting smoking improves life expectancy, wellbeing, protects children and the unborn child, reduces poverty and improves productivity. We work closely on evidence translation with the Tobacco Advisory Group of the RCP, the National Institute for Health and Care Excellence, where Britton chaired the Public Health Guideline Development Group from 2011 to 2013 and the Public Health Advisory Committee from 2013 to 2016, and was a member of the PH45 Guideline Group and Quality Standards Committee; the Medicines and Healthcare products Regulatory Agency, on e-cigarette safety, where Britton was a member of the expert working group from 2019 to 2020; on policy advocacy with Action on Smoking and Health, where Britton has been a Board member since 2000, and with the National Centre for Smoking Cessation Training, co-producing research-informed clinician guidance and training. Dissemination also occurs through Public Health England via their Tobacco Control Implementation Board, where Britton has been chair since 2014. As described in detail below, and as acknowledged by Public Health England, our research in a range of policy areas has been extremely valuable in demonstrating policy effectiveness and hence ensuring both the adoption and maintenance of new policies. The Tobacco Control Programme Lead for Public Health England stated it is no exaggeration to say that research at the University of Nottingham has had a central role in shaping the tobacco control policy and practice changes promoted by Public Health England. These impacts include the 2017 Tobacco Control Plan for England, which defines policies to reduce smoking prevalence in England and cites two RCP reports as evidence on harms to children, and on the need to help smokers with mental health problems to quit. It recommends screening for smoking in all women attending antenatal care with routine use of carbon monoxide monitors as part of a new NHS care bundle for reducing stillbirth, for which our work was underpinning evidence. It further recommends automatic provision of cessation support for women identified as smokers, requiring women to "opt-out," citing our work in justification. The 2019 NHS Long Term Plan set out commitments for future NHS services and cited the 2018 RCP report, which demonstrated the impact of implementing tobacco dependency treatment in NHS services, with a statement that the recommended model will be adopted by the NHS and thus influence future NHS tobacco dependency treatment. The Tobacco Control Programme Lead at Public Health England states that the work of John Britton, who with other Nottingham colleagues led the publication of a series of systematic reviews and economic models supporting this approach in the 2018 Royal College of Physicians report, led directly to the adoption of this approach in the 2018 NHS Long Term Plan. From 2018, Britton advised senior NHS England staff on the design and costs of service provision. The use of electronic cigarettes to reduce tobacco-related harm through partial or complete substitution for smoking is controversial and has met strong opposition in the UK, USA, Australia, and other countries. However, the UK has adopted the internationally ground-breaking policy of promoting electronic cigarettes as a tobacco harm reduction strategy, a policy substantially underpinned by our research. The Tobacco Control Programme Lead for Public Health England states that the research and evidence synthesis carried out by the Nottingham group has been fundamental to the adoption and endorsement by Public Health England of vaping as a reduced harm substitute for smoking. The 2016 RCP report advocated promotion of electronic cigarettes to replace smoking and was quoted verbatim in the Introduction section of the 2018 House of Commons Science and Technology Committee report on e-cigarettes. The UKCTAS submission to the Committee, led by Britton, which cites our research, commissioned by Public Health England and the RCP, was also quoted verbatim and cited extensively throughout the report, which concluded, as our reports and submissions advocated, that e-cigarette use should be promoted as a harm reduction strategy. The UK is now a world leader in encouraging vaping instead of smoking and has seen greater annual reductions in smoking prevalence than the USA, where vaping is not medically endorsed, or Australia, where vaping nicotine is prohibited. Our 2015 report on indoor air quality for the National Offender Management Service, which was subsequently published as a peer-reviewed article, demonstrated high levels of tobacco smoke exposure in prisons. This led to NOMS commissioning an independent organisation to repeat and corroborate our findings and to cite both sources as evidence to justify English prisons becoming smoke free since March 2016. Legislation prohibiting smoking in cars carrying children was cited by the Department of Health in their consultation on this legislation, which was enacted in 2015. Our review of smoking in pregnancy and stillbirth risk was underpinning evidence for the 2016 introduction of "opt-out" smoking cessation provision as routine within NHS antenatal care as a core element of the "Saving Babies' Lives Care Bundle" for reducing stillbirth. An independent "care bundle" evaluation found that implementation across England resulted in a decrease of 20% in stillbirths per year, approximately 600 fewer, again citing our review. Our review was further cited as core rationale for reducing smoking in pregnancy in version two of the care bundle, designed to reduce perinatal mortality. In 2012, the World Health Organisation (WHO) asked us to share findings from our then unpublished Cochrane systematic review on NRT in pregnancy, which featured prominently in WHO recommendations for the prevention and management of tobacco use and second-hand smoke exposure in pregnancy, concluding there was insufficient evidence for recommending NRT in pregnancy and calling for further research. We updated the review with new trials and showed NRT is effective, published in 2015. The impact of this work continues; our updated review now justifies pregnant women's NRT use in European guidelines as well as those from the United Kingdom, Australia, New Zealand, and Canada. In the UK, in 2015 before the updated review was finalized, influenced by efficacy data from our major NRT trial, NICE proposed an update to UK guidance on NRT in pregnancy; this update is ongoing, due in 2021. Service design and costing proposals for cessation services, summarised in the 2018 RCP report, drawing on our 2013 trial and similar models developed in Ottawa, were cited in the 2018 NHS Long Term Plan. Our work on the provision of cessation services in mental health settings and the March 2013 RCP report "Smoking and mental health" were heavily cited in PHE implementation 2015 guidance for commissioners and providers of mental health services. Our Cochrane review of NRT in pregnancy and NRT trial feature prominently in two evidence reviews for the November 2013 NICE guidance PH48 and justify the recommendation to offer NRT to pregnant women who are unable to stop smoking with the aim of reducing cravings for cigarettes during hospital stays. Our impact on NHS training is detailed in the evidence letter from the Chief Executive of the National Centre for Smoking Cessation Training. Our consensus on how to support pregnant women to stop smoking, produced in the Supporting Smokefree Pregnancies project with the NCSCT, was used to formulate the first ever NHS Standard Treatment Programme for smoking in pregnancy and to update NHS staff training for treating patients according to this. NCSCT used emergent project findings to underpin the STP and changed three NHS online training packages and a face-to-face training course by incorporating key consensus recommendations. Before being permitted to support pregnant smokers, UK health professionals must successfully complete NCSCT training modules or courses and follow the STP; hence, our work substantially underpins current NHS clinical practice. The Chief Executive of the NCSCT states that the SSP project findings have substantially helped the NCSCT to improve the quality of stop smoking support for pregnant women and that the NCSCT guidance for standard treatments and interventions delivered to pregnant women have been significantly improved based on the evidence base developed by the SSP project. The SSP project findings will continue to impact on pregnant women who are helped to stop smoking by the NHS for years to come. Training modules were revised by November 2018; before then, 33,751 NHS professionals had completed these, and by October 2020, 10,208 had completed new versions. Similarly, prior to course curriculum changes around 1,500 health professionals completed interpersonal training courses and between November 2018 and October 2020, 390 have done so. Our work demonstrating the presence of substantial tobacco imagery in television and music videos seen by young people was quoted by Kevin Barron MP as a justification for standardised packaging in a Committee stage debate of the Draft Standardised Packaging of Tobacco Products Regulations 2015, which was subsequently enacted in May 2016. Our study of smoking frequency in the 2017 ITV Love Island series received widespread media attention and was presented to Ofcom in 2017. Subsequently, depiction of smoking was removed from the 2018 series. Britton was listed in Health Services Journal top 100 clinical leaders for innovation in 2015. Coleman received a 5-year Senior Investigator Award from the National Institute for Health Research in 2017.

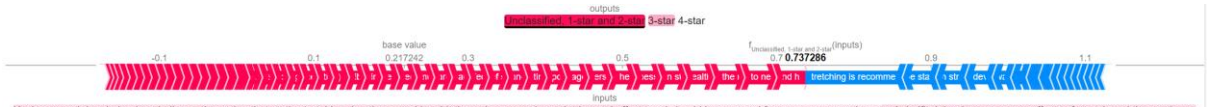
Index 35: actual: 2-star or less predicted:

Etchells' work on video games and screen time has led to a number of opportunities to promote, explain and discuss the underpinning research in a variety of public contexts, for example, through public lectures, writing for popular media, publishing a popular science book, and commenting on television and radio. This has fed directly into discussions that have shifted the narrative and general understanding of the effects that digital technology use can have, away from moral panics about solely detrimental effects and towards a more nuanced understanding of how we can best balance the risks and benefits of technology use, changing the viewpoints of journalists, parents and professional bodies. On the basis of the underpinning research, in 2018/19 Etchells wrote and published a popular science book called "Lost in a Good Game: Why We Play Games and What They Can Do For Us". This was a means to disseminate the complexities of the extant research literature and to respond to the cycle of moral panics about video games and screen time that often persist in the mainstream news media. To date approximately 6,000 copies of the book have been sold in English-speaking countries, including, but not limited to, the UK, USA, Canada, Australia, New Zealand and South Africa, and a Korean translation will follow soon. The book has received widespread positive recommendations in the press, noted as a convincing debut. Those interested in the effects of playing video games will find here much to ponder, with some cheered by a heartfelt defence of a demonised pastime, and others arguing that it is blisteringly relevant, enriching and touching, while issuing a challenge to the bad science surrounding the subject. Further coverage has highlighted that the book provides a more mature way of thinking about video games that is both useful and relevant for concerned members of the public; for example, one outlet argues that Etchells does a great job of reassuring us that, as long as we''re having fun, there isn''t any need to worry. The international press presents similar evaluations highlighting the shift in thinking that the book affords; for example, Etchells is trying to capture the current conversation around video games, while also being quite aware that he has to cater to people who are still having the old conversation; Etchells is a skilled enough writer to speak to both. The publication of "Lost in a Good Game" has led to discussions around the effects of screen time and video games with a number of key stakeholders, including journalists, parents and professional bodies. As a result of successfully communicating both the uncertainty around these effects, as well as highlighting research suggesting minimal risks, these stakeholders have expressed a change in opinion and beliefs about digital technology. For example, in an April 2019 interview in the "Sunday Times Magazine", the interviewing journalist commented that he banned his children from playing games consoles, until a pioneering psychologist changed his mind. While the journalist started from the position of someone who was against allowing children to play video games, Etchells'' explanation of the underpinning research, both through the book and subsequent interviews, made the journalist re-evaluate his views on parenting technology use, and he even went on to purchase a games console for his children. In January 2018, Etchells convened and spoke at a public engagement event funded by the British Academy, which was held at the Wellcome Collection in London and brought together scientists, journalists, clinicians, representatives of professional bodies and civil servants to communicate cutting-edge research into screen time and video games. The editor of "Psychologist" magazine wrote a report explaining how Etchells'' event and his involvement in discussions of the underpinning research allowed him to see screen time differently and develop a more positive perspective on video games and screen use. A key stakeholder in ParentZone, an organisation which provides support and information to parents, children and schools about online safety and video game effects, expressed a positive change of view about video gaming in light of discussions around Etchells'' book and underpinning research. In conclusion, dissemination of the underpinning research has resulted in a shift in the public narrative concerning video games, away from polarised narratives about absolute risks or benefits, towards a more nuanced and accurate discussion of subtler effects. Between 2014 and 2018 Etchells was the science blog network coordinator for "The Guardian", which at its peak had a monthly readership of 1.6 million unique visitors per month. He established and took a lead role in writing for the newspaper''s psychology blog, ''Head Quarters,'' and wrote regularly about both his own research and the general literature on video game and screen time effects. As a result of promoting the underpinning research on an international news platform and becoming a well-known expert in the area, since 2018 he has had the opportunity to provide evidence to parliamentary enquiries concerning the screen use and immersive and/or addictive technologies on health and wellbeing. Evidence submitted on the basis of the underpinning research and written science communication efforts featured in the 2019 House of Commons Science and Technology Committee report on the impact of social media and screen-use on young people''s health, where Etchells is directly named and quoted regarding the complex nature of conducting screen time research. In addition, the report names a 2017 open letter in "The Guardian", organised by Etchells and signed by an international group of scientists, which argues for a nuanced approach to screen time research. Etchells was an interviewee and reviewer for a Parliamentary Office of Science and Technology research briefing on screen use and health in young people, which was published in 2020. Etchells also collaborated with colleagues to submit research-based evidence to the 2019 House of Commons Digital, Culture, Media and Sport Committee report on immersive and addictive technologies. In line with this evidence, Etchells and colleagues recommended that video games companies be required to make aggregate data available to researchers, and to contribute financially to independent research. This was a key highlight of the report, and subsequently video games companies have started to become more engaged with researchers in this area. Etchells has recently joined an industry-academia working group to liaise closely with The Independent Game Developers Association, with the aim of allowing relevant researchers access to aggregated player data. More broadly, the research has made a strong contribution to parliamentary recommendations, which are in the process of driving change concerning how video games are monetised in the context of the Gambling Act 2005. Via online discussions and roundtable events, Etchells continues to be in close discussion with the DCMS to ensure that such changes are evidence-based, appropriate and effective.

Etchells' work on video games and screen time has led to a number of opportunities to promote, explain and discuss the underpinning research in a variety of public contexts, for example, through public lectures, writing for popular media, publishing a popular science book, and commenting on television and radio. This has fed directly into discussions that have shifted the narrative and general understanding of the effects that digital technology use can have, away from moral panics about solely detrimental effects and towards a more nuanced understanding of how we can best balance the risks and benefits of technology use, changing the viewpoints of journalists, parents and professional bodies. On the basis of the underpinning research, in 2018/19 Etchells wrote and published a popular science book called "Lost in a Good Game: Why We Play Games and What They Can Do For Us". This was a means to disseminate the complexities of the extant research literature and to respond to the cycle of moral panics about video games and screen time that often persist in the mainstream news media. To date approximately 6,000 copies of the book have been sold in English-speaking countries, including, but not limited to, the UK, USA, Canada, Australia, New Zealand and South Africa, and a Korean translation will follow soon. The book has received widespread positive recommendations in the press, noted as a convincing debut. Those interested in the effects of playing video games will find here much to ponder, with some cheered by a heartfelt defence of a demonised pastime, and others arguing that it is blisteringly relevant, enriching and touching, while issuing a challenge to the bad science surrounding the subject. Further coverage has highlighted that the book provides a more mature way of thinking about video games that is both useful and relevant for concerned members of the public; for example, one outlet argues that Etchells does a great job of reassuring us that, as long as we''re having fun, there isn''t any need to worry. The international press presents similar evaluations highlighting the shift in thinking that the book affords; for example, Etchells is trying to capture the current conversation around video games, while also being quite aware that he has to cater to people who are still having the old conversation; Etchells is a skilled enough writer to speak to both. The publication of "Lost in a Good Game" has led to discussions around the effects of screen time and video games with a number of key stakeholders, including journalists, parents and professional bodies. As a result of successfully communicating both the uncertainty around these effects, as well as highlighting research suggesting minimal risks, these stakeholders have expressed a change in opinion and beliefs about digital technology. For example, in an April 2019 interview in the "Sunday Times Magazine", the interviewing journalist commented that he banned his children from playing games consoles, until a pioneering psychologist changed his mind. While the journalist started from the position of someone who was against allowing children to play video games, Etchells'' explanation of the underpinning research, both through the book and subsequent interviews, made the journalist re-evaluate his views on parenting technology use, and he even went on to purchase a games console for his children. In January 2018, Etchells convened and spoke at a public engagement event funded by the British Academy, which was held at the Wellcome Collection in London and brought together scientists, journalists, clinicians, representatives of professional bodies and civil servants to communicate cutting-edge research into screen time and video games. The editor of "Psychologist" magazine wrote a report explaining how Etchells'' event and his involvement in discussions of the underpinning research allowed him to see screen time differently and develop a more positive perspective on video games and screen use. A key stakeholder in ParentZone, an organisation which provides support and information to parents, children and schools about online safety and video game effects, expressed a positive change of view about video gaming in light of discussions around Etchells'' book and underpinning research. In conclusion, dissemination of the underpinning research has resulted in a shift in the public narrative concerning video games, away from polarised narratives about absolute risks or benefits, towards a more nuanced and accurate discussion of subtler effects. Between 2014 and 2018 Etchells was the science blog network coordinator for "The Guardian", which at its peak had a monthly readership of 1.6 million unique visitors per month. He established and took a lead role in writing for the newspaper''s psychology blog, ''Head Quarters,'' and wrote regularly about both his own research and the general literature on video game and screen time effects. As a result of promoting the underpinning research on an international news platform and becoming a well-known expert in the area, since 2018 he has had the opportunity to provide evidence to parliamentary enquiries concerning the screen use and immersive and/or addictive technologies on health and wellbeing. Evidence submitted on the basis of the underpinning research and written science communication efforts featured in the 2019 House of Commons Science and Technology Committee report on the impact of social media and screen-use on young people''s health, where Etchells is directly named and quoted regarding the complex nature of conducting screen time research. In addition, the report names a 2017 open letter in "The Guardian", organised by Etchells and signed by an international group of scientists, which argues for a nuanced approach to screen time research. Etchells was an interviewee and reviewer for a Parliamentary Office of Science and Technology research briefing on screen use and health in young people, which was published in 2020. Etchells also collaborated with colleagues to submit research-based evidence to the 2019 House of Commons Digital, Culture, Media and Sport Committee report on immersive and addictive technologies. In line with this evidence, Etchells and colleagues recommended that video games companies be required to make aggregate data available to researchers, and to contribute financially to independent research. This was a key highlight of the report, and subsequently video games companies have started to become more engaged with researchers in this area. Etchells has recently joined an industry-academia working group to liaise closely with The Independent Game Developers Association, with the aim of allowing relevant researchers access to aggregated player data. More broadly, the research has made a strong contribution to parliamentary recommendations, which are in the process of driving change concerning how video games are monetised in the context of the Gambling Act 2005. Via online discussions and roundtable events, Etchells continues to be in close discussion with the DCMS to ensure that such changes are evidence-based, appropriate and effective.

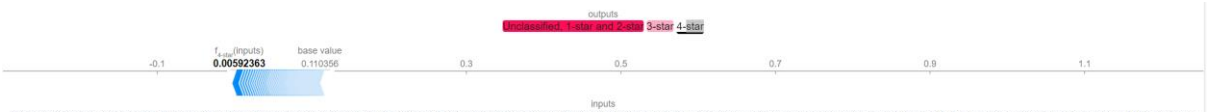
Etchells' work on video games and screen time has led to a number of opportunities to promote, explain and discuss the underpinning research in a variety of public contexts, for example, through public lectures, writing for popular media, publishing a popular science book, and commenting on television and radio. This has fed directly into discussions that have shifted the narrative and general understanding of the effects that digital technology use can have, away from moral panics about solely detrimental effects and towards a more nuanced understanding of how we can best balance the risks and benefits of technology use, changing the viewpoints of journalists, parents and professional bodies. On the basis of the underpinning research, in 2018/19 Etchells wrote and published a popular science book called "Lost in a Good Game: Why We Play Games and What They Can Do For Us". This was a means to disseminate the complexities of the extant research literature and to respond to the cycle of moral panics about video games and screen time that often persist in the mainstream news media. To date approximately 6,000 copies of the book have been sold in English-speaking countries, including, but not limited to, the UK, USA, Canada, Australia, New Zealand and South Africa, and a Korean translation will follow soon. The book has received widespread positive recommendations in the press, noted as a convincing debut. Those interested in the effects of playing video games will find here much to ponder, with some cheered by a heartfelt defence of a demonised pastime, and others arguing that it is blisteringly relevant, enriching and touching, while issuing a challenge to the bad science surrounding the subject. Further coverage has highlighted that the book provides a more mature way of thinking about video games that is both useful and relevant for concerned members of the public; for example, one outlet argues that Etchells does a great job of reassuring us that, as long as we''re having fun, there isn''t any need to worry. The international press presents similar evaluations highlighting the shift in thinking that the book affords; for example, Etchells is trying to capture the current conversation around video games, while also being quite aware that he has to cater to people who are still having the old conversation; Etchells is a skilled enough writer to speak to both. The publication of "Lost in a Good Game" has led to discussions around the effects of screen time and video games with a number of key stakeholders, including journalists, parents and professional bodies. As a result of successfully communicating both the uncertainty around these effects, as well as highlighting research suggesting minimal risks, these stakeholders have expressed a change in opinion and beliefs about digital technology. For example, in an April 2019 interview in the "Sunday Times Magazine", the interviewing journalist commented that he banned his children from playing games consoles, until a pioneering psychologist changed his mind. While the journalist started from the position of someone who was against allowing children to play video games, Etchells'' explanation of the underpinning research, both through the book and subsequent interviews, made the journalist re-evaluate his views on parenting technology use, and he even went on to purchase a games console for his children. In January 2018, Etchells convened and spoke at a public engagement event funded by the British Academy, which was held at the Wellcome Collection in London and brought together scientists, journalists, clinicians, representatives of professional bodies and civil servants to communicate cutting-edge research into screen time and video games. The editor of "Psychologist" magazine wrote a report explaining how Etchells'' event and his involvement in discussions of the underpinning research allowed him to see screen time differently and develop a more positive perspective on video games and screen use. A key stakeholder in ParentZone, an organisation which provides support and information to parents, children and schools about online safety and video game effects, expressed a positive change of view about video gaming in light of discussions around Etchells'' book and underpinning research. In conclusion, dissemination of the underpinning research has resulted in a shift in the public narrative concerning video games, away from polarised narratives about absolute risks or benefits, towards a more nuanced and accurate discussion of subtler effects. Between 2014 and 2018 Etchells was the science blog network coordinator for "The Guardian", which at its peak had a monthly readership of 1.6 million unique visitors per month. He established and took a lead role in writing for the newspaper''s psychology blog, ''Head Quarters,'' and wrote regularly about both his own research and the general literature on video game and screen time effects. As a result of promoting the underpinning research on an international news platform and becoming a well-known expert in the area, since 2018 he has had the opportunity to provide evidence to parliamentary enquiries concerning the screen use and immersive and/or addictive technologies on health and wellbeing. Evidence submitted on the basis of the underpinning research and written science communication efforts featured in the 2019 House of Commons Science and Technology Committee report on the impact of social media and screen-use on young people''s health, where Etchells is directly named and quoted regarding the complex nature of conducting screen time research. In addition, the report names a 2017 open letter in "The Guardian", organised by Etchells and signed by an international group of scientists, which argues for a nuanced approach to screen time research. Etchells was an interviewee and reviewer for a Parliamentary Office of Science and Technology research briefing on screen use and health in young people, which was published in 2020. Etchells also collaborated with colleagues to submit research-based evidence to the 2019 House of Commons Digital, Culture, Media and Sport Committee report on immersive and addictive technologies. In line with this evidence, Etchells and colleagues recommended that video games companies be required to make aggregate data available to researchers, and to contribute financially to independent research. This was a key highlight of the report, and subsequently video games companies have started to become more engaged with researchers in this area. Etchells has recently joined an industry-academia working group to liaise closely with The Independent Game Developers Association, with the aim of allowing relevant researchers access to aggregated player data. More broadly, the research has made a strong contribution to parliamentary recommendations, which are in the process of driving change concerning how video games are monetised in the context of the Gambling Act 2005. Via online discussions and roundtable events, Etchells continues to be in close discussion with the DCMS to ensure that such changes are evidence-based, appropriate and effective.

Index 136: actual: 2-star or less predicted:



Kay's research has helped to challenge the notion that static stretching durations used in athletic environments have detrimental effects and should be removed from warm-up practices, and clarified the dose-response effects of stretch and the maximum safe duration for stretching during warm-up. His research has also put to rest one of the most contentious issues and confirmed that static muscle stretching exercises can reduce the risk of muscle strain injury, the most common injury in sport. The research findings have been disseminated among key stakeholders leading to a change in both guidance and practice on the effects of stretch. These findings have also been propagated through international media to increase public awareness of the benefits of this research. The dominant media narrative until recently has focused on studies reporting the potentially detrimental effects of static stretching before exercise including negative effects on performance and no injury preventative benefit. Four of Kay's studies and his Position Stand have challenged this narrative and having been widely cited in the international media, changing media coverage of stretching practices in exercise, and increasing public understanding of the benefits of stretching and the different types of stretching. Kay's systematic review of the acute effects of several types of muscle stretching techniques on physical performance, flexibility, and injury prevention has been cited in 35 news stories from 28 international news outlets including the New York Times and The Guardian. Kay's research on PNF stretching was highlighted in Shape Magazine, bringing attention to the benefits of this type of stretching, something that was not widely known by the general public. His subsequent research clarifies that there is no negative effect of several types of muscle stretching, when used within a full, dynamic warm-up on athletic performance. This has also been widely cited by 11 international news sources including the New York Times and ABC Life. For a more comprehensive list of media coverage see. Kay clarified the misconceptions reported in the media about the negative effect of muscle stretching on muscle strength and further increased public awareness of ways to improve muscle strength through a study he conducted with the BBC 2's Trust Me I'm a Doctor programme based on his research, which had viewing figures of 2,300,000. The study looked at the possibility of increasing muscle strength simply by thinking about exercise (something called 'motor imagery'). The study directly impacted participants, who improved their muscle strength on average by 8% (with one participant increasing their strength by 33.9%). It introduced viewers to new scientific research on stretching and muscle strength in an accessible way. Kay produced two 'Think Yourself Stronger' audio-guides to complement the programme, which focused on lower and upper body strength. These are available to download free, increasing the public''s access to these techniques. Kay''s findings on the limited detrimental effects of static stretching for both athletic and general populations have helped clarify common misconceptions that were reported to the general public through the wider media. Prior to Kay''s research, some governing bodies, such as The American College of Sports Medicine, recommended the removal of static stretching in warm-up routines due to the belief that it would cause performance impairments in strength, speed, and power-related activities, and no injury preventative benefit. Kay''s research has been instrumental in identifying the limitations within the research. His systematic and meta-analytical reviews have clarified the actual effects of stretch within a warm-up. His research has corrected previous misconceptions, provided evidence-based guidance to health bodies and has clarified for governing bodies and athletes misleading information that had previously been inconsistent. This informed new Position Stands and public health guidance on previously contentious subjects. England's National Health Service (NHS) approached Kay to help develop, write, and edit their NHS Choices health advice section on stretching and flexibility. This guidance quotes Kay, "It is likely that durations of stretch used in the warm-up routines of most recreational exercisers produce negligible and transient reductions in strength." It further cites Kay''s research reporting "the reduction in performance from pre-exercise stretching has been overstated." This was the first guidance the NHS released on stretching and flexibility, and Kay''s research was key to the development of their recommendations. According to a Freedom of Information Act request, the page has been viewed 62,061 times since November 2018. The Canadian Society for Exercise Physiology (CSEP), Canada's leading authority in exercise science, focused on promoting research and evidence-based fitness, performance and health outcomes to all Canadians, approved and directly cites Kay''s research in their first position stand on the topic that was co-authored by Kay. The CSEP Chair states, "The recommendation in the CSEP Position Stand is that all components of a warm-up be included with appropriate duration of stretching. The inclusion of static, or Proprioceptive Neuromuscular Facilitation (PNF), stretching is recommended and has the potential to positively influence the standard warm-up routines of a large number of athletes." This position stand has used Kay''s extensive research to enable a change of position stand of CESP, which has informed the National Academy of Sports Medicine (NASM) in providing athletic guidance on how and why to stretch from the ACSM (American College of Sports Medicine). The National Academy of Sports Medicine (NASM) in the United States, a leading authority in personal training certification worldwide, directly cites Kay''s research in regard to best practices for stretching. Kay''s research is used to inform the narrative including "static stretches of <45 seconds can be used in a pre-exercise protocol without significant decreases in strength, power, or speed-dependent task performances" and that "static stretching <60 seconds is considered an effective method for increasing joint range of motion (ROM), and is often thought to improve performance and reduce the incidence of activity-related injuries." NASM concludes that "Static stretching can be beneficial in many ways, such as correcting muscle imbalances, decreasing muscle hypertonicity, increasing joint ROM, relieving joint stress, improving the extensibility of the musculotendinous junction, maintaining the normal functional length of a muscle (length-tension relationships), decreasing the chance of injury and in turn enhancing power and strength." These recommendations use Kay''s research to provide an evidence-based exercise prescription guide to the use of stretching within a warm-up.

Kay's research has helped to challenge the notion that static stretching durations used in athletic environments have detrimental effects and should be removed from warm-up practices, and clarified the dose-response effects of stretch and the maximum safe duration for stretching during warm-up. His research has also put to rest one of the most contentious issues and confirmed that static muscle stretching exercises can reduce the risk of muscle strain injury, the most common injury in sport. The research findings have been disseminated among key stakeholders leading to a change in both guidance and practice on the effects of stretch. These findings have also been propagated through international media to increase public awareness of the benefits of this research. The dominant media narrative until recently has focused on studies reporting the potentially detrimental effects of static stretching before exercise including negative effects on performance and no injury preventative benefit. Four of Kay's studies and his Position Stand have challenged this narrative and having been widely cited in the international media, changing media coverage of stretching practices in exercise, and increasing public understanding of the benefits of stretching and the different types of stretching. Kay's systematic review of the acute effects of several types of muscle stretching techniques on physical performance, flexibility, and injury prevention has been cited in 35 news stories from 28 international news outlets including the New York Times and The Guardian. Kay's research on PNF stretching was highlighted in Shape Magazine, bringing attention to the benefits of this type of stretching, something that was not widely known by the general public. His subsequent research clarifies that there is no negative effect of several types of muscle stretching, when used within a full, dynamic warm-up on athletic performance. This has also been widely cited by 11 international news sources including the New York Times and ABC Life. For a more comprehensive list of media coverage see. Kay clarified the misconceptions reported in the media about the negative effect of muscle stretching on muscle strength and further increased public awareness of ways to improve muscle strength through a study he conducted with the BBC 2's Trust Me I'm a Doctor programme based on his research, which had viewing figures of 2,300,000. The study looked at the possibility of increasing muscle strength simply by thinking about exercise (something called 'motor imagery'). The study directly impacted participants, who improved their muscle strength on average by 8% (with one participant increasing their strength by 33.9%). It introduced viewers to new scientific research on stretching and muscle strength in an accessible way. Kay produced two 'Think Yourself Stronger' audio-guides to complement the programme, which focused on lower and upper body strength. These are available to download free, increasing the public''s access to these techniques. Kay''s findings on the limited detrimental effects of static stretching for both athletic and general populations have helped clarify common misconceptions that were reported to the general public through the wider media. Prior to Kay''s research, some governing bodies, such as The American College of Sports Medicine, recommended the removal of static stretching in warm-up routines due to the belief that it would cause performance impairments in strength, speed, and power-related activities, and no injury preventative benefit. Kay''s research has been instrumental in identifying the limitations within the research. His systematic and meta-analytical reviews have clarified the actual effects of stretch within a warm-up. His research has corrected previous misconceptions, provided evidence-based guidance to health bodies and has clarified for governing bodies and athletes misleading information that had previously been inconsistent. This informed new Position Stands and public health guidance on previously contentious subjects. England's National Health Service (NHS) approached Kay to help develop, write, and edit their NHS Choices health advice section on stretching and flexibility. This guidance quotes Kay, "It is likely that durations of stretch used in the warm-up routines of most recreational exercisers produce negligible and transient reductions in strength." It further cites Kay''s research reporting "the reduction in performance from pre-exercise stretching has been overstated." This was the first guidance the NHS released on stretching and flexibility, and Kay''s research was key to the development of their recommendations. According to a Freedom of Information Act request, the page has been viewed 62,061 times since November 2018. The Canadian Society for Exercise Physiology (CSEP), Canada's leading authority in exercise science, focused on promoting research and evidence-based fitness, performance and health outcomes to all Canadians, approved and directly cites Kay''s research in their first position stand on the topic that was co-authored by Kay. The CSEP Chair states, "The recommendation in the CSEP Position Stand is that all components of a warm-up be included with appropriate duration of stretching. The inclusion of static, or Proprioceptive Neuromuscular Facilitation (PNF), stretching is recommended and has the potential to positively influence the standard warm-up routines of a large number of athletes." This position stand has used Kay''s extensive research to enable a change of position stand of CESP, which has informed the National Academy of Sports Medicine (NASM) in providing athletic guidance on how and why to stretch from the ACSM (American College of Sports Medicine). The National Academy of Sports Medicine (NASM) in the United States, a leading authority in personal training certification worldwide, directly cites Kay''s research in regard to best practices for stretching. Kay''s research is used to inform the narrative including "static stretches of <45 seconds can be used in a pre-exercise protocol without significant decreases in strength, power, or speed-dependent task performances" and that "static stretching <60 seconds is considered an effective method for increasing joint range of motion (ROM), and is often thought to improve performance and reduce the incidence of activity-related injuries." NASM concludes that "Static stretching can be beneficial in many ways, such as correcting muscle imbalances, decreasing muscle hypertonicity, increasing joint ROM, relieving joint stress, improving the extensibility of the musculotendinous junction, maintaining the normal functional length of a muscle (length-tension relationships), decreasing the chance of injury and in turn enhancing power and strength." These recommendations use Kay''s research to provide an evidence-based exercise prescription guide to the use of stretching within a warm-up.

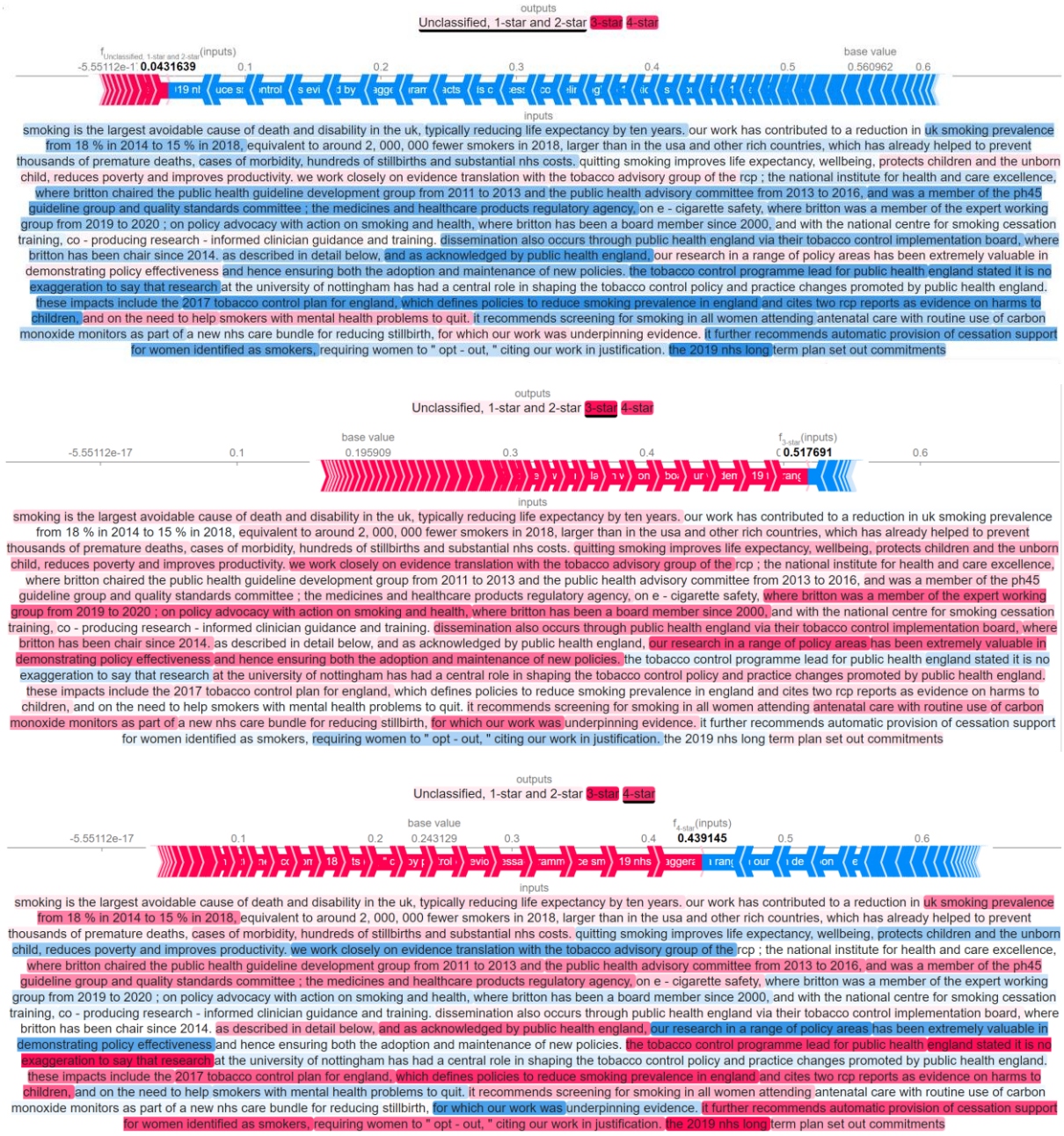


Kay's research has helped to challenge the notion that static stretching durations used in athletic environments have detrimental effects and should be removed from warm-up practices, and clarified the dose-response effects of stretch and the maximum safe duration for stretching during warm-up. His research has also put to rest one of the most contentious issues and confirmed that static muscle stretching exercises can reduce the risk of muscle strain injury, the most common injury in sport. The research findings have been disseminated among key stakeholders leading to a change in both guidance and practice on the effects of stretch. These findings have also been propagated through international media to increase public awareness of the benefits of this research. The dominant media narrative until recently has focused on studies reporting the potentially detrimental effects of static stretching before exercise including negative effects on performance and no injury preventative benefit. Four of Kay's studies and his Position Stand have challenged this narrative and having been widely cited in the international media, changing media coverage of stretching practices in exercise, and increasing public understanding of the benefits of stretching and the different types of stretching. Kay's systematic review of the acute effects of several types of muscle stretching techniques on physical performance, flexibility, and injury prevention has been cited in 35 news stories from 28 international news outlets including the New York Times and The Guardian. Kay's research on PNF stretching was highlighted in Shape Magazine, bringing attention to the benefits of this type of stretching, something that was not widely known by the general public. His subsequent research clarifies that there is no negative effect of several types of muscle stretching, when used within a full, dynamic warm-up on athletic performance. This has also been widely cited by 11 international news sources including the New York Times and ABC Life. For a more comprehensive list of media coverage see. Kay clarified the misconceptions reported in the media about the negative effect of muscle stretching on muscle strength and further increased public awareness of ways to improve muscle strength through a study he conducted with the BBC 2's Trust Me I'm a Doctor programme based on his research, which had viewing figures of 2,300,000. The study looked at the possibility of increasing muscle strength simply by thinking about exercise (something called 'motor imagery'). The study directly impacted participants, who improved their muscle strength on average by 8% (with one participant increasing their strength by 33.9%). It introduced viewers to new scientific research on stretching and muscle strength in an accessible way. Kay produced two 'Think Yourself Stronger' audio-guides to complement the programme, which focused on lower and upper body strength. These are available to download free, increasing the public''s access to these techniques. Kay''s findings on the limited detrimental effects of static stretching for both athletic and general populations have helped clarify common misconceptions that were reported to the general public through the wider media. Prior to Kay''s research, some governing bodies, such as The American College of Sports Medicine, recommended the removal of static stretching in warm-up routines due to the belief that it would cause performance impairments in strength, speed, and power-related activities, and no injury preventative benefit. Kay''s research has been instrumental in identifying the limitations within the research. His systematic and meta-analytical reviews have clarified the actual effects of stretch within a warm-up. His research has corrected previous misconceptions, provided evidence-based guidance to health bodies and has clarified for governing bodies and athletes misleading information that had previously been inconsistent. This informed new Position Stands and public health guidance on previously contentious subjects. England's National Health Service (NHS) approached Kay to help develop, write, and edit their NHS Choices health advice section on stretching and flexibility. This guidance quotes Kay, "It is likely that durations of stretch used in the warm-up routines of most recreational exercisers produce negligible and transient reductions in strength." It further cites Kay''s research reporting "the reduction in performance from pre-exercise stretching has been overstated." This was the first guidance the NHS released on stretching and flexibility, and Kay''s research was key to the development of their recommendations. According to a Freedom of Information Act request, the page has been viewed 62,061 times since November 2018. The Canadian Society for Exercise Physiology (CSEP), Canada's leading authority in exercise science, focused on promoting research and evidence-based fitness, performance and health outcomes to all Canadians, approved and directly cites Kay''s research in their first position stand on the topic that was co-authored by Kay. The CSEP Chair states, "The recommendation in the CSEP Position Stand is that all components of a warm-up be included with appropriate duration of stretching. The inclusion of static, or Proprioceptive Neuromuscular Facilitation (PNF), stretching is recommended and has the potential to positively influence the standard warm-up routines of a large number of athletes." This position stand has used Kay''s extensive research to enable a change of position stand of CESP, which has informed the National Academy of Sports Medicine (NASM) in providing athletic guidance on how and why to stretch from the ACSM (American College of Sports Medicine). The National Academy of Sports Medicine (NASM) in the United States, a leading authority in personal training certification worldwide, directly cites Kay''s research in regard to best practices for stretching. Kay''s research is used to inform the narrative including "static stretches of <45 seconds can be used in a pre-exercise protocol without significant decreases in strength, power, or speed-dependent task performances" and that "static stretching <60 seconds is considered an effective method for increasing joint range of motion (ROM), and is often thought to improve performance and reduce the incidence of activity-related injuries." NASM concludes that "Static stretching can be beneficial in many ways, such as correcting muscle imbalances, decreasing muscle hypertonicity, increasing joint ROM, relieving joint stress, improving the extensibility of the musculotendinous junction, maintaining the normal functional length of a muscle (length-tension relationships), decreasing the chance of injury and in turn enhancing power and strength." These recommendations use Kay''s research to provide an evidence-based exercise prescription guide to the use of stretching within a warm-up.

Appendix 7: BERT Base Variant for impact case studies at index 21, 32, 35, and 136

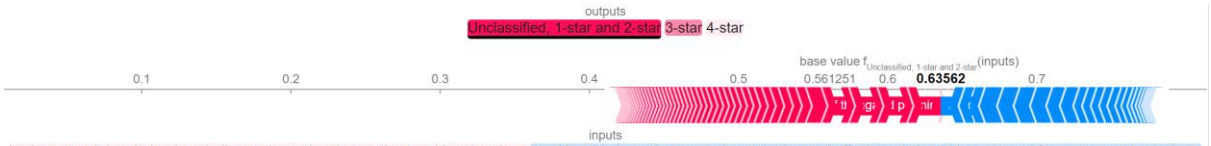
Index 21: actual: 4-star predicted: 3-star

Index 32: actual: 4-star predicted: 4-star



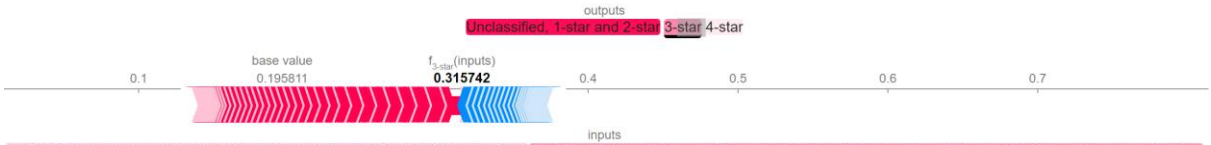
Index 35: actual: 2-star or less predicted: 2-star or less

Index 136: actual: 2-star or less predicted: 2 star or less



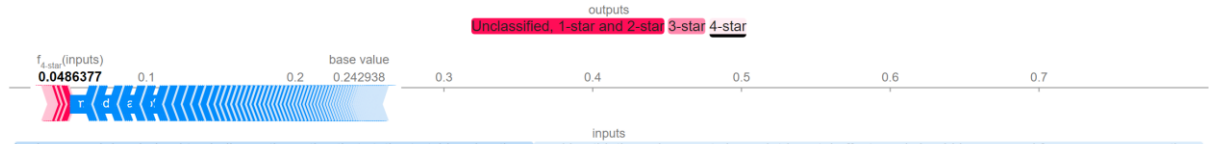
inputs

kay's research has helped to challenge the notion that static stretching durations used in athletic environments have detrimental effects and should be removed from warm - up practices, and clarified the dose - response effects of stretch and the maximum safe duration for stretching during warm - up. his research has also put to rest one of the most contentious issues and confirmed that static muscle stretching exercises can reduce the risk of muscle strain injury, the most common injury in sport. the research findings have been disseminated among key stakeholders leading to a change in both guidance and practice on the effects of stretch. these findings have also been propagated through international media to increase public awareness of the benefits of this research. the dominant media narrative until recently has focused on studies reporting the potentially detrimental effects of static stretching before exercise including negative effects on performance and no injury preventative benefit. four of kay's studies and his position stand have challenged this narrative and having been widely cited in the international media, changing media coverage of stretching practices in exercise, and increasing public understanding of the benefits of stretching and the different types of stretching. kay's systematic review of the acute effects of several types of muscle stretching techniques on physical performance, flexibility, and injury prevention has been cited in 35 news stories from 28 international news outlets including the new york times and the guardian. kay's research on pnf stretching was highlighted in shape magazine, bringing attention to the benefits of this type of stretching, something that was not widely known by the general public. his subsequent research clarifies that there is no negative effect of several types of muscle stretching, when used within a full, dynamic warm - up on athletic performance. this has also been widely cited by 11 international news sources including the new york times and abc life. for a more comprehensive list of media coverage see. kay clarified the misconceptions reported in the media about the negative effect of muscle stretching on muscle strength and further increased public awareness of ways to improve muscle strength through a study he conducted with the bbc 2's trust me i'm a doctor programme based on his research, which had viewing figures of 2, 300, 000. the study looked at the possibility of increasing muscle strength simply by thinking about exercise (something called'motor imagery'). the study directly impacted participants, who improved their muscle strength on average by 8 % (with one participant increasing their strength by 33. 9 %). it introduced viewers to new scientific research on stretching and muscle strength in an accessible way. kay produced two'think yourself stronger '



inputs

kay's research has helped to challenge the notion that static stretching durations used in athletic environments have detrimental effects and should be removed from warm - up practices, and clarified the dose - response effects of stretch and the maximum safe duration for stretching during warm - up. his research has also put to rest one of the most contentious issues and confirmed that static muscle stretching exercises can reduce the risk of muscle strain injury, the most common injury in sport. the research findings have been disseminated among key stakeholders leading to a change in both guidance and practice on the effects of stretch. these findings have also been propagated through international media to increase public awareness of the benefits of this research. the dominant media narrative until recently has focused on studies reporting the potentially detrimental effects of static stretching before exercise including negative effects on performance and no injury preventative benefit. four of kay's studies and his position stand have challenged this narrative and having been widely cited in the international media, changing media coverage of stretching practices in exercise, and increasing public understanding of the benefits of stretching and the different types of stretching. kay's systematic review of the acute effects of several types of muscle stretching techniques on physical performance, flexibility, and injury prevention has been cited in 35 news stories from 28 international news outlets including the new york times and the guardian. kay's research on pnf stretching was highlighted in shape magazine, bringing attention to the benefits of this type of stretching, something that was not widely known by the general public. his subsequent research clarifies that there is no negative effect of several types of muscle stretching, when used within a full, dynamic warm - up on athletic performance. this has also been widely cited by 11 international news sources including the new york times and abc life. for a more comprehensive list of media coverage see. kay clarified the misconceptions reported in the media about the negative effect of muscle stretching on muscle strength and further increased public awareness of ways to improve muscle strength through a study he conducted with the bbc 2's trust me i'm a doctor programme based on his research, which had viewing figures of 2, 300, 000. the study looked at the possibility of increasing muscle strength simply by thinking about exercise (something called'motor imagery'). the study directly impacted participants, who improved their muscle strength on average by 8 % (with one participant increasing their strength by 33. 9 %). it introduced viewers to new scientific research on stretching and muscle strength in an accessible way. kay produced two'think yourself stronger '



inputs

kay's research has helped to challenge the notion that static stretching durations used in athletic environments have detrimental effects and should be removed from warm - up practices, and clarified the dose - response effects of stretch and the maximum safe duration for stretching during warm - up. his research has also put to rest one of the most contentious issues and confirmed that static muscle stretching exercises can reduce the risk of muscle strain injury, the most common injury in sport. the research findings have been disseminated among key stakeholders leading to a change in both guidance and practice on the effects of stretch. these findings have also been propagated through international media to increase public awareness of the benefits of this research. the dominant media narrative until recently has focused on studies reporting the potentially detrimental effects of static stretching before exercise including negative effects on performance and no injury preventative benefit. four of kay's studies and his position stand have challenged this narrative and having been widely cited in the international media, changing media coverage of stretching practices in exercise, and increasing public understanding of the benefits of stretching and the different types of stretching. kay's systematic review of the acute effects of several types of muscle stretching techniques on physical performance, flexibility, and injury prevention has been cited in 35 news stories from 28 international news outlets including the new york times and the guardian. kay's research on pnf stretching was highlighted in shape magazine, bringing attention to the benefits of this type of stretching, something that was not widely known by the general public. his subsequent research clarifies that there is no negative effect of several types of muscle stretching, when used within a full, dynamic warm - up on athletic performance. this has also been widely cited by 11 international news sources including the new york times and abc life. for a more comprehensive list of media coverage see. kay clarified the misconceptions reported in the media about the negative effect of muscle stretching on muscle strength and further increased public awareness of ways to improve muscle strength through a study he conducted with the bbc 2's trust me i'm a doctor programme based on his research, which had viewing figures of 2, 300, 000. the study looked at the possibility of increasing muscle strength simply by thinking about exercise (something called'motor imagery'). the study directly impacted participants, who improved their muscle strength on average by 8 % (with one participant increasing their strength by 33. 9 %). it introduced viewers to new scientific research on stretching and muscle strength in an accessible way. kay produced two'think yourself stronger '