The implementation of a sentence transformer to classify UKRI awards

UK Research and Innovation

June 2024

Introduction

UK Research and Innovation (UKRI) is introducing sentence transformer methodology to help classify awards by theme, enabling targeted searches within extensive datasets.

In 2022-23, UKRI awarded 6,118 research and innovation applications totalling £3.1bn.[1] In recent years, UKRI has made significant steps to publish data to improve public understanding of its investments. For example, award level data published in Gateway to Research shows the breadth of activities covered by UKRI funding and enables users to access a range of information about individual awards and their outcomes. Gateway to Research also provides award level classifications based on funding source or programme.

However, increasingly, UKRI is being asked by policy makers, researchers and the wider public for awards to be identified and grouped by theme. Given the size of UKRI's portfolio, servicing requests for such thematic classification places considerable demands on resources. It requires prioritisation, a range of skills to identify relevant keywords and craft multiple search syntaxes to meet criteria, followed by manual review of a significant number of awards by subject matter experts to refine a final dataset.

It is, therefore, UKRI's ambition to introduce a tool that enables rapid and cost-effective thematic classification. The work developed in this paper sets out a generic application of a sentence transformer model, which processes and maps text contained in UKRI award title and the summary descriptions.

It is designed for specific use cases (see the Use cases section) and is recommended only when there is sufficient understanding of the technique. The approach was selected for its performance at semantic similarity, clustering and classification. Unlike traditional word embeddings, such as Word2Vec[2] or GloVe[3], which primarily focus on individual word meanings, sentence transformers encode entire phrases and sentences, allowing for more nuanced understanding of text.[4] Moreover, the model's proficiency in clustering and classification tasks stems from its use of BERT (Bidirectional Encoder

---

[1] UKRI Investment and Outputs Data 2022-23

[2] T Mikolov, K Chen, G Corrado, J Dean. Efficient Estimation of Word Representations in Vector Space. arXiv preprint arXiv:1301.3781, 2013. https://doi.org/10.48550/arXiv.1301.3781

[3] J Pennington, R Socher, CD Manning. GloVe: Global Vectors for Word Representation. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2014. https://nlp.stanford.edu/pubs/glove.pdf

[4] J Devlin, MW Chang, K Lee, K Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805, 2018. https://doi.org/10.48550/arXiv.1810.04805

Representations from Transformers) as a backbone, which has set new standards in the field of natural language processing.[5]

Methodology

Description

Identifying which awards fall under a particular theme – thematic classification – is achieved by following the methodological steps below:

1. Selection of a sentence transformer model

From the many available implementations of sentence transformer models, the llmrails/ember-v1 was selected because of its current inclusion in the top 10 performing classification models on the HuggingFace 'Massive Text Embedding Benchmark' (MTEB) Leaderboard[6], as well as its additional training on a broad spectrum of domains, including science and medicine. The model maps text (sentences and paragraphs) to a multi-dimensional dense vector space using a tokenisation[7] and can be used for tasks, such as clustering or semantic search. This is ideal for the type of data that describes UKRI awards: title and abstracts. This model can be found at the Hugging Face repository.[8]

2. Creation of unique digital signatures for all awards

The second step is to apply the llmrails/ember-v1 model across the database containing all award titles, abstracts and expected impact. This ensures that awards are mapped to a multi-dimensional dense vector space to create a unique digital signature for each award.

However, the model has a maximum sequence length of 512 tokens, which is currently exceeded by many awards. For example, between 2020-21 and 2022-2023 there were 19,342 (67%) awards with text exceeding that limit.

---

[5] A Vaswani, N Shazeer, N Parmar, J Uszkoreit, L Jones, AN Gomez, Ł Kaiser, I Polosukhin. Attention is all you need. Advances in neural information processing systems, 2017. https://doi.org/10.48550/arXiv.1706.03762

[6] See https://huggingface.co/spaces/mteb/leaderboard for more details.

[7] Tokenization involves breaking down a piece of text into smaller units, typically words or subwords, to facilitate analysis. Each unit is referred to as a 'token.' For more in-depth information on tokenization, please see https://huggingface.co/docs/transformers/tokenizer_summary

[8] See https://huggingface.co/llmrails/ember-v1 for more details.

Chunking into sets of 512 tokens enables all the information from an award title, abstract, technical abstract and expected impact to be included. The award's unique digital signature is the average of these constituent chunks. There is an additional challenge where small chunks, or those with limited information, contribute as much to an award's unique digital signature as large or information-heavy chunks. When this situation occurs, omitting any constituent chunk with fewer than 100 tokens ensures a more even distribution of information and allows for text to straddle multiple chunks. The decision to keep the minimum size of a constituent chunk at 100 tokens is tested in Table 1.

This step needs to be run only once for the data held. However, it will need to be updated if underlying award titles and abstracts change or new awards are added. There is also a possibility that a new version of the model is released and, if so, then the step must be run again for the full database.

3. Keyword list

The third step is to determine a list of keywords used to identify a particular theme. Ideally, this would begin with a few terms that are descriptive of the theme. It is likely that subject matter experts will be needed to expand an initial set of terms to an agreed final set of terms that better describes the theme.

Once the keyword list is determined, it is put through the llmrails/ember-v1 model as text, separating each keyword/terms by a comma, to gain a unique digital signature for the keyword list. Since digital signatures are dependent on the position of the words, the order of the keywords will impact the results (the impact of the order of keywords is tested in Table 1). However, it is recommended that efforts are focused on identifying the right keywords for the theme rather than their order. Additionally, the order is relevant for reproducibility – the same order will allow a repeat of analysis so that results can be verified.

4. Calculation of metric of similarity

Once the unique digital signatures for all awards and the keyword list have been obtained, the similarities can be determined by calculating the cosine similarities scores between the keyword list and the awards. This method was selected as it is widely used in natural language processing techniques. Its implementation in Python for this project is publicly available[9].

[9] See https://github.com/ukri-analysis/portfolio-classification for the full GitHub repository.

The output is a dataset of cosine similarity scores for each award in relation to the keyword list with scores ranging between -1 and 1, where 1 is perfect similarity and similarity scores moving away from 1 represent awards less similar to the keyword list.[10]

5. Threshold determination

To select the threshold for the cosine similarity scores to decide which awards are relevant to a theme or not, a dual approach is employed that combines stratified sampling and blind manual labelling by one or more subject matter experts.

First, the full range of cosine similarity scores is divided into 10 equal intervals where a random sample of three awards is taken from each interval (total sample size 30). This sample of awards is then labelled by the subject matter experts according to whether they fall under the theme or not. This is then regressed on the cosine similarity scores using a logistic regression. The outcome of the regression is a probability that each award falls under the theme. To determine the overall cosine similarity scores threshold, we use the following equation:

$$threshold_1 = -\frac{\beta_0}{\beta_1}$$

where β_0 and β_1 are the constant and the cosine similarity scores coefficient respectively from the logistic regression (with a probability cut-off point of 0.5).

Using $threshold_1$ we establish a new narrower range, $[threshold_1 - 0.05 \; to \; threshold_1 + 0.05]$ of cosine similarity scores. This additional range is also divided into 10 equal intervals, and a new random sample of three awards from each interval is taken (total sample size of 30).

This new sample is then labelled by the subject matter experts and regressed on the cosine similarity scores using a logistic regression in the same fashion as $threshold_{\,1}$. If any of the thresholds are outside the sample range, this is a likely indicator that a keyword list does not reflect the subject matter experts' understanding of the theme. In this instance, any misalignments must be resolved before rerunning the methodology; either adjusting the keyword list, or ensuring that the person labelling the data has agreed that the keyword list represents the theme.

The outcome of this second logistic regression is $threshold_{final}$. The impact of the selection of the sample size and the width of the interval into second step (0.1) is tested in the section below.

[10] See https://en.wikipedia.org/wiki/Cosine_similarity for more details.

6. Uncertainty

As discussed above, determining whether an award falls under a specific theme is not exact, so the final step aims at estimating the uncertainty of the threshold determined in the previous step. To estimate this uncertainty, we have opted for testing the impact of five different assumptions and subjective decisions set out by this methodology (Table 1).

Table 1: List of assumptions and subjective decisions tested and its estimated impact on the final threshold

Assumption/ Decision	Description	Test	Estimated impact on the $threshold_{final}$
Award text length	Testing the impact of changing the chunking method	Decrease the minimum chunk from 100 to 50	An increase in the $threshold_{final}$ by 0.5%
List of keywords	Testing the impact of a shorter keyword list	Removing a random number of keywords from a list (around 25% of the words)	A decrease in the $threshold_{final}$ by 1%
Order of keywords	Testing the impact of changing the order of the keyword list	Randomly change the order of the keyword list	A decrease in the $threshold_{final}$ by 0.5%
Boundary determination	Testing the impact of the range of 0.1: $[threshold_1 - 0.05 \text{ to } threshold_1 + 0.05]$	Narrowing the range to 0.05: $[threshold_1 - 0.025 \text{ to } threshold_1 + 0.025]$	An increase in the $threshold_{final}$ by 0.5%
Sample size	Size of the sample to be manually evaluated by subject matter experts	Tested sample sizes of 60 in both sampling steps	An increase in the $threshold_{final}$ by 1%

Based on the impact on the threshold described in Table 1, we estimate that the uncertainty level for $threshold_{final}$ to be plus or minus 1%.

Practical application

The following is an example of applying the sentence transformer methodology to a real-life scenario: estimating the amount UKRI funds on 'space'. (Please note, the example is purely illustrative and should not be seen as official figures for awards and expenditure on 'space' across UKRI.)

Example of classifying awards by the theme of 'space'

This request sought to provide UKRI expenditure on 'space'-related awards for three financial years: 2020-21, 2021-22 and 2022-23.

For this analysis, we used a dataset encompassing all 29,032 awards funded by UKRI that incurred expenditure during the aforementioned financial years. To create a unique digital signature for each award and map it to a multi-dimensional dense vector space, we used text from an award's title and summary, which provided a plain English description. Where available, we further enriched this text by incorporating the technical summary and text describing the expected impact of an award.

We expanded the meaning of 'space' to a list of keywords that better captured the meaning of the theme (i.e., cosmos). Table 2 shows the initial list of eight keywords that were good candidates for the theme.

Table 2: Initial list of eight keywords to capture the theme of 'space'

1. space	2. satellite	3. astronomy	4. cosmology
5. universe	6. solar	7. orbit	8. spaceflight

This keyword list was then shared with subject matter experts to review with the objective of removing or adding terms, creating a final list of keywords (Table 3). This list was then allocated a unique digital signature and ran through the llmrails/ember-v1 model on the 29,032 awards. The output gave a metric of similarity for each award to the final keyword list. Awards had metric of similarity ranging between –0.427 (award reference: BB/V018922/1) and 0.756 (award reference: ST/S000461/1).

Table 3: Final iteration of 37 keywords that capture the theme of 'space'

1. space	2. satellite	3. astronomy	4. cosmology
5. universe	6. solar	7. orbit	8. spaceflight
9. sentinels	10. gaia	11. mission	12. earth observation

13. remote sensing	14. position navigation and timing	15. on-orbit computing	16. space domain awareness
17. severe space weather	18. space surveillance and tracking	19. space operations management	20. in-orbit services
21. space transportation	22. space science	23. planetary science	24. space exploration
25. space habitation	26. micro-satellites	27. nanosatellites	28. ground station networks
29. rocket technology	30. payload systems	31. spacecraft development	32. satellite imagery
33. satellite telecommunications	34. space-based data	35. space materials	36. space tourism
37. asteroid mining			

The next step was to identify the threshold cosine similarity score where awards are no longer relevant to the topic. We took a random sample of 30 awards for metric of similarities of a possible 29,032 awards, with three samples per decile. These 30 awards were sent to subject matter experts for manual classification which gave us $threshold_1 = 0.667$. For $threshold_{final}$ the number of awards that fell between 0.617 and 0.717 were 2,289 (see Table 4).

Table 4: Number of awards for the 10 intervals for cosine similarity score between 0.617 and 0.717

Interval start	Interval end	Interval width	Number of awards	Sampled awards
0.617	0.627	0.01	775	3
0.627	0.637	0.01	504	3
0.637	0.647	0.01	327	3
0.647	0.657	0.01	184	3
0.657	0.667	0.01	136	3
0.667	0.677	0.01	154	3

Interval start	Interval end	Interval width	Number of awards	Sampled awards
0.677	0.687	0.01	104	3
0.687	0.697	0.01	46	3
0.697	0.707	0.01	37	3
0.707	0.717	0.01	22	3

The final logistic regression model was applied to the 30 manually classified awards which resulted in $threshold_{final} = 0.646$ (Figure 1).

Figure 1: Logistic regression fit of the 30 manually classified awards to the cosine similarity, 2020-21 to 2022-23

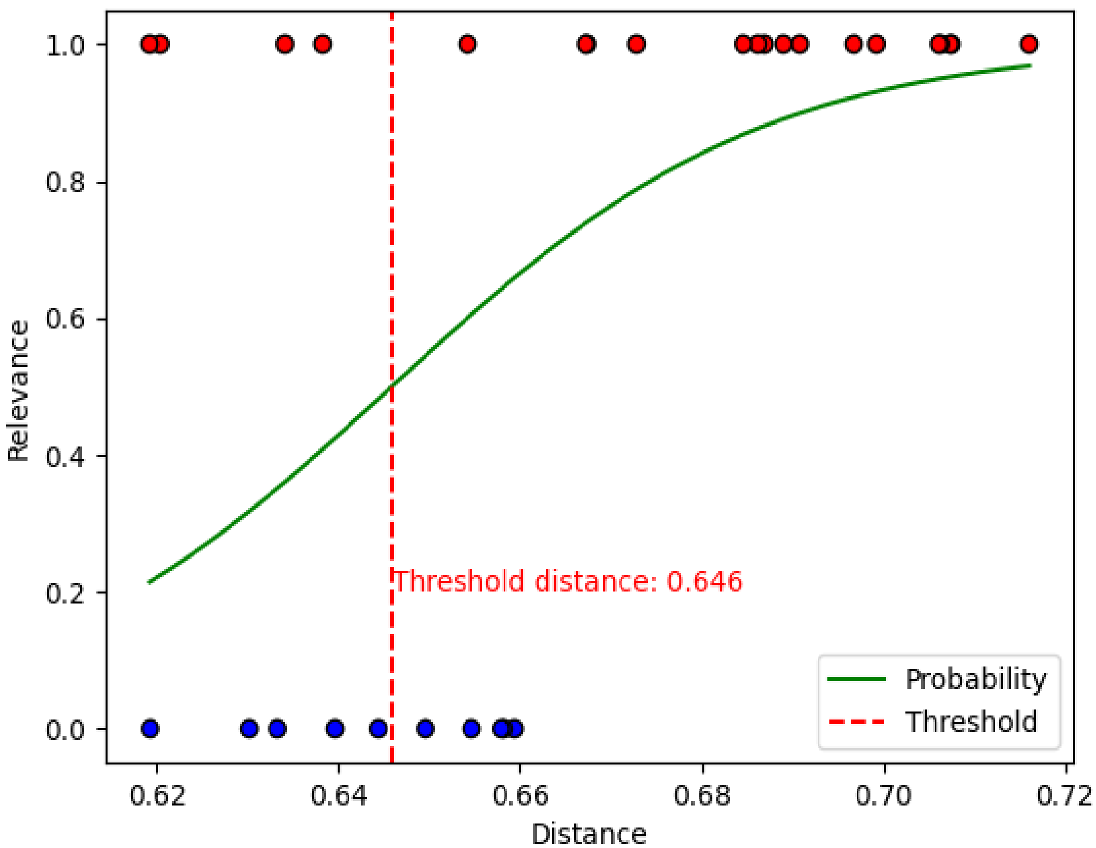

Table 5 shows that across the three financial years, from 2020-21 to 2022-23, UKRI had 740 unique awards that classified under the theme of ‘space’. Of these 744 awards, 104 (14%) contained no keywords from the keyword list (See Annex A for the full list of ‘space’-related awards). This resulted in a total expenditure of £341.2m.

Table 5: UKRI expenditure on ‘space’-related awards by financial year, 2020-21 to 2022-23

Financial year	Number of awards	Number of ‘space’ awards	Total ‘space’ expenditure
2020-21	12,587	195	£116.3m
2021-22	11,260	258	£110.3m
2022-23	9,102	320	£114.6m
Total[1]	29,032	744	£341.2m

[1] The total number of awards is not the same as the sum of all three financial years as some awards would have been active in one or more years.

Applying the uncertainty level to these figures ($threshold_{final} = 0.646$ plus or minus 1%) we have an estimated range for UKRI expenditure on ‘space’-related awards between £290.6m and £446.8m between 2020-21 to 2022-23.

Discussion

In this section, we discuss the data and methodological challenges facing the implementation of sentence transformers. Wherever possible, we provide potential ways forward that will help others facing these challenges.

Keyword list

In areas where subject matter knowledge is limited or restricted, such as a very new area of research, identifying keywords can be challenging. In this situation a ‘light touch’ use of experts to develop an initial list may be advisable. Expert engagement can also come with its own set of challenges. For example, it can take time to identify and engage with relevant experts.

Iteration to refine the process can help address these challenges. Start with a provisional list that is tested and can then be expanded using commercially available linguistic models (e.g., ChatGPT), which assess the contextual significance of terms. However, the use of expert knowledge is recommended to ensure thematic keywords are as relevant and encompassing as possible.

Data quality

There is varying quality of text data available for UKRI awards. For example, around a quarter of UKRI awards have absent or minimal descriptions (fewer than 30 words). Conversely, award abstracts can be verbose but the quality of information is not suitable for classification. In both cases, the lack of information or inconsistency in detail impacts the model's ability to uniformly assess and categorise awards, leading to potential misclassifications.

For this implementation we have retained all awards, regardless of the text quality of their title, abstract or objectives. The result is that awards with limited descriptions will not fall under any particular theme and this issue falls outside the practical implementation of this methodology.

Threshold determination

For the model to provide good discrimination between awards that fall under the theme and those that don't, the threshold determination is crucial. A two-step approach to determine the threshold has been implemented, but it is possible for a third step to further narrow the search range. This may be necessary if the logistic regression is a very poor fit to the data: the output of the llmrails/ember-v1model does help to distinguish the subject matter expert award classification.

For cases where classification is more difficult, using multiple subject matter experts to remove some of the subjective biases is recommended. Another challenge of the existing model is if a topic is too broad and the llmrails/ember-v1 model is not successfully discriminating between awards. In this situation, the classification may benefit from being broken down into subset themes with each one run independently. In our experience, this improves its performance.

Model performance

This paper does not address directly the llmrails/ember-v1 model performance against other classification methods. The main reason for not carrying out a direct comparison with other methods is the level of resources required to perform such comparison on a wide range of themes. There is also a challenge in establishing what a *true* classification under a theme really is. For example, there can be disputes even among subject matter experts on whether particular awards fall under a theme or not. To mitigate some of these issues, we have estimated an uncertainty range to provide users the ability of contextualising their results.

The goal of implementing the llmrails/ember-v1 model is to demonstrate the plausibility of the output of such a model in identifying awards that align closely with a specified theme, effectively distinguishing between highly relevant, moderately relevant, and irrelevant awards. This proficiency underscores the model's utility in facilitating targeted searches within extensive datasets. It is possible that a set of experts would dispute the inclusion or exclusion of different awards under a certain theme. This has always been a challenge in this area and every other method would face similar challenges.

Use cases

The key strength of implementing sentence transformer methodology is the ability to rapidly categorise awards according to a particular theme on a vast array of awards. This can be applied to thematic groups that currently lack a clear definition or to existing thematic groups where their meaning may change over time. The approach can be implemented across a large number of awards quickly and reliably.

Annex A – Full list references of ‘Space’-related awards between 2020-21 and 2022-23

Below is a full list of 744 awards identified by the sentence transformer methodology that fall under the theme ‘space’ between 2020-21 and 2022-23. In **bold** are the 104 awards that contain no terms in the award’s descriptors from the final keyword list.

ST/S000461/1	NE/P017274/1	ST/P000533/1	NE/T005637/1	NE/V003062/1
ST/S000348/1	NE/P017347/1	MR/S035338/1	ST/V000209/1	NE/V002554/1
ST/N000722/1	NE/P017185/1	ST/V000659/1	10046903	NE/V002511/1
ST/N000838/1	NE/P01738X/1	ST/T000384/1	ST/V001000/1	NE/V002759/1
ST/P000657/1	ST/V000977/1	ST/R000751/1	ST/P006930/1	ST/T001542/1
ST/R000824/1	ST/N000811/1	NE/S010033/1	ST/V000527/1	ST/V000799/1
ST/N000749/1	ST/R000891/1	ST/T000422/1	ST/S000402/1	ST/R002355/1
ST/N000692/1	NE/S016066/1	ST/N000900/1	ST/T00021X/1	ST/R006857/1
ST/R000921/1	NE/S016104/1	ST/T000252/1	ST/V000861/1	ST/M001334/1
ST/T002255/1	NE/S016163/1	ST/R001383/1	ST/X001970/1	ST/S000496/1
ST/S000429/1	ST/S000488/1	ST/P000584/1	ST/R00286X/1	ST/T002034/1
ST/P000622/1	ST/R000875/1	ST/T007133/1	ST/S002677/1	ST/T001747/1
ST/S000518/1	ST/S000372/1	MR/T044136/1	ST/W005387/1	ST/T001828/1
ST/R000476/1	ST/T000406/1	ST/P006299/1	NE/V00283X/1	ST/P000304/1
ST/S000364/1	ST/S001239/1	ST/V000640/1	NE/V002570/1	ST/T000198/1
ST/T000228/1	ST/V000764/1	ST/P001394/1	NE/V002708/1	ST/R000727/1
ST/V000675/1	ST/V000225/1	ST/T000341/1	NE/V002732/1	ST/S001948/1
MR/T020261/1	ST/N000919/1	ST/T00018X/1	EP/X026221/1	BB/T018909/1
ST/V000497/1	ST/S000216/1	ST/P001297/1	ST/T000295/1	NE/T014164/1
MR/W009498/1	ST/P007236/1	ST/T000287/1	ST/R000573/1	ST/S002103/1
ST/S000240/1	ST/N000854/1	NE/T005645/1	ST/P001572/1	ST/N006852/1
NE/P017061/1	ST/V000586/1	NE/T005734/1	NE/V00249X/1	ST/P003370/1

ST/S001980/1	ST/T001763/1	ST/P000312/1	ST/S001433/1	ST/P000495/1
ST/T001356/1	ST/R001758/1	NE/P017290/1	ST/S000291/1	ST/S001514/1
ST/R000506/1	ST/T00178X/1	NE/P016782/1	ST/S006559/1	ST/S005447/1
NE/N014480/1	ST/R005737/1	ST/R000395/1	ST/P002064/1	ST/S000542/1
ST/N001443/1	MR/T040726/1	ST/P00718X/1	512118	ST/N006801/1
ST/R000980/1	ST/R001405/1	**EP/S000631/1**	ST/S000550/1	ST/S004106/1
EP/R026092/1	ST/T000244/1	NE/P016863/1	ST/T002972/1	ST/P000525/1
ST/T002913/1	ST/P000649/1	NE/T00939X/1	ST/V000500/1	**NE/P017371/1**
ST/V006320/1	NE/V000748/1	NE/S016236/1	NE/S002235/1	**NE/P017045/1**
ST/R000719/1	NE/P016928/1	NE/S016244/1	ST/S002642/1	**NE/P017444/1**
ST/T000473/1	ST/R002827/1	ST/S00260X/1	ST/P006876/1	**NE/P017312/1**
ST/S000615/1	ST/P000592/1	ST/V001329/1	ST/R005265/1	**NE/P01724X/1**
ST/V000594/1	EP/S515735/1	ST/S001492/1	ST/P000541/1	**NE/P017053/1**
ST/T000333/1	ST/P003168/1	ST/R000999/1	NE/X009254/1	**NE/P017452/1**
ST/P001270/1	NE/P016758/1	EP/R51326X/1	NE/V013963/1	
ST/T000481/1	NE/P016693/1	ES/M011879/1	ST/V000284/1	NE/P002498/1
ST/V000918/1	NE/P017150/1	ST/S00033X/1	**ST/S000399/1**	ST/V00199X/1
ST/R003203/1	ST/S00047X/1	ST/R000700/1	ST/S00257X/1	ST/W00206X/1
NE/K010867/1	ST/V000276/1	NE/R015651/1	ST/T000171/1	**NE/M000427/1**
NE/K011006/1	ST/S000631/1	ST/S002634/1	ST/W004801/1	**NE/M000303/1**
NE/K011014/1	MR/S032223/1	ST/R000905/1	ST/R000697/1	**NE/M000397/1**
NE/K010794/1	ST/V000454/1	MR/S016066/1	ST/S001522/1	**NE/M000443/1**
NE/K010816/1	ST/W000989/1	ST/R000417/1	ST/S001476/1	ST/V000330/1
ST/T000414/1	NE/P017142/1	ST/T003081/1	**NE/P013643/1**	ST/T000325/1
ST/V00087X/1	NE/P016715/1	ST/T007281/1	ST/T001461/1	ST/S002561/1
ST/W000393/1	NE/P017231/1	ST/X002365/1	ST/S000747/1	ST/N005481/1

ST/N005716/1	**EP/R007187/1**	EP/P006094/1	**NE/T012684/1**	ST/T00729X/1
ST/N005406/1	ST/P000320/1	EP/S036393/1	**NE/T01248X/1**	EP/V000543/1
ST/N005422/1	ST/P000665/1	EP/W005654/1	**NE/T012633/1**	ST/W003163/1
EP/P017487/1	ST/R00045X/1	ST/N00258X/1	**NE/T012641/1**	NE/V002724/1
ST/P00041X/1	ST/M007596/1	EP/S515747/1	ST/P005764/1	NE/V003070/1
ST/R006768/1	NE/M000117/1	ST/N000757/1	NE/S00632X/1	NE/V002694/1
ST/W000938/1	NE/M000346/1	**NE/L002507/1**	ST/S006168/1	NE/V002716/1
ST/R002673/1	NE/M000370/1	ST/T000279/1	ST/V00137X/1	EP/T028270/1
EP/S023577/1	NE/M000400/1	AH/S013067/1	ST/S006117/1	ST/S006109/1
ST/S000267/1	NE/M000125/1	ST/W000547/1	ST/R000840/1	ST/T001798/1
ST/S000321/1	ST/P001998/1	NE/V002651/1	ST/V000942/1	MR/S035214/1
ST/V002279/1	ST/S00615X/1	NE/V002619/1	ST/W003449/1	ST/W00089X/1
ST/X001903/1	ST/T000163/1	NE/V002864/1	ST/T001836/1	NE/T000767/1
EP/S026347/1	MR/S03465X/1	NE/V002899/1	MR/S017216/1	NE/L010828/1
ST/V000748/1	EP/M000885/1	ST/V000780/1	ST/V005782/1	ST/T006609/1
NE/M000419/1	ST/T001399/1	ST/S006095/1	**ST/P004008/1**	ST/T006668/1
ST/P000428/1	ST/P002242/1	ST/N005430/1	ST/N000331/1	ST/T006404/1
EP/R513465/1	**NE/T008814/1**	**EP/P027482/1**	ST/N000765/1	ST/T006331/1
NE/R012288/1	ST/R000603/1	ST/P000703/1	ST/S006141/1	EP/P017436/1
NE/R01227X/1	ST/S001425/1	ST/P000673/1	160078	NE/K010824/1
ST/W001004/1	ST/P001262/1	ST/S001298/1	ST/V000888/1	NE/K010662/1
ST/S000623/1	ST/S002618/1	ST/S006192/1	ST/Y000218/1	NE/K010743/1
ST/V001337/1	EP/R030340/1	**NE/T012455/1**	ST/V000519/1	NE/K010654/1
ST/V001019/1	ST/V000624/1	**NE/T012536/1**	NE/N018559/1	NE/K010611/1
ST/P003826/1	ST/N002962/1	**NE/T012463/1**	NE/S01537X/1	ST/R00563X/1
ST/S000437/1	**EP/L014998/1**	**NE/T012501/1**	ST/R001421/1	ST/N000927/1

ST/T002964/1	ST/S001301/1	NE/N018508/1	AH/S012907/1	NE/T000937/1
ST/V001221/1	BB/T018941/1	EP/S001751/1	EP/R030855/1	NE/T000295/1
ST/X001741/1	EP/N007565/1	ST/T001755/1	ST/R00062X/1	ST/T00049X/1
EP/M019284/1	EP/T024216/1	EP/R043167/1	ST/R001480/1	AH/T002239/1
MR/T041218/1	NE/N018303/1	**EP/S034587/1**	EP/T023139/1	EP/P021859/1
MR/T020784/1	ST/W001136/1	ST/R003009/1	ST/R002096/1	ST/R000484/1
ST/R000425/1	NE/T012331/1	ST/W00254X/1	EP/P026133/1	EP/T007346/1
ST/S006176/1	EP/V002910/1	ST/V002678/1	ST/S006206/1	ST/N005805/1
EP/P008046/1	ST/W001179/1	ST/P001300/1	ST/R002770/1	EP/T017570/1
EP/W028921/1	ST/V001167/1	ST/N006798/1	ST/V000837/1	**EP/S032347/1**
ES/X00645X/1	EP/X036405/1	**NE/R01423X/1**	**NE/T001615/1**	**EP/S033483/1**
ST/N001494/1	NE/V002686/1	**EP/P013570/1**	EP/S030301/1	ST/T007176/1
ES/T005238/1	NE/R001332/1	**EP/L016796/1**	ST/W00075X/1	ST/S000666/1
ST/R000514/1	ST/P002099/1	ST/S002421/1	ST/N004981/1	EP/T017287/1
ST/P004474/1	EP/S035761/1	ST/R00059X/1	ST/W000857/1	NE/P001556/1
EP/S001182/1	NE/V002643/1	ST/S002952/1	MR/T020989/1	ST/T001666/1
ST/P001254/1	NE/V002791/1	**EP/N027736/1**	ST/S000879/1	ST/W002957/1
NE/S013970/1	NE/V002597/1	EP/R041431/1	ST/S001484/1	NE/N011791/1
ST/V000551/1	NE/V002678/1	ST/W001128/1	**EP/T026111/1**	ST/R002878/1
ST/R006571/1	NE/R006768/1	ST/N002571/1	ST/W007037/1	ST/S00145X/1
NE/S009167/1	NE/R007241/1	**NE/T000430/1**	ST/W006960/1	ST/S001417/1
ST/S006060/1	NE/R00675X/1	ST/R000662/1	NE/T012323/1	**NE/T012595/1**
EP/P002331/1	NE/R006776/1	ST/S000305/1	10012626	75238
ST/W006022/1	NE/R006822/1	NE/T000228/1	EP/X034895/1	ST/X001946/1
ST/R001502/1	NE/R006849/1	MR/S016449/1	73055	ST/V000713/1
ST/R000638/1	NE/N018478/1	EP/S000976/1	ST/W001195/1	NE/L011514/1

ST/T003146/1	ST/T001593/1	EP/R009260/1	ST/W001098/1	**EP/R033439/1**
NE/S015736/1	ST/V000268/1	**EP/P510270/1**	ST/W00108X/1	ST/X001814/1
BB/N015894/1	MR/T043733/1	**EP/R02331X/1**	ST/V00235X/1	ST/V002120/1
NE/S01134X/1	ST/S001891/1	MR/T042230/1	NE/S015167/1	**EP/T019018/1**
ST/N003748/1	ST/R000549/1	EP/V026887/1	NE/R013144/1	**EP/S032207/1**
NE/S015817/1	NE/V001183/1	NE/T004835/1	BB/T017511/1	ST/T005629/1
NE/S015795/1	ST/V003224/1	**MC_UU_00031**/9	ST/R005761/1	EP/T026952/1
NE/S015655/1	ST/V001590/1	**NE/T01234X/1**	ST/T002069/1	ST/W001071/1
NE/S015728/1	**EP/M020355/1**	ST/T002026/1	**NE/V000411/1**	**NE/T010983/1**
NE/S015612/1	**ST/V00106X/1**	EP/S033238/1	**EP/N035437/1**	ST/X002624/1
EP/T01735X/1	**EP/L015897/1**	ST/V004883/1	**EP/L015315/1**	ST/T003472/1
EP/P005896/1	**NE/P013651/1**	ST/K003240/1	511880	ST/V006371/1
ST/S005951/1	NE/R000425/1	EP/T031441/1	NE/S015590/1	ST/W000865/1
MR/S018859/1	EP/T517926/1	MR/S01795X/1	ST/S000593/1	EP/S007903/1
511984	NE/T005564/1	**EP/S023925/1**	**EP/P02839X/1**	EP/P024289/1
NE/P013678/1	NE/T012307/1	ST/V000233/1	EP/S023305/1	**EP/T015403/1**
NE/P013627/1	ST/T00147X/1	ST/V000632/1	ST/T001623/1	ST/R005125/1
NE/P013775/1	BB/N009150/1	ST/V000535/1	EP/W008718/1	ES/S013601/1
ST/V003194/1	EP/W027011/1	ST/V000543/1	ST/R001413/1	ST/X002241/1
EP/T001046/1	ST/R000743/1	ST/V000462/1	ST/R00143X/1	EP/P008550/1
ST/T002573/1	EP/S001727/1	**NE/S007415/1**	ST/S00307X/1	ST/T001402/1
EP/S019375/1	NE/R015546/1	**EP/S023445/1**	EP/R035393/1	ST/X001377/1
ST/X002152/1	ST/T001739/1	**EP/P030181/1**	ST/S001379/1	10031944
NE/X019071/1	ST/T002956/1	EP/P016294/1	ST/T00715X/1	ST/W001810/1
ST/W000997/1	ST/R001103/1	EP/N022513/1	ST/R006555/1	ST/R000972/1
ST/R005184/1	NE/R004935/1	ST/V005979/1	ST/V000721/1	ST/T000449/1

BB/T010525/1	**EP/R025819/1**	EP/R030073/1	**EP/L026120/1**	ST/W002817/1
ST/T002204/1	EP/P002285/1	**NE/S009094/1**	ST/S001506/1	NE/W005530/1
EP/W023024/1	EP/M022498/1	EP/R044902/1	**EP/L015374/1**	**ST/X002357/1**
ST/T001488/1	NE/P001378/1	ST/R001928/1	**EP/M024385/1**	**NE/R010196/1**
EP/S020357/1	MR/S03241X/1	**NE/S009116/1**	**NE/R002134/1**	ST/N001117/1
ES/S008241/1	NE/M017893/1	ST/X001822/1	ST/T001313/1	**EP/S023291/1**
ST/T000694/1	ST/P003850/1	ST/X001830/1	NE/S001018/1	EP/S001018/1
NE/W003074/1	**ST/R002908/1**	ST/X001873/1	ST/R001367/1	**EP/S027513/1**
EP/T005734/1	EP/N031938/1	**NE/S009140/1**	NE/M018288/1	**EP/S009647/1**
ST/M001652/1	NE/P002331/1	**NE/S009175/1**	MR/S016929/1	ST/X001911/1
MR/S035141/1	MR/W007657/1	**NE/S009124/1**	ST/V002007/1	10030792
ST/V000691/1	EP/L016613/1	**EP/M010619/1**	ST/R006598/1	
MR/T042575/1	NE/V001388/1	**BB/S013024/1**	NE/T006749/1	
NE/R000751/1	ST/L006189/1	ST/N002830/1	ST/W001020/1	
NE/T014822/1	NE/S000518/1	ST/P000754/1	ST/S002456/1	
EP/S000550/1	**NE/M011372/1**	EP/T003553/1	ST/R001898/1	
EP/L015927/1	BB/T01895X/1	**NE/S007334/1**	ST/R001243/1	
EP/V521802/1	ST/V001035/1	EP/T022574/1	**NE/S01067X/1**	
NE/T010924/1	ST/V001051/1	NE/W003198/1	**NE/X005941/1**	
ST/T000864/1	NE/P008615/1	NE/W002981/1	ST/V005243/1	
EP/V002449/1	NE/P012426/1	NE/W003147/1	NE/P013104/1	
NE/N012070/1	EP/T025611/1	NE/W003015/1	ST/W003015/1	
ST/P002803/1	**EP/R034826/1**	NE/W003066/1	NE/T006102/1	
EP/N012372/1	ST/X002616/1	NE/W003104/1	EP/M00158X/1	
ST/T005882/1	ST/W000830/1	NE/W003007/1	**10008317**	
EP/S021892/1	**AH/T013362/1**	NE/W002914/1	ST/R001499/1	