SOST30071 Quantitative text analysis in social sciences

Lecturer: Dr. Yan Wang (yan.wang-5@manchester.ac.uk)

Office: G.20 HBS Building

Office hour: By appointment

GTA: Ragdad Cani Miranti (ragdad.miranti@postgrad.manchester.ac.uk)
Office hour: By appointment

Course Description

Aims
The availability of text data has increased exponentially in recent years, alongside a growing demand for its analysis. This course introduces students to the quantitative analysis of text from a social science perspective, with a wide coverage of applications in economics, sociology & communication, and political science. The course adopts an applied approach: while theoretical aspects will be addressed, the primary objective is to equip students with the skills to formulate research questions that can be explored through text data and to understand the methodologies required to answer them. To this end, we begin by explaining how text can be conceptualized and modelled quantitatively, examining methods for comparing textual data. Following this, we delve into both supervised and unsupervised techniques in considerable depth, before addressing several specialized topics pertinent to social science research. Ultimately, the course aims to enable students to undertake their own research projects using text as data, providing a foundation for more advanced and technical investigations.

Prerequisites:
Students are expected to have completed at least one course in statistics or inference prior to enrolling. Foundational knowledge in probability, probability densities, distributions, statistical testing, hypothesis testing, linear models, generalized linear models, and maximum likelihood estimation is assumed. Additionally, students should possess a working knowledge of the R or Python programming language, as assignments will require this. Students should also have basic computer skills, be familiar with navigating their computer's file system, and feel comfortable using a command-line interface. While this course addresses techniques originating in computer science and statistics, it is not a computer science course. Traditional topics in Natural Language Processing (e.g., machine translation, optical character recognition, parts of speech tagging) will be covered minimally, as these are comprehensively addressed in other courses. Likewise, this course will not focus extensively on data acquisition methods, such as web-scraping, as these topics are well-covered in other offerings.

Intended Learning Outcomes
The primary objective of this course is to familiarize students with machine learning methods and contemporary quantitative text analysis techniques, equipping them with the skills needed to apply these methods in their own research. In pursuit of this objective, students will also engage with foundational concepts in machine learning and statistics, cultivating skills that are applicable to a broad range of data and inference challenges. Additionally, students will have the opportunity to enhance their programming competencies and develop an original research project.
Knowledge and Understanding:

- Demonstrate a theoretical understanding of content analysis approaches and machine learning techniques
- Analyse and interpret outputs from applied textual data analysis

Intellectual Skills:

- Prepare datasets for text analysis using content analysis and statistical software
- Visualise, describe, and critically assess quantitative text analysis in R/Python, utilizing advanced methods
- Reflect on potential biases in data processing, modelling, and presentation of text analysis results

Practical Skills:

- Produce reports for academic and non-academic audiences
- Present inferences derived from quantitative text analysis outputs

Transferable Skills and Personal Qualities:

- Design and execute small-scale projects applying machine learning to social science research questions using text data
- Effectively communicate academic research findings to a broader audience

Core Textbook:

- Grimmer, J., Roberts, M.E. and Stewart, B.M., 2022. *Text as data: A new framework for machine learning and the social sciences*. Princeton, NJ: Princeton University Press. This textbook is a recent survey of quantitative text analysis as used in the social sciences.

Supplementary Texts:

- Jurafsky, D. and Martin, J.H., 2024. *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition with language models*. 3rd ed. Online manuscript released 20 August 2024. Available at: https://web.stanford.edu/~jurafsky/slp3. This is a great reference book for the more technical aspects of quantitative text analysis.
- Van Atteveldt, W., Trilling, D. and Calderon, C.A., 2022. *Computational analysis of communication*. Chichester: John Wiley & Sons. Available at: https://cssbook.net/ with codes and practices.
- Goldberg, Y., 2017. *Neural network methods in natural language processing*. San Rafael, CA: Morgan & Claypool Publishers. Available at: https://link.springer.com/book/10.1007/978-3-031-02165-7

Lecture and Practical Session Schedule

10 sessions of 2-hour lectures and biweekly 2-hour computer lab sessions

Assignments (Not marked)

Students will complete five homework assignments designed to reinforce lecture material and develop practical analytical skills applicable to their own research. Solutions will be provided before the following seminars. Students are encouraged to use these assignments to build knowledge and skills for the final essay, keeping their specific research interests in mind. Assignments will be completed in R/R Markdown and require installation of the *knitr* package. RStudio, an integrated development environment for R, is recommended for creating R Markdown documents. Python code will also be provided for students who prefer to work in a Python environment.

Final Project (Academic essay or technical report, 100% of grade)

As the culmination of the course, students will undertake a final project involving either an original research study using text data or a technical report on applied work using text as data. Ideally, this project will contribute to their dissertation, field paper, or other ongoing research. Detailed guidelines for the final paper will be discussed in class to ensure it aligns with students' research interests and academic or career objectives. Maximum 2,000 words. Reproducible code is expected with the final submission.

Course content

Week 1 (Lecture 1): Introduction to Quantitative Text Analysis

Introduction, logistics and course arrangement. Overview of the field, its applications in social sciences, and fundamental principles of text as data. We will also be discussing topics such as vector space model of a document, feature choices/representation, text data pre-processing, bag of words assumptions and alternatives.

Key readings:

Grimmer, J. and Stewart, B.M., 2013. Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political analysis*, *21*(3), pp.267-297.

Additional readings:

Denny, M.J. and Spirling, A., 2017. *Text preprocessing for unsupervised learning: Why it matters, when it misleads, and what to do about it*. SSRN Working Paper. Available at: https://doi.org/10.2139/ssrn.2849145.

Gentzkow, M. and Shapiro, J.M., 2010. What drives media slant? Evidence from U.S. daily newspapers. *Econometrica*, 78(1), pp.35–71. https://doi.org/10.3982/ECTA7195.

Gentzkow, M., Kelly, B. and Taddy, M., 2019. Text as data. *Journal of Economic Literature*, 57(3), pp.535–574. https://doi.org/10.1257/jel.20181020.

Hassan, T.A., Hollander, S., van Lent, L. and Tahoun, A., 2019. Firm-level political risk: Measurement and effects. *Quarterly Journal of Economics*, 134(4), pp.2135–2202. https://doi.org/10.1093/qje/qjz021.

Week 2 (Lecture 2): Descriptive Statistical Methods for Text Analysis

Exploration of foundational descriptive statistics in text analysis, focusing on word frequency, term-document matrices, co-occurrence, collocations and phrasemes; key words in context; lexical diversity; sampling distributions for estimates.

Key readings:

Grimmer, Roberts and Stewart (2022, chs. 3-9)

Additional readings:

Benoit, K., Munger, K. and Spirling, A., 2017. *Measuring and explaining political sophistication through textual complexity*. SSRN. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3062061.

Däubler, T., Benoit, K., Mikhaylov, S. and Laver, M., 2012. Natural sentences as valid units for coded political texts. *British Journal of Political Science*, 42(4), pp.937–951. https://doi.org/10.1017/S0007123412000105.

DuBay, W., 2004. *The principles of readability*. Costa Mesa, CA: Impact Information. Available at: https://files.eric.ed.gov/fulltext/ED490073.pdf.

Hassan, T.A., Hollander, S., van Lent, L. and Tahoun, A., 2019. Firm-level political risk: Measurement and effects. *Quarterly Journal of Economics*, 134(4), pp.2135–2202. https://doi.org/10.1093/qje/qjz021.

Hengel, E., 2017. *Publishing while female: Are women held to higher standards? Evidence from peer review*. Available at: http://www.erinhengel.com/research/publishing_female.pdf.

Huang, L., Perry, P. and Spirling, A., 2020. A general model of author "style" with application to the UK House of Commons, 1935–2018. *Political Analysis*, 28(3), pp.412–438. https://doi.org/10.1017/pan.2019.49.

Week 3 (Lecture 3): Supervised Techniques with Text Data I
Dictionary-based approaches, including sentiment analysis and the application of content dictionaries. Document classification, including precision and recall as evaluation metrics, the role of crowdsourcing in supervised learning,

Key readings:
Grimmer, Roberts and Stewart (2022, chs. 11 and 16)

Additional readings:

Advani, A., Ash, E., Cai, M. and Rasul, I., 2021. *Race-related research in economics and other social sciences*. Working Paper. Available at https://warwick.ac.uk/fac/soc/economics/research/centres/cage/manage/publications/wp565.2021.pdf.

Pang, B. and Lee, L., 2008. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1–2), pp.1–135. https://doi.org/10.1561/1500000011.

Laver, M. and Garry, J., 2000. Estimating policy positions from political texts. *American Journal of Political Science*, 44(3), pp.619–634. https://doi.org/10.2307/2669268.

Tausczik, Y.R. and Pennebaker, J.W., 2010. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, 29(1), pp.24–54. https://doi.org/10.1177/0261927X09351676.

Week 4 (Lecture 4): Supervised Techniques with Text Data II
Introduction of various commonly used classifiers: naïve bayes; regularised regression; cross-validations.

Key readings:
Grimmer, Roberts and Stewart (2022, chps. 17-20)

Additional readings:

Benoit, K., Conway, D., Lauderdale, B.E., Laver, M. and Mikhaylov, S., 2015. Crowd-sourced text analysis: Reproducible and agile production of political data. *American Political Science Review*, 109(2), pp.278–295. https://doi.org/10.1017/S0003055416000058.

Gentzkow, M., Shapiro, J.M. and Taddy, M., 2019. Measuring group differences in high-dimensional choices: Method and application to congressional speech. *Econometrica*, 87(4), pp.1307–1340. https://doi.org/10.3982/ECTA16566.

Hopkins, D.J. and King, G., 2010. A method of automated nonparametric content analysis for social science. *American Journal of Political Science*, 54(1), pp.229–247. https://doi.org/10.1111/j.1540-5907.2009.00428.x.

Laver, M., Benoit, K. and Garry, J., 2003. Extracting policy positions from political texts using words as data. *American Political Science Review*, 97(2), pp.311–331. https://doi.org/10.1017/S0003055403000698.

Lowe, W., 2008. Understanding Wordscores. *Political Analysis*, 16(4), pp.356–371. https://doi.org/10.1093/pan/mpn004.

Osnabrügge, M., Ash, E. and Morelli, M., 2023. Cross-domain topic classification for political texts. *Political Analysis*, 31(1), pp.59–80. https://doi.org/10.1017/pan.2021.37.

Peterson, A. and Spirling, A., 2018. Classification accuracy as a substantive quantity of interest: Measuring polarization in Westminster systems. *Political Analysis*, 26(1), pp.120–128. https://doi.org/10.1017/pan.2017.39.

Widmer, P., Galletta, S. and Ash, E., 2022. Media slant is contagious. *arXiv preprint* arXiv:2202.07269. Available at: https://arxiv.org/abs/2202.07269.

Week 5 (Lecture 5): Transition from Supervised to Unsupervised Techniques
More about machine learning fundamentals, covering support vector machines, k-nearest neighbours, random forests, tree-based methods, and ensemble models.

Key readings:
Grimmer, Roberts and Stewart (2022, chps. 16 and 21)

Additional readings:

Bingler, J., Kraus, M. and Leippold, M., 2021. Cheap talk and cherry-picking: What ClimateBERT has to say on corporate climate risk disclosures. SSRN Working Paper. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3796152.
D'Orazio, V., Landis, S., Palmer, G. and Schrodt, P., 2014. Separating the wheat from the chaff: Applications of automated document classification using support vector machines. *Political Analysis*, 22(2), pp.224–242. https://doi.org/10.1093/pan/mpt030.
Hillard, D., Purpura, S. and Wilkerson, J., 2008. Computer assisted topic classification for mixed methods social science research. *Journal of Information Technology and Politics*, 4(4), pp.31–46. https://doi.org/10.1080/19331680801975367.
Hoberg, G. and Phillips, G., 2016. Text-based network industries and endogenous product differentiation. *Journal of Political Economy*, 124(5), pp.1423–1465. https://doi.org/10.1086/688176.
Kelly, B., Papanikolaou, D., Seru, A. and Taddy, M., 2021. Measuring technological innovation over the long run. *American Economic Review: Insights*, 3(3), pp.303–320. https://doi.org/10.1257/aeri.20190499.
Pei, J. and Jurgens, D., 2020. Quantifying intimacy in language. *arXiv preprint* arXiv:2011.03020. Available at: https://arxiv.org/abs/2011.03020.
Siroky, D.S., 2009. Navigating random forests and related advances in algorithmic modeling. *Statistical Surveys*, 3, pp.147–163. https://doi.org/10.1214/07-SS033.

Week 6: Reading week

Week 7 (Lecture 6): Unsupervised Techniques with Text Data I
Basics of unsupervised learning, with a focus on dimensionality reduction methods, including principal component analysis and singular value decomposition, clustering and latent sematic scaling.

Key readings:
Grimmer, Roberts and Stewart (2022, chps. 7 and 12)

Additional readings:

Rose, B., n.d. Document clustering in Python: A practical guide. Available at: http://brandonrose.org/clustering.
Gillis, N., 2014. The why and how of nonnegative matrix factorization. *arXiv preprint* arXiv:1401.5226. Available at: https://arxiv.org/pdf/1401.5226.
Hoberg, G. and Phillips, G., 2016. Text-based network industries and endogenous product differentiation. *Journal of Political Economy*, 124(5), pp.1423–1465.
Grimmer, J. and King, G., 2010. General purpose computer-assisted clustering and conceptualization. *Proceedings of the National Academy of Sciences*, 108(7), pp.2643–2650. https://doi.org/10.1073/pnas.1018067108.
Kelly, B., Papanikolaou, D., Seru, A. and Taddy, M., 2021. Measuring technological innovation over the long run. *American Economic Review: Insights*, 3(3), pp.303–320. http://www.nber.org/papers/w25266.

Lee, M.D., Pincombe, B. and Welsh, M., 2005. An empirical evaluation of models of text document similarity. In: *Proceedings of the Annual Meeting of the Cognitive Science Society*, 27(27). https://escholarship.org/content/qt48g155nq/qt48g155nq.pdf

Jackman, S., 2000. Estimation and inference via Bayesian simulation: An introduction to Markov chain Monte Carlo. *American Journal of Political Science*, 44(2), pp.375–404. https://doi.org/10.2307/2669318.

Slapin, J.B. and Proksch, S.-O., 2008. A scaling model for estimating time-series party positions from texts. *American Journal of Political Science*, 52(3), pp.705–722. https://doi.org/10.1111/j.1540-5907.2008.00338.x.

Landauer, T.K., Foltz, P.W. and Laham, D., 1998. An introduction to latent semantic analysis. *Discourse Processes*, 25(2–3), pp.259–284. https://doi.org/10.1080/01638539809545028.

<u>Week 8 (Lecture 7): Unsupervised Techniques with Text Data II</u>

Various topic modelling approaches (e.g., Latent Dirichlet Allocation, Structural Topic Modelling).

Key readings:

Grimmer, Roberts and Stewart (2022, chps. 13)

Additional readings:

Ash, E., Morelli, M. and Vannoni, M., 2025. More laws, more growth? Evidence from US states. *Journal of Political Economy*, 133(7), pp.000–000.

Barron, A.T., Huang, J., Spang, R.L. and DeDeo, S., 2018. Individuals, institutions, and innovation in the debates of the French Revolution. *Proceedings of the National Academy of Sciences*, 115(18), pp.4607–4612.

Blei, D.M. and Jordan, M.I., 2006. Variational inference for Dirichlet process mixtures. *Bayesian Analysis*, 1(1), pp.121–143.

Blei, D.M. and Lafferty, J.D., 2007. A correlated topic model of science. *Annals of Applied Statistics*, 1(1), pp.17–35.

Blei, D.M., Ng, A.Y. and Jordan, M.I., 2003. Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3(Jan), pp.993–1022.

Fong, C. and Grimmer, J., 2016. Discovery of treatments from text corpora. In: *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Berlin: Association for Computational Linguistics, pp.1600–1609.

Grimmer, J., 2010. A Bayesian hierarchical topic model for political texts: Measuring expressed agendas in Senate press releases. *Political Analysis*, 18(1), pp.1–35.

Hansen, S., McMahon, M. and Prat, A., 2018. Transparency and deliberation within the FOMC: A computational linguistics approach. *Quarterly Journal of Economics*, 133(2), pp.801–870.

Quinn, K.M., Monroe, B.L., Colaresi, M., Crespin, M.H. and Radev, D.R., 2010. How to analyze political attention with minimal assumptions and costs. *American Journal of Political Science*, 54(1), pp.209–228.

Roberts, M.E., Stewart, B.M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S.K., Albertson, B. and Rand, D.G., 2014. Structural topic models for open-ended survey responses. *American Journal of Political Science*, 58(4), pp.1064–1082.

Wallach, H.M., Murray, I., Salakhutdinov, R. and Mimno, D., 2009. Evaluation methods for topic models. In: *Proceedings of the 26th Annual International Conference on Machine Learning*. Montreal: ACM, pp.1105–1112.

<u>Week 9 (Lecture 8): Word Embeddings</u>

Examination of word embeddings for semantic analysis, covering methods such as Word2Vec, GloVe, and embeddings derived from language models.

Key readings:
Grimmer, Roberts and Stewart (2022, chps. 8)

Additional readings:

Allen, C. and Hospedales, T., 2019. Analogies explained: Towards understanding word embeddings. In: *International Conference on Machine Learning*. PMLR, pp.223–231.

Antoniak, M. and Mimno, D., 2018. Evaluating the stability of embedding-based word similarities. *Transactions of the Association for Computational Linguistics*, 6, pp.107–119.

Arora, S., Li, Y., Liang, Y., Ma, T. and Risteski, A., 2018. Linear algebraic structure of word senses, with applications to polysemy. *Transactions of the Association for Computational Linguistics*, 6, pp.483–495.

Ash, E., Chen, D.L. and Ornaghi, A., 2024. Gender attitudes in the judiciary: Evidence from US circuit courts. *American Economic Journal: Applied Economics*, 16(1), pp.314–350.

Bhatia, S., Lau, J.H. and Baldwin, T., 2016. Automatic labelling of topics with neural embeddings. *arXiv preprint* arXiv:1612.05340.

Bojanowski, P., Grave, E., Joulin, A. and Mikolov, T., 2017. Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, 5, pp.135–146.

Goldberg, Y. and Levy, O., 2014. word2vec explained: Deriving Mikolov et al.'s negative-sampling word-embedding method. *arXiv preprint* arXiv:1402.3722.

Hamilton, W.L., Clark, K., Leskovec, J. and Jurafsky, D., 2016. Inducing domain-specific sentiment lexicons from unlabeled corpora. In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp.595–605.

Kusner, M., Sun, Y., Kolkin, N. and Weinberger, K., 2015. From word embeddings to document distances. In: *International Conference on Machine Learning*. PMLR, pp.957–966.

Meyer, D., 2016. How exactly does word2vec work. Uoregon.edu, Brocade.com, pp.1–18. Available at: http://www.1-4-5.net/~dmm/ml/how_does_word2vec_work.pdf.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S. and Dean, J., 2013. Distributed representations of words and phrases and their compositionality. In: *Advances in Neural Information Processing Systems*, 26.

Peters, M.E., Ruder, S. and Smith, N.A., 2019. To tune or not to tune? Adapting pretrained representations to diverse tasks. *arXiv preprint* arXiv:1903.05987.

Molino, P., 2017. Word embeddings: Past, present, and future. Available at https://www.youtube.com/watch?v=AsGf8cV4hqg.

Rodriguez, P.L. and Spirling, A., 2022. Word embeddings: What works, what doesn't, and how to tell the difference for applied research. *Journal of Politics*, 84(1), pp.101–115.

Ruder, S., 2017. Approximating the softmax. Available at https://opendatascience.com/on-word-embeddings-part-2-approximating-the-softmax/.

Rudolph, M., Ruiz, F., Athey, S. and Blei, D., 2017. Structured embedding models for grouped data. In: *Advances in Neural Information Processing Systems*, 30.

Week 10 (Lecture 9): Neural Network-Based Models

Introduction to neural networks for text analysis, with a focus on recurrent neural networks, convolutional neural networks, and transformer architectures.

Key readings:
Goldberg, Y. (2017). *Neural network methods in natural language processing.* Morgan & Claypool Publishers. Chps. 3-5, 8 (accessible via university library)

Additional readings:

Ash, E., Durante, R., Grebenshchikova, M. and Schwarz, C., 2021. Visual representation and stereotypes in news media. *SSRN Working Paper.*

Iyyer, M., Enns, P., Boyd-Graber, J. and Resnik, P., 2014. Political ideology detection using recursive neural networks. In: *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Baltimore, MD: ACL, pp.1113–1122.

Joulin, A., Grave, E., Bojanowski, P. and Mikolov, T., 2016. Bag of tricks for efficient text classification. *arXiv preprint* arXiv:1607.01759.

Meursault, V., 2019. The language of earnings announcements. *SSRN Working Paper*.

Stammbach, D., Antoniak, M. and Ash, E., 2022. Heroes, villains, and victims, and GPT-3: Automated extraction of character roles without training data. *arXiv preprint* arXiv:2205.07557.

Stammbach, D. and Ash, E., 2020. e-fever: Explanations and summaries for automated fact checking. In: *Proceedings of the 2020 Truth and Trust Online (TTO 2020)*. pp.32–43.

Vamossy, D.F., 2021. Investor emotions and earnings announcements. *Journal of Behavioral and Experimental Finance*, 30, 100474.

Webersinke, N., Kraus, M., Bingler, J.A. and Leippold, M., 2021. ClimateBERT: A pretrained language model for climate-related text. *arXiv preprint* arXiv:2110.12010.

<u>Week 11 (Lecture 10): Advanced Applications of Large Language Models (LLMs)</u>

Exploration of recent developments in LLMs, with an emphasis on their applications, limitations, and ethical considerations in text analysis.

Key readings:

Goldberg, Y. (2017). *Neural network methods in natural language processing*. Morgan & Claypool Publishers. Chp. 9 & 17

Additional readings:

Bingler, J.A., Kraus, M., Leippold, M. and Webersinke, N., 2022. Cheap talk and cherry-picking: What ClimateBERT has to say on corporate climate risk disclosures. *Finance Research Letters*, 47, 102776.

Bloem, P., 2019. Transformers from scratch. Available at: https://peterbloem.nl/blog/transformers.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J.D., Dhariwal, P., ... and Amodei, D., 2020. Language models are few-shot learners. In: *Advances in Neural Information Processing Systems*, 33, pp.1877–1901.

Devlin, J., Chang, M.W., Lee, K. and Toutanova, K., 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long and Short Papers)*. Minneapolis, MN: ACL, pp.4171–4186.

He, K., Zhang, X., Ren, S. and Sun, J., 2016. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, pp.770–778.

Nie, A., Bennett, E. and Goodman, N., 2019. DisSent: Learning sentence representations from explicit discourse relations. In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Florence: ACL, pp.4497–4510.

Pei, J. and Jurgens, D., 2020. Quantifying intimacy in language. *arXiv preprint* arXiv:2011.03020.

Shree, P., 2020. The journey of OpenAI GPT models. Available at: https://medium.com/walmartglobaltech/the-journey-of-open-ai-gpt-models-32d95b7b7fb2.

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D. and Sutskever, I., 2019. Language models are unsupervised multitask learners. *OpenAI Blog*, 1(8), p.9.

Reimers, N. and Gurevych, I., 2019. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. *arXiv preprint* arXiv:1908.10084.

Sutton, R.S. and Barto, A.G., 2020. *Reinforcement learning: An introduction*. 2nd ed. Cambridge, MA: MIT Press.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. and Polosukhin, I., 2017. Attention is all you need. In: *Advances in Neural Information Processing Systems*, 30.