Course:
Computational Analysis of Political Text, Audio, and Images

Fall 2023

Time	⌚:	Fall 2023, Wednesdays 09 🕘 – 12 🕛
Location	🏫:	1330-038
Instructor	👨‍🏫:	Mathias Rask (📧 mathiasrask@ps.au.dk)
Office Hour	🕰️:	Friday 09.15-10.15
Exam	📜:	7-day take-home (December 12 at 09:00 am)
Course language	🗣️:	Danish 🇩🇰

Course Schedule – Short

Class	Date	Main Topic		Topic
1	September 6	Python	I	Introduction to course + Python
2	September 7		II	Python lab session
3	September 13		I	Learning from data
4	September 20	ML Basics	II	ML lab session
5	September 27		III	Neural nets
6	October 4		I	Text basics
7	October 11	Text	II	Topic models and dictionaries
8	October 25		III	Embeddings
9	November 1		I	Audio basics
10	November 1	Audio	II	Audio measurement
11	November 8		III	Speech and speaker recognition
12	November 15		I	Image basics
13	November 22	Images	II	Object detection and face recognition
14	November 29		III	Image classification

* Wednesday, October 18 does not feature any class due to the Autumn break in week 42.

Course Schedule – Detailed

🔗 https://kursuskatalog.au.dk/en/course/119313/222E23-Computational-Methods-and-Analysis-of-Political-Text-Audio-and-Image

Readings marked by a (∗) are in the curriculum. Supplemental readings are marked by a (▷).

Main Topic 0: Python

Class 1: Course Introduction + Python 🐍

Date: 6, September 2023, 09-12 am
Location: 1330-038

Lecture 👨‍🏫

1. Course introduction (structure, classes, exam ...)
2. Defining and locating computational social science/computational analysis
3. Good research questions using text, audio, and image data
4. Computational measurement of social science concepts
5. Promises ✔️ and pitfalls ❌ of computational social science
6. Python

Coding Tour 🤓+🛤️

- Getting started with 🐍: local setup (Anaconda), virtual environments, and modules
- Syntax
- Workflow
- Operators
- Data types: numbers, strings, lists, and dictionaries
- Functions: arguments (keyword, positional, and default) and scope
- Control flow: if/else, continue, while, break, pass, ...)

Lab 🥼+💻

1. Implementation of a rock, scissor, and paper game

Readings 📕

∗ J. Y. Kim and Y. M. M. Ng, "Teaching computational social science for all," *PS: Political Science & Politics*, vol. 55, no. 3, pp. 605–609, 2022 (number of pages: 4)

* Y. Theocharis and A. Jungherr, “Computational social science and the study of political communication,” *Political Communication*, vol. 38, no. 1-2, pp. 1–22, 2021 (number of pages: 22)

* J. Grimmer and B. M. Stewart, “Text as data: The promise and pitfalls of automatic content analysis methods for political texts,” *Political analysis*, vol. 21, no. 3, pp. 267–297, 2013 (number of pages: 3)

  – **Section 2**

* G. Lin and C. Lucas, “An introduction to neural networks for the social sciences,” in *The Oxford Handbook of Methodological Pluralism in Political Science*, Oxford University Press, forthcoming. [Online]. Available: http://christopherlucas.org/files/PDFs/nn_chapter.pdf (number of pages: 4)

  – **Section 1-2**

* W. McKinney, *Python for data analysis: Data wrangling with Pandas, NumPy, and IPython*, 2nd ed. ” O’Reilly Media, Inc.”, 2017. Accessible through AUL: https://soeg.kb.dk/permalink/45KBDK_KGL/1f0go08/cdi_askewsholts_vlebooks_9781491957639 (number of pages: 52)

  – **Section 1.4**, **Section 2.1**, **Section 2.2 (p. 17-20)**, **Section 2.3 (p. 30-38, 46-50)**, **Chapter 3**

→ Total number of pages: 85

▷ C. A. Bail, “Can generative ai improve social science?,” 2023

Resources 🔗

– https://www.youtube.com/watch?v=rfscVS0vtbw

– https://github.com/Akuli/python-tutorial/tree/master

Class 2: Lab

Date: 7, September 2023, 12-15 pm
Location: 1323-118

Lecture 👨‍🏫

1. Class 1 continued ...

Coding Tour 🤓+🛤️

- Loops and comprehensions
- Classes
- Errors
- Reading and writing files
- Go-to 🐍 modules (e.g. NumPy, Pandas)

Lab 🥼+💻

1. Reading and loading files
2. Classes and methods
3. NumPy and Pandas exercises

Readings 📕

* W. McKinney, *Python for data analysis: Data wrangling with Pandas, NumPy, and IPython*, 2nd ed. " O'Reilly Media, Inc.", 2017. Accessible through AUL: https://soeg.kb.dk/permalink/45KBDK_KGL/1f0go08/cdi_askewsholts_vlebooks_9781491957639 (number of pages: 63)
    - **Chapter 4, 6**

→ Total number of pages: 63

Resources 🔗

- https://www.dataquest.io/blog/using-classes-in-python/
- https://www.kaggle.com/code/themlphdstudent/learn-numpy-numpy-50-exercises-and-solution
- https://github.com/guipsamora/pandas_exercises/tree/master
- https://www.kaggle.com/code/icarofreire/pandas-24-useful-exercises-with-solutions

Main Topic 1: ML Basics

Class 3: Learning From Data

Date: 6, September 2023, 09-12 am
Location: 1330-038

Lecture 👨‍🏫

1. Learning types: supervised, unsupervised, and transfer learning
2. Model validation
   - Model capacity: Overfitting and underfitting
   - Hyperparameters and validation sets
   - Loss functions and metrics
3. Measurement validation
4. Logistic regression and Support Vector Machines (SVMs)
5. Principal Component Analysis (PCA) and k-means clustering

Coding Tour 🤓+🛤️

- Overview of the `scikit-learn` module

Lab 🥼+💻

- No lab for this class

Readings 📕

∗ M. Laurer *et al.*, "Less annotating, more classifying: Addressing the data scarcity issue of supervised machine learning with deep transfer learning and bert-nli," *Political Analysis*, pp. 1–33, 2022 (number of pages: 3)

  – **Section 2.1-2.2**

∗ Z. Terechshenko *et al.*, "A comparison of methods in political science text classification: Transfer learning language models for politics," *Available at SSRN 3724644*, 2020. DOI: http://dx.doi.org/10.2139/ssrn.3724644 (number of pages: 1)

  – **Section 2**

∗ I. Goodfellow *et al.*, *Deep learning*. MIT press, 2016. [Online]. Available: https://www.deeplearningbook.org/ Freely accessible at: https://www.deeplearningbook.org/ (number of pages: 39)

  – **Section 5.1-5.3, 5.7-5.8, 5.10**

→ Total number of pages: 40

Resources 🔗

– https://setosa.io/ev/principal-component-analysis/

Class 4: Lab

Date: 20, September 2023, 09-12 am
Location: 1330-038

Lecture 👨‍🏫

1. Class 3 continued ...

Coding Tour 🤓+🛤️

- Class 3 continued ...

Lab 🥼+💻

- Building your own classifier using an algorithm of your choice from `scikit-learn`
- Implement PCA and k-means with `scikit-learn`

Readings 📕

* Same as Class 3

Resources 🔗

– Same as Class 3

Class 5: Neural Networks

Date: 27, September 2023, 09-12 am
Location: 1330-038

Lecture 👨‍🏫

1. From machine – to deep learning
2. From logistic regression to neural networks
3. How do networks learn?
4. A primer on architectures: CNN, LTSM, RNN, Transformers

Coding Tour 🤓+🛤️

- Introduction to PyTorch

Lab 🥼+💻

- Implementation of a vanilla neural network in PyTorch

Readings 📕

* M. A. Nielsen, *Neural networks and deep learning*. Determination press San Francisco, CA, USA, 2015, vol. 25 Freely accessible at: http://neuralnetworksanddeeplearning.com/ (number of pages: 35)
  - **Chapter 1 p. 1-35**
* G. Lin and C. Lucas, "An introduction to neural networks for the social sciences," in *The Oxford Handbook of Methodological Pluralism in Political Science*, Oxford University Press, forthcoming. [Online]. Available: http://christopherlucas.org/files/PDFs/nn_chapter.pdf (number of pages: 7)
  - **Section 3**

→ Total number of pages: 42

Resources 🔗

- https://towardsdatascience.com/beginners-ask-how-many-hidden-layers-neurons-to-use-in-artificial-neural-networks-51466afa0d3e
- https://analyticsindiamag.com/xor-problem-with-neural-networks-an-explanation-for-beginners/
- https://playground.tensorflow.org/
- 3Blue1Brown video on the theory neural networks https://www.youtube.com/watch?v=aircAruvnKk&t=1s
- 3Blue1Brown video on how neural networks learn https://www.youtube.com/watch?v=IHZwWFHWa-w&list=PLZHQObOWTQDNU6R1_67000Dx_ZCJB-3pi&index=2

Main Topic 2: Text 📜

Class 6: Text Basics

Date: 4, October 2023, 09-12 am
Location: 1330-038

Lecture 👨‍🏫

1. Vectorization: Representing text as numbers
2. Preprocessing: Why do we need it?

Coding Tour 🤓+🛤️

- To be announced.

Lab 🥼+💻

- Vectorize and preprocess raw text
  - Vectorize: bag-of-words and tf-idf
  - What is a "good" numerical representation of text?
  - Preprocessing: Tokenization, stopwords, special characters, numbers, stemming, lemmatization, lower casing, removing features
  - Word descriptives

Readings 📕

* J. Grimmer and B. M. Stewart, "Text as data: The promise and pitfalls of automatic content analysis methods for political texts," *Political analysis*, vol. 21, no. 3, pp. 267–297, 2013 (number of pages: 2)
  - **Section 4**
* M. J. Denny and A. Spirling, "Text preprocessing for unsupervised learning: Why it matters, when it misleads, and what to do about it," *Political Analysis*, vol. 26, no. 2, pp. 168–189, 2018 (number of pages: 12)
  - **Section 1-5**
* S. Akuma *et al.*, "Comparing bag of words and tf-idf with different models for hate speech detection from live tweets," *International Journal of Information Technology*, vol. 14, no. 7, pp. 3629–3635, 2022 (number of pages: 1)
  - **Section 3.3**

→ Total number of pages: 15

Resources 🔗

- https://towardsdatascience.com/basics-of-countvectorizer-e26677900f9c
- https://okan.cloud/posts/2021-04-08-text-vectorization-using-python-term-document-matrix/

Class 7: Topic Models and Dictionaries

Date: 11, October 2023, 09-12 am
Location: 1330-038

Lecture 👨‍🏫

1. What are (unsupervised) topic models and how do they work?
2. What assumptions underlie topic models? What are their strength/weaknesses?
3. What is the core idea of dictionary-based methods?
4. Off-the-shelf vs. creating your own dictionary

Coding Tour 🤓+🛤️

- Introduction to `Gensim`

Lab 🥼+💻

- Implement unsupervised LDA model on a text corpus
- Create and apply a dummy dictionary

Readings 📕

∗ D. M. Blei, "Probabilistic topic models," *Communications of the ACM*, vol. 55, no. 4, pp. 77–84, 2012. DOI: `http://doi.acm.org/10.1145/2133806.2133826`

∗ C. Wratil *et al.*, "Government rhetoric and the representation of public opinion in international negotiations," *American Political Science Review*, pp. 1–18, 2022

∗ B. C. Silva and S.-O. Proksch, "Politicians unleashed? political communication on twitter and in parliament in western europe," *Political science research and methods*, vol. 10, no. 4, pp. 776–792, 2022

→ Total number of pages: 40

Resources 🔗

– `https://rpubs.com/chelseyhill/672546`

Class 8: Embeddings

Date: 25, October 2023, 09-12 am
Location: 1330-038

Lecture 👨‍🏫

1. What is the general idea about word embeddings?
2. Word semantics and the distributional hypothesis
3. How do they improve upon bag-of-words and tf-idf representations of text?
4. Using word embeddings as features vs. a measure

Coding Tour 🤓+🛤️

- Word2vec and Doc2vec
- Accessing pretrained language models in Python using transfer learning
- Word and document similarity

Lab 🥼+💻

- Measuring polarization with word embeddings

Readings 📕

∗ P. L. Rodriguez and A. Spirling, "Word embeddings: What works, what doesn't, and how to tell the difference for applied research," *The Journal of Politics*, vol. 84, no. 1, pp. 101–115, 2022 (number of pages: 14)

∗ L. Rheault and C. Cochrane, "Word embeddings for the analysis of ideological placement in parliamentary corpora," *Political Analysis*, vol. 28, no. 1, pp. 112–133, 2020 (number of pages: 21)

→ Total number of pages: 35

▷ E. Rudkowsky *et al.*, "More than bags of words: Sentiment analysis with word embeddings," *Communication Methods and Measures*, vol. 12, no. 2-3, pp. 140–157, 2018

▷ T. Mikolov *et al.*, "Distributed representations of words and phrases and their compositionality," *Advances in neural information processing systems*, vol. 26, 2013

▷ A. C. Kozlowski *et al.*, "The geometry of culture: Analyzing the meanings of class through word embeddings," *American Sociological Review*, vol. 84, no. 5, pp. 905–949, 2019

▷ P. L. Rodriguez *et al.*, "Multilanguage word embeddings for social scientists: Estimation, inference and validation resources for 157 languages,"

▷ P. L. Rodriguez *et al.*, "Embedding regression: Models for context-specific description and inference," *American Political Science Review*, pp. 1–20, 2023

▷ C. Barrie *et al.*, "Measuring media freedom," 2023

Resources 🔗

–

Main Topic 3: Audio 🔊

Class 9: Audio Basics

Date: 1, November 2023, 09-12 am
Location: 1441-210

Lecture 👨‍🏫

1. Audio vs. text
2. Sound theory: What do we actually measure?
3. Basic concepts: Sampling rate, amplitude, frequency, ...
4. Digital signal processing fundamentals
5. Audio representations: waveform and spectrogram
6. Audio software

Coding Tour 🤓+🛤️

- To be announced.

Lab 🥼+💻

- Visualization of the waveform
- Convert an audio file from the time domain to the frequency domain

Readings 📕

* L. Rheault and S. Borwein, "Audio as data," in *Elgar Encyclopedia of Technology and Politics*, Edward Elgar Publishing, pp. 86–90 Available: https://lrheault.github.io/downloads/AudioAsData_RheaultBorwein.pdf (number of pages: 7)
* F. Camastra and A. Vinciarelli, *Machine learning for audio, image and video analysis: theory and applications*. Springer, 2015 Accessible through AUL: https://soeg.kb.dk/discovery/fulldisplay?docid=alma99123011568605763&context=L&vid=45KBDK_KGL:KGL&lang=da&search_scope=MyInst_and_CI&adaptor=Local%20Search%20Engine (number of pages: 42)

→ Total number of pages: 49

Resources 🔗

- https://github.com/YannickJadoul/Parselmouth/tree/stable

Class 10: Audio Measurement

Date: 1, November 2023, 12-15 am
Location: 1441-210

Lecture 👨‍🏫

1. Supervised vs. unsupervised measurement
2. Time - vs. frequency features
3. Measurement error
4. Acoustic analysis
5. Audio classification

Coding Tour 🤓+🛤️

- Introduction of `Parselmouth` (Pythonic `Praat` binding)

Lab 🥼+💻

- Pitch estimation with and without speaker standardization

Readings 📕

∗ B. J. Dietrich *et al.*, “Pitch perfect: Vocal pitch and the emotional intensity of congressional speech,” *American Political Science Review*, vol. 113, no. 4, pp. 941–962, 2019 (number of pages: 21)

∗ D. Knox and C. Lucas, “A dynamic model of speech for the social sciences,” *American Political Science Review*, vol. 115, no. 2, pp. 649–666, 2021 (number of pages: 17)

∗ L. Rheault and S. Borwein, “Multimodal techniques for the study of a ect in political videos,” Working Paper, Tech. Rep., 2019 Available: `https://polmeth.mit.edu/sites/default/files/documents/RheaultBorwein_PolMeth2019.pdf` (number of pages: 32)

∗ M. Rask and F. Hjorth, “Nonverbal-based measures of elite conflict and polarization,” *Working Paper*, pp. 1–25, 2023 (number of pages: 25)

∗ M. Vainio *et al.*, “The power of prosody and prosody of power: An acoustic analysis of finnish parliamentary speech,” *arXiv preprint arXiv:2305.16040*, 2023 (number of pages: 4)

→ Total number of pages: 99

Resources 🔗

- No links available

Class 11: Speech and Speaker Recognition

Date: 8, November 2023, 09-12 am
Location: 1330-038

Lecture 👨‍🏫

1. What are automatic speech recognition and diarization?
2. Why is alignment crucial to computational audio analysis?
3. What information is contained in speaker embeddings?

Coding Tour 🤓+🛤️

- Introduce neural diarization with `pyannote.audio`
- ASR using `WhisperX` or `Faster Whisper` or plain `Whisper`

Lab 🥼+💻

- Apply diarization and ASR on a political debate
- Visualize speaker embeddings using dimensionality reduction

Readings 📕

* S.-O. Proksch *et al.*, "Testing the validity of automatic speech recognition for political text analysis," *Political Analysis*, vol. 27, no. 3, pp. 339–359, 2019 (number of pages: 20)
* A. Tarr *et al.*, "Automated coding of political campaign advertisement videos: An empirical validation study," *Political Analysis*, pp. 1–21, 2022 (number of pages: 3)
  - **Section 3.2**
* M. Neumann, "Hooked with phonetics: The strategic use of style-shifting in political rhetoric," in *Annual Meeting of the American Political Science Association. Washington, DC*, 2019 Available: `https://markusneumann.github.io/files/Neumann_APSA.pdf` (number of pages: 44)
* M. Rask, "Automated annotation of political speech recordings," *Working Paper*, pp. 1–20, 2023 (number of pages: 20)

→ Total number of pages: 87

Resources 🔗

- `https://github.com/resemble-ai/Resemblyzer`
- `https://huggingface.co/pyannote/speaker-diarization`

Main Topic 4: Images 🖼️

Class 12: Image Basics

Date: 15, November 2023, 09-12 am
Location: 1330-038

Lecture 👨‍🏫

1. Why study images as a social scientist?
2. Representing images (i.e. pixels) as a matrix
3. Image channels and color spaces
4. Software (`OpenCV`, `scikit-image`)
5. Basic image operations

Coding Tour 🤓+🛤️

- Introduction to `OpenCV`

Lab 🥼+💻

- Conversion of an image to a matrix
- Display an image
- Image processing

Readings 📕

∗ N. Webb Williams, "What type of data are images?" *Available at SSRN 4012789*, 2023 Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4012789 (number of pages: 21)

∗ N. W. Williams *et al.*, *Images as data for social science research: An introduction to convolutional neural nets for image classification*. Cambridge University Press, 2020 (number of pages: 15)

  – **Section 1, Section 4.2**

∗ M. Torres and F. Cantú, "Learning to see: Convolutional neural networks for the analysis of social science data," *Political Analysis*, vol. 30, no. 1, pp. 113–131, 2022 (number of pages: 1)

  – **Section 2.1**

→ Total number of pages: 16

▷ E. P. Bucy, "Politics through machine eyes: What computer vision allows us to see," *Journal of Visual Political Communication*, vol. 10, no. 1, pp. 59–68, 2023

Resources 🔗

- https://scikit-image.org/docs/stable/
- https://docs.opencv.org/4.x/d6/d00/tutorial_py_root.html
- https://github.com/yilangpeng/athec/tree/main

Class 13: Object Detection and Face Recognition

Date: 22, November 2023, 09-12 am
Location: 1330-038

Lecture 👨‍🏫

1. Introduction to convolutional neural networks (CNNs)
2. Objectives of object detection and face recognition

Coding Tour 🤓+🛤️

- Introduction to `face_recognition` and `py-feat`

Lab 🥼+💻

- Face detection + facial expressions using `py-feat`
- Face recognition using `face_recognition`

Readings 📕

∗ M. Torres and F. Cantú, "Learning to see: Convolutional neural networks for the analysis of social science data," *Political Analysis*, vol. 30, no. 1, pp. 113–131, 2022 (number of pages: 11)

  – **Section 2-4**

∗ M. A. Nielsen, *Neural networks and deep learning*. Determination press San Francisco, CA, USA, 2015, vol. 25 (number of pages: 18)

  – **Section 6.0-6.2**

∗ C. Boussalis *et al.*, "Mad, sad, but mostly glad: How men and women in politics communicate using emotions in images," 2022 (number of pages: 18)

∗ S. H. R. Rasmussen *et al.*, "Using deep learning to predict ideology from facial photographs: Expressions, beauty, and extra-facial information," *Scientific Reports*, vol. 13, no. 1, p. 5257, 2023 (number of pages: 7)

→ Total number of pages: 54

▷ B. J. Dietrich and M. L. Sands, "Seeing racial avoidance on new york city streets," *Nature human behaviour*, pp. 1–7, 2023

▷ B. J. Dietrich, "Using motion detection to measure social polarization in the us house of representatives," *Political Analysis*, vol. 29, no. 2, pp. 250–259, 2021

Resources 🔗

- https://github.com/ageitgey/face_recognition
- https://towardsdatascience.com/face-detection-in-2-minutes-using-opencv-python-90f89d7c0f81
- https://py-feat.org/pages/intro.html

Class 14: Image Classification

Date: 29, November 2023, 09-12 am
Location: 1330-038

Lecture 👨‍🏫

1. Image classification overview
2. Transfer learning and CNNs
3. Popular classifiers: AlexNet, VGG-16, GoogleNet, and Resnet

Coding Tour 🤓+🛤️

- Fine-tuning of CNN classifier using PyTorch

Lab 🥼+💻

- Classification of hand-written digits using the MNIST dataset

Readings 📕

* G. Lin and C. Lucas, "An introduction to neural networks for the social sciences," in *The Oxford Handbook of Methodological Pluralism in Political Science*, Oxford University Press, forthcoming. [Online]. Available: http://christopherlucas.org/files/PDFs/nn_chapter.pdf (number of pages: 3)
  - **Section 4.1**
* M. Torres and F. Cantú, "Learning to see: Convolutional neural networks for the analysis of social science data," *Political Analysis*, vol. 30, no. 1, pp. 113–131, 2022 (number of pages: 4)
  - **Section 5**
* N. Xi *et al.*, "Understanding the political ideology of legislators from social media images," in *Proceedings of the international aaai conference on web and social media*, vol. 14, 2020, pp. 726–737 (number of pages: 9)

→ Total number of pages: 16

Resources 🔗

- https://colab.research.google.com/drive/1KFHwz8wjDdcFfsTmXfo-gwkKc-itN3MS#scrollTo=WXF1RoZFwF0j

References

[1] J. Y. Kim and Y. M. M. Ng, "Teaching computational social science for all," *PS: Political Science & Politics*, vol. 55, no. 3, pp. 605–609, 2022.

[2] Y. Theocharis and A. Jungherr, "Computational social science and the study of political communication," *Political Communication*, vol. 38, no. 1-2, pp. 1–22, 2021.

[3] J. Grimmer and B. M. Stewart, "Text as data: The promise and pitfalls of automatic content analysis methods for political texts," *Political analysis*, vol. 21, no. 3, pp. 267–297, 2013.

[4] G. Lin and C. Lucas, "An introduction to neural networks for the social sciences," in *The Oxford Handbook of Methodological Pluralism in Political Science*, Oxford University Press, forthcoming. [Online]. Available: http://christopherlucas.org/files/PDFs/nn_chapter.pdf.

[5] W. McKinney, *Python for data analysis: Data wrangling with Pandas, NumPy, and IPython*, 2nd ed. " O'Reilly Media, Inc.", 2017.

[6] C. A. Bail, "Can generative ai improve social science?," 2023.

[7] M. Laurer, W. van Atteveldt, A. Casas, and K. Welbers, "Less annotating, more classifying: Addressing the data scarcity issue of supervised machine learning with deep transfer learning and bert-nli," *Political Analysis*, pp. 1–33, 2022.

[8] Z. Terechshenko *et al.*, "A comparison of methods in political science text classification: Transfer learning language models for politics," *Available at SSRN 3724644*, 2020. DOI: http://dx.doi.org/10.2139/ssrn.3724644.

[9] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016. [Online]. Available: https://www.deeplearningbook.org/.

[10] M. A. Nielsen, *Neural networks and deep learning*. Determination press San Francisco, CA, USA, 2015, vol. 25.

[11] M. J. Denny and A. Spirling, "Text preprocessing for unsupervised learning: Why it matters, when it misleads, and what to do about it," *Political Analysis*, vol. 26, no. 2, pp. 168–189, 2018.

[12] S. Akuma, T. Lubem, and I. T. Adom, "Comparing bag of words and tf-idf with different models for hate speech detection from live tweets," *International Journal of Information Technology*, vol. 14, no. 7, pp. 3629–3635, 2022.

[13] D. M. Blei, "Probabilistic topic models," *Communications of the ACM*, vol. 55, no. 4, pp. 77–84, 2012. DOI: http://doi.acm.org/10.1145/2133806.2133826.

[14] C. Wratil, J. Waeckerle, and S.-O. Proksch, "Government rhetoric and the representation of public opinion in international negotiations," *American Political Science Review*, pp. 1–18, 2022.

[15] B. C. Silva and S.-O. Proksch, "Politicians unleashed? political communication on twitter and in parliament in western europe," *Political science research and methods*, vol. 10, no. 4, pp. 776–792, 2022.

[16] P. L. Rodriguez and A. Spirling, "Word embeddings: What works, what doesn't, and how to tell the difference for applied research," *The Journal of Politics*, vol. 84, no. 1, pp. 101–115, 2022.

[17] L. Rheault and C. Cochrane, "Word embeddings for the analysis of ideological placement in parliamentary corpora," *Political Analysis*, vol. 28, no. 1, pp. 112–133, 2020.

[18] E. Rudkowsky, M. Haselmayer, M. Wastian, M. Jenny, Š. Emrich, and M. Sedlmair, "More than bags of words: Sentiment analysis with word embeddings," *Communication Methods and Measures*, vol. 12, no. 2-3, pp. 140–157, 2018.

[19] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," *Advances in neural information processing systems*, vol. 26, 2013.

[20] A. C. Kozlowski, M. Taddy, and J. A. Evans, "The geometry of culture: Analyzing the meanings of class through word embeddings," *American Sociological Review*, vol. 84, no. 5, pp. 905–949, 2019.

[21] P. L. Rodriguez, A. Spirling, B. M. Stewart, and E. M. Wirsching, "Multilanguage word embeddings for social scientists: Estimation, inference and validation resources for 157 languages,"

[22] P. L. Rodriguez, A. Spirling, and B. M. Stewart, "Embedding regression: Models for context-specific description and inference," *American Political Science Review*, pp. 1–20, 2023.

[23] C. Barrie, N. Ketchley, A. Siegel, and M. Bagdouri, "Measuring media freedom," 2023.

[24] L. Rheault and S. Borwein, "Audio as data," in *Elgar Encyclopedia of Technology and Politics*, Edward Elgar Publishing, pp. 86–90.

[25] F. Camastra and A. Vinciarelli, *Machine learning for audio, image and video analysis: theory and applications*. Springer, 2015.

[26] B. J. Dietrich, M. Hayes, and D. Z. O'Brien, "Pitch perfect: Vocal pitch and the emotional intensity of congressional speech," *American Political Science Review*, vol. 113, no. 4, pp. 941–962, 2019.

[27] D. Knox and C. Lucas, "A dynamic model of speech for the social sciences," *American Political Science Review*, vol. 115, no. 2, pp. 649–666, 2021.

[28] L. Rheault and S. Borwein, "Multimodal techniques for the study of a ect in political videos," Working Paper, Tech. Rep., 2019.

[29] M. Rask and F. Hjorth, "Nonverbal-based measures of elite conflict and polarization," *Working Paper*, pp. 1–25, 2023.

[30] M. Vainio, A. Suni, J. Šimko, and S. Kakouros, "The power of prosody and prosody of power: An acoustic analysis of finnish parliamentary speech," *arXiv preprint arXiv:2305.16040*, 2023.

[31] S.-O. Proksch, C. Wratil, and J. Wäckerle, "Testing the validity of automatic speech recognition for political text analysis," *Political Analysis*, vol. 27, no. 3, pp. 339–359, 2019.

[32] A. Tarr, J. Hwang, and K. Imai, "Automated coding of political campaign advertisement videos: An empirical validation study," *Political Analysis*, pp. 1–21, 2022.

[33] M. Neumann, "Hooked with phonetics: The strategic use of style-shifting in political rhetoric," in *Annual Meeting of the American Political Science Association. Washington, DC*, 2019.

[34] M. Rask, “Automated annotation of political speech recordings,” *Working Paper*, pp. 1–20, 2023.

[35] N. Webb Williams, “What type of data are images?” *Available at SSRN 4012789*, 2023.

[36] N. W. Williams, A. Casas, and J. D. Wilkerson, *Images as data for social science research: An introduction to convolutional neural nets for image classification.* Cambridge University Press, 2020.

[37] M. Torres and F. Cantú, “Learning to see: Convolutional neural networks for the analysis of social science data,” *Political Analysis*, vol. 30, no. 1, pp. 113–131, 2022.

[38] E. P. Bucy, “Politics through machine eyes: What computer vision allows us to see,” *Journal of Visual Political Communication*, vol. 10, no. 1, pp. 59–68, 2023.

[39] C. Boussalis, T. Coan, and M. Holman, “Mad, sad, but mostly glad: How men and women in politics communicate using emotions in images,” 2022.

[40] S. H. R. Rasmussen, S. G. Ludeke, and R. Klemmensen, “Using deep learning to predict ideology from facial photographs: Expressions, beauty, and extra-facial information,” *Scientific Reports*, vol. 13, no. 1, p. 5257, 2023.

[41] B. J. Dietrich and M. L. Sands, “Seeing racial avoidance on new york city streets,” *Nature human behaviour*, pp. 1–7, 2023.

[42] B. J. Dietrich, “Using motion detection to measure social polarization in the us house of representatives,” *Political Analysis*, vol. 29, no. 2, pp. 250–259, 2021.

[43] N. Xi, D. Ma, M. Liou, Z. C. Steinert-Threlkeld, J. Anastasopoulos, and J. Joo, “Understanding the political ideology of legislators from social media images,” in *Proceedings of the international aaai conference on web and social media*, vol. 14, 2020, pp. 726–737.