

- [1] Wei Qi, Lie Gu, Hao Jiang, Xiang-Rong Chen and Hong-Jiang Zhang, "Integrating visual, audio and text analysis for news video," Proceedings 2000 International Conference on Image Processing (Cat. No.00CH37101), Vancouver, BC, Canada, 2000, pp. 520-523 vol.3, doi: 10.1109/ICIP.2000.899482. keywords: Text analysis;TV;Data mining;Optical character recognition software;Video recording;Information analysis;Speech analysis;Image analysis;Image segmentation;Gunshot detection systems,
- [2] J. Liedke and L. Wang, "News platform fact sheet," Pew Research Center's Journalism Project, <https://www.pewresearch.org/journalism/fact-sheet/news-platform-fact-sheet/> (accessed Feb. 13, 2024).
- [3] L. Galan, J. Osserman, T. Parker, and M. Taylor, "How Young People Consume News and the Implications for Mainstream Media", Creative Commons 4.0, Oxford, United Kingdom, rep., 2019
- [4] T. M. I. Project, "The news consumption habits of 16- to 40-year-olds," American Press Institute, <https://americanpressinstitute.org/the-news-consumption-habits-of-16-to-40-year-olds/> (accessed Feb. 13, 2024).
- [5] N. Newman, R. Fletcher, K. Eddy, C. T. Robertson, and R. K. Nielsen, "Digital News Report 2023" rep., 2023
- [6] J. Berger and K. L. Milkman, "What makes online content viral?," Journal of Marketing Research, vol. 49, no. 2, pp. 192–205, Apr. 2012. doi:10.1509/jmr.10.0353
- [7] B. Rimé, C. Finkenauer, O. Luminet, E. Zech, and P. Philippot, "Social sharing of emotion: New evidence and new questions," European Review of Social Psychology, vol. 9, no. 1, pp. 145–189, Jan. 1998. doi:10.1080/14792779843000072
- [8] D. Zeng, H. Chen, R. Lusch and S. -H. Li, "Social Media Analytics and Intelligence," in IEEE Intelligent Systems, vol. 25, no. 6, pp. 13-16, Nov.-Dec. 2010, doi: 10.1109/MIS.2010.151.
- [9] Wu He, Gongjun Yan, Mining Blogs And Forums To Understand the Use of Social Media in Customer Co-creation, The Computer Journal, Volume 58, Issue 9, September 2015, Pages 1909–1920,
- [10] M. Taboada, "Sentiment analysis: An overview from linguistics," Annual Review of Linguistics, vol. 2, no. 1, pp. 325–347, Jan. 2016. doi:10.1146/annurev-linguistics-011415-040518
- [11] H. J. Khasawneh, Z. A. Ghazal, W. M. Al-Khatib, A. M. Al-Hadi, and Z. M. Arabiyat, "Creating optimized machine learning pipelines for PV power generation forecasting using the grid search and tree-based pipeline optimization tool," Cogent Engineering, vol. 11, no. 1, p. 2, Mar. 2024. doi:10.1080/23311916.2024.2323818
- [12] J. Langseth, N. Vivatrat, and G. Sohn, "SCHEMA AND ETL TOOLS FOR STRUCTURED AND UNSTRUCTURED DATA," Dec. 7, 2010

- [13] D. Seenivasan, “ETL (extract, transform, load) best practices,” *International Journal of Computer Trends and Technology*, vol. 71, no. 1, pp. 40–44, Jan. 2023. doi:10.14445/22312803/ijctt-v71i1p106
- [14] Bylund, A., *Data Pipeline Design for Audit Analytics: Data Ingestion Tools Evaluation: Proof of Concept (Dissertation)*. 2023. Retrieved from <https://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-210572>
- [15] A. Radford, J. W. Kim, T. Xu, C. McLeavey, and I. Sutskever, “Robust Speech Recognition via Large-Scale Weak Supervision.” Dec. 6, 2022
- [16] A. Rao, *Transcribing Educational Videos Using Whisper: A preliminary study on using AI for transcribing educational videos*, Jul. 2023. doi: <https://doi.org/10.48550/arXiv.2307.03200>
- [17] R. Yakubovskiy and Y. Morozov, “Speech models training technologies comparison using word error rate,” *Advances in Cyber-Physical Systems*, vol. 8, no. 1, pp. 74–80, May 2023. doi:10.23939/acps2023.01.074
- [18] A. Trajkovska, T. Dimovski, R. Markoska, and Z. Kotevski, “13th International Conference on Applied Internet and Information Technologies AIIT2023, October 13th 2023, Bitola, Republic of North Macedonia,” in *Proceedings of the 13th International Conference on Applied Internet and Information Technologies AIIT 2023 13 October, 2023, Bitola, Republic of North Macedonia*, 2023, pp. 212–220
- [19] F. Junaedi, “Berita dan Jurnalisme,” in *Jurnalisme Penyiaran dan Reportase Televisi*, 1st ed, Jakarta, Rawamangun: PRENADAMEDIA GROUP, 2014, pp. 6–7
- [20] M. W. Van Alstyne, G. G. Parker, and S. P. Choudary, “Pipelines, platforms, and the new rules of strategy,” *Harvard business review*, vol. 94, no. 4, pp. 54–62, 2016.
- [21] A. Raj, J. Bosch, H. H. Olsson and T. J. Wang, "Modelling Data Pipelines," 2020 46th Euromicro Conference on Software Engineering and Advanced Applications (SEAA), Portoroz, Slovenia, 2020, pp. 13-20, doi: 10.1109/SEAA51224.2020.00014.
- [22] J. Densmore, “Data pipelines pocket reference.” O’Reilly Media, 2021.
- [23] H. Sun, S. Hu, S. McIntosh, and Y. Cao, “Big data trip classification on the new york city taxi and uber sensor network,” *Journal of Internet Technology*, vol. 19, no. 2, pp. 591–598, 2018
- [24] B. Deepa and K. Ramesh, "Production Level Data Pipeline Environment for Machine Learning Models," 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2021, pp. 404-407, doi: 10.1109/ICACCS51430.2021.9442035.
- [25] K. Kramer, *TOWARDS EVOLUTION CAPABILITIES IN DATA PIPELINES*, Aug. 2023. doi: <https://doi.org/10.48550/arXiv.2308.14591>
- [26] “Proof of concept (POC) vs. proof of value (POV): What do they mean for your business?,” Tenable®, <https://www.tenable.com/blog/proof-of-concept-poc-vs-proof-of-value-pov-what-do-they-mean-for-your-business> (accessed Jul. 3, 2024).

- [28] H. Zhong and X. Wang, "An empirical study on API usages from code search engine and local library," *Empirical Software Engineering*, vol. 28, no. 3, Apr. 2023. doi:10.1007/s10664-023-10304-z
- [29] I. Sriram and A. Khajeh-Hosseini, "Research Agenda in Cloud Technologies", *CoRR*, vol. abs/1001.3259, 2010.
- [30] Manish Mukhija, "A Resourceful Technique for virtual Machine Migration in Fog Computing", *International Journal of Innovative Science and Research Technology*, vol-6, issue-6, pp. 167-170, 2016.
- [31] NIST, <http://www.nist.gov/itl/cloud/index.cfm>
- [32] D. ZEYTIKAYA, "A didactic proposal for translator training with an educational learning theory 'connectivism': From cloud computing to cloud based Translation Systems," *JOURNAL OF ACADEMIC SOCIAL RESOURCES*, vol. 6, no. 27, pp. 1127–1134, Jan. 2021. doi:10.31569/asrjournal.274
- [33] S. Furui, T. Kikuchi, Y. Shinnaka, and C. Hori, "Speech-to-text and speech-to-speech summarization of spontaneous speech," *IEEE Transactions on Speech and Audio Processing*, vol. 12, no. 4, pp. 401–408, Jul. 2004. doi:10.1109/tsa.2004.828699
- [34] C. Park et al., "BTS: Back transcription for speech-to-text post-processor using text-to-speech-to-text," *Proceedings of the 8th Workshop on Asian Translation (WAT2021)*, 2021. doi:10.18653/v1/2021.wat-1.10
- [35] V. M. Reddy, T. Vaishnavi and K. P. Kumar, "Speech-to-Text and Text-to-Speech Recognition Using Deep Learning," *2023 2nd International Conference on Edge Computing and Applications (ICECAA)*, Namakkal, India, 2023, pp. 657-666, doi: 10.1109/ICECAA58104.2023.10212222.
- [36] "Learn how google improves speech models," *Google Assistant Help*, <https://support.google.com/assistant/answer/11140942?hl=en#zippy=%%202Cconventional-learning>. (accessed May 5, 2024).
- [37] K. Kreics, "Quality of analytics management of data pipelines for retail forecasting" thesis, Otaniemi, 2019
- [38] Apache. *Apache Airflow Documentation*. Accessed 3.04.2024. Accessible at: <https://airflow.apache.org/>
- [39] Luigi. *Luigi repository*. Accessed 5.04.2024. Accessible at: <https://github.com/spotify/luigi>
- [40] S. Paulakis, V. Tsetsos and S. Hadjiefthymiades, "Enterprise Job Scheduling for Clustered Environments," *10th IEEE International Symposium on Object and Component-Oriented Real-Time Distributed Computing (ISORC'07)*, Santorini, Greece, 2007, pp. 282-290, doi: 10.1109/ISORC.2007.34.

- [41] V. M. Ionescu, "The analysis of the performance of RabbitMQ and ActiveMQ," 2015 14th RoEduNet International Conference - Networking in Education and Research (RoEduNet NER), Craiova, Romania, 2015, pp. 132-137, doi: 10.1109/RoEduNet.2015.7311982.
- [42] Alexis Richardson, RabbitMQ - An open source message broker that just work, QCon 2009. Available at: <https://www.rabbitmq.com/resources/RabbitMQcon.pdf>
- [43] David Dossot, RabbitMQ Essentials, Packt Publishing, April 2014.
- [44] S. Shaik and N. Rao, "A Review of ElasticSearch: Performance Metrics and Challenges," International Journal on Recent and Innovation Trends in Computing and Communication, vol. 5, no. 11, pp. 222–229. doi:<https://www.academia.edu/download/56805561/1512628680-07-12-2017.pdf>
- [45] A. Ali and S. Renals, "Word error rate estimation without ASR output: E-WER2," Interspeech 2020, Oct. 2020. doi:10.21437/interspeech.2020-2357
- [46] J. Rugayan, T. Svendsen, and G. Salvi, "Semantically meaningful metrics for Norwegian ASR systems," Interspeech 2022, Sep. 2022. doi:10.21437/interspeech.2022-817
- [47] Andrew Cameron Morris, Viktoria Maier, and Phil Green. From WER and RIL to MER and WIL: improved evaluation measures for connected speech recognition. In Eighth International Conference on Spoken Language Processing, 2004.
- [48] A. K. Sheshadri, A. Rao Vijjini, and S. Kharbanda, "Wer-bert: Automatic wer estimation with Bert in a balanced ordinal classification paradigm," Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, 2021. doi:10.18653/v1/2021.eacl-main.320
- [49] E. Blessing and H. Klaus, 'Pipeline Automation: Techniques to create efficient and automated data processing pipelines for ML models', 12 2023.
- [50] J. Cao et al., 'A Comparative Analysis of Automatic Speech Recognition Errors in Small Group Classroom Discourse', 06 2023, pp. 250–262.
- [51] Gregory Dyke, Iris Howley, David Adamson, Rohit Kumar, and Carolyn Penstein Rosé. [n. d.]. Towards academically productive talk supported by conversational agents. In Proceedings of the International Conference on Intelligent Tutoring Systems. Springer, 459–476.
- [52] Michael Flor, Su-Youn Yoon, Jiangang Hao, Lei Liu, and Alina von Davier. 2016. Automated classification of collaborative problem solving interactions in simulated science tasks. In Proceedings of the 11th workshop on innovative use of NLP for building educational applications. 31–41.
- [53] Jiangang Hao, Lei Chen, Michael Flor, Lei Liu, and Alina A von Davier. 2017. CPS-Rater: Automated sequential annotation for conversations in collaborative problem-solving activities. ETS Research Report Series 2017, 1 (2017), 1–9.



[54] Sharon Goldwater, Dan Jurafsky, and Christopher D Manning. 2010. Which words are hard to recognize? Prosodic, lexical, and disfluency factors that increase speech recognition error rates. *Speech Communication* 52, 3 (2010), 181–200.

[55] Linda Liu, Yile Gu, Aditya Gourav, Ankur Gandhe, Shashank Kalmane, Denis Filimonov, Ariya Rastrow, and Ivan Bulyko. 2021. Domain-aware neural language models for speech recognition. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 7373–7377.