Attribution Card for "Bridging the Data Provenance Gap Across Text, Speech and Video"

Anonymous

This document provides detailed information about the licenses of each data collection and its constituent datasets, and cites all of the papers (455 in all) which introduced datasets we consider. Text datasets are laid out in Table 1, audio datasets in Table 2, and video datasets in Table 3.

Table 1: **References and licenses for alignment-tuning (text)** dataset collections presented in this paper. Collections containing material under more than three distinct licenses are marked as having "Various" licenses, and we refer readers to our raw data for the full details. Datasets are sorted alphabetically for ease of dataset lookup.

Collection	Licenses	Cite
10k Prompt Ranked	Unspecified	–
AgentInstruct	Unspecified, CC BY 4.0, MIT License	[1]–[5]
Aya	Apache License 2.0	[6]
Bactrian-X	CC BY-SA 3.0, CC BY-NC 4.0	[7]
COBRA Frames	BigScience OpenRAIL-M	[8]
COIG	Various	[9], [10]
Capybara	Various	–
ChatDoctor	Unspecified	[11]
ChatbotArena	CC BY 4.0, CC BY-NC 4.0	[12]
Cidar	CC BY-NC 4.0	[13]
CollectiveCognition	MIT License	–
Conifer	Apache License 2.0	[14]
Deita 10K	Apache License 2.0, CC BY-NC 4.0	[15]
DialogStudio	Various	[16]–[85]
Dynosaur	Various	[35], [39], [42], [48], [53], [54], [61], [70], [78], [81], [84], [86]–[236]
EverythingLM	MIT License	–

Continued on next page

Table 1: **References and licenses for alignment-tuning (text)** dataset collections presented in this paper. Collections containing material under more than three distinct licenses are marked as having "Various" licenses, and we refer readers to our raw data for the full details. Datasets are sorted alphabetically for ease of dataset lookup.

Collection	Licenses	Cite
ExpertQA	MIT License	[237]
Feedback Coll.	MIT License	[238]
Glaive Code Asst.	Apache License 2.0	–
Gretel Text-to-SQL	Apache License 2.0	–
HelpSteer	CC BY 4.0	[239]
Indic-Instr.	Various	[240]
InstAr	Various	[241]–[256]
KIWI	CC BY-SA 4.0	[257]
LMSYS-Chat-1M	LMSYS-Chat-1M Dataset License, Anthropic, Llama 2	[258]
Llama2-MedTuned-Instr.	CC BY-NC 4.0	[259]
LongAlign-10k	Anthropic, Apache License 2.0	[260]
Magpie-Pro	Meta Llama3 Community License	[261]
MathDial	CC BY-SA 4.0, MIT License	[262]
MathInstr.	MIT License	[263]
MedInstr.	Unspecified	[264]
Medical Meadow	Various	[265]–[268]
MegaWika	CC BY-SA 4.0	[269]
MetaMathQA	MIT License	[270]
Nectar	Various	–
No Robots	CC BY-NC 4.0	–
Open Asst. v2	Apache License 2.0	[271]
Open-Platypus	Various	[272]–[278]
OpenGPT Healthcare	Unspecified, OGL 3.0	–
OpenMathInstr.-1	Custom, MIT License, Apache License 2.0	[279]
Orca-Math	Various	[280]
PII-Masking-200k	Non Commercial	–
PMC-LLaMA Instr.	Unspecified, Apache License 2.0	[147], [281]
Pure-Dove	Apache License 2.0	–

Continued on next page

Table 1: **References and licenses for alignment-tuning (text)** dataset collections presented in this paper. Collections containing material under more than three distinct licenses are marked as having "Various" licenses, and we refer readers to our raw data for the full details. Datasets are sorted alphabetically for ease of dataset lookup.

Collection	Licenses	Cite
PygmalionAI-PIPPA	Apache License 2.0	[282]
Reasoning	Apache License 2.0	–
RiddleSense	MIT License	[283]
SeaBench	Apache License 2.0	[284]
SelFee	MIT License	[285]
Synth.-GSM8K-Refl.	Meta Llama3 Community License	–
Thai Gen AI	Various	–
UltraChat-200k	CC BY-NC 4.0	[286]
UltraFeedback Argilla	Various	–
WildChat	AI2 ImpACT License - Low Risk	[287]

Table 2: **References and licenses for audio** dataset collections presented in this paper. Collections containing material under more than three distinct licenses are marked as having "Various" licenses, and we refer readers to our raw data for the full details. Datasets are sorted alphabetically for ease of dataset lookup.

Collection	Licenses	Cite
1111 Hours Hindi	Custom	[288]
120h Spanish Speech	CC0 1.0	–
AFRISPEECH-200	CC BY-NC-SA 4.0	[289]
AISHELL-1	Apache 2.0	[290]
AISHELL-2	Unspecified	[291]
AISHELL-4	CC BY-SA 4.0	[292]
ALLSSTAR	CC BY 4.0	[293]
AMI	CC BY 4.0	[294]
Aalto Finnish Parl.	Custom	[295]
African Acc. French	Apache 2.0	–
Basq., Cat. and Gal.	CC BY-SA 4.0	[296]
Bloom Speech	Various	[297]
Bud500	Apache 2.0, CC BY-NC-SA 4.0	–
CSJ	Custom	[298]
CSLU 1.2	CSLU Agreement	[299]
CSLU 22 Langs.	CSLU Agreement	[300]
ClarinPL	CC BY 4.0	[301]
CoNASE	Custom	[302]
CoVoST-2	CC0 1.0	[303]
Common Voice 17	CC0 1.0	[304]
Crowd Sourced Speech	CC BY-SA 4.0	[305]
Czech Parliament	CC BY 4.0	[306]
DiDiSpeech	Unspecified	[307]
Earnings-22	Unspecified	[308]
EdAcc	CC BY-SA 4.0	[309]
Eng. Acc. in Brit. Isles	CC BY-SA 4.0	[310]
English-Vietnamese	CC BY-NC-ND 4.0	[311]
FT SPEECH	Custom	[312]
Fisher	LDC User Agreement	[313]
Fleurs	CC BY 4.0	[314]
GigaSpeech	Apache 2.0	[315]

Continued on next page

Table 2: **References and licenses for audio** dataset collections presented in this paper. Collections containing material under more than three distinct licenses are marked as having "Various" licenses, and we refer readers to our raw data for the full details. Datasets are sorted alphabetically for ease of dataset lookup.

Collection	Licenses	Cite
Golos	Custom	[316]
Hebrew Coursera	Unspecified	–
Hebrew Kan	Unspecified	–
Highland Puebla Nahuatl	CC BY-NC-SA 3.0	[317]
JTUBESPEECH	Unspecified	[318]
Japanese Anime Speech	CC0 1.0	–
KSC	CC BY 4.0	[319]
Kathbath	CC0 1.0	[320]
KeSpeech	Custom	[321]
KsponSpeech	Unspecified	[322]
LJSpeech	Public Domain	[323]
LaboroTVSpeech	Custom	[324]
LibriSpeech	CC BY 4.0	[325]
M-AILABS	Custom	[326]
M2ASR	Unspecified	[327]–[330]
MAGICDATA	CC BY-NC-ND 4.0	–
MASC	CC BY 4.0	[331]
MaSS	Unspecified	[332]
MagicData-RAMC	CC BY-NC-ND 4.0	[333]
MediaSpeech	CC BY 4.0	[334]
Minds14	CC BY 4.0	[335]
MuST-C	CC BY-NC-ND 4.0	[336]
Multiling. LibriSpeech	CC BY 4.0	[337]
Multiling. TEDx	CC BY-NC-ND 4.0	[338]
NST Danish	CC0 1.0	–
NST Norwegian	CC0 1.0	–
NST Swedish	CC0 1.0	–
Nigerian English	CC BY-SA 4.0	–
Norwegian Parl.	CC0 1.0	[339]
Norwegian Parl. Speech	CC0 1.0	[339]
OLKAVS	Custom	[340]
OpenSTT	CC BY-NC 4.0	[341]

Continued on next page

Table 2: **References and licenses for audio** dataset collections presented in this paper. Collections containing material under more than three distinct licenses are marked as having "Various" licenses, and we refer readers to our raw data for the full details. Datasets are sorted alphabetically for ease of dataset lookup.

Collection	Licenses	Cite
People's Speech	Various	[342]
QASR	Unspecified	[343]
RTVE	Custom	–
ReazonSpeech	CDLA-Sharing-1.0	[344]
Regional Af. Am. Lang.	CC BY-NC-SA 4.0	–
RixVox	CC BY 4.0	–
SDS-200	Custom	[345]
SPGISpeech	Custom	[346]
Samromur	CC BY 4.0	[347]
Samromur Children	CC BY 4.0	[348]
Samromur Milljon	CC BY 4.0	–
Shrutilipi	CC0 1.0	[349]
Snow Mountain	CC BY-SA 4.0	[350]
Spoken Wikipedia	CC BY-SA 4.0	[351]
Switchboard	LDC User Agreement	[352]
TED-LIUM3	CC BY-NC-ND 3.0	[353]
THCHS-30	Apache 2.0	[354]
THUYG-20	Apache 2.0	[355]
TIMIT	LDC User Agreement	[356]
VCTK	CC BY 4.0	–
VibraVox	CC BY 4.0	–
VoxPopuli	CC0 1.0	[357]
Vystadial	CC BY-SA 3.0	[358]
WenetSpeech	CC BY 4.0	[359]
West Afr. Radio	CC BY-SA 4.0	[360]
West Afr. Virt. Asst.	CC BY-SA 4.0	[360]
YODAS	CC BY 3.0	[361]
Zeroth-Korean	CC BY 4.0	–
aidatatang	CC BY-NC-ND 4.0	–

Table 3: **References and licenses for video** dataset collections presented in this paper. Collections containing material under more than three distinct licenses are marked as having "Various" licenses, and we refer readers to our raw data for the full details. Datasets are sorted alphabetically for ease of dataset lookup.

Collection	Licenses	Cite
100DOH	Custom	[362]
20BN-SOMETHING	Custom	[363]
20BN-jester	Custom	[364]
50 Salads	CC BY-NC-SA 4.0	[365]
ActivityNet	MIT License	[366]
Apes	Unspecified	[367]
Ava	CC BY 4.0	[368], [369]
Breakfast	CC BY 4.0	[370]
CDAD	Unspecified	[371]
COIN	Custom	[372]
Charades	Custom	[373]
Charades-Ego	Custom	[374]
Collective	Unspecified	[375]
Condensed Movies	CC BY 4.0	[376]
CrossTask	Unspecified	[377]
Davis	Custom	[378]
DiDeMo	BSD 2-Clause License	[379]
Disney Vid. Gen.	Apache 2.0	–
EEV	CC BY 4.0	[380]
EPIC-KITCHENS	CC BY-NC 4.0	[381]
Ego4D	Custom, MIT License	[382]
FERV39k	CC BY-NC 4.0	[383]
FineGym	CC BY-NC 4.0	[384]
FineVideo	CC BY 4.0	–
HAA500	Unspecified	[385]
HACS	Custom	[386]
HD-VILA-100M	Custom	[387]
HMDB	CC BY 4.0	[388]
HOLLYWOOD2	Unspecified	[389]
HOMAGE	Unspecified	[390]
HVU	Custom	[391]

Continued on next page

Table 3: **References and licenses for video** dataset collections presented in this paper. Collections containing material under more than three distinct licenses are marked as having "Various" licenses, and we refer readers to our raw data for the full details. Datasets are sorted alphabetically for ease of dataset lookup.

Collection	Licenses	Cite
Hollywood Ext.	MIT License	[392]
How2	Various	[393]
HowTo100M	Unspecified	[394]
ImageNet-Vid	CC BY-NC 4.0	[395]
Kinetics	Unspecified	[396]–[398]
LEMMA	Unspecified	[399]
LSMDC	Custom, MIT License	[400], [401]
M-MiT	Unspecified	[402]
MAD	Custom	[403]
MMAct	Custom	[404]
MPII	Unspecified, Custom	[405], [406]
MSA	Unspecified	[407]
MSR-VTT	Unspecified	[408]
MVBench	MIT License	[409]
Mars	Unspecified	[410]
Mimetics	Unspecified	[411]
Moments in Time	Custom	[412]
Movie-Net	Unspecified	[413]
MovieGraphs	Custom	[414]
MovieQA	Unspecified	[415]
MovieScenes	Unspecified	[416]
MultiTHUMOS	CC BY 4.0	[417]
NTU RGB+D	Custom	[418]
Narrated Instr. Vid.	MIT License	[419]
OmniSource-Web	Apache License 2.0	[420]
Oops!	CC BY-NC-SA 4.0	[421]
OpenVid-1M	CC-BY-4.0	[422]
PKU-MMD	Unspecified	[423]
Project-Aria	Apache License 2.0	[424], [425]
QFVS	Unspecified	[426]
QuerYD	Unspecified	[427]
RareAct	Unspecified	[428]

Continued on next page

Table 3: **References and licenses for video** dataset collections presented in this paper. Collections containing material under more than three distinct licenses are marked as having "Various" licenses, and we refer readers to our raw data for the full details. Datasets are sorted alphabetically for ease of dataset lookup.

Collection	Licenses	Cite
SOA	Unspecified	[391]
ShareGPT4Video	Attribution-NonCommercial 4.0 International	[429]
Spoken Moments	Custom	[430]
Sports-1M	CC BY 3.0	[431]
StoryGraphs	Unspecified	[432]
SumMe	Unspecified	[433]
TGIF	Custom	[434]
THUMOS	Custom	[435]
TITAN	Non Commercial	[436]
TRECVid	CC BY-NC-SA 4.0	[437]
TVSum	CC BY 3.0	[438]
TinyVIRAT	Unspecified	[439]
Toyota Smarthome	Custom	[440]
UAV-Human	Custom	[441]
UCF101	Unspecified	[442]
VIOLIN	Unspecified	[443]
VLOG	Custom	[444]
VTW	Unspecified	[445]
VaTeX	CC BY 4.0	[446]
VidProm	CC-BY-NC 4.0	[447]
VideoLT	Non Commercial	[448]
VideoStory	Unspecified	[449]
Volleyball	Unspecified	[450]
VoxCeleb	Custom	[451]
WebVid	Custom	[452]
YouCook	Unspecified	[453]
YouCook2	MIT License	[454]
Youtube-8M	Unspecified	[455]

References

[1] M. Shridhar, X. Yuan, M.-A. Côté, Y. Bisk, A. Trischler, and M. Hausknecht, *ALFWorld: Aligning text and embodied environments for interactive learning*, 2021. arXiv: 2010.03768[cs].

[2] S. Yao, H. Chen, J. Yang, and K. Narasimhan, *WebShop: Towards scalable real-world web interaction with grounded language agents*, 2023. arXiv: 2207.01206[cs].

[3] X. Liu, H. Yu, H. Zhang, *et al.*, *AgentBench: Evaluating LLMs as agents*, 2023. arXiv: 2308.03688[cs].

[4] A. Zeng, M. Liu, R. Lu, *et al.*, *AgentTuning: Enabling generalized agent abilities for LLMs*, 2023. arXiv: 2310.12823[cs].

[5] X. Deng, Y. Gu, B. Zheng, *et al.*, *Mind2web: Towards a generalist agent for the web*, 2023. arXiv: 2306.06070[cs].

[6] S. Singh, F. Vargus, D. Dsouza, *et al.*, *Aya dataset: An open-access collection for multilingual instruction tuning*, 2024. arXiv: 2402.06619[cs].

[7] H. Li, F. Koto, M. Wu, A. F. Aji, and T. Baldwin, *Bactrian-x: Multilingual replicable instruction-following models with low-rank adaptation*, 2023. arXiv: 2305.15011[cs].

[8] X. Zhou, H. Zhu, A. Yerukola, *et al.*, *COBRA frames: Contextual reasoning about effects and harms of offensive statements*, 2023. arXiv: 2306.01985[cs].

[9] G. Zhang, Y. Shi, R. Liu, *et al.*, *Chinese open instruction generalist: A preliminary release*, 2023. arXiv: 2304.07987[cs].

[10] Y. Bai, X. Du, Y. Liang, *et al.*, *COIG-CQIA: Quality is all you need for chinese instruction fine-tuning*, version: 1, 2024. arXiv: 2403.18058[cs].

[11] Y. Li, Z. Li, K. Zhang, R. Dan, S. Jiang, and Y. Zhang, *ChatDoctor: A medical chat model fine-tuned on a large language model meta-AI (LLaMA) using medical domain knowledge*, 2023. arXiv: 2303.14070[cs].

[12] L. Zheng, W.-L. Chiang, Y. Sheng, *et al.*, *Judging LLM-as-a-judge with MT-bench and chatbot arena*, 2023. arXiv: 2306.05685[cs].

[13] Z. Alyafeai, K. Almubarak, A. Ashraf, *et al.*, *CIDAR: Culturally relevant instruction dataset for arabic*, 2024. arXiv: 2402.03177[cs].

[14] H. Sun, L. Liu, J. Li, *et al.*, *Conifer: Improving complex constrained instruction-following ability of large language models*, 2024. arXiv: 2404.02823[cs].

[15] W. Liu, W. Zeng, K. He, Y. Jiang, and J. He, *What makes good data for alignment? a comprehensive study of automatic data selection in instruction tuning*, 2024. arXiv: 2312.15685[cs].

[16] D. Chen, H. Chen, Y. Yang, A. Lin, and Z. Yu, "Action-based conversations dataset: A corpus for building more in-depth task-oriented dialogue systems," in *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, K. Toutanova, A. Rumshisky, L. Zettlemoyer, *et al.*, Eds., Online: Association for Computational Linguistics, 2021, pp. 3002–3017. DOI: 10/gtsqxt.

[17] W. Wei, Q. Le, A. Dai, and J. Li, "AirDialogue: An environment for goal-oriented dialogue research," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, E. Riloff, D. Chiang, J. Hockenmaier, and J. Tsujii, Eds., Brussels, Belgium: Association for Computational Linguistics, 2018, pp. 3844–3854. DOI: 10/gf6gq2.

[18] Z. Lin, A. Madotto, G. I. Winata, *et al.*, *BiToD: A bilingual multi-domain dataset for task-oriented dialogue modeling*, 2021. arXiv: 2106.02787[cs].

[19] K. Chawla, J. Ramirez, R. Clever, G. Lucas, J. May, and J. Gratch, "CaSiNo: A corpus of campsite negotiation dialogues for automatic negotiation systems," in *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, K. Toutanova, A. Rumshisky, L. Zettlemoyer, *et al.*, Eds., Online: Association for Computational Linguistics, 2021, pp. 3167–3185. DOI: 10/gtsqxv.

[20] H. He, D. Chen, A. Balakrishnan, and P. Liang, *Decoupling strategy and generation in negotiation dialogues*, 2018. arXiv: 1808.09637[cs].

[21] N. Mrkšić, D. Ó. Séaghdha, T.-H. Wen, B. Thomson, and S. Young, *Neural belief tracker: Data-driven dialogue state tracking*, 2017. arXiv: 1606.03777[cs].

[22] K. Qian, S. Kottur, A. Beirami, *et al.*, "Database search results disambiguation for task-oriented dialog systems," in *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, M. Carpuat, M.-C. de Marneffe, and I. V. Meza Ruiz, Eds., Seattle, United States: Association for Computational Linguistics, 2022, pp. 1158–1173. DOI: 10/gtsqx2.

[23] Z. Liu, H. Wang, Z.-Y. Niu, H. Wu, and W. Che, "DuRecDial 2.0: A bilingual parallel corpus for conversational recommendation," in *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, M.-F. Moens, X. Huang, L. Specia, and S. W.-t. Yih, Eds., Online and Punta Cana, Dominican Republic: Association for Computational Linguistics, 2021, pp. 4335–4347. DOI: 10/gtsqxr.

[24] L. El Asri, H. Schulz, S. Sharma, *et al.*, "Frames: A corpus for adding memory to goal-oriented dialogue systems," in *Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue*, K. Jokinen, M. Stede, D. DeVault, and A. Louis, Eds., Saarbrücken, Germany: Association for Computational Linguistics, 2017, pp. 207–219. DOI: 10/gtsqx4.

[25] J. Quan, D. Xiong, B. Webber, and C. Hu, "GECOR: An end-to-end generative ellipsis and co-reference resolution model for task-oriented dialogue," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, K. Inui, J. Jiang, V. Ng, and X. Wan, Eds., Hong Kong, China: Association for Computational Linguistics, 2019, pp. 4547–4557. DOI: 10/gk3btq.

[26] W. Chen, J. Chen, P. Qin, X. Yan, and W. Y. Wang, *Semantically conditioned dialog response generation via hierarchical disentangled self-attention*, 2019. arXiv: 1905.12866[cs].

[27] Z. Chen, B. Liu, S. Moon, C. Sankar, P. Crook, and W. Y. Wang, *KETOD: Knowledge-enriched task-oriented dialogue*, 2022. arXiv: 2205.05589[cs].

[28] M. Eric and C. D. Manning, *Key-value retrieval networks for task-oriented dialogue*, 2017. arXiv: 1705.05414[cs].

[29] X. Zang, A. Rastogi, S. Sunkara, R. Gupta, J. Zhang, and J. Chen, *MultiWOZ 2.2 : A dialogue dataset with additional annotation corrections and state tracking baselines*, 2020. arXiv: 2007.12720[cs].

[30] I. Shalyminov, S. Lee, A. Eshghi, and O. Lemon, *Few-shot dialogue generation without annotated data: A transfer learning approach*, 2019. arXiv: 1908.05854[cs].

[31] S. Martin, S. Poddar, and K. Upasani, "MuDoCo: Corpus for multidomain coreference resolution and referring expression generation," in *Proceedings of the Twelfth Language Resources and Evaluation Conference*, N. Calzolari, F. Béchet, P. Blache, *et al.*, Eds., Marseille, France: European Language Resources Association, 2020, pp. 104–111, ISBN: 979-10-95546-34-4. [Online]. Available: https://aclanthology.org/2020.lrec-1.13 (visited on 05/01/2024).

[32] D. Peskov, N. Clarke, J. Krone, *et al.*, "Multi-domain goal-oriented dialogues (MultiDoGO): Strategies toward curating and annotating large scale dialogue data," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, K. Inui, J. Jiang, V. Ng, and X. Wan, Eds., Hong Kong, China: Association for Computational Linguistics, 2019, pp. 4526–4536. DOI: 10/gkr9hj.

[33] M. Eric, R. Goel, S. Paul, *et al.*, *MultiWOZ 2.1: A consolidated multi-domain dialogue dataset with state corrections and state tracking baselines*, 2019. arXiv: 1907.01669[cs].

[34] S. Moon, P. Shah, A. Kumar, and R. Subba, "OpenDialKG: Explainable conversational reasoning with attention-based walks over knowledge graphs," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, A. Korhonen, D. Traum, and L. Màrquez, Eds., Florence, Italy: Association for Computational Linguistics, 2019, pp. 845–854. DOI: 10/ggt4wm.

[35] A. Rastogi, X. Zang, S. Sunkara, R. Gupta, and P. Khaitan, *Towards scalable multi-domain conversational agents: The schema-guided dialogue dataset*, 2020. arXiv: 1909.05855[cs].

[36] J. E. M. Mosig, S. Mehri, and T. Kober, *STAR: A schema-guided dialog dataset for transfer learning*, 2020. arXiv: 2010.11853[cs].

[37] S. Chiu, M. Li, Y.-T. Lin, and Y.-N. Chen, "SalesBot: Transitioning from chit-chat to task-oriented dialogues," in *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, S. Muresan, P. Nakov, and A. Villavicencio, Eds., Dublin, Ireland: Association for Computational Linguistics, 2022, pp. 6143–6158. DOI: 10/gtsqxw.

[38] P. Shah, D. Hakkani-Tür, G. Tür, *et al.*, *Building a conversational agent overnight with dialogue self-play*, 2018. arXiv: 1801.04871[cs].

[39] B. Byrne, K. Krishnamoorthi, C. Sankar, *et al.*, *Taskmaster-1: Toward a realistic and diverse dialog dataset*, 2019. arXiv: 1909.05358[cs].

[40] N. Mrkšić and I. Vulić, *Fully statistical neural belief tracking*, 2018. arXiv: 1805.11350[cs].

[41] G. Shang, W. Ding, Z. Zhang, *et al.*, "Unsupervised abstractive meeting summarization with multi-sentence compression and budgeted submodular maximization," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, I. Gurevych and Y. Miyao, Eds., Melbourne, Australia: Association for Computational Linguistics, 2018, pp. 664–674. DOI: 10/gm8jvb.

[42] R. Rameshkumar and P. Bailey, "Storytelling with dialogue: A critical role dungeons and dragons dataset," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, D. Jurafsky, J. Chai, N. Schluter, and J. Tetreault, Eds., Online: Association for Computational Linguistics, 2020, pp. 5121–5134. DOI: 10/gtsqxp.

[43] A. Fabbri, F. Rahman, I. Rizvi, *et al.*, "ConvoSumm: Conversation summarization benchmark and improved abstractive summarization with argument mining," in *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, C. Zong, F. Xia, W. Li, and R. Navigli, Eds., Online: Association for Computational Linguistics, 2021, pp. 6866–6880. DOI: 10/gmf9qs.

[44] Y. Chen, Y. Liu, L. Chen, and Y. Zhang, "DialogSum: A real-life scenario dialogue summarization dataset," in *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, C. Zong, F. Xia, W. Li, and R. Navigli, Eds., Online: Association for Computational Linguistics, 2021, pp. 5062–5074. DOI: 10/gtsqxs.

[45] R. Mukherjee, A. Bohra, A. Banerjee, *et al.*, "ECTSum: A new benchmark dataset for bullet point summarization of long earnings call transcripts," in *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, Y. Goldberg, Z. Kozareva, and Y. Zhang, Eds., Abu Dhabi, United Arab Emirates: Association for Computational Linguistics, 2022, pp. 10 893–10 906. DOI: 10/gtsqxz.

[46] C. Zhu, Y. Liu, J. Mei, and M. Zeng, *MediaSum: A large-scale media interview dataset for dialogue summarization*, 2021. arXiv: 2103.06410[cs].

[47] M. Zhong, D. Yin, T. Yu, *et al.*, *QMSum: A new benchmark for query-based multi-domain meeting summarization*, 2021. arXiv: 2104.05938[cs].

[48] B. Gliwa, I. Mochol, M. Biesek, and A. Wawer, "SAMSum corpus: A human-annotated dialogue dataset for abstractive summarization," in *Proceedings of the 2nd Workshop on New Frontiers in Summarization*, L. Wang, J. C. K. Cheung, G. Carenini, and F. Liu, Eds., Hong Kong, China: Association for Computational Linguistics, 2019, pp. 70–79. DOI: 10/gmjqgr.

[49] M. Chen, Z. Chu, S. Wiseman, and K. Gimpel, "SummScreen: A dataset for abstractive screenplay summarization," in *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, S. Muresan, P. Nakov, and A. Villavicencio, Eds., Dublin, Ireland: Association for Computational Linguistics, 2022, pp. 8602–8615. DOI: 10/gtsqxx.

[50] G. Feigenblat, C. Gunasekara, B. Sznajder, S. Joshi, D. Konopnicki, and R. Aharonov, *TWEETSUMM – a dialog summarization dataset for customer service*, 2021. arXiv: 2111.11894[cs].

[51] Y. Li, K. Qian, W. Shi, and Z. Yu, *End-to-end trainable non-collaborative dialog system*, 2019. arXiv: 1911.10742[cs].

[52] E. Dinan, V. Logacheva, V. Malykh, *et al.*, *The second conversational intelligence challenge (ConvAI2)*, 2019. arXiv: 1902.00098[cs].

[53] H. Rashkin, E. M. Smith, M. Li, and Y.-L. Boureau, *Towards empathetic open-domain conversation models: A new benchmark and dataset*, 2019. arXiv: 1811.00207[cs].

[54] Y. Bai, A. Jones, K. Ndousse, *et al.*, *Training a helpful and harmless assistant with reinforcement learning from human feedback*, 2022. arXiv: 2204.05862[cs].

[55] M. Chen, A. Papangelis, C. Tao, *et al.*, *PLACES: Prompting language models for social conversation synthesis*, 2023. arXiv: 2302.03269[cs].

[56] H. Kim, Y. Yu, L. Jiang, *et al.*, *ProsocialDialog: A prosocial backbone for conversational agents*, 2022. arXiv: 2205.12688[cs].

[57] W. Myers, T. Etchart, and N. Fulda, "Conversational scaffolding: An analogy-based approach to response prioritization in open-domain dialogs:" in *Proceedings of the 12th International Conference on Agents and Artificial Intelligence*, Valletta, Malta: SCITEPRESS - Science and Technology Publications, 2020, pp. 69–78. DOI: 10/gtsq86.

[58] S. Reddy, D. Chen, and C. D. Manning, *CoQA: A conversational question answering challenge*, 2019. arXiv: 1808.07042[cs].

[59] T. Yu, R. Zhang, H. Y. Er, *et al.*, *CoSQL: A conversational text-to-SQL challenge towards cross-domain natural language interfaces to databases*, 2019. arXiv: 1909.05378[cs].

[60] A. Talmor and J. Berant, "The web as a knowledge-base for answering complex questions," in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, M. Walker, H. Ji, and A. Stent, Eds., New Orleans, Louisiana: Association for Computational Linguistics, 2018, pp. 641–651. DOI: 10/gkz2k6.

[61] L. Nan, D. Radev, R. Zhang, *et al.*, "DART: Open-domain structured data record to text generation," in *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, K. Toutanova, A. Rumshisky, L. Zettlemoyer, *et al.*, Eds., Online: Association for Computational Linguistics, 2021, pp. 432–447. DOI: 10/gnh49f.

[62] L. Nan, C. Hsieh, Z. Mao, *et al.*, "FeTaQA: Free-form table question answering," *Transactions of the Association for Computational Linguistics*, vol. 10, B. Roark and A. Nenkova, Eds., pp. 35–49, 2022, Place: Cambridge, MA Publisher: MIT Press. DOI: 10/gtsqx3.

[63] Y. Gu, S. Kase, M. Vanni, *et al.*, "Beyond i.i.d.: Three levels of generalization for question answering on knowledge bases," in *Proceedings of the Web Conference 2021*, 2021, pp. 3477–3488. DOI: 10/gnnfbt.

[64] W. Chen, H. Zha, Z. Chen, W. Xiong, H. Wang, and W. Y. Wang, "HybridQA: A dataset of multi-hop question answering over tabular and textual data," in *Findings of the Association for Computational Linguistics: EMNLP 2020*, T. Cohn, Y. He, and Y. Liu, Eds., Online: Association for Computational Linguistics, 2020, pp. 1026–1036. DOI: 10/gpmd4x.

[65] D. Gupta, S. Kumari, A. Ekbal, and P. Bhattacharyya, "MMQA: A multi-domain multilingual question-answering framework for english and hindi," in *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, N. Calzolari, K. Choukri, C. Cieri, *et al.*, Eds., Miyazaki, Japan: European Language Resources Association (ELRA), 2018. [Online]. Available: https://aclanthology.org/L18-1440 (visited on 05/01/2024).

[66] H. Li, A. Arora, S. Chen, A. Gupta, S. Gupta, and Y. Mehdad, "MTOP: A comprehensive multilingual task-oriented semantic parsing benchmark," in *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, P. Merlo, J. Tiedemann, and R. Tsarfaty, Eds., Online: Association for Computational Linguistics, 2021, pp. 2950–2962. DOI: 10/gtsqxq.

[67] A. Talmor, O. Yoran, A. Catav, *et al.*, *MultiModalQA: Complex question answering over text, tables and images*, 2021. arXiv: 2104.06039[cs].

[68] T. Yu, R. Zhang, M. Yasunaga, *et al.*, "SParC: Cross-domain semantic parsing in context," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, A. Korhonen, D. Traum, and L. Màrquez, Eds., Florence, Italy: Association for Computational Linguistics, 2019, pp. 4511–4523. DOI: 10/gj4fwd.

[69] M. Iyyer, W.-t. Yih, and M.-W. Chang, "Search-based neural structured learning for sequential question answering," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, R. Barzilay and M.-Y. Kan, Eds., Vancouver, Canada: Association for Computational Linguistics, 2017, pp. 1821–1831. DOI: 10/gf6nx8.

[70] T. Yu, R. Zhang, K. Yang, *et al.*, *Spider: A large-scale human-labeled dataset for complex and cross-domain semantic parsing and text-to-SQL task*, 2019. arXiv: 1809.08887[cs].

[71] A. Parikh, X. Wang, S. Gehrmann, *et al.*, "ToTTo: A controlled table-to-text generation dataset," in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, B. Webber, T. Cohn, Y. He, and Y. Liu, Eds., Online: Association for Computational Linguistics, 2020, pp. 1173–1186. DOI: 10/gm3nmg.

[72] W.-t. Yih, M. Richardson, C. Meek, M.-W. Chang, and J. Suh, "The value of semantic parse labeling for knowledge base question answering," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, K. Erk and N. A. Smith, Eds., Berlin, Germany: Association for Computational Linguistics, 2016, pp. 201–206. DOI: 10/gkz3hq.

[73] V. Zhong, C. Xiong, and R. Socher, *Seq2sql: Generating structured queries from natural language using reinforcement learning*, 2017. arXiv: 1709.00103[cs].

[74] P. Pasupat and P. Liang, "Compositional semantic parsing on semi-structured tables," in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, C. Zong and M. Strube, Eds., Beijing, China: Association for Computational Linguistics, 2015, pp. 1470–1480. DOI: 10/gfz98s.

[75] M. Komeili, K. Shuster, and J. Weston, "Internet-augmented dialogue generation," in *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, S. Muresan, P. Nakov, and A. Villavicencio, Eds., Dublin, Ireland: Association for Computational Linguistics, 2022, pp. 8460–8478. DOI: 10/gr75db.

[76] E. Dinan, S. Roller, K. Shuster, A. Fan, M. Auli, and J. Weston, *Wizard of wikipedia: Knowledge-powered conversational agents*, 2019. arXiv: 1811.01241[cs].

[77] C. T. Hemphill, J. J. Godfrey, and G. R. Doddington, "The ATIS spoken language systems pilot corpus," in *Speech and Natural Language: Proceedings of a Workshop Held at Hidden Valley, Pennsylvania, June 24-27,1990*, 1990. DOI: 10/cz3442.

[78] I. Casanueva, T. Temčinas, D. Gerz, M. Henderson, and I. Vulić, "Efficient intent detection with dual sentence encoders," in *Proceedings of the 2nd Workshop on Natural Language Processing for Conversational AI*, T.-H. Wen, A. Celikyilmaz, Z. Yu, *et al.*, Eds., Online: Association for Computational Linguistics, 2020, pp. 38–45. DOI: 10/gjhzzs.

[79] J. Zhang, K. Hashimoto, Y. Wan, *et al.*, *Are pretrained transformers robust in intent classification? a missing ingredient in evaluation of out-of-scope intent detection*, 2022. arXiv: 2106.04564[cs].

[80] S. Larson, A. Mahendran, J. J. Peper, *et al.*, "An evaluation dataset for intent classification and out-of-scope prediction," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, K. Inui, J. Jiang, V. Ng, and X. Wan, Eds., Hong Kong, China: Association for Computational Linguistics, 2019, pp. 1311–1316. DOI: 10/gqrfvv.

[81] X. Liu, A. Eshghi, P. Swietojanski, and V. Rieser, *Benchmarking natural language understanding services for building conversational agents*, 2019. arXiv: 1903.05566[cs].

[82] J. Liu, P. Pasupat, S. Cyphers, and J. Glass, "Asgard: A portable architecture for multilingual dialogue systems," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, Vancouver, BC, Canada: IEEE, 2013, pp. 8386–8390. DOI: 10/gtsqxn.

[83] S. Coope, T. Farghly, D. Gerz, I. Vulić, and M. Henderson, *Span-ConveRT: Few-shot span extraction for dialog with pretrained conversational representations*, 2020. arXiv: 2005.08866[cs].

[84] A. Coucke, A. Saade, A. Ball, *et al.*, *Snips voice platform: An embedded spoken language understanding system for private-by-design voice interfaces*, 2018. arXiv: 1805.10190[cs].

[85] S. Gupta, R. Shah, M. Mohit, A. Kumar, and M. Lewis, *Semantic parsing for task oriented dialog using hierarchical representations*, 2018. arXiv: 1810.07942[cs].

[86] V. Adlakha, S. Dhuliawala, K. Suleman, H. de Vries, and S. Reddy, *TopiOCQA: Open-domain conversational question answering with topic switching*, 2022. arXiv: 2110.00768[cs].

[87] O. Agarwal, H. Ge, S. Shakeri, and R. Al-Rfou, *Knowledge graph based synthetic corpus generation for knowledge-enhanced language model pre-training*, 2021. arXiv: 2010.12688[cs].

[88] E. Akyürek, T. Bolukbasi, F. Liu, *et al.*, *Towards tracing factual knowledge in language models back to the training data*, 2022. arXiv: 2205.11482[cs].

[89] A. Amini, S. Gabriel, P. Lin, R. Koncel-Kedziorski, Y. Choi, and H. Hajishirzi, *MathQA: Towards interpretable math word problem solving with operation-based formalisms*, 2019. arXiv: 1905.13319[cs].

[90] M. C. Ardanuy, F. Nanni, K. Beelen, *et al.*, *Living machines: A study of atypical animacy*, 2020. arXiv: 2005.11140[cs].

[91] J. Austin, A. Odena, M. Nye, *et al.*, *Program synthesis with large language models*, 2021. arXiv: 2108.07732[cs].

[92] Z. Azerbayev, B. Piotrowski, H. Schoelkopf, E. W. Ayers, D. Radev, and J. Avigad, *ProofNet: Autoformalizing and formally proving undergraduate-level mathematics*, 2023. arXiv: 2302.12433[cs].

[93] P. Bajaj, D. Campos, N. Craswell, *et al.*, *MS MARCO: A human generated MAchine reading COmprehension dataset*, 2018. arXiv: 1611.09268[cs].

[94] A. Balakrishnan, J. Rao, K. Upasani, M. White, and R. Subba, *Constrained decoding for neural NLG from compositional representations in task-oriented dialogue*, 2019. arXiv: 1906.07220[cs].

[95] M. Bartolo, A. Roberts, J. Welbl, S. Riedel, and P. Stenetorp, "Beat the AI: Investigating adversarial human annotation for reading comprehension," *Transactions of the Association for Computational Linguistics*, vol. 8, pp. 662–678, 2020. DOI: 10/gjzgwj.

[96] Y. Bisk, R. Zellers, R. L. Bras, J. Gao, and Y. Choi, *PIQA: Reasoning about physical commonsense in natural language*, 2019. arXiv: 1911.11641[cs].

[97] M. Boratko, X. L. Li, R. Das, T. O'Gorman, D. Le, and A. McCallum, *ProtoQA: A question answering dataset for prototypical common-sense reasoning*, 2020. arXiv: 2005.00771[cs].

[98] J. A. Botha, M. Faruqui, J. Alex, J. Baldridge, and D. Das, *Learning to split and rephrase from wikipedia edit history*, 2018. arXiv: 1808.09468[cs].

[99] F. Boudin and Y. Gallina, "Redefining absent keyphrases and their effect on retrieval effectiveness," in *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, K. Toutanova, A. Rumshisky, L. Zettlemoyer, *et al.*, Eds., Online: Association for Computational Linguistics, 2021, pp. 4185–4193. DOI: 10/g6k64d.

[100] À. Bravo, J. Piñero, N. Queralt-Rosinach, M. Rautschka, and L. I. Furlong, "Extraction of relations between genes and diseases from text and large-scale data analysis: Implications for translational research," *BMC Bioinformatics*, vol. 16, no. 1, p. 55, 2015. DOI: 10/f7kn8s.

[101] T. B. Brown, B. Mann, N. Ryder, *et al.*, *Language models are few-shot learners*, 2020. arXiv: 2005.14165[cs].

[102] S. Cao and L. Wang, *Controllable open-ended question generation with a new question type ontology*, 2021. arXiv: 2107.00152[cs].

[103] S. Cao, J. Shi, L. Pan, *et al.*, *KQA pro: A dataset with explicit compositional programs for complex question answering over knowledge base*, 2022. arXiv: 2007.03875[cs].

[104] A. Cetoli, M. Akbari, S. Bragaglia, A. D. O'Harney, and M. Sloan, "Named entity disambiguation using deep learning on graphs," in vol. 11438, 2019, pp. 78–86. DOI: 10.1007/978-3-030-15719-7_10.

[105] I. Chalkidis, M. Fergadiotis, P. Malakasiotis, and I. Androutsopoulos, *Large-scale multi-label text classification on EU legislation*, 2019. arXiv: 1906.02192[cs].

[106] I. Chalkidis, I. Androutsopoulos, and N. Aletras, *Neural legal judgment prediction in english*, 2019. arXiv: 1906.02059[cs].

[107] I. Chalkidis, M. Fergadiotis, N. Manginas, E. Katakalou, and P. Malakasiotis, *Regulatory compliance through doc2doc information retrieval: A case study in EU/UK legislation where text similarity has limitations*, 2021. arXiv: 2101.10726[cs].

[108] J. S. Chan, M. Pieler, J. Jao, J. Scheurer, and E. Perez, *Few-shot adaptation works with UnpredicTable data*, 2022. arXiv: 2208.01009[cs].

[109] E. Chapuis, P. Colombo, M. Manica, M. Labeau, and C. Clavel, *Hierarchical pre-training for sequence labelling in spoken dialog*, 2021. arXiv: 2009.11152[cs].

[110] H. Chen, Y. Ji, and D. Evans, *Finding friends and flipping frenemies: Automatic paraphrase dataset augmentation using graph theory*, 2020. arXiv: 2011.01856[cs].

[111] Z. Cheng, H. Dong, Z. Wang, *et al.*, *HiTab: A hierarchical table dataset for question answering and natural language generation*, 2022. arXiv: 2108.06712[cs].

[112] A. Chouldechova, *Fair prediction with disparate impact: A study of bias in recidivism prediction instruments*, 2017. arXiv: 1703.00056[cs,stat].

[113] P. Christmann, R. S. Roy, A. Abujabal, J. Singh, and G. Weikum, "Look before you hop: Conversational question answering over knowledge graphs using judicious context expansion," in *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, 2019, pp. 729–738. DOI: 10/gkz233.

[114] C. Clark, K. Lee, M.-W. Chang, T. Kwiatkowski, M. Collins, and K. Toutanova, *BoolQ: Exploring the surprising difficulty of natural yes/no questions*, 2019. arXiv: 1905.10044[cs].

[115] P. Clark, I. Cowhey, O. Etzioni, *et al.*, *Think you have solved question answering? try ARC, the AI2 reasoning challenge*, 2018. arXiv: 1803.05457[cs].

[116] K. Cobbe, V. Kosaraju, M. Bavarian, *et al.*, *Training verifiers to solve math word problems*, 2021. arXiv: 2110.14168[cs].

[117] V. Dankers, C. Lucas, and I. Titov, "Can transformer be too compositional? analysing idiom processing in neural machine translation," in *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, S. Muresan, P. Nakov, and A. Villavicencio, Eds., Dublin, Ireland: Association for Computational Linguistics, 2022, pp. 3608–3626. DOI: 10/g6k6xn.

[118] P. Dasigi, N. F. Liu, A. Marasović, N. A. Smith, and M. Gardner, *Quoref: A reading comprehension dataset with questions requiring coreferential reasoning*, 2019. arXiv: 1908.05803[cs].

[119] A. Devaraj, I. J. Marshall, B. C. Wallace, and J. J. Li, *Paragraph-level simplification of medical texts*, 2021. arXiv: 2104.05767[cs].

[120] J. DeYoung, I. Beltagy, M. van Zuylen, B. Kuehl, and L. L. Wang, *MS2: Multi-document summarization of medical studies*, 2021. arXiv: 2104.06486[cs].

[121] T. Diggelmann, J. Boyd-Graber, J. Bulian, M. Ciaramita, and M. Leippold, *CLIMATE-FEVER: A dataset for verification of real-world climate claims*, 2021. arXiv: 2012.00614[cs].

[122] D. Emelin, R. L. Bras, J. D. Hwang, M. Forbes, and Y. Choi, *Moral stories: Situated reasoning about norms, intents, actions, and their consequences*, 2020. arXiv: 2012.15738[cs].

[123] A. R. Fabbri, I. Li, T. She, S. Li, and D. R. Radev, *Multi-news: A large-scale multi-document summarization dataset and abstractive hierarchical model*, 2019. arXiv: 1906.01749[cs].

[124] M. Faruqui and D. Das, *Identifying well-formed natural language questions*, 2018. arXiv: 1808.09419[cs].

[125] S. Feng, S. S. Patel, H. Wan, and S. Joshi, "MultiDoc2dial: Modeling dialogues grounded in multiple documents," in *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 2021, pp. 6162–6176. DOI: 10/g6k938.

[126] Y. Gallina, F. Boudin, and B. Daille, "KPTimes: A large-scale dataset for keyphrase generation on news documents," in *Proceedings of the 12th International Conference on Natural Language Generation*, K. van Deemter, C. Lin, and H. Takamura, Eds., Tokyo, Japan: Association for Computational Linguistics, 2019, pp. 130–135. DOI: 10/g6k64k.

[127] K. Ganesan, C. Zhai, and J. Han, "Opinosis: A graph based approach to abstractive summarization of highly redundant opinions," in *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, C.-R. Huang and D. Jurafsky, Eds., Beijing, China: Coling 2010 Organizing Committee, 2010, pp. 340–348. [Online]. Available: https://aclanthology.org/C10-1039 (visited on 10/02/2024).

[128] M. G. Gazzola, S. E. Leal, and S. M. Aluísio, "Predição da complexidade textual de recursos educacionais abertos em português," *Symposium in Information and Human Language Technology - STIL*, 2019. [Online]. Available: https://repositorio.usp.br/item/002971271 (visited on 10/02/2024).

[129] E. J. George and R. Mamidi, *Conversational implicatures in english dialogue: Annotated dataset*, 2019. arXiv: 1911.10704[cs].

[130] M. Geva, E. Malmi, I. Szpektor, and J. Berant, *DiscoFuse: A large-scale dataset for discourse-based sentence fusion*, 2019. arXiv: 1902.10526[cs].

[131] G. Gorrell, K. Bontcheva, L. Derczynski, E. Kochkina, M. Liakata, and A. Zubiaga, *RumourEval 2019: Determining rumour veracity and support for rumours*, 2018. arXiv: 1809.06683[cs].

[132] Y. Gu, R. Tinn, H. Cheng, *et al.*, "Domain-specific language model pretraining for biomedical natural language processing," *ACM Transactions on Computing for Healthcare*, vol. 3, no. 1, pp. 1–23, 2022. DOI: 10/gnmkjx.

[133] A. Gupta, J. Xu, S. Upadhyay, D. Yang, and M. Faruqui, *Disfl-QA: A benchmark dataset for understanding disfluencies in question answering*, 2021. arXiv: 2106.04016[cs].

[134] D. Ha and D. Eck, *A neural representation of sketch drawings*, 2017. arXiv: 1704.03477[cs, stat].

[135] H. Haagsma, J. Bos, and M. Nissim, "MAGPIE: A large corpus of potentially idiomatic expressions," in *Proceedings of the Twelfth Language Resources and Evaluation Conference*, N. Calzolari, F. Béchet, P. Blache, *et al.*, Eds., Marseille, France: European Language Resources Association, 2020, pp. 279–287, ISBN: 979-10-95546-34-4. [Online]. Available: https://aclanthology.org/2020.lrec-1.35 (visited on 10/02/2024).

[136] M. Hazoom, V. Malik, and B. Bogin, *Text-to-SQL in the wild: A naturally-occurring dataset based on stack exchange data*, 2021. arXiv: 2106.05006[cs].

[137] P. Henderson, M. S. Krass, L. Zheng, *et al.*, *Pile of law: Learning responsible data filtering from the law and a 256gb open-source legal dataset*, 2022. arXiv: 2207.00220[cs].

[138] D. Hendrycks, C. Burns, A. Chen, and S. Ball, *CUAD: An expert-annotated NLP dataset for legal contract review*, 2021. arXiv: 2103.06268[cs].

[139] L. Huang, R. L. Bras, C. Bhagavatula, and Y. Choi, *Cosmos QA: Machine reading comprehension with contextual commonsense reasoning*, 2019. arXiv: 1909.00277[cs].

[140] L. Huang, S. Cao, N. Parulian, H. Ji, and L. Wang, *Efficient attentions for long document summarization*, 2021. arXiv: 2104.02112[cs].

[141] A. Huang, *'tis but thy name: Semantic question answering evaluation with 11m names for 1m entities*, 2022. arXiv: 2202.13581[cs].

[142] J. J. Irwin, K. G. Tang, J. Young, *et al.*, "ZINC20—a free ultralarge-scale chemical database for ligand discovery," *Journal of Chemical Information and Modeling*, vol. 60, no. 12, pp. 6065–6073, 2020, Publisher: American Chemical Society. DOI: 10/gmjg8b.

[143] M. Ivgi, U. Shaham, and J. Berant, *Efficient long-text understanding with short-text models*, 2022. arXiv: 2208.00748[cs].

[144] S. Iyer, I. Konstas, A. Cheung, J. Krishnamurthy, and L. Zettlemoyer, *Learning a neural semantic parser from user feedback*, 2017. arXiv: 1704.08760[cs].

[145] Y. Jiang, S. Bordia, Z. Zhong, C. Dognin, M. Singh, and M. Bansal, *HoVer: A dataset for many-hop fact extraction and claim verification*, 2020. arXiv: 2011.03088[cs].

[146] C. Jiang, M. Maddela, W. Lan, Y. Zhong, and W. Xu, *Neural CRF model for sentence alignment in text simplification*, 2021. arXiv: 2005.02324[cs].

[147] Q. Jin, B. Dhingra, Z. Liu, W. W. Cohen, and X. Lu, *PubMedQA: A dataset for biomedical research question answering*, 2019. arXiv: 1909.06146[cs,q-bio].

[148] M. Joshi, E. Choi, D. S. Weld, and L. Zettlemoyer, *TriviaQA: A large scale distantly supervised challenge dataset for reading comprehension*, 2017. arXiv: 1705.03551[cs].

[149] J. Juraska, K. K. Bowden, and M. Walker, *ViGGO: A video game corpus for data-to-text generation in open-domain conversation*, 2019. arXiv: 1910.12129[cs].

[150] T. Jurczyk, M. Zhai, and J. D. Choi, *SelQA: A new benchmark for selection-based question answering*, 2016. arXiv: 1606.08513[cs].

[151] A. Kanade, P. Maniatis, G. Balakrishnan, and K. Shi, *Learning and evaluating contextual embedding of source code*, 2020. arXiv: 2001.00059[cs].

[152] D. Kaushik, E. Hovy, and Z. C. Lipton, *Learning the difference that makes a difference with counterfactually-augmented data*, 2020. arXiv: 1909.12434[cs,stat].

[153] T. Khot, P. Clark, M. Guerquin, P. Jansen, and A. Sabharwal, *QASC: A dataset for question answering via sentence composition*, 2020. arXiv: 1910.11473[cs].

[154] T. Khot, A. Sabharwal, and P. Clark, "SciTaiL: A textual entailment dataset from science question answering," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, no. 1, 2018. DOI: 10/grm22d.

[155] S. Kim, I. Kang, and N. Kwak, *Semantic sentence matching with densely-connected recurrent and co-attentive information*, 2018. arXiv: 1805.11360[cs].

[156] A. Kornilova and V. Eidelman, "BillSum: A corpus for automatic summarization of US legislation," in *Proceedings of the 2nd Workshop on New Frontiers in Summarization*, 2019, pp. 48–56. DOI: 10/gtwwtd.

[157] F. Kury, A. Butler, C. Yuan, *et al.*, "Chia, a large annotated corpus of clinical trial eligibility criteria," *Scientific Data*, vol. 7, no. 1, p. 281, 2020, Publisher: Nature Publishing Group. DOI: 10/gr4ftn.

[158] G. Lai, Q. Xie, H. Liu, Y. Yang, and E. Hovy, *RACE: Large-scale ReAding comprehension dataset from examinations*, 2017. arXiv: 1704.04683[cs].

[159] B. M. Lake and M. Baroni, *Generalization without systematicity: On the compositional skills of sequence-to-sequence recurrent networks*, 2018. arXiv: 1711.00350[cs].

[160] R. Lebret, D. Grangier, and M. Auli, *Neural text generation from structured data with application to the biography domain*, 2016. arXiv: 1603.07771[cs].

[161] M. Lewis, D. Yarats, Y. N. Dauphin, D. Parikh, and D. Batra, *Deal or no deal? end-to-end learning for negotiation dialogues*, 2017. arXiv: 1706.05125[cs].

[162] Y. Li, D. Choi, J. Chung, *et al.*, "Competition-level code generation with AlphaCode," *Science*, vol. 378, no. 6624, pp. 1092–1097, 2022. DOI: 10/grggxf.

[163] H. Li, S. Kim, and S. Chandra, *Neural code search evaluation dataset*, 2019. arXiv: 1908.09804[cs].

[164] B. Y. Lin, S. Lee, R. Khanna, and X. Ren, *Birds have four legs?! NumerSense: Probing numerical commonsense knowledge of pre-trained language models*, 2020. arXiv: 2005.00683[cs].

[165] B. Y. Lin, W. Zhou, M. Shen, *et al.*, *CommonGen: A constrained text generation challenge for generative commonsense reasoning*, 2020. arXiv: 1911.03705[cs].

[166] K. Lin, O. Tafjord, P. Clark, and M. Gardner, *Reasoning over paragraph effects in situations*, 2019. arXiv: 1908.05852[cs].

[167] S. Lin, J. Hilton, and O. Evans, *TruthfulQA: Measuring how models mimic human falsehoods*, 2022. arXiv: 2109.07958[cs].

[168] W. Ling, D. Yogatama, C. Dyer, and P. Blunsom, *Program induction by rationale generation : Learning to solve and explain algebraic word problems*, 2017. arXiv: 1705.04146[cs].

[169] A. Louis, D. Roth, and F. Radlinski, *"i'd rather just go to bed": Understanding indirect answers*, 2020. arXiv: 2010.03450[cs].

[170] R. Lowe, N. Pow, I. Serban, and J. Pineau, *The ubuntu dialogue corpus: A large dataset for research in unstructured multi-turn dialogue systems*, 2016. arXiv: 1506.08909[cs].

[171] P. Malo, A. Sinha, P. Takala, P. Korhonen, and J. Wallenius, *Good debt or bad debt: Detecting semantic orientations in economic texts*, 2013. arXiv: 1307.5336[cs,q-fin].

[172] L. J. Martin, P. Ammanabrolu, X. Wang, *et al.*, "Event representations for automated story generation with deep neural nets," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, no. 1, 2018. DOI: 10/g6k72p.

[173] S. Merity, C. Xiong, J. Bradbury, and R. Socher, *Pointer sentinel mixture models*, 2016. arXiv: 1609.07843[cs].

[174] T. Mihaylov, P. Clark, T. Khot, and A. Sabharwal, *Can a suit of armor conduct electricity? a new dataset for open book question answering*, 2018. arXiv: 1809.02789[cs].

[175] S. Mishra, M. Finlayson, P. Lu, *et al.*, *Lila: A unified benchmark for mathematical reasoning*, 2023. arXiv: 2210.17517[cs].

[176] N. Moniz and L. Torgo, *Multi-source social feedback of online news feeds*, 2018. arXiv: 1801.07055[cs].

[177] N. Mostafazadeh, A. Kalyanpur, L. Moon, *et al.*, *GLUCOSE: GeneraLized and COntextualized story explanations*, 2020. arXiv: 2009.07758[cs].

[178] S. Narayan, S. B. Cohen, and M. Lapata, *Don't give me the details, just the summary! topic-aware convolutional neural networks for extreme summarization*, 2018. arXiv: 1808.08745[cs].

[179] T.-P. Nguyen, S. Razniewski, and G. Weikum, "Advanced semantics for commonsense knowledge extraction," in *Proceedings of the Web Conference 2021*, 2021, pp. 2636–2647. DOI: 10/gnnffn.

[180] Y. Nie, A. Williams, E. Dinan, M. Bansal, J. Weston, and D. Kiela, *Adversarial NLI: A new benchmark for natural language understanding*, 2020. arXiv: 1910.14599[cs].

[181] J. Novikova, O. Dušek, and V. Rieser, *The e2e dataset: New challenges for end-to-end generation*, 2017. arXiv: 1706.09254[cs].

[182] C. Paik, S. Aroca-Ouellette, A. Roncone, and K. Kann, *The world of an octopus: How reporting bias influences a language model's perception of color*, 2021. arXiv: 2110.08182[cs].

[183] S. Pakhomov, B. McInnes, T. Adam, Y. Liu, T. Pedersen, and G. B. Melton, "Semantic similarity and relatedness between clinical terms: An experimental study," *AMIA Annual Symposium Proceedings*, vol. 2010, pp. 572–576, 2010. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3041430/ (visited on 10/02/2024).

[184] B. Pang and L. Lee, *Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales*, 2005. arXiv: cs/0506075.

[185] N. Pavlichenko, I. Stelmakh, and D. Ustalov, *CrowdSpeech and VoxDIY: Benchmark datasets for crowdsourced audio transcription*, 2021. arXiv: 2107.01091[cs,eess].

[186] T. Pedersen, S. V. S. Pakhomov, S. Patwardhan, and C. G. Chute, "Measures of semantic similarity and relatedness in the biomedical domain," *Journal of Biomedical Informatics*, vol. 40, no. 3, pp. 288–299, 2007. DOI: 10/fghjwr.

[187] L. Perez-Beltrachini, Y. Liu, and M. Lapata, *Generating summaries with topic templates and structured convolutional decoders*, 2019. arXiv: 1906.04687[cs].

[188] F. Petroni, T. Rocktäschel, P. Lewis, *et al.*, *Language models as knowledge bases?* 2019. arXiv: 1909.01066[cs].

[189] T. M. Pham, S. Yoon, T. Bui, and A. Nguyen, *PiC: A phrase-in-context dataset for phrase understanding and semantic search*, 2023. arXiv: 2207.09068[cs].

[190] N. F. Rajani, B. McCann, C. Xiong, and R. Socher, *Explain yourself! leveraging language models for commonsense reasoning*, 2019. arXiv: 1906.02361[cs].

[191] P. Rajpurkar, J. Zhang, K. Lopyrev, and P. Liang, *SQuAD: 100,000+ questions for machine comprehension of text*, 2016. arXiv: 1606.05250[cs].

[192] A. Royer, K. Bousmalis, S. Gouws, *et al.*, *XGAN: Unsupervised image-to-image translation for many-to-many mappings*, 2018. arXiv: 1711.05139[cs].

[193] A. M. Rush, S. Chopra, and J. Weston, *A neural attention model for abstractive sentence summarization*, 2015. arXiv: 1509.00685[cs].

[194] P. Rust, J. F. Lotz, E. Bugliarello, E. Salesky, M. de Lhoneux, and D. Elliott, *Language modelling with pixels*, 2023. arXiv: 2207.06991[cs].

[195] M. Saeidi, M. Bartolo, P. Lewis, *et al.*, *Interpretation of natural language rules in conversational machine reading*, 2018. arXiv: 1809.01494[cs,stat].

[196] A. Saha, R. Aralikatte, M. M. Khapra, and K. Sankaranarayanan, *DuoRC: Towards complex language understanding with paraphrased reading comprehension*, 2018. arXiv: 1804.07927[cs].

[197] K. Sakaguchi, R. L. Bras, C. Bhagavatula, and Y. Choi, *WinoGrande: An adversarial winograd schema challenge at scale*, 2019. arXiv: 1907.10641[cs].

[198] V. Sanh, A. Webson, C. Raffel, *et al.*, *Multitask prompted training enables zero-shot task generalization*, 2022. arXiv: 2110.08207[cs].

[199] M. Sap, H. Rashkin, D. Chen, R. LeBras, and Y. Choi, *SocialIQA: Commonsense reasoning about social interactions*, 2019. arXiv: 1904.09728[cs].

[200] C. Schulz, J. Levy-Kramer, C. Van Assel, M. Kepes, and N. Hammerla, *Biomedical concept relatedness – a large EHR-based benchmark*, 2020. arXiv: 2010.16218[cs].

[201] A. See, P. J. Liu, and C. D. Manning, *Get to the point: Summarization with pointer-generator networks*, 2017. arXiv: 1704.04368[cs].

[202] E. Sharma, C. Li, and L. Wang, *BIGPATENT: A large-scale dataset for abstractive and coherent summarization*, 2019. arXiv: 1906.03741[cs].

[203] E. Shriberg, R. Bates, A. Stolcke, *et al.*, "Can prosody aid the automatic classification of dialog acts in conversational speech?" *Language and Speech*, vol. 41 (Pt 3-4), pp. 443–492, 1998. DOI: 10.1177/002383099804100410.

[204] D. Sileo and M.-F. Moens, *Probing neural language models for understanding of words of estimative probability*, 2023. arXiv: 2211.03358[cs].

[205] A. Soleimani, C. Monz, and M. Worring, "NLQuAD: A non-factoid long question answering data set," in *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, Online: Association for Computational Linguistics, 2021, pp. 1245–1255. DOI: 10/g6k6th.

[206] A. Stolcke, K. Ries, N. Coccaro, *et al.*, "Dialogue act modeling for automatic tagging and recognition of conversational speech," *Computational Linguistics*, vol. 26, no. 3, pp. 339–373, 2000. DOI: 10/dqmv4j.

[207] O. Tafjord, M. Gardner, K. Lin, and P. Clark, *QuaRTz: An open-domain dataset of qualitative relationship questions*, 2019. arXiv: 1909.03553[cs].

[208] O. Tafjord, P. Clark, M. Gardner, W.-t. Yih, and A. Sabharwal, *QuaRel: A dataset and models for answering questions about qualitative relationships*, 2018. arXiv: 1811.08048[cs].

[209] A. Talmor, J. Herzig, N. Lourie, and J. Berant, *CommonsenseQA: A question answering challenge targeting commonsense knowledge*, 2019. arXiv: 1811.00937[cs].

[210] N. Tandon, B. D. Mishra, K. Sakaguchi, A. Bosselut, and P. Clark, *WIQA: A dataset for "what if..." reasoning over procedural text*, 2019. arXiv: 1909.04739[cs].

[211] R. Tang, R. Nogueira, E. Zhang, *et al.*, *Rapidly bootstrapping a question answering dataset for COVID-19*, 2020. arXiv: 2004.11339[cs].

[212] A. Thawani, J. Pujara, and F. Ilievski, "Numeracy enhances the literacy of language models," in *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, M.-F. Moens, X. Huang, L. Specia, and S. W.-t. Yih, Eds., Online and Punta Cana, Dominican Republic: Association for Computational Linguistics, 2021, pp. 6960–6967. DOI: 10/g6k6w5.

[213] J. Thorne, A. Vlachos, C. Christodoulopoulos, and A. Mittal, *FEVER: A large-scale dataset for fact extraction and VERification*, 2018. arXiv: 1803.05355[cs].

[214] R. Tyleček and R. Šára, "Spatial pattern templates for recognition of objects with regular structure," in *Pattern Recognition*, J. Weickert, M. Hein, and B. Schiele, Eds., Berlin, Heidelberg: Springer, 2013, pp. 364–374. DOI: 10/ggwb5g.

[215] H. Ullrich, J. Drchal, M. Rýpar, H. Vincourová, and V. Moravec, "CsFEVER and CTKFacts: Acquiring czech data for fact verification," *Language Resources and Evaluation*, vol. 57, no. 4, pp. 1571–1605, 2023. DOI: 10/g6k95h.

[216] A. Ushio, F. Alva-Manchego, and J. Camacho-Collados, *Generative language models for paragraph-level question generation*, 2023. arXiv: 2210.03992[cs].

[217] B. Wang, C. Xu, S. Wang, *et al.*, *Adversarial GLUE: A multi-task benchmark for robustness evaluation of language models*, 2022. arXiv: 2111.02840[cs].

[218] W. Wang, G. Li, B. Ma, X. Xia, and Z. Jin, *Detecting code clones with graph neural networkand flow-augmented abstract syntax tree*, 2020. arXiv: 2002.08653[cs].

[219] A. Wang, A. Singh, J. Michael, F. Hill, O. Levy, and S. R. Bowman, *GLUE: A multi-task benchmark and analysis platform for natural language understanding*, 2019. arXiv: 1804.07461[cs].

[220] Y. Wang, Y. Kordi, S. Mishra, *et al.*, *Self-instruct: Aligning language models with self-generated instructions*, 2023. arXiv: 2212.10560[cs].

[221] A. Warstadt, A. Parrish, H. Liu, *et al.*, *BLiMP: The benchmark of linguistic minimal pairs for english*, 2023. arXiv: 1912.00582[cs].

[222] J. Welbl, P. Stenetorp, and S. Riedel, *Constructing datasets for multi-hop reading comprehension across documents*, 2018. arXiv: 1710.06481[cs].

[223] J. Welbl, N. F. Liu, and M. Gardner, *Crowdsourcing multiple choice science questions*, 2017. arXiv: 1707.06209[cs,stat].

[224] O. Weller, N. Lourie, M. Gardner, and M. E. Peters, *Learning from task descriptions*, 2020. arXiv: 2011.08115[cs].

[225] J. Weston, A. Bordes, S. Chopra, *et al.*, *Towards AI-complete question answering: A set of prerequisite toy tasks*, 2015. arXiv: 1502.05698[cs,stat].

[226] A. Williams, T. Thrush, and D. Kiela, *ANLIzing the adversarial natural language inference dataset*, 2020. arXiv: 2010.12729[cs].

[227] Z. Wu, B. Ramsundar, E. N. Feinberg, *et al.*, *MoleculeNet: A benchmark for molecular machine learning*, 2018. arXiv: 1703.00564[physics,stat].

[228] W. Xiong, J. Wu, H. Wang, *et al.*, *TWEETQA: A social media focused question answering dataset*, 2019. arXiv: 1907.06292[cs].

[229] Z. Yang, P. Qi, S. Zhang, *et al.*, *HotpotQA: A dataset for diverse, explainable multi-hop question answering*, 2018. arXiv: 1809.09600[cs].

[230] R. Zellers, A. Holtzman, Y. Bisk, A. Farhadi, and Y. Choi, *HellaSwag: Can a machine really finish your sentence?* 2019. arXiv: 1905.07830[cs].

[231] X. Zhang, J. Zhao, and Y. LeCun, *Character-level convolutional networks for text classification*, 2016. arXiv: 1509.01626[cs].

[232] J. Zhang, R. Gan, J. Wang, *et al.*, *Fengshenbang 1.0: Being the foundation of chinese cognitive intelligence*, 2023. arXiv: 2209.02970[cs].

[233] Y. Zhang, J. Baldridge, and L. He, *PAWS: Paraphrase adversaries from word scrambling*, 2019. arXiv: 1904.01130[cs].

[234] Y. Zhou, S. Liu, J. Siow, X. Du, and Y. Liu, *Devign: Effective vulnerability identification by learning comprehensive program semantics via graph neural networks*, 2019. arXiv: 1909.03496[cs,stat].

[235] S. Zhou, U. Alon, F. F. Xu, Z. Wang, Z. Jiang, and G. Neubig, *DocPrompting: Generating code by retrieving the docs*, 2023. arXiv: 2207.05987[cs].

[236] M. Zhu, A. Jain, K. Suresh, R. Ravindran, S. Tipirneni, and C. K. Reddy, *XLCoST: A benchmark dataset for cross-lingual code intelligence*, 2022. arXiv: 2206.08474[cs].

[237] C. Malaviya, S. Lee, S. Chen, E. Sieber, M. Yatskar, and D. Roth, *ExpertQA: Expert-curated questions and attributed answers*, 2024. arXiv: 2309.07852[cs].

[238] S. Kim, J. Shin, Y. Cho, *et al.*, *Prometheus: Inducing fine-grained evaluation capability in language models*, 2024. arXiv: 2310.08491[cs].

[239] Z. Wang, Y. Dong, J. Zeng, *et al.*, *HelpSteer: Multi-attribute helpfulness dataset for SteerLM*, 2023. arXiv: 2311.09528[cs].

[240] J. Gala, T. Jayakumar, J. A. Husain, *et al.*, *Airavata: Introducing hindi instruction-tuned LLM*, 2024. arXiv: 2401.15006[cs].

[241] J. Hu, S. Ruder, A. Siddhant, G. Neubig, O. Firat, and M. Johnson, *XTREME: A massively multilingual multi-task benchmark for evaluating cross-lingual generalization*, 2020. arXiv: 2003.11080[cs].

[242] O. Einea, A. Elnagar, and R. Al Debsi, "SANAD: Single-label arabic news articles dataset for automatic text categorization," *Data in Brief*, vol. 25, p. 104 076, 2019. DOI: 10/g6k6cn.

[243] H. Mozannar, K. E. Hajal, E. Maamary, and H. Hajj, *Neural arabic question answering*, 2019. arXiv: 1906.05394[cs].

[244] A. Pratapa, R. Gupta, and T. Mitamura, *Multilingual event linking to wikidata*, 2022. arXiv: 2204.06535[cs].

[245] H. Chouikhi, M. Aloui, C. B. Hammou, G. Chaabane, H. Kchaou, and C. Dhaouadi, *GemmAr: Enhancing LLMs through arabic instruction-tuning*, 2024. arXiv: 2407.02147[cs].

[246] M. Abbas, K. Smaïli, and D. Berkani, "Evaluation of topic identification methods on arabic corpora," *Journal of Digital Information Management*, vol. 9, no. 5, 2011. [Online]. Available: https://www.dline.info/fpaper/jdim/v9i5/1.pdf (visited on 10/02/2024).

[247] A. Abdelghany, H. Abdelaal, A. Kamr, and P. Elkafrawy, "Doc2vec: An approach to identify hadith similarities," *Australian Journal of Basic and Applied Sciences*, pp. 46–53, 2020. DOI: 10.22587/ajbas.2020.14.12.5.

[248] M. Orabi, H. E. Rifai, and A. Elnagar, "Classical arabic poetry: Classification based on era," in *2020 IEEE/ACS 17th International Conference on Computer Systems and Applications (AICCSA)*, Antalya, Turkey: IEEE, 2020, pp. 1–6. DOI: 10/g6k6b2.

[249] H. ElSahar and S. R. El-Beltagy, "Building large arabic multi-domain resources for sentiment analysis," in *Computational Linguistics and Intelligent Text Processing*, A. Gelbukh, Ed., Cham: Springer International Publishing, 2015, pp. 23–34. DOI: 10/g6k58r.

[250] A. Elnagar and O. Einea, "BRAD 1.0: Book reviews in arabic dataset," in *2016 IEEE/ACS 13th International Conference of Computer Systems and Applications (AICCSA)*, Agadir, Morocco: IEEE, 2016, pp. 1–8. DOI: 10/g6k6jm.

[251] S. Pieri, S. S. Mullappilly, F. S. Khan, *et al.*, *BiMediX: Bilingual medical mixture of experts LLM*, 2024. arXiv: 2402.13253[cs].

[252] R. Alghamdi, Z. Liang, and X. Zhang, "ArMATH: A dataset for solving arabic math word problems," in *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, N. Calzolari, F. Béchet, P. Blache, *et al.*, Eds., Marseille, France: European Language Resources Association, 2022, pp. 351–362. [Online]. Available: https://aclanthology.org/2022.lrec-1.37 (visited on 10/02/2024).

[253] A. Abdallah, M. Kasem, M. Abdalla, *et al.*, *ArabicaQA: A comprehensive dataset for arabic question answering*, 2024. arXiv: 2403.17848[cs].

[254] M. Biltawi, A. Awajan, and S. Tedmori, "Arabic reading comprehension benchmarks created semiautomatically," in *2020 21st International Arab Conference on Information Technology (ACIT)*, Giza, Egypt: IEEE, 2020, pp. 1–6. DOI: 10/g6k6b8.

[255] M. Aloui, H. Chouikhi, G. Chaabane, H. Kchaou, and C. Dhaouadi, *101 billion arabic words dataset*, 2024. arXiv: 2405.01590[cs].

[256] I. A. El-khair, *1.5 billion words arabic corpus*, 2016. arXiv: 1611.04033[cs].

[257] F. Xu, K. Lo, L. Soldaini, B. Kuehl, E. Choi, and D. Wadden, *KIWI: A dataset of knowledge-intensive writing instructions for answering research questions*, 2024. arXiv: 2403.03866[cs].

[258] L. Zheng, W.-L. Chiang, Y. Sheng, *et al.*, *LMSYS-chat-1m: A large-scale real-world LLM conversation dataset*, 2024. arXiv: 2309.11998[cs].

[259] O. Rohanian, M. Nouriborji, and D. A. Clifton, *Exploring the effectiveness of instruction tuning in biomedical language processing*, 2023. arXiv: 2401.00579[cs].

[260] Y. Bai, X. Lv, J. Zhang, *et al.*, *LongAlign: A recipe for long context alignment of large language models*, 2024. arXiv: 2401.18058[cs].

[261] Z. Xu, F. Jiang, L. Niu, *et al.*, *Magpie: Alignment data synthesis from scratch by prompting aligned LLMs with nothing*, 2024. DOI: 10.48550/arXiv.2406.08464.

[262] J. Macina, N. Daheim, S. P. Chowdhury, *et al.*, *MathDial: A dialogue tutoring dataset with rich pedagogical properties grounded in math reasoning problems*, 2023. arXiv: 2305.14536[cs].

[263] X. Yue, X. Qu, G. Zhang, *et al.*, *MAmmoTH: Building math generalist models through hybrid instruction tuning*, 2023. arXiv: 2309.05653[cs].

[264] X. Zhang, C. Tian, X. Yang, L. Chen, Z. Li, and L. R. Petzold, *AlpaCare:instruction-tuned large language models for medical application*, 2024. arXiv: 2310.14558[cs].

[265] T. Han, L. C. Adams, J.-M. Papaioannou, *et al.*, *MedAlpaca – an open-source collection of medical conversational AI models and training data*, 2023. arXiv: 2304.08247[cs].

[266] L. L. Wang, K. Lo, Y. Chandrasekhar, *et al.*, *CORD-19: The COVID-19 open research dataset*, 2020. arXiv: 2004.10706[cs].

[267] D. Jin, E. Pan, N. Oufattole, W.-H. Weng, H. Fang, and P. Szolovits, *What disease does this patient have? a large-scale open domain question answering dataset from medical exams*, 2020. arXiv: 2009.13081[cs].

[268] M. Savery, A. B. Abacha, S. Gayen, and D. Demner-Fushman, *Question-driven summarization of answers to consumer health questions*, 2020. arXiv: 2005.09067[cs].

[269] S. Barham, O. Weller, M. Yuan, *et al.*, *MegaWika: Millions of reports and their sources across 50 diverse languages*, 2023. arXiv: 2307.07049[cs].

[270] L. Yu, W. Jiang, H. Shi, *et al.*, *MetaMath: Bootstrap your own mathematical questions for large language models*, 2023. arXiv: 2309.12284[cs].

[271] A. Köpf, Y. Kilcher, D. von Rütte, *et al.*, *OpenAssistant conversations – democratizing large language model alignment*, 2023. arXiv: 2304.07327[cs].

[272] T. Sawada, D. Paleka, A. Havrilla, *et al.*, *ARB: Advanced reasoning benchmark for large language models*, 2023. arXiv: 2307.13692[cs].

[273] T. Dettmers, A. Pagnoni, A. Holtzman, and L. Zettlemoyer, *QLoRA: Efficient finetuning of quantized LLMs*, 2023. arXiv: 2305.14314[cs].

[274] H. Lightman, V. Kosaraju, Y. Burda, *et al.*, *Let's verify step by step*, 2023. arXiv: 2305.20050[cs].

[275] W. Yu, Z. Jiang, Y. Dong, and J. Feng, *ReClor: A reading comprehension dataset requiring logical reasoning*, 2020. arXiv: 2002.04326[cs].

[276] X. Wang, Z. Hu, P. Lu, *et al.*, *SciBench: Evaluating college-level scientific problem-solving abilities of large language models*, 2024. arXiv: 2307.10635[cs].

[277] P. Lu, S. Mishra, T. Xia, *et al.*, *Learn to explain: Multimodal reasoning via thought chains for science question answering*, 2022. arXiv: 2209.09513[cs].

[278] W. Chen, M. Yin, M. Ku, *et al.*, *TheoremQA: A theorem-driven question answering dataset*, 2023. arXiv: 2305.12524[cs].

[279] S. Toshniwal, I. Moshkov, S. Narenthiran, D. Gitman, F. Jia, and I. Gitman, *OpenMathInstruct-1: A 1.8 million math instruction tuning dataset*, 2024. arXiv: 2402.10176[cs].

[280] A. Mitra, H. Khanpour, C. Rosset, and A. Awadallah, *Orca-math: Unlocking the potential of SLMs in grade school math*, 2024. arXiv: 2402.14830[cs].

[281] C. Wu, W. Lin, X. Zhang, Y. Zhang, Y. Wang, and W. Xie, *PMC-LLaMA: Towards building open-source language models for medicine*, 2023. arXiv: 2304.14454[cs].

[282] T. Gosling, A. Dale, and Y. Zheng, *PIPPA: A partially synthetic conversational dataset*, 2023. arXiv: 2308.05884[cs].

[283] B. Y. Lin, Z. Wu, Y. Yang, D.-H. Lee, and X. Ren, *RiddleSense: Reasoning about riddle questions featuring linguistic creativity and commonsense knowledge*, 2021. arXiv: 2101.00376[cs].

[284] X.-P. Nguyen, W. Zhang, X. Li, *et al.*, *SeaLLMs – large language models for southeast asia*, 2023. arXiv: 2312.00738[cs].

[285] S. Ye, Y. Jo, D. Kim, S. Kim, H. Hwang, and M. Seo. "SelFee: Iterative self-revising LLM empowered by self-feedback generation," LK Lab. (2023), [Online]. Available: https://kaistai.github.io/SelFee/ (visited on 05/01/2024).

[286] N. Ding, Y. Chen, B. Xu, *et al.*, *Enhancing chat language models by scaling high-quality instructional conversations*, 2023. arXiv: 2305.14233[cs].

[287] W. Zhao, X. Ren, J. Hessel, C. Cardie, Y. Choi, and Y. Deng, "WildChat: 1m ChatGPT interaction logs in the wild," in *Proceedings of the Twelfth International Conference on Learning Representations*, 2023. [Online]. Available: https://openreview.net/forum?id=Bl8u7ZRlbM (visited on 05/01/2024).

[288] A. Bhanushali, G. Bridgman, D. G, *et al.*, "Gram vaani ASR challenge on spontaneous telephone speech recordings in regional variations of hindi," in *Interspeech 2022*, ISCA, 2022, pp. 3548–3552. DOI: 10/gtsqzn.

[289] T. Olatunji, T. Afonja, A. Yadavalli, *et al.*, *AfriSpeech-200: Pan-african accented speech dataset for clinical and general domain ASR*, 2023. arXiv: 2310.00274[cs].

[290] H. Bu, J. Du, X. Na, B. Wu, and H. Zheng, *AISHELL-1: An open-source mandarin speech corpus and a speech recognition baseline*, 2017. arXiv: 1709.05522[cs].

[291] J. Du, X. Na, X. Liu, and H. Bu, *AISHELL-2: Transforming mandarin ASR research into industrial scale*, 2018. arXiv: 1808.10583[cs].

[292] Y. Fu, L. Cheng, S. Lv, *et al.*, *AISHELL-4: An open source dataset for speech enhancement, separation, recognition and speaker diarization in conference scenario*, 2021. arXiv: 2104.03603[cs,eess].

[293] A. Bradlow, *ALLSSTAR: Archive of l1 and l2 scripted and spontaneous transcripts and recordings*, 2010. [Online]. Available: https://speechbox.linguistics.northwestern.edu/#!/?goto=allsstar (visited on 05/01/2024).

[294] J. Carletta, S. Ashby, S. Bourban, *et al.*, "The AMI meeting corpus: A pre-announcement," in *Machine Learning for Multimodal Interaction*, S. Renals and S. Bengio, Eds., vol. 3869, Series Title: Lecture Notes in Computer Science, Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 28–39. DOI: 10.1007/11677482_3.

[295] A. Virkkunen, A. Rouhe, N. Phan, and M. Kurimo, *Finnish parliament ASR corpus - analysis, benchmarks and statistics*, 2022. arXiv: 2203.14876[cs,eess].

[296] O. Kjartansson, A. Gutkin, A. Butryna, I. Demirsahin, and C. Rivera, "Open-source high quality speech datasets for basque, catalan and galician," in *Proceedings of the 1st Joint Workshop on Spoken Language Technologies for Under-resourced languages (SLTU) and Collaboration and Computing for Under-Resourced Languages (CCURL)*, D. Beermann, L. Besacier, S. Sakti, and C. Soria, Eds., Marseille, France: European Language Resources association, 2020, pp. 21–27, ISBN: 979-10-95546-35-1. [Online]. Available: https://aclanthology.org/2020.sltu-1.3 (visited on 05/29/2024).

[297] C. Leong, J. Nemecek, J. Mansdorfer, A. Filighera, A. Owodunni, and D. Whitenack, *Bloom library: Multimodal datasets in 300+ languages for a variety of downstream tasks*, 2022. arXiv: 2210.14712[cs].

[298] K. Maekawa, "Corpus of spontaneous japanese: Its design and evaluation," in *Proceedings of the ISCA/IEEE Workshop on Spontaneous Speech Processing and Recognition*, 2003, paper MMO2. [Online]. Available: https://www.isca-archive.org/sspr_2003/maekawa03_sspr.html (visited on 05/29/2024).

[299] Lander, T, *CSLU: Foreign accented english release 1.2*, Artwork Size: 1468006 KB Pages: 1468006 KB, 2007. DOI: 10.35111/0VWP-XN48.

[300] Lander, T, *CSLU: 22 languages corpus*, 2005. DOI: 10.35111/ZKN2-5X88.

[301] D. Koržinek, K. Marasek, Ł. Brocki, and K. Wołk, *Polish read speech corpus for speech tools and services*, 2017. arXiv: 1706.00245[cs].

[302] S. Coats, "A corpus of regional american language from YouTube," in *Proceedings of the Digital Humanities in the Nordic Countries 5th Conference*, 2019. [Online]. Available: https://www.semanticscholar.org/paper/A-Corpus-of-Regional-American-Language-from-YouTube-Coats/bc428db824d261794a7e081a53c4315b8e02f855 (visited on 05/29/2024).

[303] C. Wang, A. Wu, and J. Pino, *CoVoST 2 and massively multilingual speech-to-text translation*, 2020. arXiv: 2007.10310[cs,eess].

[304] R. Ardila, M. Branson, K. Davis, *et al.*, *Common voice: A massively-multilingual speech corpus*, 2020. arXiv: 1912.06670[cs].

[305] O. Kjartansson, S. Sarin, K. Pipatsrisawat, M. Jansche, and L. Ha, "Crowd-sourced speech corpora for javanese, sundanese, sinhala, nepali, and bangladeshi bengali," in *6th Workshop on Spoken Language Technologies for Under-Resourced Languages (SLTU 2018)*, ISCA, 2018, pp. 52–55. DOI: 10/gtwwbs.

[306] J. Kratochvil, P. Polák, and O. Bojar, "Large corpus of czech parliament plenary hearings," in *Proceedings of the Twelfth Language Resources and Evaluation Conference*, N. Calzolari, F. Béchet, P. Blache, *et al.*, Eds., Marseille, France: European Language Resources Association, 2020, pp. 6363–6367, ISBN: 979-10-95546-34-4. [Online]. Available: https://aclanthology.org/2020.lrec-1.781 (visited on 05/29/2024).

[307] T. Guo, C. Wen, D. Jiang, *et al.*, *DiDiSpeech: A large scale mandarin speech corpus*, 2021. arXiv: 2010.09275[eess].

[308] M. Del Rio, P. Ha, Q. McNamara, C. Miller, and S. Chandra, *Earnings-22: A practical benchmark for accents in the wild*, 2022. arXiv: 2203.15591[cs].

[309] R. Sanabria, N. Bogoychev, N. Markl, A. Carmantini, O. Klejch, and P. Bell, *The edinburgh international accents of english corpus: Towards the democratization of english ASR*, 2023. arXiv: 2303.18110[cs,eess].

[310] I. Demirsahin, O. Kjartansson, A. Gutkin, and C. Rivera, "Open-source multi-speaker corpora of the english accents in the british isles," in *Proceedings of the Twelfth Language Resources and Evaluation Conference*, N. Calzolari, F. Béchet, P. Blache, *et al.*, Eds., Marseille, France: European Language Resources Association, 2020, pp. 6532–6541, ISBN: 979-10-95546-34-4. [Online]. Available: https://aclanthology.org/2020.lrec-1.804 (visited on 05/29/2024).

[311] L. T. Nguyen, N. L. Tran, L. Doan, M. Luong, and D. Q. Nguyen, *A high-quality and large-scale dataset for english-vietnamese speech translation*, 2022. arXiv: 2208.04243[cs].

[312] A. Kirkedal, M. Stepanović, and B. Plank, *FT speech: Danish parliament speech corpus*, 2020. DOI: 10.21437/Interspeech.2020-3164.

[313] C. Cieri, D. Miller, and K. Walker, "The fisher corpus: A resource for the next generations of speech-to-text," in *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04)*, M. T. Lino, M. F. Xavier, F. Ferreira, R. Costa, and R. Silva, Eds., Lisbon, Portugal: European Language Resources Association (ELRA), 2004. [Online]. Available: http://www.lrec-conf.org/proceedings/lrec2004/pdf/767.pdf (visited on 05/01/2024).

[314] A. Conneau, M. Ma, S. Khanuja, *et al.*, *FLEURS: Few-shot learning evaluation of universal representations of speech*, version: 1, 2022. arXiv: 2205.12446[cs,eess].

[315] G. Chen, S. Chai, G. Wang, *et al.*, *GigaSpeech: An evolving, multi-domain ASR corpus with 10,000 hours of transcribed audio*, 2021. arXiv: 2106.06909[cs,eess].

[316] N. Karpov, A. Denisenko, and F. Minkin, *Golos: Russian dataset for speech research*, 2021. arXiv: 2106.10161[eess].

[317] J. Shi, J. D. Amith, X. Chang, S. Dalmia, B. Yan, and S. Watanabe, "Highland puebla nahuatl speech translation corpus for endangered language documentation," in *Proceedings of the First Workshop on Natural Language Processing for Indigenous Languages of the Americas*, M. Mager, A. Oncevay, A. Rios, *et al.*, Eds., Online: Association for Computational Linguistics, 2021, pp. 53–63. DOI: 10/gtwwcd.

[318] S. Takamichi, L. Kürzinger, T. Saeki, S. Shiota, and S. Watanabe, *JTubeSpeech: Corpus of japanese speech collected from YouTube for speech recognition and speaker verification*, 2021. arXiv: 2112.09323[cs,eess].

[319] Y. Khassanov, S. Mussakhojayeva, A. Mirzakhmetov, A. Adiyev, M. Nurpeiissov, and H. A. Varol, *A crowdsourced open-source kazakh speech corpus and initial speech recognition baseline*, 2021. arXiv: 2009.10334[cs,eess].

[320] T. Javed, K. S. Bhogale, A. Raman, A. Kunchukuttan, P. Kumar, and M. M. Khapra, *IndicSUPERB: A speech processing universal performance benchmark for indian languages*, 2022. arXiv: 2208.11761[cs,eess].

[321] Z. Tang, D. Wang, Y. Xu, *et al.*, "KeSpeech: An open source speech dataset of mandarin and its eight subdialects," in *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks*, vol. 1, 2021. [Online]. Available: https://datasets-benchmarks-proceedings.neurips.cc/paper/2021/hash/0336dcbab05b9d5ad24f4333c7658a0e-Abstract-round2.html (visited on 05/29/2024).

[322] J.-U. Bang, S. Yun, S.-H. Kim, *et al.*, "KsponSpeech: Korean spontaneous speech corpus for automatic speech recognition," *Applied Sciences*, vol. 10, no. 19, p. 6936, 2020, Number: 19 Publisher: Multidisciplinary Digital Publishing Institute. DOI: 10/gtwwck.

[323] K. Ito and L. Johnson, *The LJ speech dataset*, 2017. [Online]. Available: https://keithito.com/LJ-Speech-Dataset (visited on 05/01/2024).

[324] S. Ando and H. Fujihara, *Construction of a large-scale japanese ASR corpus on TV recordings*, 2021. arXiv: 2103.14736[cs,eess].

[325] V. Panayotov, G. Chen, D. Povey, and S. Khudanpur, "Librispeech: An ASR corpus based on public domain audio books," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, South Brisbane, Queensland, Australia: IEEE, 2015, pp. 5206–5210. DOI: 10/gfv84w.

[326] I. Solak, *The m-AILABS speech dataset – caito*, 2024. [Online]. Available: https://www.caito.de/2019/01/03/the-m-ailabs-speech-dataset/ (visited on 05/01/2024).

[327] Y. Shi, A. Hamdullah, Z. Tang, D. Wang, and T. F. Zheng, "A free kazakh speech database and a speech recognition baseline," in *2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, Kuala Lumpur: IEEE, 2017, pp. 745–748. DOI: 10/gtsqzf.

[328] I. Mamtimin, W. Du, and A. Hamdulla, "M2asr-KIRGHIZ: A free kirghiz speech database and accompanied baselines," *Information*, vol. 14, no. 1, p. 55, 2023. DOI: 10/gtsqzm.

[329] T. Zhi, Y. Shi, W. Du, G. Li, and D. Wang, "M2asr-MONGO: A free mongolian speech database and accompanied baselines," in *2021 24th Conference of the Oriental COCOSDA International Committee for the Co-ordination and Standardisation of Speech Databases and Assessment Techniques (O-COCOSDA)*, Singapore, Singapore: IEEE, 2021, pp. 140–145. DOI: 10/gtsqzg.

[330] G. Li, H. Yu, T. F. Zheng, J. Yan, and S. Xu, "Free linguistic and speech resources for tibetan," in *2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, Kuala Lumpur: IEEE, 2017, pp. 733–736. DOI: 10/gtsqzh.

[331] M. Al-Fetyani, M. Al-Barham, G. Abandah, A. Alsharkawi, and M. Dawas, "MASC: Massive arabic speech corpus," in *2022 IEEE Spoken Language Technology Workshop (SLT)*, Doha, Qatar: IEEE, 2023, pp. 1006–1013. DOI: 10/gtsqzj.

[332] M. Z. Boito, W. N. Havard, M. Garnerin, É. L. Ferrand, and L. Besacier, *MaSS: A large and clean multilingual corpus of sentence-aligned spoken utterances extracted from the bible*, 2020. arXiv: 1907.12895[cs].

[333] Z. Yang, Y. Chen, L. Luo, *et al.*, *Open source MagicData-RAMC: A rich annotated mandarin conversational(RAMC) speech dataset*, 2022. arXiv: 2203.16844[cs,eess].

[334] R. Kolobov, O. Okhapkina, O. Omelchishina, *et al.*, *MediaSpeech: Multilanguage ASR benchmark and dataset*, 2021. arXiv: 2103.16193[cs,eess].

[335] D. Gerz, P.-H. Su, R. Kusztos, *et al.*, *Multilingual and cross-lingual intent detection from spoken data*, 2021. arXiv: 2104.08524[cs].

[336] M. A. Di Gangi, R. Cattoni, L. Bentivogli, M. Negri, and M. Turchi, "MuST-c: A multilingual speech translation corpus," 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)*, J. Burstein, C. Doran, and T. Solorio, Eds., Minneapolis, Minnesota: Association for Computational Linguistics, 2019, pp. 2012–2017. DOI: 10/gtsqzk.

[337] V. Pratap, Q. Xu, A. Sriram, G. Synnaeve, and R. Collobert, "MLS: A large-scale multilingual dataset for speech research," in *Interspeech 2020*, 2020, pp. 2757–2761. DOI: 10/grk6mp.

[338] E. Salesky, M. Wiesner, J. Bremerman, *et al.*, *The multilingual TEDx corpus for speech recognition and translation*, 2021. arXiv: 2102.01757[cs].

[339] P. E. Solberg and P. Ortiz, "The norwegian parliamentary speech corpus," in *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, N. Calzolari, F. Béchet, P. Blache, *et al.*, Eds., Marseille, France: European Language Resources Association, 2022, pp. 1003–1008. [Online]. Available: https://aclanthology.org/2022.lrec-1.106 (visited on 05/29/2024).

[340] J. Park, J.-W. Hwang, K. Choi, *et al.*, *OLKAVS: An open large-scale korean audio-visual speech dataset*, 2023. arXiv: 2301.06375[cs].

[341] A. Andrusenko, A. Laptev, and I. Medennikov, "Exploration of end-to-end ASR for OpenSTT – russian open speech-to-text dataset," in *Lecture Notes in Computer Science*, vol. 12335, Springer, Cham, 2020, pp. 35–44. DOI: 10.1007/978-3-030-60276-5_4.

[342] D. Galvez, G. Diamos, J. Ciro, *et al.*, *The people's speech: A large-scale diverse english speech recognition dataset for commercial usage*, 2021. arXiv: 2111.09344[cs,stat].

[343] H. Mubarak, A. Hussein, S. A. Chowdhury, and A. Ali, *QASR: QCRI aljazeera speech resource – a large scale annotated arabic speech corpus*, 2021. arXiv: 2106.13000[cs,eess].

[344] Y. Yin, D. Mori, and S. Fujimoto, *ReazonSpeech: A free and massive corpus for japanese ASR*, 2023. [Online]. Available: https://research.reazon.jp/_static/reazonspeech_nlp2023.pdf.

[345] M. Plüss, M. Hürlimann, M. Cuny, *et al.*, *SDS-200: A swiss german speech to standard german text corpus*, 2022. arXiv: 2205.09501[cs].

[346] P. K. O'Neill, V. Lavrukhin, S. Majumdar, *et al.*, *SPGISpeech: 5,000 hours of transcribed financial audio for fully formatted end-to-end speech recognition*, 2021. arXiv: 2104.02014[cs,eess].

[347] D. E. Mollberg, Ó. H. Jónsson, S. Þorsteinsdóttir, S. Steingrímsson, E. H. Magnúsdóttir, and J. Gudnason, "Samrómur: Crowd-sourcing data collection for icelandic speech recognition," in *Proceedings of the Twelfth Language Resources and Evaluation Conference*, N. Calzolari, F. Béchet, P. Blache, *et al.*, Eds., Marseille, France: European Language Resources Association, 2020, pp. 3463–3467, ISBN: 979-10-95546-34-4. [Online]. Available: https://aclanthology.org/2020.lrec-1.425 (visited on 05/01/2024).

[348] C. D. Hernandez Mena, D. E. Mollberg, M. Borský, and J. Guðnason, "Samrómur children: An icelandic speech corpus," in *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, N. Calzolari, F. Béchet, P. Blache, *et al.*, Eds., Marseille, France: European Language Resources Association, 2022, pp. 995–1002. [Online]. Available: https://aclanthology.org/2022.lrec-1.105 (visited on 05/01/2024).

[349] K. S. Bhogale, A. Raman, T. Javed, *et al.*, *Effectiveness of mining audio and text pairs from public data for improving ASR systems for low-resource languages*, 2022. arXiv: 2208.12666[cs,eess].

[350] K. Raju, A. V, R. Lish, and J. Mathew, *Snow mountain: Dataset of audio recordings of the bible in low resource languages*, 2023. arXiv: 2206.01205[cs,eess].

[351] T. Baumann, A. Köhn, and F. Hennig, "The spoken wikipedia corpus collection: Harvesting, alignment and an application to hyperlistening," *Language Resources and Evaluation*, vol. 53, no. 2, pp. 303–329, 2019. DOI: 10/gq5xdf.

[352] J. Godfrey, E. Holliman, and J. McDaniel, "SWITCHBOARD: Telephone speech corpus for research and development," in *[Proceedings] ICASSP-92: 1992 IEEE International Conference on Acoustics, Speech, and Signal Processing*, San Francisco, CA, USA: IEEE, 1992, 517–520 vol.1. DOI: 10/fp48kw.

[353] F. Hernandez, V. Nguyen, S. Ghannay, N. Tomashenko, and Y. Estève, "TED-LIUM 3: Twice as much data and corpus repartition for experiments on speaker adaptation," in *Lecture Notes in Computer Science*, vol. 11096, Springer, Cham, 2018, pp. 198–208. DOI: 10.1007/978-3-319-99579-3_21.

[354] D. Wang and X. Zhang, *THCHS-30 : A free chinese speech corpus*, 2015. arXiv: 1512.01882[cs].

[355] A. Rozi, Dong Wang, Zhiyong Zhang, and T. F. Zheng, "An open/free database and benchmark for uyghur speaker recognition," in *2015 International Conference Oriental COCOSDA held jointly with 2015 Conference on Asian Spoken Language Research and Evaluation (O-COCOSDA/CASLRE)*, Shanghai, China: IEEE, 2015, pp. 81–85. DOI: 10/grh5rd.

[356] Garofolo, John S., Lamel, Lori F., Fisher, William M., *et al.*, *TIMIT acoustic-phonetic continuous speech corpus*, Artwork Size: 715776 KB Pages: 715776 KB, 1993. DOI: 10.35111/17GK-BN40.

[357] C. Wang, M. Rivière, A. Lee, *et al.*, *VoxPopuli: A large-scale multilingual speech corpus for representation learning, semi-supervised learning and interpretation*, 2021. arXiv: 2101.00390[cs,eess].

[358] M. Korvas, O. Plátek, O. Dušek, L. Žilka, and F. Jurčíček, "Free english and czech telephone speech corpus shared under the CC-BY-SA 3.0 license," in *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, N. Calzolari, K. Choukri, T. Declerck, *et al.*, Eds., Reykjavik, Iceland: European Language Resources Association (ELRA), 2014, pp. 4423–4428. [Online]. Available: http://www.lrec-conf.org/proceedings/lrec2014/pdf/535_Paper.pdf (visited on 05/01/2024).

[359] B. Zhang, H. Lv, P. Guo, *et al.*, *WenetSpeech: A 10000+ hours multi-domain mandarin corpus for speech recognition*, 2022. arXiv: 2110.03370[cs].

[360] M. Doumbouya, L. Einstein, and C. Piech, *Using radio archives for low-resource speech recognition: Towards an intelligent virtual assistant for illiterate users*, 2021. arXiv: 2104.13083[cs].

[361] X. Li, S. Takamichi, T. Saeki, W. Chen, S. Shiota, and S. Watanabe, "Yodas: Youtube-oriented dataset for audio and speech," in *2023 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, Taipei, Taiwan: IEEE, 2023, pp. 1–8. DOI: 10/gtsqzc.

[362] D. Shan, J. Geng, M. Shu, and D. F. Fouhey, *Understanding human hands in contact at internet scale*, 2020. arXiv: 2006.06669[cs].

[363] R. Goyal, S. E. Kahou, V. Michalski, *et al.*, *The "something something" video database for learning and evaluating visual common sense*, 2017. arXiv: 1706.04261[cs].

[364] J. Materzynska, G. Berger, I. Bax, and R. Memisevic, "The jester dataset: A large-scale video dataset of human gestures," in *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, Seoul, Korea (South): IEEE, 2019, pp. 2874–2882. DOI: 10/gh5k47.

[365] S. Stein and S. J. McKenna, "Combining embedded accelerometers with computer vision for recognizing food preparation activities," in *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*, ser. UbiComp '13, New York, NY, USA: Association for Computing Machinery, 2013, pp. 729–738. DOI: 10/gtwv9t.

[366] F. C. Heilbron, V. Escorcia, B. Ghanem, and J. C. Niebles, "ActivityNet: A large-scale video benchmark for human activity understanding," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA: IEEE, 2015, pp. 961–970. DOI: 10/gfsvdw.

[367] J. L. Alcazar, L. Mai, F. Perazzi, *et al.*, *APES: Audiovisual person search in untrimmed video*, 2021. arXiv: 2106.01667[cs].

[368] J. Roth, S. Chaudhuri, O. Klejch, *et al.*, *AVA-ActiveSpeaker: An audio-visual dataset for active speaker detection*, 2019. arXiv: 1901.01342[cs,eess].

[369] C. Gu, C. Sun, D. A. Ross, *et al.*, *AVA: A video dataset of spatio-temporally localized atomic visual actions*, 2018. arXiv: 1705.08421[cs].

[370] H. Kuehne, A. Arslan, and T. Serre, "The language of actions: Recovering the syntax and semantics of goal-directed human activities," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA: IEEE, 2014, pp. 780–787. DOI: 10/gqdc3v.

[371] W. Xiang, C. Li, K. Li, B. Wang, X.-S. Hua, and L. Zhang, "CDAD: A common daily action dataset with collected hard negative samples," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, New Orleans, LA, USA: IEEE, 2022, pp. 3920–3929. DOI: 10/gtsqzp.

[372] Y. Tang, D. Ding, Y. Rao, *et al.*, *COIN: A large-scale dataset for comprehensive instructional video analysis*, 2019. arXiv: 1903.02874[cs].

[373] G. A. Sigurdsson, G. Varol, X. Wang, A. Farhadi, I. Laptev, and A. Gupta, *Hollywood in homes: Crowdsourcing data collection for activity understanding*, 2016. arXiv: 1604.01753[cs].

[374] G. A. Sigurdsson, A. Gupta, C. Schmid, A. Farhadi, and K. Alahari, *Actor and observer: Joint modeling of first and third-person videos*, 2018. arXiv: 1804.09627[cs].

[375] Wongun Choi, K. Shahid, and S. Savarese, "What are they doing? : Collective activity classification using spatio-temporal relationship among people," in *2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops*, Kyoto, Japan: IEEE, 2009, pp. 1282–1289. DOI: 10/fv4tmg.

[376] M. Bain, A. Nagrani, A. Brown, and A. Zisserman, *Condensed movies: Story based retrieval with contextual embeddings*, 2020. arXiv: 2005.04208[cs].

[377] D. Zhukov, J.-B. Alayrac, R. G. Cinbis, D. Fouhey, I. Laptev, and J. Sivic, *Cross-task weakly supervised learning from instructional videos*, 2019. arXiv: 1903.08225[cs].

[378] F. Perazzi, J. Pont-Tuset, B. McWilliams, L. Van Gool, M. Gross, and A. Sorkine-Hornung, "A benchmark dataset and evaluation methodology for video object segmentation," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, ISSN: 1063-6919, 2016, pp. 724–732. DOI: 10/ggdmmw.

[379] L. A. Hendricks, O. Wang, E. Shechtman, J. Sivic, T. Darrell, and B. Russell, "Localizing moments in video with temporal language," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, E. Riloff, D. Chiang, J. Hockenmaier, and J. Tsujii, Eds., Brussels, Belgium: Association for Computational Linguistics, 2018, pp. 1380–1390. DOI: 10/gtsqzq.

[380] J. J. Sun, T. Liu, A. S. Cowen, F. Schroff, H. Adam, and G. Prasad, *EEV: A large-scale dataset for studying evoked expressions from video*, 2021. arXiv: 2001.05488[cs].

[381] D. Damen, H. Doughty, G. M. Farinella, *et al.*, *Scaling egocentric vision: The EPIC-KITCHENS dataset*, 2018. arXiv: 1804.02748[cs].

[382] K. Grauman, A. Westbury, E. Byrne, *et al.*, *Ego4d: Around the world in 3,000 hours of egocentric video*, 2022. arXiv: 2110.07058[cs].

[383] Y. Wang, Y. Sun, Y. Huang, *et al.*, *FERV39k: A large-scale multi-scene dataset for facial expression recognition in videos*, 2022. arXiv: 2203.09463[cs].

[384] D. Shao, Y. Zhao, B. Dai, and D. Lin, *FineGym: A hierarchical video dataset for fine-grained action understanding*, 2020. arXiv: 2004.06704[cs].

[385] J. Chung, C.-h. Wuu, H.-r. Yang, Y.-W. Tai, and C.-K. Tang, *HAA500: Human-centric atomic action dataset with curated videos*, 2021. arXiv: 2009.05224[cs,eess].

[386] H. Zhao, A. Torralba, L. Torresani, and Z. Yan, *HACS: Human action clips and segments dataset for recognition and temporal localization*, 2019. arXiv: 1712.09374[cs].

[387] H. Xue, T. Hang, Y. Zeng, *et al.*, *Advancing high-resolution video-language representation with large-scale video transcriptions*, 2022. arXiv: 2111.10337[cs].

[388] H. Kuehne, H. Jhuang, E. Garrote, T. Poggio, and T. Serre, "HMDB: A large video database for human motion recognition," in *2011 International Conference on Computer Vision*, Barcelona, Spain: IEEE, 2011, pp. 2556–2563. DOI: 10/fxpf8k.

[389] M. Marszalek, I. Laptev, and C. Schmid, "Actions in context," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, Miami, FL: IEEE, 2009, pp. 2929–2936. DOI: 10/d5bs7p.

[390] N. Rai, H. Chen, J. Ji, *et al.*, *Home action genome: Cooperative compositional action understanding*, 2021. arXiv: 2105.05226[cs].

[391] A. Diba, M. Fayyaz, V. Sharma, *et al.*, *Large scale holistic video understanding*, 2020. arXiv: 1904.11451[cs].

[392] P. Bojanowski, R. Lajugie, F. Bach, *et al.*, *Weakly supervised action labeling in videos under ordering constraints*, 2014. arXiv: 1407.1208[cs].

[393] R. Sanabria, O. Caglayan, S. Palaskar, *et al.*, *How2: A large-scale dataset for multimodal language understanding*, 2018. arXiv: 1811.00347[cs].

[394] A. Miech, D. Zhukov, J.-B. Alayrac, M. Tapaswi, I. Laptev, and J. Sivic, *HowTo100m: Learning a text-video embedding by watching hundred million narrated video clips*, 2019. arXiv: 1906.03327[cs].

[395] O. Russakovsky, J. Deng, H. Su, *et al.*, "ImageNet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015. DOI: 10/gcgk7w.

[396] W. Kay, J. Carreira, K. Simonyan, *et al.*, *The kinetics human action video dataset*, 2017. arXiv: 1705.06950[cs].

[397] J. Carreira, E. Noland, A. Banki-Horvath, C. Hillier, and A. Zisserman, *A short note about kinetics-600*, 2018. arXiv: 1808.01340[cs].

[398] L. Smaira, J. Carreira, E. Noland, E. Clancy, A. Wu, and A. Zisserman, *A short note on the kinetics-700-2020 human action dataset*, 2020. arXiv: 2010.10864[cs].

[399] B. Jia, Y. Chen, S. Huang, Y. Zhu, and S.-c. Zhu, *LEMMA: A multi-view dataset for learning multi-agent multi-task activities*, 2020. arXiv: 2007.15781[cs].

[400] A. Rohrbach, A. Torabi, M. Rohrbach, *et al.*, *Movie description*, 2016. arXiv: 1605.03705[cs].

[401] V. Sharma, M. Tapaswi, and R. Stiefelhagen, *Deep multimodal feature encoding for video ordering*, 2020. arXiv: 2004.02205[cs].

[402] M. Monfort, B. Pan, K. Ramakrishnan, *et al.*, *Multi-moments in time: Learning and interpreting models for multi-action video understanding*, 2021. arXiv: 1911.00232[cs,eess].

[403] M. Soldan, A. Pardo, J. L. Alcázar, *et al.*, *MAD: A scalable dataset for language grounding in videos from movie audio descriptions*, 2022. arXiv: 2112.00431[cs].

[404] Q. Kong, Z. Wu, Z. Deng, M. Klinkigt, B. Tong, and T. Murakami, "MMAct: A large-scale dataset for cross modal human action understanding," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, ISSN: 2380-7504, 2019, pp. 8657–8666. DOI: 10/ghfhxx.

[405] M. Rohrbach, A. Rohrbach, M. Regneri, *et al.*, "Recognizing fine-grained and composite activities using hand-centric features and script data," *International Journal of Computer Vision*, vol. 119, no. 3, pp. 346–373, 2016. DOI: 10/f8w6kp.

[406] A. Rohrbach, M. Rohrbach, N. Tandon, and B. Schiele, *A dataset for movie description*, 2015. arXiv: 1501.02530[cs].

[407] Y. Xiong, Q. Huang, L. Guo, H. Zhou, B. Zhou, and D. Lin, *A graph-based framework to bridge movies and synopses*, 2019. arXiv: 1910.11009[cs].

[408] J. Xu, T. Mei, T. Yao, and Y. Rui, "MSR-VTT: A large video description dataset for bridging video and language," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, ISSN: 1063-6919, 2016, pp. 5288–5296. DOI: 10/ggv9gj.

[409] K. Li, Y. Wang, Y. He, *et al.*, *MVBench: A comprehensive multi-modal video understanding benchmark*, 2024. arXiv: 2311.17005[cs].

[410] L. Zheng, Z. Bie, Y. Sun, *et al.*, "MARS: A video benchmark for large-scale person re-identification," in *Computer Vision – ECCV 2016*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds., Cham: Springer International Publishing, 2016, pp. 868–884. DOI: 10/gtwv9w.

[411] P. Weinzaepfel and G. Rogez, *Mimetics: Towards understanding human actions out of context*, 2021. arXiv: 1912.07249[cs].

[412] M. Monfort, A. Andonian, B. Zhou, *et al.*, *Moments in time dataset: One million videos for event understanding*, 2019. arXiv: 1801.03150[cs].

[413] Q. Huang, Y. Xiong, A. Rao, J. Wang, and D. Lin, *MovieNet: A holistic dataset for movie understanding*, 2020. arXiv: 2007.10937[cs].

[414] P. Vicol, M. Tapaswi, L. Castrejon, and S. Fidler, *MovieGraphs: Towards understanding human-centric situations from videos*, 2018. arXiv: 1712.06761[cs].

[415] M. Tapaswi, Y. Zhu, R. Stiefelhagen, A. Torralba, R. Urtasun, and S. Fidler, *MovieQA: Understanding stories in movies through question-answering*, 2016. arXiv: 1512.02902[cs].

[416] A. Rao, L. Xu, Y. Xiong, *et al.*, *A local-to-global approach to multi-modal movie scene segmentation*, 2020. arXiv: 2004.02678[cs].

[417] S. Yeung, O. Russakovsky, N. Jin, M. Andriluka, G. Mori, and L. Fei-Fei, *Every moment counts: Dense detailed labeling of actions in complex videos*, 2017. arXiv: 1507.05738[cs].

[418] A. Shahroudy, J. Liu, T.-T. Ng, and G. Wang, *NTU RGB+d: A large scale dataset for 3d human activity analysis*, 2016. arXiv: 1604.02808[cs].

[419] J.-B. Alayrac, P. Bojanowski, N. Agrawal, J. Sivic, I. Laptev, and S. Lacoste-Julien, *Unsupervised learning from narrated instruction videos*, 2016. arXiv: 1506.09215[cs].

[420] H. Duan, Y. Zhao, Y. Xiong, W. Liu, and D. Lin, *Omni-sourced webly-supervised learning for video recognition*, 2020. arXiv: 2003.13042[cs].

[421] D. Epstein, B. Chen, and C. Vondrick, "Oops! predicting unintentional action in video," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA: IEEE, 2020, pp. 916–926. DOI: 10/ghbckh.

[422] K. Nan, R. Xie, P. Zhou, *et al.*, *OpenVid-1m: A large-scale high-quality dataset for text-to-video generation*, 2024. arXiv: 2407.02371[cs].

[423] C. Liu, Y. Hu, Y. Li, S. Song, and J. Liu, *PKU-MMD: A large scale benchmark for continuous multi-modal human action understanding*, 2017. arXiv: 1703.07475[cs].

[424] X. Pan, N. Charron, Y. Yang, *et al.*, *Aria digital twin: A new benchmark dataset for egocentric 3d machine perception*, 2023. arXiv: 2306.06362[cs].

[425] Z. Lv, N. Charron, P. Moulon, *et al.*, *Aria everyday activities dataset*, 2024. arXiv: 2402.13349[cs].

[426] A. Sharghi, J. S. Laurel, and B. Gong, *Query-focused video summarization: Dataset, evaluation, and a memory network based approach*, 2017. arXiv: 1707.04960[cs].

[427] A.-M. Oncescu, J. F. Henriques, Y. Liu, A. Zisserman, and S. Albanie, *QuerYD: A video dataset with high-quality text and audio narrations*, 2021. arXiv: 2011.11071[cs].

[428] A. Miech, J.-B. Alayrac, I. Laptev, J. Sivic, and A. Zisserman, *RareAct: A video dataset of unusual interactions*, 2020. arXiv: 2008.01018[cs].

[429] L. Chen, X. Wei, J. Li, *et al.*, *ShareGPT4video: Improving video understanding and generation with better captions*, 2024. arXiv: 2406.04325[cs].

[430] M. Monfort, S. Jin, A. Liu, *et al.*, *Spoken moments: Learning joint audio-visual representations from video descriptions*, 2021. arXiv: 2105.04489[cs,eess].

[431] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei, "Large-scale video classification with convolutional neural networks," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA: IEEE, 2014, pp. 1725–1732. DOI: 10/gf4hdn.

[432] M. Tapaswi, M. Bauml, and R. Stiefelhagen, "StoryGraphs: Visualizing character interactions as a timeline," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 827–834. [Online]. Available: https://openaccess.thecvf.com/content_cvpr_2014/html/Tapaswi_StoryGraphs_Visualizing_Character_2014_CVPR_paper.html (visited on 05/29/2024).

[433] M. Gygli, H. Grabner, H. Riemenschneider, and L. Van Gool, "Creating summaries from user videos," in *Computer Vision – ECCV 2014*, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds., vol. 8695, Series Title: Lecture Notes in Computer Science, Cham: Springer International Publishing, 2014, pp. 505–520. DOI: 10.1007/978-3-319-10584-0_33.

[434] Y. Li, Y. Song, L. Cao, *et al.*, *TGIF: A new dataset and benchmark on animated GIF description*, 2016. arXiv: 1604.02748[cs].

[435] H. Idrees, A. R. Zamir, Y.-G. Jiang, *et al.*, "The THUMOS challenge on action recognition for videos "in the wild"," *Computer Vision and Image Understanding*, vol. 155, pp. 1–23, 2017. DOI: 10/f9rwnr.

[436] S. Malla, B. Dariush, and C. Choi, "TITAN: Future forecast using action priors," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA: IEEE, 2020, pp. 11 183–11 193. DOI: 10/gg99rg.

[437] G. Awad, A. A. Butt, K. Curtis, *et al.*, *TRECVID 2019: An evaluation campaign to benchmark video activity detection, video captioning and matching, and video search & retrieval*, 2020. arXiv: 2009.09984[cs].

[438] Yale Song, J. Vallmitjana, A. Stent, and A. Jaimes, "TVSum: Summarizing web videos using titles," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA: IEEE, 2015, pp. 5179–5187. DOI: 10/gfsj74.

[439] U. Demir, Y. S. Rawat, and M. Shah, *TinyVIRAT: Low-resolution video action recognition*, 2020. arXiv: 2007.07355[cs,eess].

[440] S. Das, R. Dai, M. Koperski, *et al.*, "Toyota smarthome: Real-world activities of daily living," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, Korea (South): IEEE, 2019, pp. 833–842. DOI: 10/ghfjc7.

[441] T. Li, J. Liu, W. Zhang, Y. Ni, W. Wang, and Z. Li, *UAV-human: A large benchmark for human behavior understanding with unmanned aerial vehicles*, 2021. arXiv: 2104.00946[cs].

[442] K. Soomro, A. R. Zamir, and M. Shah, *UCF101: A dataset of 101 human actions classes from videos in the wild*, 2012. arXiv: 1212.0402[cs].

[443] J. Liu, W. Chen, Y. Cheng, *et al.*, *VIOLIN: A large-scale dataset for video-and-language inference*, 2020. arXiv: 2003.11618[cs].

[444] D. F. Fouhey, W.-c. Kuo, A. A. Efros, and J. Malik, *From lifestyle vlogs to everyday interactions*, 2017. arXiv: 1712.02310[cs].

[445] K.-H. Zeng, T.-H. Chen, J. C. Niebles, and M. Sun, *Title generation for user generated videos*, 2016. arXiv: 1608.07068[cs].

[446] X. Wang, J. Wu, J. Chen, L. Li, Y.-F. Wang, and W. Y. Wang, *VATEX: A large-scale, high-quality multilingual dataset for video-and-language research*, 2020. arXiv: 1904.03493[cs].

[447] W. Wang and Y. Yang, *VidProM: A million-scale real prompt-gallery dataset for text-to-video diffusion models*, 2024. arXiv: 2403.06098[cs].

[448] X. Zhang, Z. Wu, Z. Weng, *et al.*, *VideoLT: Large-scale long-tailed video recognition*, 2021. arXiv: 2105.02668[cs].

[449] A. Habibian, T. Mensink, and C. G. Snoek, "VideoStory: A new multimedia embedding for few-example recognition and translation of events," in *Proceedings of the 22nd ACM international conference on Multimedia*, Orlando Florida USA: ACM, 2014, pp. 17–26. DOI: 10/ggs25n.

[450] M. Ibrahim, S. Muralidharan, Z. Deng, A. Vahdat, and G. Mori, *A hierarchical deep temporal model for group activity recognition*, 2016. arXiv: 1511.06040[cs].

[451] A. Nagrani, J. S. Chung, and A. Zisserman, *VoxCeleb: A large-scale speaker identification dataset*, 2018. DOI: 10.21437/Interspeech.2017-950.

[452] M. Bain, A. Nagrani, G. Varol, and A. Zisserman, *Frozen in time: A joint video and image encoder for end-to-end retrieval*, 2022. DOI: 10.48550/arXiv.2104.00650.

[453] P. Das, C. Xu, R. F. Doell, and J. J. Corso, "A thousand frames in just a few words: Lingual description of videos through latent topics and sparse object stitching," in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, Portland, OR, USA: IEEE, 2013, pp. 2634–2641. DOI: 10/gtsqzr.

[454] L. Zhou, C. Xu, and J. J. Corso, *Towards automatic learning of procedures from web instructional videos*, 2017. arXiv: 1703.09788[cs].

[455] S. Abu-El-Haija, N. Kothari, J. Lee, *et al.*, *YouTube-8m: A large-scale video classification benchmark*, 2016. arXiv: 1609.08675[cs].