Ratish Puduppully

rapu@itu.dk
+45 71554702
https://ratishsp.github.io
http://www.linkedin.com/in/ratishsp/

Research Interests

My research interests span four areas: **(1) Long-Context Modeling and Planning**: improving LLMs' ability to reason over long inputs through neural content planning and long-context architectures; **(2) Multilinguality and Interpretability**: developing methods for low-resource and non-Roman script languages via romanization and studying latent representational patterns in multilingual models; **(3) Reasoning**: investigating mathematical reasoning in LLMs, such as identifying and enforcing unit consistency in word problems; **(4) Domain-Specific Applications**: building language models tailored to specialized domains, including genomics, using strategies like task-specific self-pretraining.

Education

Sep 2017 - Nov 2021	PhD in Informatics, University of Edinburgh. Awarded Best Thesis in Informatics in Scotland by SICSA. Advisor: Prof. Mirella Lapata
Dec 2014 - Feb 2017	MS in CSE by Research, IIIT Hyderabad Advisor: Prof. Manish Shrivastava Grade: 9.33 CGPA
Jul 2001 – Jun 2005	B.E in Electronics and Telecommunications. Mumbai University. Grade: equiv. to 9 CGPA.

Research Positions

Sep 2024 - Present	Assistant Professor at IT University of Copenhagen in the NLP group.
Sep 2022 - Aug 2024	Senior Scientist I, A*STAR Research
Dec 2021 - July 2022	Research Associate, University of Edinburgh.
Jun - Oct 2019	Intern, Google Research London.
Mar - Aug 2017	Research Assistant, Singapore University of Technology and Design.
May - Dec 2014	Research Engineer, Center for Indian Language Technology (CFILT) lab at IIT Bombay.

Research Publications

Journal Articles

1. Gala, J. P., Chitale, P. A., AK, R., Gumma, V., Doddapaneni, S., M., A. K., … Kunchukuttan, A. (2023). Indictrans2: Towards high-quality and accessible machine translation models for all 22 scheduled indian languages. *Trans. Mach. Learn. Res., 2023*. Retrieved from https://openreview.net/forum?id=vfT4YuzAYA

2. **Puduppully**, **R.**, Fu, Y., & Lapata, M. (2022). Data-to-text generation with variational sequential planning. *Transactions of the Association for Computational Linguistics, 10*, 697–715. doi:10.1162/tacl_a_00484

3. **Puduppully**, **R.**, & Lapata, M. (2021). Data-to-text generation with macro planning. *Transactions of the Association for Computational Linguistics, 9*, 510–527. doi:10.1162/tacl_a_00381

Conference Proceedings

1. Saji, A., Husain, J. A., Jayakumar, T., Dabre, R., Kunchukuttan, A., & **Puduppully**, **R.** (2025, July). Romanlens: The role of latent romanization in multilinguality in llms. In *Findings of the 63rd annual meeting of the association for computational linguistics (to appear)*. arXiv: 2502.07424 [cs.CL]. Retrieved from https://arxiv.org/abs/2502.07424

2. Mundra, N., Khandavally, A. N. K., Dabre, R., **Puduppully**, **R.**, Kunchukuttan, A., & Khapra, M. M. (2024, November). An empirical comparison of vocabulary expansion and initialization approaches for language models. In L. Barak & M. Alikhani (Eds.), *Proceedings of the 28th conference on computational natural language learning* (pp. 84–104). doi:10.18653/v1/2024.conll-1.8

3. J, J., Dabre, R., M, A., Gala, J., Jayakumar, T., **Puduppully**, **R.**, & Kunchukuttan, A. (2024, August). RomanSetu: Efficiently unlocking multilingual capabilities of large language models via Romanization. In L.-W. Ku, A. Martins, & V. Srikumar (Eds.), *Proceedings of the 62nd annual meeting of the association for computational linguistics (volume 1: Long papers)* (pp. 15593–15615). doi:10.18653/v1/2024.acl-long.833

4. Singh, A., Sai, A. B., Dabre, R., **Puduppully**, **R.**, Kunchukuttan, A., & Khapra, M. M. (2024, July). How good is zero-shot mt evaluation for low resource indian languages? In *Proceedings of the 62nd annual meeting of the association for computational linguistics (short papers)*.

5. Hwang, S., Lahoti, A., **Puduppully**, **R.**, Dao, T., & Gu, A. (2024). Hydra: Bidirectional state space models through generalized matrix mixers. In *The thirty-eighth annual conference on neural information processing systems*. Retrieved from https://openreview.net/forum?id=preo49P1VY

6. Kumar, A., Kunchukuttan, A., **Puduppully**, **R.**, & Dabre, R. (2023, December). In-context example selection for machine translation using multiple features. In *Findings of the 2023 conference on empirical methods in natural language processing*, Singapore: Association for Computational Linguistics. eprint: 2305.14105

7. **Puduppully**, **R.**, Kunchukuttan, A., Dabre, R., Aw, A. T., & Chen, N. (2023, December). DecoMT: Decomposed prompting for machine translation between related languages using large language models. In H. Bouamor, J. Pino, & K. Bali (Eds.), *Proceedings of the 2023 conference on empirical methods in natural language processing* (pp. 4586–4602). doi:10.18653/v1/2023.emnlp-main.279

8. **Puduppully**, **R.**, Jain, P., Chen, N., & Steedman, M. (2023, July). Multi-document summarization with centroid-based pretraining. In A. Rogers, J. Boyd-Graber, & N. Okazaki (Eds.), *Proceedings of the 61st annual meeting of the association for computational linguistics (volume 2: Short papers)* (pp. 128–138). doi:10.18653/v1/2023.acl-short.13

9. Mundra, N., Doddapaneni, S., Dabre, R., Kunchukuttan, A., **Puduppully**, **R.**, & Khapra, M. M. (2023, January). A comprehensive analysis of adapter efficiency. In *Proceedings of the 2024 cods-comad conference*, Bengaluru, India. eprint: 2305.07491

10. Kumar, A., Shrotriya, H., Sahu, P., Mishra, A., Dabre, R., **Puduppully**, **R.**, … Kumar, P. (2022, December). IndicNLG benchmark: Multilingual datasets for diverse NLG tasks in Indic languages. In *Proceedings of the 2022 conference on empirical methods in natural language processing* (pp. 5363–5394). Abu Dhabi, United Arab Emirates: Association for Computational Linguistics. Retrieved from https://aclanthology.org/2022.emnlp-main.360

11. Dabre, R., Shrotriya, H., Kunchukuttan, A., **Puduppully**, **R.**, Khapra, M., & Kumar, P. (2022, May). IndicBART: A pre-trained model for indic natural language generation. In *Findings of the association for computational linguistics: Acl 2022* (pp. 1849–1863). doi:10.18653/v1/2022.findings-acl.145

12. **Puduppully**, **R.**, Dong, L., & Lapata, M. (2019a, July). Data-to-text generation with entity modeling. In *Proceedings of the 57th annual meeting of the association for computational linguistics* (pp. 2023–2035). doi:10.18653/v1/P19-1195

13. **Puduppully**, **R.**, Dong, L., & Lapata, M. (2019b). Data-to-text generation with content selection and planning. In *Proceedings of the 33rd AAAI Conference on Artificial Intelligence*. doi:https://doi.org/10.1609/aaai.v33i01.33016908

14. **Puduppully**, **R.**, Zhang, Y., & Shrivastava, M. (2017, April). Transition-based deep input linearization. In *Proceedings of the 15th conference of the European chapter of the association for computational linguistics: Volume 1, long papers* (pp. 643–654). Valencia, Spain: Association for Computational Linguistics. Retrieved from https://www.aclweb.org/anthology/E17-1061

15. **Puduppully**, **R.**, Zhang, Y., & Shrivastava, M. (2016, June). Transition-based syntactic linearization with lookahead features. In *Proceedings of the 2016 conference of the north American chapter of the association for computational linguistics: Human language technologies* (pp. 488–493). doi:10.18653/v1/N16-1058

Workshop/ Demonstrations Proceedings

1. Lahoti, A., Marwah, T., **Puduppully**, **R.**, & Gu, A. (2025). *Chimera: State space models beyond sequences*. Retrieved from https://openreview.net/forum?id=uneMbnwmW8

2. Mupparapu, S., Krishnamurthy, P., & **Puduppully**, **R.** (2025). *Improving genomic models via task-specific self-pretraining*. arXiv: 2506.17766 [q-bio.GN]. Retrieved from https://arxiv.org/abs/2506.17766

3. Tang, Y., **Puduppully**, **R.**, Liu, Z., & Chen, N. (2023, December). *In-context learning of large language models for controlled dialogue summarization: A holistic benchmark and empirical analysis* (Y. Dong, W. Xiao, L. Wang, F. Liu, & G. Carenini, Eds.). doi:10.18653/v1/2023.newsum-1.6

4. Toh, V., **Puduppully**, **R.**, & Chen, N. F. (2023). *Veritymath: Advancing mathematical reasoning by self-verification through unit consistency*. AI4MATH Workshop at ICML 2024. arXiv: 2311.07172 [cs.CL]

5. Gehrmann, S., Bhattacharjee, A., Mahendiran, A., Wang, A., Papangelis, A., Madaan, A., … Hou, Y. (2022, December). *GEMv2: Multilingual NLG benchmarking in a single line of code*. Abu Dhabi, UAE: Association for Computational Linguistics. Retrieved from https://aclanthology.org/2022.emnlp-demos.27

6. **Puduppully**, **R.**, Mallinson, J., & Lapata, M. (2019, November). *University of Edinburgh's submission to the document-level generation and translation shared task*. doi:10.18653/v1/D19-5630

7. Kunchukuttan, A., **Puduppully**, **R.**, & Bhattacharyya, P. (2015, June). *Brahmi-net: A transliteration and script conversion system for languages of the Indian subcontinent*. doi:10.3115/v1/N15-3017

8. Bhingardive, S., **Puduppully**, **R.**, Singh, D., & Bhattacharyya, P. (2014, December). *Merging verb senses of Hindi WordNet using word embeddings*. Goa, India: NLP Association of India. Retrieved from https://www.aclweb.org/anthology/W14-5148

Preprint

1. Gala, J., Jayakumar, T., Husain, J. A., M, A. K., Khan, M. S. U. R., Kanojia, D., … Kunchukuttan, A. (2024). *Airavata: Introducing hindi instruction-tuned llm*. arXiv: 2401.15006 [cs.CL]

Patents

2016 Method and system for sharing content. US Patent 9,256,695. Willis, B. Natraj,S., Shinde, S., Agarwal,T., **Puduppully, R.**, Santhi Pulagala S. and Chang S.

Teaching Experience

- 2021 Tutor, Demonstrator and Marker for Natural Language Understanding, Generation, and Machine Translation
- 2018 Tutor for Accelerated Natural Language Processing

Mentoring Experience

- 2024 Jaavid Akhtar, Research Intern at IIT Madras on "RomanSetu: Efficiently unlocking multilingual capabilities of Large Language Models models via Romanization"
- 2023 Nandini Mundhra, M.S. by Research student from IIT Madras on "A Comprehensive Analysis of Adapter Efficiency"
- Aswanth Kumar, MTech student from IIT Madras on "In-context Example Selection for Machine Translation Using Multiple Features"
- Yuting Tang, undergraduate student from NTU Singapore on LLM evaluation
- Vernon Toh, undergraduate student from SUTD Singapore on Math Language Models
- 2022 Himani Shrotriya, Aman Kumar and Prachi Sahu, MTech students from IIT Madras on the project "IndicNLG Benchmark: Multilingual Datasets for Diverse NLG Tasks in Indic Languages"

Employment History

- 2005 – 2008 Software Engineer, Infosys Technologies Ltd.
- 2008 – 2014 Technical Architect, R&D Division, Saba Software, Mumbai.

Skills

- Languages: Strong reading, writing and speaking competencies for English, Hindi, Marathi and Malayalam.
- Programming Languages: Regular Use: Python. Less recent use: Java, C++, Javascript
- Machine Learning Tools: Pytorch.

Miscellaneous Experience

Awards and Achievements

- 2022 Best PhD Dissertation in Scotland award from SICSA Scotland
- 2017 Edinburgh Global Research Scholarship and Principal's Career Development Scholarship for pursuing PhD studies at University of Edinburgh.
- Travel Grant from ACM India for presenting paper at EACL conference in Valencia, Spain.
- 2016 Travel Grant from Microsoft for presenting paper at NAACL conference in San Diego, US
- 2002-04 JRD Tata Scholarship for Academic Excellence for undergraduate studies.

Service

- 2018-Present **Reviewer**. ACL, EMNLP, NAACL, EACL
- 2020 **Reviewer**. EMNLP: Outstanding reviewer

Miscellaneous Experience (continued)

Volunteering

2018-2020 — Digital Ambassador at University of Edinburgh.
Volunteered as a Digital Ambassador to help improve digital literacy amongst people, mainly elderly persons in community. The project won the 2019 University of Edinburgh Social Responsibility and Sustainability Community Partnership Award.