

Jaydeep Sen

IBM Research AI, Bangalore, India

+91 9686209251

jaydeepsen18@gmail.com

Summary

- Jaydeep is a hands-on AI practitioner with 10+ years of experience in designing and implementing AI solutions to practical business problems. He is an expert in Question Answering, Neural Retrieval and related NLP applications and has been leading the research and development efforts for building a cutting-edge neural retrieval model which has been released as WatsonX.Slate125m.rtrvr family of models. Previously he has lead efforts to build a research asset called "NLQ" (Natural Language Querying), which aims towards neural retriever model which has been released with democratizing data access and data-based insights to non-technical users.
- NLQ powers some of IBM's most prominent QA and NL application portfolio, most recent of which is Watson Orchestrate(<https://www.ibm.com/en/cloud/automation/watson-orchestrate>), that won CES innovation award'2022(<https://www.ces.tech/Innovation-Awards/Honorees/2022/Honorees/W/Watson-Orchestrate.aspx>).
- Jaydeep has received Outstanding Technical Achievement Award (OTAA) twice (2018,2020,2022,2023), which is a pan-IBM award meant to honour real world business impact through research innovations.
- With a keen interest in innovation, Jaydeep has 30+ patents granted/filed/in process of being filed in USPTO (as of Oct 2024) and 15+ research papers in top conferences. He has been awarded

Work Experience

IBM Research: Senior Research Scientist and Manager (June 2022-till date)

Key Responsibility: Research lead of retrieval technologies and Manager of NLP technologies in Aconversational AI.

Highlights:

- Lead a team of 5 researchers to collaborate with US team and deliver a cost effective and accurate neural retrieval model.
- Slate.125m.rtrvr models released with WatsonX.
- Built PrimeQA: An open source toolkit to promote reusable components in advancing state-of-the-art research in QA.

IBM Research: Advisory Research Scientist (June 2020-till date)

Key Responsibility: Research lead and Architect of QA solutions for emerging use-cases

Highlights:

- NLQ as a skill in Watson Orchestrate'2021 (<https://www.ibm.com/en/cloud/automation/watson-orchestrate>) : Architect of "NLQ as a skill" design and leading a cross-continent research team to develop and build NLQ as a generic NL querying skill across different types of data and artifacts to automate a sequence of AI skills from natural language task descriptions.
 - Used ElasticSearch, Document Store as backend.
 - Watson Orchestrate won CES innovation award'2022
- Leading a research team for building a robust retriever-reader based QA model on hybrid (table+text) context.

IBM Research: Advisory Research Software Engineer (Sep 2018-May 2020)

Key Responsibility: Build a research agenda for NL querying in business automation and related client applications.

Highlights:

- IBM delivering Retail-insights to Aditya Birla Fashion Retail Limited '2018: Enabling business executives to access large volume of business transaction data and derive real time insights through natural language interface.
 - Used Teradata engine over RDBMS as backend.
- BAI conversational assistant tech preview in IBM CloudPak for Data 2020 (<https://www.ibm.com/support/pages/technology-preview-bai-conversational-assistant-business-performance-center>): Enabling creation of dynamic dashboards and an automated conversational assistant over streaming process automation data for real-time insight generation and querying by non-technical users.
 - Used Elastic Search as data backend.

IBM Research: Staff Research Software Engineer (Dec 2015-Aug 2018)

Key Responsibility: Main Developer for building a NL querying pipeline over structured data.

Highlights:

- Built an ontology based robust NL querying research asset called NLQ (Natural Language Querying) which could interpret NL queries in a backend agnostic way and produce executable queries across different (semi) structured data backends such as IBM DB2, DashDB, RDF datastores, Watson Discovery.
- IBM Micromedex with Watson (<https://www.ibm.com/products/micromedex-with-watson>)2017: NLQ based dialog framework that allows any user to query a vast medical knowledge graph and get real time answers and guidance.
 - Used combination of DashDB and RDF as backend.

Qualcomm: Engineer (June 2013-Nov 2015)

Key Responsibility: Developer profile assigned to work on extensive C++/C coding in Android multimedia, Linux kernel.

Highlights:

- Worked on Android MultiMedia Framework. Exposed new APIs in Java Layer to support additional MM features and enhance functionality as offered to an end user
- Worked on core kernel team to develop driver modules for kernel clock framework to suit chipset specific requirements.
- Awarded Qualstar for 2 consecutive cycles for my contribution in Qualcomms latest MSM launches.

M.Tech, CSE, Indian Institute of Technology Kanpur (2011-2013)

Highlights:

- Worked on TECHNIQUES FOR NONLINEAR DIMENSIONALITY REDUCTION THROUGH RANDOM PROJECTIONS AND NEURAL COMPUTATIONS in Machine Learning as M.Tech. Thesis under Prof.. Harish Karnick, Head of the department, CSE,IIT K
- Research paper: Informed Weighted Random Projection for Dimension Reduction, Jaydeep Sen, Harish Karnick, ADMA '2013
- Ranked #1 in M.Tech and PhD entrant batch, received Academic Excellence Award for 2011-12

Patents:

Highlights:

- Have 30+ patents FILED (/to be filed) in USPTO focused on the area of Natural Language Understanding and Answering , Semantic Parsing and Ontology

Publications:

Summary:

- Published research papers in top conferences like VLDB, SIGMOD, IJCAI, EMNLP, COLING which pushed the state-of-the-art of QA systems in academics, introduced novel applications through new datasets and proposed intelligent system designs in NLP.
- Best Paper Award at CODS COMAD 2019.
- Gave a tutorial on "Natural Language Interfaces to Data" in SIGMOD'2020.
- Gave a tutorial on “Reproducible and Efficient Multi-modal Open Retrieval Question Answering” at IJCAI 2023.

Details:

- **New State-of-the-art**

- 1. ATHENA++: Natural Language Querying for Complex Nested SQL Queries (VLDB 2020), Jaydeep Sen, Chuan Lei, Abdul Quamar, Fatma Ozcan, Diptikalyan Saha, Karthik Sankaranarayanan : First NLIDB system to handle complex nested queries, most commonly seen in Enterprise settings. SOTA performance on Spider and FIBEN Dataset.
- 2. T3QA: Topic Transferable Table Question Answering (EMNLP 2021): Saneem A. Chemmengath, Vishwajeet Kumar, Samarth Bharadwaj, Jaydeep Sen, Soumen Chakrabarti, Karthik Sankaranarayanan et al: First system to propose evaluating TableQA performance on unseen topics. Achieves state-of-the-art performance in two topic transfer datasets for WikiTableQuestions and WikiSQL.
- 3. Multi-row, multi-span distant supervision for Table+ Text question answering(ACL 2023): Vishwajeet Kumar, Yash Gupta, Saneem Chemmengath, Jaydeep Sen, Soumen Chakrabarti, Samarth Bharadwaj, Feifei Pan, : proposed unique training regime to use linked row as the retrieval unit to obtain best performance on table and text QA.

- **Novel Applications**

- 4. Bootstrapping Natural Language Querying on Process Automation Data(IEEE SCC 2020) Xue Han, Lianxue Hu, Jaydeep Sen, Yabin Dang et al. : First system to propose natural language querying on process automation BAI data.
- 4. Schema Aware Semantic Reasoning for Interpreting Natural Language Queries in Enterprise Settings (COLING 2020): Jaydeep Sen, Tanaya Babiwale, Kanishk Saxena, Yash Butala, Sumit Bhatia, Karthik Sankaranarayanan: First system to propose a ontology reasoning framework around data schema for accurate NL interpretation.
- 6. Functional Partitioning of Ontologies for Natural Language Query Completion in Question Answering Systems,(IJCAI 2018): Jaydeep Sen, Ashish R. Mittal, Diptikalyan Saha, Karthik Sankaranarayanan: First system to propose the concept of "Functional Partitioning" of ontologies to optimize and localize query suggestions intelligently.
- 7. An Ontology based Dialog Interface to Database (SIGMOD 2018): Ashish R. Mittal, Jaydeep Sen, Diptikalyan Saha, Karthik Sankaranarayanan: First system to propose usage of ontologies to understand dialog context.

- **Intelligent Optimization**
 - 8. Optimizing Interpretation Generation in Natural Language Query Answering for Real Time End Users (CODS-COMAD'2021): Jaydeep Sen, Diptikalyan Saha, Ashish R. Mittal, Karthik Sankaranarayanan: Proposed a novel architecture of NL interpretation engine to optimize interpretation time by a set of intelligent ambiguity resolution techniques.
 - 9. Goal-based Ontology Creation for Natural Language Querying in SAP-ERP Platform(CODS-COMAD 2019): Neelamadhav Gantayat, Diptikalyan Saha, Jaydeep Sen, Senthil Mani: Proposed intelligent algorithms to produce an optimized ontology for QA over large scale SAP data. Won the BEST paper Award.
- **System Design**
 - 10. Creation and Interaction with Large-scale Domain-Specific Knowledge Bases VLDB'17 : Shreyas Bharadwaj, Laura Chiticariu, Marina Danilevsky, Jaydeep Sen et al: An end-to-end system for QA from documents by extracting KG as an intermediate artifact.
 - 11. Tooling Framework for Instantiating Natural Language Querying System VLDB'18: Manasa Jammi, Jaydeep Sen, Diptikalyan Saha, Karthik Sankaranarayanan et al: An end-to-end automated solution for instantiating a NL querying engine on user provided backend data.
 - 12. Anu Question Answering System, ISWC'20 : Balaji Ganesan, Avirup Saha, Jaydeep Sen et al: An end-to-end QA system backend on Ontology based KG extraction and entity resolution.
 - 13. Bootstrapping Chatbot Interfaces to Databases. COMAD/CODS 2021: Ashish R. Mittal, Diptikalyan Saha, Parag Jain, Jaydeep Sen et al :An end-to-end system for automatic chatbot creation on databases.
- **New Datasets**
 - 14. AIT-QA : Yannis Katsis, Saneem A. Chemmengath, Jaydeep Sen, Karthik Sankaranarayanan, Soumen Chakrabarti et al: A new benchmark for QA over complex documents including tables, numeric quantity definitions and linked free text descriptions.
 - 15.FIBEN : Jaydeep Sen, Chuan Lei, Abdul Quamar, Fatma Ozcan, Diptikalyan Saha, Karthik Sankaranarayanan : A new benchmark for complex Business Intelligence queries on Finance Databases.

Awards and Achievements

- IBM Outstanding Technical Achievement Award in in 2023 for building PrimeQA an open source reusable toolkit hosting state-of-the-art models for QA research.
- IBM Outstanding Technical Achievement Award in 2022 for integrating NLQ in Watson Orchestrate delivery.
- IBM Outstanding Technical Achievement Award in 2021 for NLQ on BAI tech preview.
- IBM Outstanding Technical Achievement Award in 2019 for delivering NLQ in ABFRL client environment
- Received IBM equity awards (given only to 5-10% employees) for the year 2018
- Received IBM equity awards (given only to 5-10% employees) for the year 2021
- IBM Manager's choice award in IBM for "Dare to Create Original Idea" 2016, "Unite to get it done"2018.
- Qualcomm Qualstar Award in Qualcomm for two consecutive cycles Oct'2014 & Apr'2015 for my contribution in Qualcomm's latest MSM(mobile Station Modem) launches.
- Academic Excellence award in 2013 from IIT Kanpur.

Educational Qualifications

Year of Passing	Degree	Name of Institution	Result	Rank
2013	Master of Technology (Department of Computer Science and Engineering)	Indian Institute of Technology, Kanpur	9.33/10 (cgpa)	01/39
2011	Bachelor of Engineering (Department of Computer Science and Technology)	Bengal Engineering and Science University, Shibpur	9.25/10 (cgpa)	

Academic Services:

Served as Program Committee member in :

- AAAI - American Association for AI National Conf
- SIGMOD - Intl. Conference on Management of Data
- IJCAI - Intl. Joint Conf on Artificial Intelligence
- EMNLP – Empirical Methods in Natural Language Processing
- ICDE - Intl. Conference on Data Engineering
- ICMLA- International Conference on Machine Learning Applications
- NAACL: North American Chapter of the Association for Computational Linguistics

Tutorials and Invited Talks

- Tutorial on Reproducible and Efficient Multi-modal Open Retrieval Question Answering” at IJCAI 2023.
- Tutorial on Natural Language Interfaces to Data: the state-of-the-art and open challenges” at SIGMOD 2020
- Invited talk at MSRIT for Faculty Day program to talk on “Evolution and State-of-the-art of of LLMs for Retrieval and QA”, Aug 2024.
- Invited talk at IIT Patna on “Natural Language Interfaces to Data: Reviewing Ontology based Approaches” on IEEE Day’ June 2020.
- Invited talk at MSRIT on “ Recent Trends in Natural Language Question Answering” on MSRIT Research Day’ 2017