Name: Kuntal Dey   **Last Updated:** 29-Aug-2023
Email ID: kd@kuntaldey.com; u2ckuntal@gmail.com;   **Mobile No.:** +91-9871388275
Gender: Male   **Date of Birth:** 23rd May, 1979

Work Experience: January 2003 – Till Date.
Present Designation: Technical R&D Senior Principal (Sr. Manager)
Highest Academic Degree: PhD (IIT, Delhi)
Current Employer: Application/Software Engineering Team, Accenture Labs, Accenture, Bangalore
Honorary Role: Adjunct Faculty (Rank: Professor), Indian Institute of Information Technology (IIIT), Guwahati

I have worked across research and product development and implementation of various system, analytics, business intelligence and artificial intelligence applications. My expertise is in out-of-the-box technical thinking for strong business impact, and new product incubation. My deep cross-domain expertise is well-demonstrated via the 160+ granted US utility (technical) patents where I have been an inventor.

Presently: I am a Tech R&D Sr. Principal (Sr. Manager) in Accenture Tech Labs, which is the R/R&D wing of Accenture. I presently lead a team of 7 people at the moment, including managers, researchers and software developers. I own the following areas for R&D:
- Frugal Large Language Models
- AI for Cloud Continuum Control Planes
- TinyML for Cloud Continuum

Further, for Application Engineering R&D of the Accenture Labs Bangalore, I own the academic relationships between Accenture and IISc, as well as, Accenture and IIIT Guwahati, from the Labs side. I am also an honorary Adjunct Faculty at IIIT Guwahati with a Professor rank.

Career Summary: My work experience spans across R&D (VERITAS Software / Symantec Corporation), Development (Microsoft) and Research (IBM Research, Accenture Labs). My academic background is a combination of Natural Language Processing (NLP) / Artificial Intelligence (AI) and Social Network Analysis (SNA). However, in work, along with numerous projects with varying nuances in the technical areas of AI, NLP and SNA, I have cross-cut into Cloud Technologies, Software Engineering (both from Research and Development perspectives), Sustainable AI in Cloud, Networking / Messenger and Availability. And I have led the incubation of multiple products by leading and contributing in the process of ideation, development and operationalization.

My innovations have often stood out as significant value-differentiators, throughout my career. I have 200+ technology patents filed / to be filed in the US patent office among which 189 have been granted till date and the rest in process, making me the **4th-most prolific patent inventor ever in history among resident Indian citizen inventors in terms of the number of granted US Patents**, to the best of my knowledge. I had been selected for the title IBM Master Inventor, which is a prestigious title in IBM, only given to very few (between 1%-2%) of the researchers with strong invention track record with significant impact sustained over a long duration of time.

Further, I have 80+ research publications in top academic journals and conferences, program committee membership of every single top-tier NLP conference (ACL, NAACL, EMNLP, Coling, EACL, LREC and others) as well as other top-tier AI conferences (AAAI, IJCAI, ICPR, ECIR) among others, demonstrate my excellence in research and innovation. Sustained streamlining of the innovation as I kept working in industrial research / R&D, working with teams that take those innovations forward to business, demonstrate my capability in and track-record of translating innovation into IT Business.

WORK EXPERIENCE SUMMARY
Accenture (Labs), as Technology R&D Senior Principal (Senior Manager)
June 2020 – Till Date
- **Functionality:** As a Senior Manager and R&D Senior Principal in Accenture Labs, the main corporate technical research lab of Accenture, my responsibility has been to create and seed a specialized R&D team with focus in certain technical areas. At present, I am in the process of ideating & incubating three key R&D areas. (i) Frugal Large Language Models, that ensures performances of LLMs while improving frugality (operating costs), making them more viable. (ii) Applying AI for cloud continuum planes, wherein, AI is used for improving workload distribution and management of hybrid cloud applications, in an operating ecosystem that is largely diverse in terms of hardware, systems, control and capabilities, with limited central control. (iii) TinyML: Breaking AI models (such as the first few layers of given CNNs) such that some of the nodes can operate in front-end resource-limited setups, which some other layers (the other layers of that given CNN, for example) can work at the back end mainstream cloud server. I am also working on adopting AI via knowledge graphs (and knowledge models), thereby operationalizing & monitoring distributed AI & general applications on cloud continuum.

  In the past, in Accenture Labs, I founded and led the Privacy Preserving Computing research here from the standpoint of software systems, application architecture and incorporation of machine learning (ML) / artificial intelligence (AI) within such architecture, spanning over edge, via cloud, to secure enclave-based computing, ensuring privacy (in parallel to security) of data and models at the levels of collection /

building, storing, transfer (network) and execution (processing). I led the incubation of digital twins for hybrid clouds and applications running on such clouds. I led in creating and managing distributed workloads on hybrid clouds, from the viewpoints of disposition and ongoing re-optimization, while jointly optimizing for performance, complying with regulatory requirements (e.g., GDPR etc.), tuning for privacy, and remaining sensitive to carbon impact (sustainability) as well as client-provided business constraints. I also ideated and led the implementation of ground-level technology for green, sustainable AI, using container lifecycle management techniques.

- **Technical directions:** Presently, I lead my team to apply advanced mechanisms in AI, for making AI more operationalizable and practicable, especially adopting it to the modern-day cloud continuum technology.

    In the past in Accenture Labs, I had led my team over a multitude of privacy-preservation techniques and technologies, including but not limited to edge and hybrid clouds, distributed clouds, federated machine learning, secure enclaves, container-based code and data transfers, and so on. I had led my team to use several hybrid cloud stacks for implementation of hybrid clouds. I had led my team to certain AI approaches for creating distributed AI/ML with privacy preservation. I had led my team towards certain green/sustainable AI techniques.

Indian Institute of Information Technology (IIIT), Guwahati
January 2021 – Till Date
- **Functionality:** I am also appointed as an Adjunct Faculty in IIIT Guwahati, on an honorary basis, with a **Professor** rank.
- **Coverage:** Here, I have engaged in academic activities for the subject Natural Language Processing (NLP), and teach a course on NLP in the spring semesters. I am also advising 3 PhD research scholars at the moment, for the institute.

IBM (IBM Research), as Senior Software Engineer & IBM Master Inventor
May 2007 – June 2020
- **Functionality:** I used to conceptualize novel technology ideas, architecting and implementing them with the rest of the team, and delivering them to the clients (internal and external, as applicable). I had worked in the areas of social network analysis, mobile analytics, mobile money and application performance scaling during my tenure at IBM. I had technically led several projects, spanning across mobile analytics solutions, human eye gaze estimation & tracking, social network analysis, natural language processing and some computer visions. Sometimes, the projects have involved members across IBM business units and across continental boundaries.
- **Technologies:** My work in IBM Research had spanned across a number of tools, technologies, computer scientific specializations and programming platforms. Working on Linux and at times Windows, I had used C, C++, Java, Linux shell scripts, Keras (for deep learning), Microsoft Powershell, awk, sed, Python and PHP for programing, and DB2 and MySQL for databases. Some tools that I used, among many, are Mallet (natural language processing), Weka (machine learning), graph processing tools (social network analysis involves a lot of graph theory), exposure to Android mobile app platforms and WebSphere, mobile notification services and many other tools.

Microsoft (India Design Center), as Software Design Engineer (Developer)
October 2005 – May 2007
- **Functionality:** Design, architecture and implementation of MSN Live Communication Server, which was a server-side part of the then-popular MSN Messenger. This component was a pain-point that my colleagues in the company were hesitant to take upon. I took the challenge, redesigned the original architecture, implemented the system that made the process 700% faster than the original, and it led to a massively positive impact on the dollar income of Microsoft MSN Messenger.
- **Technologies:** The product involved programming in C# on DotNET, and used MS SQL at the back end and IIS at the front end.

VERITAS Software (later merged with Symantec Corporation) India, as Software Engineer
April 2003 – October 2005
- **Background:** This was the job that I had on-campus from IIT Bombay. I had stood first and had landed the first official A-grade job in entire IIT Bombay in my batch, across every department, and that was this job. So I was landed in a highly specialized innovation group.
- **Functionality:** I was assigned to a specialized micro-team under the VERITAS Cluster Server, with the name New Products Initiative. As a rather unconventional and very highly regarded job in the company, with the project lead, one of the senior-most technical directors across all of VERITAS over the world and the senior-most person in the entire India division of VERITAS, which was the #5 company in terms of turnover at that time, and only two other team members, my job was to define and explore new and innovative product paths and prototypes, and create roadmaps for the entire VERITAS Cluster Server (VCS) team to move forward with.
- **Technologies:** My work in VERITAS had spanned across a number of OS platforms including Linux, Windows, Solaris, HP_UX and a bit of IBM AIX, for providing high-availability failover and instant application switchover. As part of competition exploration, I had programmed VCS to embrace and

extend Microsoft's Cluster Server (MSCS). I had single-handedly created an end-to-end C++ object-oriented API interface for the C-based VCS to work on APIs. In the final part of my tenure, I and had made contributions to the Linux and Solaris kernels for stateful application migration, a concept that was almost entirely novel in the industry back in 2005.

Tata Consultancy Services (TCS) Research Lab, IIT Bombay, as Research Associate
January 2003 – March 2003
- This was a 3-month contract with TCS Research Lab in IIT Bombay, as a Research Associate, researching natural language processing problems. I had completed my M. Tech requirements earlier than scheduled, and hence my Prof (Guide) had honored me with this job to extend my stay in IIT Bombay, before joining my campus offer that I had received from the industry.

Visiting/Honorary Teaching Roles and Works:
I have held Visiting Faculty roles in several universities/institutes in the past, notable ones being Symbiosis Institute of Computer Studies and Research Pune (2003-05), EMPI Management Institute Delhi (2009-13), South Asian University Delhi (2013) and my current Adjunct Faculty (Professor) role in IIIT Guwahati (Jan 2021 – till date). I have delivered numerous lectures and have conducted other activities in many universities, including IIT Bombay (as a TA), IIT Kanpur, IIT Mandi, IIITM Gwalior, IIIT Guwahati (before I was an Adjunct Faculty member there), several NITs and a large number of other universities.

EDUCATION
- PhD in Information Technology in IIT Delhi, October 2019 (done in parallel to my office work, while I was working in IBM Research). Thesis title: *Modeling Correlation between Social Connections, Topics and Information Diffusion on Social Media*
- M. Tech, Computer Science and Engineering, IIT Bombay, January 2003 (class topper for some part of the course)
- B. E., Computer Science and Engineering, Jadavpur University Kolkata
- M. S. in Psychology (Counseling and Therapy) – IBMS Chittoor 2004 (done in parallel to my office work, when I was working in VERITAS Software India). Specialization (Thesis): Industrial Psychology.

APPENDIX

PROJECTS UNDERTAKEN

Project Name	Frugal Large Language Models [May 2023 – Till Date]

- The aim of this project is to ensures performances of LLMs while improving frugality (operating costs), making them more viable.
- Multiple mechanisms have been proposed and are under development, over and beyond the Stanford-provided baseline mechanisms.
- *At present, this is under design and initial prototyping.*

Project Name	TinyML for Cloud Continuum [September 2022 – Till Date]

- The aim of this project is to engineer AI models to function over multiple nodes in a computer network of different capacities, that comprise a cloud continuum.
- The mechanism is to break AI models (such as the first few layers of given CNNs) such that some of the nodes can operate in front-end resource-limited setups, which some other layers (the other layers of that given CNN, for example) can work at the back end mainstream cloud server.
- *At present, this is under exploration with an aim to formalize for industry-scale use.*

Project Name	AI for Cloud Continuum Control Planes [September 2022 – Till Date]

- The aim of this project is to use generative AI to create solution architectures, and subsequently use AI to distribute workload on cloud continuum, and also monitor and manage the operations of such applications over the continuum, in an operating ecosystem that is largely diverse in terms of hardware, systems, control and capabilities, with limited central control.
- *At present, the different components of this asset are in the research prototyping stage.*

Project Name	Digital Twins for Testing and Analyzing Hybrid Cloud Applications [September 2021 – February 2023]

- The aim of this project was to create digital twins with succinct and yet appropriate data and meta-data for testing and analyzing the operations of hybrid cloud applications and predicting their near-term future health.
- The project spanned across (a) understanding application characteristics and identifying data and meta-data that can be extracted, (b) collating the data and performing log analytics, as well as correlating data logs with specialized meta-data logs, and (c) providing predictions of potential health issues, as well as, prescriptive advisory accordingly.

Project Name	Cloud Enablement for Sustainable (Green) AI [October 2021 – August 2022]

- The aim of this project was to intelligently place containers that would ensure green (sustainable) operations to the extent possible, based upon estimated data-center power sources, while respecting regulatory policy frameworks (such as GDPR), and while aware of (non-agnostic towards) data source, consumption geo-pattern of the application, etc.
- The project spanned across (a) understanding and mapping the sustainability of given cloud data centers with external pragmatic knowledge, (b) devising a mechanism of container activation for maximizing sustainability, while adhering to regulatory policies as well as customer needs (of response time/speed etc.) and (c) finally, activating and executing the containers as per the mechanism obtained.

Project Name	Multifactor Optimization for Cloud Workload Placements [April 2021 – August 2022]

- The aim of this project was to optimize hybrid cloud application solutions to respect privacy, sustainability and regulatory requirements, and thereby provide architectural guidance on how to architect the solution in terms of placing components on-premises / in private clouds, on the edge and in the public cloud.

- The project spanned across (a) factoring in client's privacy and other business requirements, regulatory policies provided by authorities that require IT compliance and cloud solution provider's as well as client premises' energy (power) source maps as inputs, (b) programmatically formulating an intelligent application architecture that would optimize across all these input factors and (c) enabling deployment of the solution architecture thus obtained.
- I co-led the team, proposed some of the key components of the project, ideated for my team and co-guided towards the goal of the work.

Project Name	Programmatic Creation and Migration of Privacy-Preserving Containers [January 2021 - July 2021]

- The aim of this project was to programmatically assess the heterogeneous privacy requirements of different clients using the same hybrid cloud solution, and based upon the assessment, intelligently create custom containers that would enable fine-grained custom privacy for each client's business needs. Each custom container would be migrated to the clients that were the right audience, while covering all the clients over different custom containers tailor-created for them. The migrated containers would execute either immediately, or on-demand, according to the adaptation modality of the underlying policy framework created as part of this project.
- The project spanned across (a) programmatic creation of fine-grain-customized hybrid cloud containers, (b) shipping custom client-side containers to the client premises (edge, on-premise/private clouds) and (c) executing those containers with immediately or on-demand, or both, according to the policy adoptable carried out. I proposed, ideated and designed the project and led my team to the work.

Project Name	Privacy-Preserving Data Cooperatives with Distributed Intelligence [June 2020 – August 2021]

- The aim of this project was to enable knowledge-sharing and intelligence-sharing across multiple clients, without data-sharing, as clients would not be interested to share data due to business mandates and privacy of their customers, but would be helped if served using richer business intelligence that can be created out of the data by extracting meta-models and combining those without leaking any data.
- The project spanned across (a) creating a data cooperative with distributed machine learning (federated machine learning was used), (b) combining the smaller (less rich) ML models obtained at individual clients to obtain a bigger (richer) ML model at the back end, and (c) using this combined model for inferencing for the consumers of the model. I co-led and co-ideated in the project, and ideated and led my team for specific components within the work.

Project Name	AI for Regulatory Compliance of IT [April 2019 – June 2020]

- The aim of this project was to understand regulatory obligation documents such as NIST-800-53, PCI-DSS, HIPAA, GDPR etc. and translate that into IT execution specifications for organizations to comply.
- The project spanned across (a) understanding the documents with appropriate knowledge extraction including identifying change of compliance requirements using NLP techniques and (b) mapping this extracted knowledge into IT infrastructural component and executing the components. I acted as the technical lead for part (a) of the project mentioned above.

Project Name	Bias and Fairness in Artificial Intelligence [Jan 2018 – March 2019]

- I worked in machine learning fairness, compliance of unstructured documents with targeted software code meta-descriptions, and other problems in the space of ML-based AI and data.
- I led the IP (patent) generation for a large part of our department, across the all the areas such as data, AI, compliance, fairness of ML, data on hybrid clouds, AI on hybrid clouds and the related.
- The work led to release of IBM AI Open Scale as well as AIF360, an open source bias/fairness tool.

Project Name	Eye Gaze Estimation, Tracking & Applications [April 2015 –Dec 2017] – Computer Vision, Machine Learning, Natural Language Processing

- I envisioned estimating and tracking the human eye gaze, in a movement-invariant mode, using low-end hardware cameras such as mobile front cameras and laptop webcams, using computer vision techniques, and led the design and implementation of the project, making IBM Research the first industrial organization to use in-computer software with inbuilt hardware to attempt to solve this problem.

- I further envisioned the multiplicity of applications of the capability, and played a lead role in demonstrating the effectiveness of the technology in applications such as natural language processing, developing non-trivial research impact.
- This project was highlighted by IBM as a company, as well as IBM Research India, as a flagship, that was demonstrated to several of its highly visible events.

Project Name	Mobile Analytics Platform [January 2014 – June 2015] – Java, Python, DB2, TeaLeaf, WebSphere, Android, IBM X2020

- In this project, I was the technical lead of the mobile analytics research services team, spanning across geographies, complex job roles and different levels of seniority of project members.
- The goal of this project is to perform end-to-end analytics of mobile applications, spanning across client, server and network.
- I was responsible for the end-to-end analytics architecture and design, coding and quality.

Project Name	Mobile Notifications [January 2013 – December 2013] - Core Java, Android, Networks

- In this project, I owned and developed two out of the three key components, namely a scalability benchmarking system and a notification performance cache subsystem. In addition, I had ideated most of the core functionalities of the end-client management GUI.
- The goal of the project was to provide a corporate interface to manage the notification lifecycle of corporate clients on Google's GCM and Apple's APNS.
- I was responsible for architecture, design, coding and testing of the said components.

Project Name	MAPM [January 2013 – December 2013] - IBM X2020, Python, DB2, WebSphere

- In this project, I envisioned, architected and developed (coded and tested) the primary differentiating factor, namely event correlation across multiple subsystems (such as WebSphere, DB2 and operating system).
- The goal of this project was to provide a backend datacenter insight tool, that would enable datacenter admins and operations teams a quick and easy interface to identify issues, and easily drill down to root causes, across multiple components such as application servers, database servers and operating systems.
- I was responsible for architecture, design, coding and testing of the correlation component that I had envisioned and proposed to IBM for development.

Project Name	Mobile Money [September 2011 – December 2012] – Java, Android, QR codes, Java servlets

- IBM undertook this project based on a key observation that I had made: mobile phones can be key to making secure monetary transactions on digitized tokens such as QR codes and value-added transaction services (which is obvious today, but was a fundamental and leading-edge thought back in March 2011, when I had first moved the IBM management – in fact, QR code was a term almost nobody had heard of back then, when I had made this proposal).
- The goal of this project was to provide an ecosystem to make mobile payments, using digitized bills and digitized tokens, and provide value-added services such as providing customer risk profile based instant micro-loans.
- I was responsible for not only making the 0th level technical propositions, but also had to convince technical leaders of IBM the immense value that mobile could bring for monetary ecosystems, and the project eventually morphed in IBM. I ended up architecting and coding a significant portion of the project, involving around the core mobile-based digital billing and token subsystem. An advanced version of the project that I had originally proposed is still playing a key role in IBM's massively highlighted Africa initiatives.

Project Name	Mobile Application Performance Optimizer [January 2012 – December 2012] – Java, Android, Large-scale multithreading

- This project was a small part of a broad exploration by IBM, aiming to provide a platform for a generic assessment of mobile application performances, and performance bottleneck identification.
- I created a workload generator in this project that would run a massive number of threads (1500+ parallel threads per server) and collate function-level performance reports. These reports would be subsequently analyzed to find application performance bottlenecks, which in turn would be investigated deeper for performance improvement.
- I was responsible for architecture, design, implementation and testing of the project, along with my team members.

Project Name	SNAzzy - Social Network Analyses [May 2007 – September 2011] – Java, graph algorithms and tools, AI tools like WEKA, Mallet etc

- This project was shortlisted by NASSCOM (http://www.nasscom.in/ibm-research-india) as one of the most innovative projects in India.
- In this project, I was a key contributor to the ideation, design and implementation of a large number of advanced graph-theoretic algorithms (that led to the publication of many research papers in some of the top academic conferences and journal in the world).
- The goal of the project was to explore and analyze the graph-theoretical and event-based inter-person analyses, and provide deep business intelligence beyond the capabilities of standard machine learning algorithms.
- I played key role in implementing many of the myriads of analytics algorithms within this massive-scale analytics toolbox. I led the ideation, design and implementation of many of the algorithms and components.

Project Name	Projects at Microsoft, VERITAS (Symantec) and TCS Research Lab

- **Microsoft, Live Communication Server (October 2005 – May 2007):** This project involved onboarding premium corporate clients on MSN Messenger Server with the server being their office-to-client communication backbone. In this project, I took the responsibility to clear up a dirty prior implementation (completely messed up) that none else in the team were keen to work on. I re-architected and re-programmed the solution to make it 700% more efficient. The tangible dollar value to Microsoft was extremely high, and I received multiple awards and internal recognition for my performance.
- **Projects in VERITAS (Symantec) (April 2003 – October 2005):** In VERITAS (and Symantec, as it got later taken over), I was a part of the VERITAS Cluster Server (VCS) New Products Initiative (NPI) team. The NPI team was tiny and unique, and tightly focused on creating innovation roadmap for the rest of the department by aggressively exploring the product innovation space and implementing prototypes for the rest of the team to follow subsequently for product development. As part of this, I had been involved in (a) creating object-oriented C++ based VCS APIs that was earlier completely written in procedural C, (b) creating VCS API to embrace and extend Microsoft Cluster Server (MSCS) engine and (c) engineering the OS (Solaris and Linux) kernels to enable stateful process migration across machines (I was involved only slightly in this third project, as it had started at the last part of my tenure and I had transitioned out to my next job much before this product got fully developed).
- **Project in TCS Lab (January 2003 – March 2003):** I was researching with several semantic-level problems as part of a natural language processing (artificial intelligence) project, as a research associate in TCS Lab at IIT Bombay, on a 3-month contract. I had been contracted after I completed my M. Tech somewhat ahead of schedule, on recommendation of my guide (Prof. Pushpak Bhattacharyya), on the basis of honorarium-based payments, motivated by my outstanding research skills as demonstrated during my M. Tech work just before this contract.