Detailed Project Report (DPR)

National Mission on Interdisciplinary Cyber-Physical Systems (NM-ICPS)

iHUB ANUBHUTI – IIITD FOUNDATION
IIIT D, Delhi, India

Executive Summary of iHub Anubhuti-IIITD Foundation

A TiH in the area of "Cognitive Computing & Social Sensing"

Welcome to the age of cognitive computing and social sensing (CC&SS), where intelligent machines simulate human brain capabilities to help solve society's most vexing problems. Early adopters in many industries are already realizing the values of CC&SS, and its enormous potential to transform the industry. Cognitive computing overlaps with AI and involves many of the same underlying technologies to power cognitive applications, including Expert Systems, Neural Networks, Robotics, Augmented Reality (AR) and Virtual Reality. Here we summarize our major efforts in consolidating the problem areas that we plan to work on.

Grand Challenges (GC): iHub Anubhuti-IIITD Foundation focuses on the development of key CC&SS technologies to develop technology systems and different applications to solve society's most vexing problems. We have identified two grand challenges after discussing them with various ministries and PSUs.

GC1: Designing cognitive computing solutions for transforming the education system in India (Possible collaboration areas addressed by the Ministry of Education):

(a) NDEAR (National Digital Education Architecture) is a unifying national digital infrastructure to energise and catalyse the education ecosystem. NDEAR is federated, unbundled, interoperable, inclusive, accessible, and evolving that aims to create and deliver diverse, relevant, contextual, innovative solutions that benefit students, teachers, parents, communities, administrators and result in timely implementation of policy goals. We will work closely with the Ministry of Education to develop this unifying national digital infrastructure.

(b) DIKSHA (Digital Infrastructure for School Education) is an initiative of the National Council of Educational Research and Training (Ministry of Education, Govt of India). It is a national platform for school education. DIKSHA was developed based on the core principles of open architecture, open access, open licensing diversity, choice and autonomy as outlined in the Strategy and Approach Paper for the National Teacher Platform. We will work on the DIKSHA platform with the objective to make the policies and tools robust enabling the education ecosystem (educationist, experts, organizations, institutions - government, autonomous institutions, non-govt and private organizations) to participate, contribute and leverage a common platform to achieve learning goals at scale for the country.

(c) UDISE+: Timely and accurate data is the basis of sound and effective planning and decision-making. Towards this end, the establishment of a well-functioning and sustainable educational management information system is of utmost importance today. UDISE+ is an updated and improved version of UDISE. The entire system is now online and has been collecting data in real-time since 2018-19. However, the data that is being captured is not in a form that can have a meaningful interpretation for any actionable insights. We will work on this platform to (a) add on the various layers of data which is needed to gauge the progress of the various educational institutes, and (b) work on the analytics to determine almost a real-time performance of all the educational institutes on various parameters. This will also help in determining which interventions are leading to what all benefits.

GC2: Designing CC&SS solutions for healthcare:

Sustainable Development Goals 3 (Good Health and Well-being) mentions providing efficient and effective Public Health, especially maternal and child health with an aim to reduce the mortality rate of mothers and newborns. The efforts to improve Public Health by the Indian government have been successful in bringing the under-5 child mortality rate (per 1000 birth) to 34.3 and neonatal mortality rate to 22. However, there is still a long way to achieve the neonatal rate of 12 and under-5 child mortality rate of 25 by the year 2030. The current healthcare efforts are primarily manually driven and only make minimal use of technology. We envision a CC&SS enabled platform to support the Indian healthcare system. We envision technological solutions, in the space of CC&SS, for the following (non-exhaustive) list of problems:

(a) Healthcare Data Collection, Curation, and Modelling: Efficient methods to collect, store and analyze on a multitude of devices, e.g., mobile phones, tablets, laptops, are required. A good amount of data collection is manual and through paper forms. This leads to delay and human errors. Moreover, the data is then further needed to be stored in a digital format leading to more delays and errors. There is also no standardization of data collection and thus making it almost impossible for further analysis. New technologies to directly collect data in digital form, check for errors, verify the authenticity (e.g., with GPS triangulation), and store it in a standard format making it available for sharing, analysis, and report generation is essential.
(b) Analysis of Public Health Data: Analytical methods for public health data need to be developed to generate insights into different public interventions, their effectiveness, and also for deciding future directions to further improve public health.
(c) Intelligent Training and Learning Solutions for Community Health Workers: The ASHA network is primarily made of women recruited from the community. While this provides easy access to the community, the poor training of ASHAs has been singled out as a major reason for the inefficient execution of services on the field. The inadequate training is a result of a lack of infrastructure, mobility, and lack of experts. Computing solutions to provide distance-based training to community health workers are needed to support their growth and efficient execution on the ground. There is a need to develop personalized solutions that are available anywhere, anytime to provide the training. The use of Chatbots in native languages, or other AI-enabled solutions that can run on mobile devices may help in filling the current training gaps.

Other problems: Apart from the two aforementioned grand challenges, we have decided to work on legal information processing systems to support the need for the ministry of law and justice and other related legal farms. To this end, we have started the discussion with DAKSH, a civil society organization that undertakes research and activities to promote accountability and better governance in India, and have built a tripartite collaboration with DAKSH, IIT Delhi and iHub Anubhuti to build a centre for excellence in law and technology. We also have started working on cognitive computing-driven cyber-security, particularly for social media. The problems include combating online abusive content such as fake news and misinformation, hate speech, cyberbullying, fraud activities, etc. The aim is to build an intelligent system that can aid to technological advancements of police and cybercrime departments.

Overall aims & objectives: One of the goals of iHub Anubhuti is to understand the fundamental principles that will enable us to create a cognitive entity management platform for distributed collaborative analytics. Particularly, we aim to address the following research themes, but not limited to:
1. Build a cognitive entity store which not only stores the entities, but also the contextual information about the data and its relationship with other data entities, how these entities can be leveraged in collaborative data analytics.
2. Address numerous data curation challenges in handling entity identification from unstructured data and images, integrating with structured data, handling different forms of data and variations in the schemas, and uncertainty introduced by the extraction and integration process.
3. Develop entity analytics techniques for entity disambiguation, remediation, discovery, correlation, linking based on ML to handle noise and uncertainty.
4. Develop a novel, context-oriented, loosely coupled integration of structured and unstructured data entities through symbiotic consolidation of related information (1) to enhance structured data retrieval by associating additional documents relevant to the user context with the query result, and (2) to enhance document contents by associating additional data entities derived from structured data.
5. Develop methods for correlating events (arrive from various sources – sensors, streaming data, social media and applications) with data entities. Events are matched based on entities that relate to the events and sometimes enrichment is required from reference data to perform the match. The matching of events is done in many cases according to contexts that may be temporal, spatial and segmentation oriented based on entities.
6. Development of scalable and effective algorithms for knowledge discovery to uncover patterns, correlations, clusters, outliers, and abnormal structures in symbiotic data entities, and

incremental knowledge discovery as data entities are evolved over time.

Current progress: So far, we have approached a total of 32 Ministries and 41 PSUs. We have already started conversations and identified problems with some of the organizations like NALCO, MCL, IIPE, etc. The interactions with some others like DBT, Ayush, DRDO, etc. are also in pipeline. Internally, we have started the process of identifying teams of experts who will be working on these problems. We have sanctioned 28 research projects. These projects majorly focus on healthcare, legal informatics and educations. We have also announced the call for the Chanakya fellowships for UG and PG students.
We already have a regular process of having regular BOD and HGB meetings wherein the entire strategy and operational initiatives are discussed. With respect to Statutory Compliances and Regulatory adherence, we have also taken a Company Secretary on board. We have appointed the Statutory auditors and the work on closing the books for last year has been done. Work on registering under GST, Sec 80 G, etc. has also started.
Through outreach activities, we regularly disseminate our updates that have helped in fastening the collaboration with various external stakeholders. Press releases about this DST initiative and our hub's vision were covered by 15 national newspapers and publications. We have also started to communicate via blogs on various forums like LinkedIn and Medium.
We are about to release a call for startups/incubations. The startups we plan to nurture, could be using cognitive computing/AI/ML/HCI to solve problems in the areas of Healthcare, Legal, Education and Sustainability.

CERTIFICATE

Name of the TIH: iHUB Anubhuti-IIITD Foundation

Technology Vertical: Cognitive Computing and Social Sensing

1. This is to certify that the Detailed Project Report (DPR) on the Technology Vertical Cognitive Computing and Social Sensing is prepared and submitted to Mission Office, NM-ICPS, DST is as part of implementation of Technology Innovation Hub (TIH) at Indraprastha Institute of Information Technology, Delhi Okhla Industrial Estate, Phase III, (Near Govind Puri Metro Station), New Delhi, India - 110020 under National Mission on Interdisciplinary Cyber-Physical System (NM-ICPS).

2. This is to certify that this DPR has been checked for plagiarism and the contents are original and not copied/taken from any one or from any other sources. If some content was taken from certain sources, it is duly acknowledged and referenced accordingly.

3. The DPR will be implemented as per the Terms, Reference and Clauses stated in Tripartite Agreement signed on 01-03-2021 between Mission Office, DST, Indraprastha Institute of Information Technology, Delhi and IHUB Anubhuti-IIITD Foundation.

Date: 25/9/21
Place: New Delhi

Pushpendra Singh (Lead Project Director and member of HGB)	**Vikram Goyal (co-Project Director)**	**Tanmoy Chakraborty (co-Project Director)**

Name(s) and Signature(s) of Project Director (s)

Endorsement from the Head of the Institution

1. Certified that the Institute welcomes participation of Prof. Pushpendra Singh (Lead P.D.), Prof. Vikram Goyal (co-P.D.), Dr. Tanmoy Chakraborty (co-P.D.) as the Project Director(s)/Co- Principal Director for the Technology Innovation Hub (TIH) and that in the unforeseen event of discontinuance by the Project Director, the Indraprastha Institute of Information Technology, Delhi will identify and place a suitable faculty as Project Director for fruitful completion of the TIH activities.

2. Certified that the Host Institute shall provide basic facilities, faculty support and such other administrative facilities as per Terms and Conditions of the award of TIH, will be extended to TIH.

3. As per Tri-partite Agreement, the Host Institute (HI) shall play its role and fulfill its responsibilities for the success of TIH.

Date: 25/9/21
Place: New Delhi

Ranjan Bose

(Prof. Ranjan Bose
Director, IIIT-Delhi)

Name and signature of Head of Institution

Contents

Topic	Page No.
1. Vision of the Hub	07
2. Context/ Background	98
3. Problems to be Addressed	12
4. Aims and Objectives	46
5. Strategy	48
6. Target Beneficiaries	52
7. Legal Framework	56
8. Environmental Impact	58
9. Technology	60
10. Management	62
11. Financial Budgets	66
12. Time Frame	76
13. Cost-Benefit Analysis	81
14. Risk Analysis	85
15. Outcomes	88
16. Evaluation	90
17. Collaboration with other TiHs	92
18. Granting of Research Projects	94
20. Hub Policies	173

1. Vision of the HUB

To establish itself as a hub of Research, Entrepreneurship, and Innovation in the area of Cognitive Computing & Social Sensing

Build a nationwide shared and distributed Cognitive Computing & Social Sensing facility for public research and commercialization

CONTEXT/ BACKGROUND

2. Context/Background

Welcome to the age of cognitive computing and social sensing, where intelligent machines simulate human brain capabilities to help solve society's most vexing problems. Early adopters in many industries are already realizing significant value from this innovative technology, and its potential to transform the industry is enormous. Cognitive computing overlaps with AI and involves many of the same underlying technologies to power cognitive applications, including Expert Systems, Neural Networks, Robotics, Augmented Reality (AR) and Virtual Reality.

The last 10 years wave characterized by the confluence of social, mobile and cloud technologies have empowered the digital experiences and has brought new disruptions through innovations in IT operations and delivery, business models and markets. Moreover, the rise of Big Data and the new kinds of analytics over big data created the value in creating new scientific and technical directions, business models, making better and insightful decisions, developing new products and services, and optimizing process and operational efficiency. These trends will continue to evolve in further personalizing the offerings to meet the end-users' demand and differentiating the products and services in the marketplace.

Industry 4.0 which is the upcoming standard for industries, where the business is expected to focus primarily on automation, interconnection, information transparency, straight through processing, and decentralized decisions. It also focuses on environmentally sustainable manufacturing by having GREEN manufacturing processes, green supply chains and green products. These objectives can only be met by digitalization and integration of vertical and horizontal value chains, digitization of products and services offerings, and informationcentric business models and customer access. The recent technologies, such as AI, ML, AR/VR, Blockchain, etc. are the catalyst to design thinkable and actionable Cognitive Systems. These technologies are disruptive to build the cognitive systems that automate the data curation process, create a knowledge driven ring-fencing data entity, automatic (federated) entity resolution/information processing for collaborative analytics and cognitive analytics.

Cognitive Computing

Where do we start with cognitive? This is a question many leaders across industries have been trying to answer. Firstly, we need to design intelligent solutions to compile real time information from society and users. Secondly, we need to understand the fundamental principles that will enable us to create a 'thinkable and actionable' Cognitive Information System which can proactively understand the incoming data (from a diverse set of sources), situations arise in the system and react automatically by learning from patterns and these situations. These systems leverage cognitive capabilities (machine learning and AI algorithms) inclusive of understanding, learning and contextual awareness to simplify the ingestion and characterization of data to make the data/system ready for AI, and to collect and manage critical semantics information needed for integration, query processing, policy enforcement and auditing. Immense amount of personalization, precise profile mapping, accurate trend analysis and future prediction, more than ever exact results from unalterably large and unstructured database – all these are going to become the new normal in the business

landscape, going by the current trajectory of development and deployment of these cognitive systems. Thirdly, we need to understand the psychology of end-users and how they are going to consume the information. We need to develop the HCI driven personalized visualization tools, cognitive methods to Compose recommended interactions, use context to deliver a point of action, and adapt the user behavior and interaction patterns to improve context.

Social Sensing

Due to the ubiquity of Internet connectivity, smart devices and IoT technologies, social sensing is emerging as a dynamic AI-driven sensing paradigm to extract real-time observations from online users and society. Social sensing is rapidly progressing as a pervasive sensing paradigm where humans and internet-enabled devices are used as sensors to attain situational awareness about the physical world. Examples of social sensing applications include predicting poverty in developing countries, studying human mobility/migration across regions, identifying traffic abnormalities, monitoring the air quality, tracking social unrest and disasters, and detecting wildfire. The social sensing concept is motivated by the following observations: 1) people actively share their information via online social media, 2) sensing technologies have become prevalent and cheaper to deploy, 3) applications need timely access to the data to perform analytics and take timely decisions.

The TiH-Anubhuti envisions unprecedented opportunities to leverage data generated through social sensing for pertinent applications. Specifically, The TiH would focus on questions: How to efficiently collect and curate sensors data prevailing with veracity problems? How to manage and share the collected data and information efficiently and timely to the end applications? How to leverage the computational power on edge devices to construct fully integrated edge-based social sensing platforms?

Challenges with the data collection include locating and scrapping relevant data from social networks for example, identifying relevant query terms/ handles in case of Twitter, intelligently working with limits to data access posed by an infrastructure in terms of number of API calls/ data velocity, addressing issues related to data life-time for example in Twitter tweets can be deleted due to various reasons, missing data due to network failures, and establishing veracity of data. The second challenge is related to the data modality as the sensed data can be of different variety or types such as text, image, location, audio, and video. Moreover, each type can further encompass different dimensionality. The third main challenge is the scalability aspect in sensing where we need AI algorithms that can be readily deployed across the edge devices in order to reduce latency and bandwidth consumption, and yield faster information extraction.

In this TIH-Anubhuti, we focus on the development of key cognitive and social sensing technologies to develop Cognitive Computing and Social Sensing systems and different applications to solve society's most vexing problems.

PROBLEMS TO BE ADDRESSED

3. Problems to be Addressed

In this TIH-Anubhuti, we focus on the development of key cognitive and social sensing technologies to develop Cognitive Computing and Social Sensing systems and different applications to solve society's most vexing problems. We have decided to work on two grand problems along with the relevant theoretical and applied challenging applications in the broad area of Cognitive Computing and Social Sensing. The following are the problems in the broad area of Cognitive Computing and Social Sensing which we aim to address as part of this TiH.

AI driven Data Curation

Artificial Intelligence (AI) has already received a lot of buzz in recent years for good (like helping us to make our daily life better) or bad reasons (like potential job losses due to automation). AI refers to computer systems that are built to mimic human intelligence and perform tasks such as data understanding (like recognition of images, speech or patterns), reasoning from that data to drive inferences, and decision making. Moreover, the size of information integration problems is increasing rapidly with the ever-growing quantities of data that enterprises must deal with nowadays, making human-centric solutions to curating data almost impossible in future integrated information systems.
Different types of Machine Learning algorithms can be used to build models for performing these tasks to discover patterns, insights from data, and to automate most decisions in the pipeline of data curation, with human intervention only when necessary. This approach involves hard problems such as identifying attributes, clustering related attributes, entity consolidation, transforming input data and deduplication. In addition, a machine-driven data curator should also adopt a continuous learning model and perform entity enrichment and context enhancement to incrementally improve the quality of the curated inventory of enterprise data. AI can do these tasks faster and more accurately than humans if the underlying system is continuously trained on large data. However, these models may be biased in taking the right decision since decision making algorithms run on the models which are built on training data and this training data set might have not seen all cases. Therefore, one of the core issues in AI is how to detect the biases from the outcome, if there is any, and explain how the models have reached the conclusion.

Intellective Information Systems

The resurgence of artificial intelligence (AI) is powering new generations of smart devices, consumer applications and enterprise systems. The proliferation of sensors and devices and the explosive growth in structured and unstructured data are causing information and contextual overload. With the increasing affordability and sophistication of smart devices, new opportunities exist to provide contextually aware and personalized services based on user views, desires, preferences and location, delivered just-in-time. Today's enterprises are looking for systems that allow them to streamline the flow of data throughout its entire life cycle with minimal human intervention. The next generation of information systems are what we have begun to call intellective or cognitive systems as shown in Figure 1. These systems leverage cognitive capabilities (machine learning and AI algorithms) inclusive of

understanding, learning and contextual awareness to simplify the ingestion and characterization of data to make the data/system ready for AI, and to collect and manage critical semantics information needed for integration, query processing, policy enforcement and auditing. Immense amount of personalization, precise profile mapping, accurate trend analysis and future prediction, more than ever exact results from unalterably large and unstructured database – all these are going to become the new normal in the business landscape, going by the current trajectory of development and deployment of these cognitive systems.

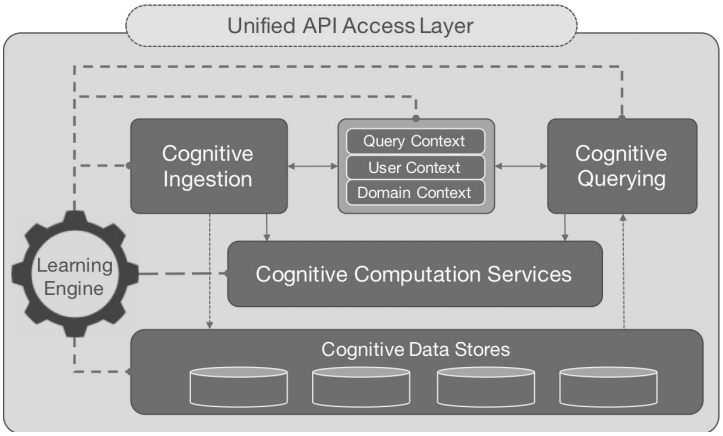


Figure 1: Intellective Information System

The information composition flow in an intellective information system is shown below in Figure 2.

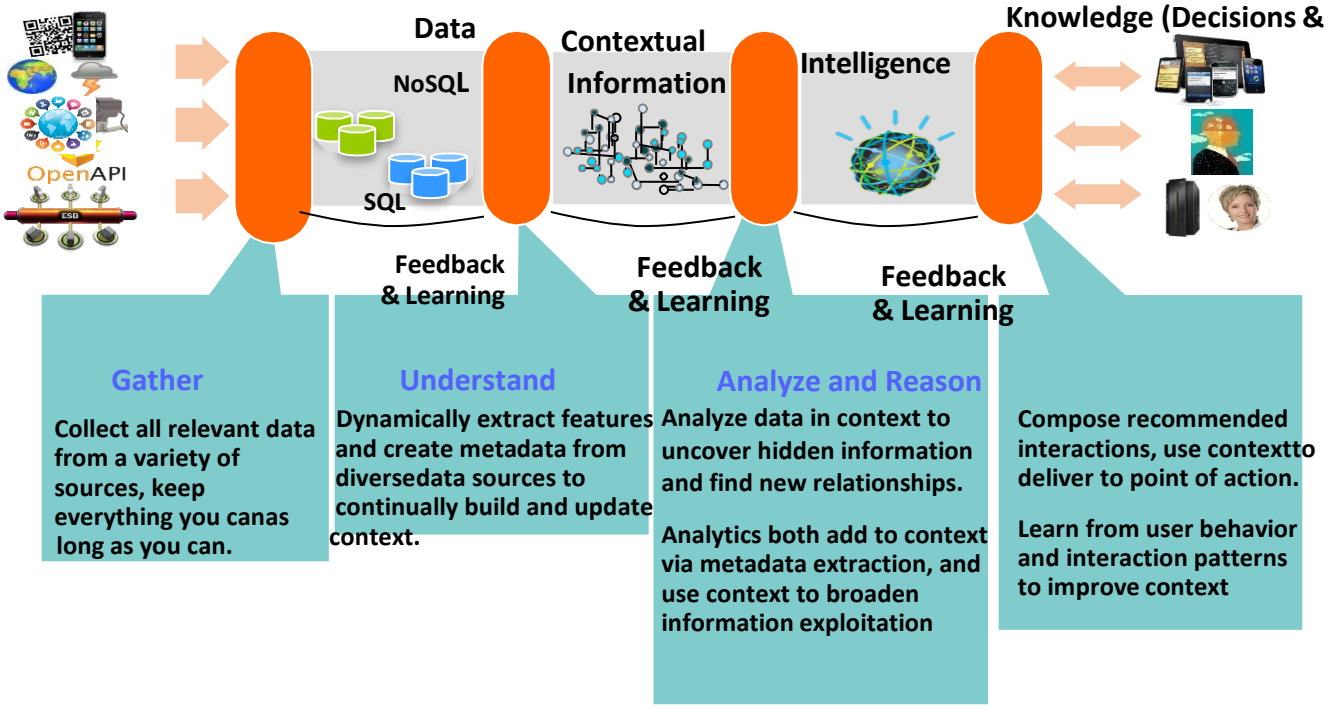


Figure 2: Information Composition Flow in Cognitive System

Understanding: Traditional systems are largely governed by well-defined structures, rules, and schemas and are programmed specifically to handle those constructs. Therefore, the system is prepared for and can accept only a certain kind of data, and once it is set up for that kind of data, it assumes everything that is fed to it with a particular structure has the same semantics and equivalent importance to the application. In essence, they are deterministic. In contrast, large cognitive systems deal with real world signals that come as varied combinations of structured and unstructured data. Their goal is to move the effort of understanding incoming data from the application developer using the system to the system itself. For such systems, it is fundamental to treat different inputs/data differently, depending on their content and context. Before accepting and storing the data, the system must able to determine the relevance of the incoming data to business use cases. Rather than moving the data to a structured store specifically tailored for that input, the system needs to understand and analyse the data significantly before routing both the explicitly supplied data and additional inferred data to appropriate data stores, often splitting it into targeted chunks in the process. These transformations go well beyond the realm of traditional ETL, into the more complex domains of natural language understanding, semantic linking, noise elimination and other such transformations.

Contextual awareness: To achieve the level of understanding described above, systems will need to leverage all kinds of information they have come across and recognize semantic relationships between data items. The queries that are posed to these systems are not intended to simply retrieve known information. They require probabilistic approaches that evaluate multiple features, include relevant unstructured data on the fly, and provide facet analysis. In fact, contextual awareness is embedded into the system in multiple layers. Context is attached to the input during querying which helps infer the query better. Context is inferred from past interactions and the relationships among data items and leveraged by the algorithms processing the input. Context is attached to the results of the algorithm in order to explain the results during presentation. Lastly, context is added to the resulting entities, which can allow the user to explore beyond the results. All of these should be automatically and smartly done by the system rather than relying on interventions from the user.

Continuous learning: A common pitfall for the engineering community is to deem a system which uses a machine learning model to make a classification or regression decision to be a “learning” system. Nevertheless, a system that relies on a static model is not one that truly learns. We stress the aspect of continuous learning, also referred to as online learning, in which the models which are being applied are updated on a continuous basis as data is being ingested and analysed. An extension of a truly smart online learning system is a system which constantly evaluates its knowledge of the problem domain and can distinguish between what it knows and what it does not know. Through interaction with the user at the presentation layer or through some automated crawling techniques, it can then explicitly seek training data/evidence which would help fill the gaps that were identified. These systems are called “active learning” systems. The idea of continuous learning can be applied to other aspects of the system as well. Systems should also be able to learn about users and interact with them based on their preferences and usage profiles. On a larger scale, systems should also be able to learn the application workload of the enterprise and tune performance accordingly.

Mobilizing information for Cognitive Intelligence

Businesses demand superior understanding of data entities ('data at rest' and 'data in motion') and mastery of information, where information as well as the underlying data entities are heterogeneous, distributed, interrelated and stored in silos in structured and unstructured data repositories. Moreover, the way the structured and unstructured data (such as Text, images, video) are managed and queried creates an artificial separation between the two, which is unfortunate since they are complementary in terms of information content. Thereby, it poses great challenges for effective discovery and linkages of data entities across data, content and images for integrated intelligence analysis, operation, and adaptation. (Note: hereafter 'Data' refers to either structured data and/or unstructured data including video, image)

Since the entity related information exists across a variety of classified and open sources, and today personnel across departments spend a lot of time to "connect the dots" across various related pieces of information, and still may skip vital artifacts to discover the complete entity. Given the ever-increasing size of data being collected by various departments, it is essential that decision-makers are provided intelligent tools that can automatically extract new relevant "data from data" without being explicitly asked, leading to actionable intelligence and making these data entities semantically and contextually discoverable.
One of the goals of the proposed TIH is to understand the fundamental principles that will enable us to create a Cognitive Entity Management platform for distributed collaborative analytics. Particularly, we aim to address the following research themes, but not limited to

1. Build a cognitive entity store which not only stores the entities, but also the contextual information about the data and its relationship with other data entities, how these entities can be leveraged in collaborative data analytics.
2. Address numerous data curation challenges in handling entity identification from unstructured data and images, integrating with structured data, handling different forms of data and variations in the schemas, and uncertainty introduced by the extraction and integration process.
3. Develop entity analytics techniques for entity disambiguation, remediation, discovery, correlation, linking based on ML to handle noise and uncertainty.
4. Develop a novel, context-oriented, loosely coupled integration of structured and unstructured data entities through symbiotic consolidation of related information (1) to enhance structured data retrieval by associating additional documents relevant to the user context with the query result, and (2) to enhance document contents by associating additional data entities derived from structured data.
5. Develop methods for correlating events (arrive from various sources – sensors, streaming data, social media and applications) with data entities. Events are matched based on entities that relate to the events and sometimes enrichment is required from reference data to perform the match. The matching of events is done in many cases according to contexts that may be temporal, spatial and segmentation oriented based on entities.
6. Development of scalable and effective algorithms for knowledge discovery to uncover patterns, correlations, clusters, outliers, and abnormal structures in symbiotic data entities, and incremental knowledge discovery as data entities are evolved over time.

Cognitive Education

To set the tone for a self-reliant future, Prime Minister Modi has recently dedicated the dream of Atma Nirbhar Bharat to the second largest populist country. Subsequently, the release of the new National Education Policy, a comprehensive revision and upgradation of the 34 year-old policy, gave high quality contemporary education a necessary impetus. We envision that in near future India will foresee an exponential growth in developing and providing world-class education. Since education plays a critical role in the development of individual and societal success, a major transformation is happening in this industry due to technology immersion (AI, AR/VR, Blockchain, IoT) and societal factors forcing to deliver personalized learning needs of an individual. These needs are causing institutions and individuals worldwide to re-invent educational methods and institutions. Some of the forcing functions to bring these changes are:

- Technology Immersion: digitization of educational content, proliferation of personal devices, online delivery for both institutional and self-learning through mechanisms such as Massive Open Online Courses (MOOCs). AI can improve and enrich learning content across all aspects of education.
- Personalized Education: build detailed learning models for every individual, personalized education pathways can be created or adapted to suit our needs and goals by analyzing the massive learning related, just as our consumer retail experience can be tailored to our tastes and preferences.
- Inclusive Learning: this includes vocational education training, education for students with special needs, etc by leveraging technology.
- Industry Needs and Economic Alignment: the desire to change how we engage with our education environment and to tie the needs of employers more tightly to the educational curriculum, including hands-on internships to avoid on the job training by the employers, and broad skills to enable lifelong learning.
- Engaging Interactions: the desire to change how we engage with our education environment by developing the cognitive virtual agents trained on specific subject matter.
- Global Integration: the expectation that educational institutions play an important role in regional competitiveness as sources and hosts of innovation, and are accountable to deliver value and satisfactory outcomes.

The increasing adoption of AI in education opens many challenges both to educationists and computer scientists. These challenges include adapting breakthroughs in cognitive data analytics, natural language processing, deep learning, machine learning, computer vision, and human-computer-interaction to develop (1) learning models using big data analytics methods, (2) methods for generating self-describing learning content and on-demand aggregations of learning content, including generating questions and practice tests based on learning patterns, (3) recommendation engine for how to read/understand/adapt the given learning content with augmented reality and/or virtual reality, (4) conversation agents for producing the engaging interactions as part of the education experience, (5) research tools to assist in understanding literature and discovering answers from millions of documents, (6) ethical issues related to cognitive computing systems that behave and act like humans and

that are modelled on human beings pose significant opportunities and challenges. Super-intelligent systems can overcome the limitations of human imagination and thought and achieve outcomes hitherto inconceivable. The development also poses significant challenges for society: would these systems enable or control human beings? Would the decisions taken by such systems be fair and not reflect bias and prejudicial behavior? Would they threaten human freedom or enable it? Would the decision procedures and the algorithms that enable them be accountable to human beings? Would they be predictable to ensure legal scrutiny? Would they respect the user's privacy? Who would be responsible in case harm is caused to human beings? Similarly, social sensing raises questions of profiling, manipulation, harm, privacy, ownership and control over data. Slightly less than half the population in the world has access to smartphones today. In India, the number stands around one third, but the growth rate of smartphone usage is above 10% in our country. Recent statistics show that the number of smartphone users in rural India (with few pre-schools) exceeds that of urban India by a significant margin (more than 10%). These numbers show that remote learning is a possibility in the sub-continent and around the world.

Researchers need to solve many challenging problems in these areas, including vocational education training, inclusive education for students with special needs (e.g. students with autism, learning disability, intellectual disability, attention deficit hyperactivity disorder [ADHD] etc.), efficient access to digital education etc. Globally 15% school going children are diagnosed as Children with Special Education Needs (SEN). These include ADHD, Autism, learning disabilities, Intellectual disability, cerebral palsy. children have trouble paying attention, controlling impulsive behaviours, and/or hyperactive. Worldwide, the condition of SEN children 3 has been exacerbated due to COVID 19 when education delivery shifted online. iHub-Anubhuti envisions to research and develop cognitive computing and social sensing technologies to provide solutions to above mentioned problems for the benefit of the learners, educators, educational institutions and the educational regulatory agencies.

Cognitive Computing and Social Sensing for Smart Cities

Developing sustainable solutions to manage our environment is an urgent need. In an era when new cities are rapidly growing unmanaged and food security is a growing concern, we need technology assisted indigenous solutions. Within this vertical of the proposal, the four main emerging themes are Intelligent Transportation Systems, pollution and waste management, disaster management, and natural resource management.

ITS or Intelligent Transportation Systems involves a wide range of technology-based solutions to manage transportation backbone of cities and traffic. Traffic congestion is a pressing problem in most big cities and using AI based approaches for analysis can offer automation for management of traffic. With such systems in-place, it will be possible to reduce manual interference which is error prone. Such visual traffic analytics solutions will strengthen the camera based infrastructure that is already present in many big Indian cities. Existing network of traffic cameras will also enable development of intelligent systems where re-identification is possible across cameras with the help of AI. Such a system can also be extended for person re-identification which is also a requirement in secure zones like airports, and railway/bus/metro stations. In addition, Internet of Things (IoT) based smart solutions for optimal transportation will provide better routing, ticketing, and congestion control on Indian roads. ITS goes hand-in-hand with good road infrastructure. Creating decision support tools for maintaining good infrastructure will keep a check on the health of public infrastructure. This will also enable civil infrastructure managers to better plan and minimize impacts of failure and take effective and proactive repair and replacement actions.

Solid waste and air pollution management are two of the most required actions we need to make to restore and protect our natural surroundings. In spite of wet waste policy there are gaps in the process of implementation with the huge public dealing. Technology embedded AI driven solutions can help reduce dumping in landfill sites. This would include designing indigenous hardware for cleaning wet waste bins and segregating waste. There are no easy to use tools available that provide real-time, local information, as well as a short-term forecast, of air quality or its expected health impact to the public. Many cities have launched air quality apps, but do not provide any tools to guide the city towards improving air quality. Leveraging IoT technology to collect and process crowd-sourced data is a feasible way forward to build dense information on air pollution. This information can then be processed with cognitive computing tools to contextualize and model various scenarios to facilitate municipal authorities and policy-makers towards shaping public policy.

Disaster management and emergency response requires special initiatives. Natural and man-made disasters destroy environments and cause difficulty for relief workers in terms of accessing the affected areas and thereby providing assistance. Technologies such as UAV based surveillance systems will provide rapid and accurate assessment of situations and help in planning relief responses. Combined with better human resource allocation and tools to collaborate will enhance the emergency response.

India’s natural resources are rich, however their management requires urgent action from authorities. This includes developing AI driven tools and techniques to monitor, manage, and conserve all forms of natural resources. Managing the health of soil is one such natural

resource on which our food security is heavily dependent. The soil health card scheme, launched in 2015, aims at shifting this paradigm toward precision farming. These cards provide farm-level information about nutrient deficiency, fertilizer application recommendations, and crop-suitability of soils. Cognitive computing has the potential to save lots of manual efforts involved and provide comprehensive recommendations in real-time. Deep learning based methods can provide precise analysis and therefore help us in building not only soil mapping models but also excellent recommender systems.

With growing emphasis on Smart Cities initiative of the Govt. of India, new technologies need to be designed, developed and deployed. The following problems related to Smart Cities would be addressed at iHub-Anubhuti:

Technologies for Intelligent Transportation Systems:

An efficient transportation system is vital and Central for any Smart City. At iHub-Anubhuti, we plan to solve following problems:

1. Smart optimal transportation system for smart cities across India: There is a need to develop smart transportation system that collects data from sensors, cameras, and other transportation infrastructure. The data could then be used to provide personalized transportation solution, e.g., finding less crowded routes, or cheapest path. The system would also help authorities to plan transportation infrastructure and services.
2. Visual Traffic Analytics: Projects to provide solutions for problems related to object tracking, reidentification and counting. This may include multi-camera target tracking (MCTT) which involves tracking objects across multiple cameras along with re-identification. Subsequently, other important tasks of vehicle reidentification and vehicle counting need to be addressed.
3. Personalized Applications for Supporting Tourism: With India becoming a global hub and Delhi being the capital of India, there is a need to develop technologies for supporting tourism. These goals include solutions to enhance and augment touristic planning by providing tourists with hyperlocal information about tourist attractions. To achieve this, there is a need to cognitively capture user preference and context information by means of social sensing and crowdsourcing, and deliver snippets of POI/event spatio-temporal information to the tourists in a cognitive manner based on their preference and context (including location). Second, for tourism authorities, there is a need to have solutions that capture statistical information using social sensing and provide actionable insights and what-if capabilities for improving touristic foot traffic. The solutions will need to provide technology for contextualizing statistical tourist data at different locations to better understand how to improve touristic foot traffic, e.g., the location of the touristic attraction could be unsafe, the routes to the location may have bad roads or lack public transportation, or admission ticket prices may be too high.

Air Quality Monitoring and Assessment Platforms

With pollution increasing everyday, there is a need to develop smart solutions that can provide effective solutions. Currently, there is paucity of air quality data. There are only 10

cities in India that are connected to the National Network where air quality data is transmitted to the web portal and shared as well as stored. The number of stations is limited, only 573 nationwide, of which only 200 are continuous air monitoring stations (as of September 2019). Researchers estimate the need for "4,000 continuous monitoring stations, 2,800 in the urban areas and 1,200 in the rural areas of the districts, as per 2011 census) to spatially, temporally, and statistically represent the PM2.5 pollution in the urban and the rural areas of India". The sparse data limits the ability to connect air pollution with causal factors. There would be challenges in the collection of air quality data. Since social sensing and crowdsourcing will be a major source of local and hyperlocal data on air quality, temperature and built-environment morphology for our proposed approach, research challenges will include the design of incentive mechanisms for users to contribute data to our system, developing mobile crowdsensing systems, efficient algorithms for sensing task distribution etc. In addition, using the air quality data, there is a need to develop models for forecasting air pollution levels in the short-term and medium-term.

Smart solutions for Agriculture

Agriculture continues to remain the biggest sector in India for employment with about 70% of Indian population dependent on it. With new agricultural techniques and in light of new agricultural policy of the Govt. of India, there is a need for developing smart cognitive computing and social sensing based solutions for the agriculture. Some of the challenges that we aim to address in the TiH-Anubhuti are:

1. Solutions for Comprehensive Real-time Soil-health Data: Currently the soil-health cards are used country-wide to manage the soil data. Adding attributes like comprehensiveness and liveliness to the existing soil-health cards may help in providing precise recommendations for Indian farms. Moreover, the existing health cards are limited to soil's chemical property based fertilizer recommendation only. There is a need to explore satellite, drones and usefulness of AI to make the health cards as comprehensive as possible with respect to agriculture, irrigation and climate-based decisions. In addition, the dynamic nature of soil needs to be taken into consideration via timely remote data collection compared to the existing approach of soil-testing. Technological solutions that make use of such comprehensive and temporal information along with recent advances in AI will be supported by the TiH to make precise recommendations.

2. Smart Sensing and Learning for Agro Analytics: Instead of walking down the field periodically and scanning every square inch of the field, today farmers can leverage cameras mounted on commercially available drones to capture images of different areas of a field and apply AI-based image recognition tools to monitor their crops faster and more accurately. Automatic detection of plant diseases is an important research topic as it enables regular monitoring of large fields of crops, which, in turn, results in early detection of symptoms of diseases from the analysis of images of plant leaves. A unified framework that brings together the libraries for controlling the flight path of drones with popular computer vision and ML libraries is required. The existing state-of-the-art work fails to address the following issues. (i) Assimilation of large data that is collected from sensors of UAV Drones. (ii) A model that generalizes the solution

of crop monitoring with analysis. (iii) Enable an end-user who does not have any knowledge of this system to operate the drones.

Cognitive Computing & Smart Sensing for Law Enforcement and Security

Today, we live in a digitally enabled world, where we individuals are nodes in a complex network of information flows. With social media becoming an important channel through which people communicate with each other across the world, it goes without saying that social media has a tremendous impact on every aspect of our lives. A second important complementary technology that has seen an exponential growth in the recent years is the Internet of Things (IoT)– billions of physical devices, including sensor systems, satellites, AI-enabled devices, cell phones, and wearable devices, that are connected to the Internet. These physical systems collect, share, and communicate real-time data without human involvement.

The development and proliferation of these technologies has led to the accumulation of vast amounts of data from different sources. It has also opened up opportunities for strategically leveraging this to inform important decisions. On the other hand it also opens up a number of threats that can come from circulation of fake information through these social media networks, and security threats that come from accumulation of information through the interaction between humans and computing devices on the IoT. Within the Law Enforcement and Security Vertical in TiH-Anubhuti, the proposed projects focus on two main themes with the rubric of cognitive computing and social sensing.

The first set of projects will provide the competitive advantage in terms of applying technologies such as multimodal knowledge graphs, combining structured and unstructured data to develop enhanced abilities to monitor and analyse social media flows, building systems that help validate fake or inaccurate information, detecting and mitigating security threats from rumours and disinformation, and countering algorithmic injustice in AI-based decision support systems. Within this theme, creating systems that leverage the vast amount of digitized information flows to predict near future, and distant future threats to both the nation's security and the nation's space assets, and develop strategies to pre-empt and mitigate such threats is another priority in these projects.

The second set of projects will focus on developing technologies and hardware for the development of energy efficient, highly secure, AI-based mobile applications and improving the efficiency and security of such applications by leveraging advanced cryptographic mechanisms and technologies such as Blockchain, Homomorphic Secret Sharing Schemes (HSS), and mobile computing paradigms such as TinyML algorithms, Edge AI, and in-memory computing. Such mobile computing platforms find extensive use in domains such health, education, and e-commerce, and several other fields.

We plan to address problems that involve applications of secure data transfer, privacy and security, and law enforcement, which can be solved using cognitive computing and social sensing. The applications will cover societal challenges that need immediate intervention. Some such problems are listed below:

1. Enabling Smart-sensors via Novel Edge-AI and In-memory Compute Paradigms: From Design, Prototype to Fabrication we aim to develop an FPGA prototype of the proposed work followed by on-chip fabrication on exclusive state-of-the-art technology. The FPGA prototypes enable significant time and cost savings by reducing the number of fabrication stages and have significant commercial potential in industry and academia.

2. Multimodal Knowledge Graphs (KGs) for Law Enforcement and Healthcare: In almost all the multimodal KG construction techniques, images and videos are included as part of the KG, i.e., nodes in the KG refer to images and videos. We propose to enrich the existing KGs with the information available from other modalities such as images, videos, and audio. There are several challenges here -- 1. Extraction of structured data in the form of triples (subject, predicate or relation, and object) from the multiple modalities such as image, video, and audio. This is commonly referred to as extraction of scene graphs from the visual source. 2. After the scene graph is extracted, identify whether some or all of the information in this graph falls into one of the following categories -- a. Complement the information in the existing KG. b. Corroborate with the information in the existing KG. c. Contradict the information in the existing KG.

3. Tools to Detect Privacy Leaks: Deep learning (DL) models have shown their great potential due to their high accuracy. On the other hand, these models work like a black box and it is difficult to interpret them. Hence, it raises questions regarding the integrity, bias and fairness of these models. To address these questions, we aim to study different attacks on Deep learning models and develop a tool that would score ML models in different aspects. The tool would help an end user agency to take an informative decision regarding the deployment of models in real settings. The tool fits perfectly in the cognitive computing settings as DL models are one of the main building blocks for Cognitive computing and we aim to investigate privacy issues in DL models.

4. Development of Efficient Blockchain Protocols and Homomorphic Secret Sharing Schemes for Applications in Public Health and Education: We wish to innovate on efficient blockchain protocols that optimize throughput, latency, storage, ease of new miners, and energy efficiency. In order to account for addition of new miners, efficient coding-theory based protocols will be investigated. To reduce the computation burden to solve the proof-of-work (PoW) puzzles, an approach called sharding has been proposed. Sharding has been used widely in IoT blockchains. This work will consider sharded block chains and apply coding theoretic techniques to reduce the storage cost as well as the bootstrap cost. We will investigate secure regenerating codes that are not only storage efficient but also bandwidth efficient while repairing single node failures. We aim to draw the equivalence between the process of bootstrapping a node and repairing a failed node. This equivalence can make the bootstrap cost low as compared to uncoded sharding.

5. A Study of Adversarial Opinion Collusion on the Web - Detection, Characterization and Impact: Online discussion forums have become an important component of the social fabric, thanks to the rapid growth of online communities like Reddit, Twitter, Facebook, online comments on news and blogs, and so on. People

discuss events, policies, leaders, celebrities and organizations, forming and influencing opinions on topics ranging from socio-political issues, to which smartphone has the best user experience. Hence it is a lucrative space for malicious groups to promote a product or ideology. They often collude and form syndicates that hijack conversations and control their sentiment. Malicious online social network activity is alleged to have influenced the outcome of the 2016 electionin the USA, and is known to have caused lethal violence in India. Given the importance of online communities in shaping public opinion, an effort to detect such malicious individuals and groups is much needed. We seek to model and detect covertcollusion in social media.

6. Combating Online Hostile Posts in Regional Languages: The increasing accessibility of the Internet has dramatically changed the way we consume information. The ease of social media usage not only encourages individuals to freely express their opinion (freedom of speech) but also provides content polluters with ecosystems to spread hostile posts (hate speech, fake news, cyberbullying, etc.). Such hostile activities are expected to increase manifold during emergencies such as the 2021 US presidential election, COVID-19 pandemic spreading. Most of such hostile posts are written in regional languages, and therefore can easily evade online surveillance engines, the majority of which are trained on the posts written in resource-rich languages such as English and Chinese. Therefore, regions such as Asia, Africa, South America, where low-resource regional languages are used for day-to-day communication, suffer due to the lack of tools, benchmark datasets and learning techniques. Other developing countries such as Italy, Spain, where the used languages (pseudo-low-resource) are not as equipped with sophisticated computational resources as English, might also be facing the same issues. This project will emphasise on three major points: Regional language: The offensive posts under inspection may be written in low-resource regional languages (e.g., Tamil, Urdu, Bangali, Polish, Czech, Lithuanian, etc.). Emergency situation: The proposed solutions should be able to tackle misinformation during emergency situations where due to the lack of enough historical data, learning models need to adopt additional intelligence to handle emerging and novel posts. Early detection: Since the effect of misinformation during emergency situations is highly detrimental for society (e.g., health-related misadvice during a pandemic may take human's life). The solutions would be able to detect hostile posts as early as possible after their appearance on social media.

7. Fair and Ethical Social Sensing: Within the broad umbrella of fair and ethical social sensing, the following is a non-comprehensive list of problems that would be addressed within the project:

    a. Fair and Ethical Hotspot Detection: We will develop algorithmic ways of correcting for and attenuating data and process biases in hotspot detection to better ensure that the fruits of the technology be more equitably spread across demographic groups.
    b. Fair and Ethical Social Media Profiling: We will consider various technological tools for social media profiling, analyze the bias embedded within them, and develop technologies that will "correct" such biases.

In short, this project will explore the "softer" sides of social sensing technologies, with a view of making the social sensing stack more aligned with modern notion underlying the democratic world order.

8. Designing Automated Conversational Tools for Intent Learning, Curated Information Presenting, and Fake News Alerting – Application in COVID-19 and Beyond: This project aims to develop a user-friendly conversational AI system for social media platforms where any news/message can be fact-checked via simple message forwarding. Moreover, it will extract the useful information from online databases related to a query by the user. Initially, we plan to design the system as a Telegram chatbot for Covid-19 as the application area. It will be further extended to other messaging platforms such as WhatsApp, Facebook Messenger, Instagram, etc. focusing on other social issues.

Designing Cognitive Computing Solutions for Legal Information Processing and Management

Along various databases curated and analyzed for efficient information processing, legal documents have been most perplexing, opaque, and tremendously hard to lay hold of by common citizens due to the lack of appropriate domain knowledge. The current National Litigation Policy is ineffective to a large extent due to ambiguity. Moreover, due to the abundance of legal cases and lawsuits, it is very tedious for a lawyer, solicitor or an attorney to read, analyze and annotate volumes of legal information and evidence. We solicit efficient technological solutions building on the legal information databases to aid the requirements of various departments and administrative authorities for effective handling of court matters. Solutions are welcome to address the following (non-exhaustive) problems:

1. Legal data curation and modeling: Modeling legal data (both structured and unstructured) and building repositories of data by gathering and curating information related to various departments, tribunals, and categorizing them into groups of customized management information systems that can be accessed through a user-friendly interface. This involves right from data understanding to reasoning from that data to drive inferences and decision making. This expedites different levels of administration by sending timely alerts for catalyzing actions by concerned functionaries in a given case.
2. Search engine for legal document retrieval: Building an efficient search engine that, given the abstract of a legal case, retrieves similar relevant cases from the historical data and presents snippets of cases for quick inferences.
3. Summarization of legal judgement: Legal documents are generally written in a comprehensive manner. An automated summarization system will play a decisive role in reading and digesting legal documents within short time frames. It would also support the search engine in recommending snippets of legal documents.
4. Legal reasoning and inferences: Legal reasoning models investigate formal and computational theories of how legal experts analyze problems, create arguments, and make decisions. Two essential cognitive abilities for legal reasoning models include: case-based reasoning and adversarial reasoning. We solicit solutions to build legal ontologies, knowledge bases and automated methods to infer latent knowledge using

entity linking, inferencing links and high order deductive reasoning.

5. Legal document simplification: Legal documents are generally prepared in the most complicated and obscure language for the average person to understand. A language simplification model can convert them to easy-to-understand form by minimizing the use of legal jargon, foreign languages, and archaic words, breaking up long sentences into shorter ones with simpler language, and carefully analyzing the content to include only what is absolutely necessary for the document's particular use.
6. Legal question answering: Traditional search engines are inadequate in providing contextual responses against legal queries due to lack of domain expertise. Moreover, documents returned by search engines are lengthy and can not provide exact solutions to the user's needs. We solicit automated Q&A tools to comprehend the legal context and provide jurisdictionally relevant answers by mining legal documents, knowledge graphs and ontologies.
7. Automatic legal document drafting: Legal document drafting requires extreme manual intervention, which often leads to high chance of error when typing and retyping information. We seek solutions to help drafting legal documents with partial human intervention. The solutions may involve intent understanding, slot filling, answer retrieval and answer positioning.
8. Chatbot for legal conversation: Citizens often approach lawyers even for simple legal issues. We seek solutions to build chatbots that can chat with users online, initiating an online legal chat with a virtual legal assistant to quickly get all the help they need. The bot-lawyer automates the administrative and financial tasks of expert lawyers and provides some other basic legal virtual assistant services.

Here, we describe some of the problems in more detailed fashion:

An Automated Complaint to First Information Report (FIR) Drafting

Local police stations are the gateways to initiate a legal investigation proceedings by reporting any crime or a misfortune event. The first information report (FIR) is a legal document containing essential information about the event. It is a fairly technical document and the use of legal jargons are common. However, a majority of the public are not used to such language. Therefore, while reporting an event they use the most commonly used terms to explain the events, and an expert (a police officer) translates their narratives into an FIR.

In recent years, many state police departments (e.g., Delhi Police, Bihar Police, etc.) have started the service of registering online complaints, where the complainants are asked to fill-in some predefined fields, such as name, contact details, etc., along with the description of events. These platforms have two fundamental issues:

1. Similar to the offline setup, the manual intervention is required to prepare an FIR from the complainant-provided description. This often limits the throughput to handle a long list of online complaints.
2. Another critical issue is the absence of essential information in the description, which requires a series of back-and-forth communication between the complainant and the officer to finalize the FIR, thus a delay in processing the complaint is undeniable.

To this end, we propose, SAHAYAK, an automated system that analyzes the complaint

description, highlights the essential-but-missing information to the complainant, and generates the corresponding FIR once all essential information is acquired. More specifically, the system will operate in two phases:

1. Analysis:
    a. Event identification: The first task in the analysis phase is the identification of the event/crime reported by the complainant. This phase is crucial as different types of events require different essential information. For example, a report for theft of an automobile may require the vehicle identification number, make and model of the vehicle, etc.
    b. Entity extraction and Relationship identification between entities: This module will be responsible for extracting all the entities from the description and establishing relationships among the identified entities. Moreover, it will also extract the anaphoric and cataphoric relationship between the entities.
    c. Semantic-role labeling with the event: Once the entities and relationships are identified, the semantic-role labelling will map the entities to the event as per their role in the event. For example, a report for a violence and associated injury needs the information of any equipment that was used; therefore, relating guns, sticks, etc. to the event as equipment is an important piece of information. The resulting output (a graph) will be matched with the predefined template of the identified event and any missing information will be reported back to the complainant and request for revising the description.

2. Generation Phase:
    a. FIR Draft Generation: After collecting all essential information, the generation module will generate a cohesive FIR, which will be used for further legal proceedings.

Languages: In Indian context, the description can belong to different regional languages. Initially, we propose to develop the above modules to handle three language setups -- English, Hindi, and code-mixed Hinglish.

An Empathetic Knowledge Grounded Conversational System for Mental Health Counseling and Legal Assistance

Aim & Objectives: The proposal aims at developing an Empathetic Knowledge grounded Chatbot for Mental Health Counseling and Legal Assistance. The Chatbot will be multilingual in nature (English and Code-mixed Hinglish), and will exploit background knowledge present in the knowledge graph and/or knowledge base, and conditioned on user-specific information, domain-specific information and empathy. This is the very first attempt to make a conversational agent in this particular domain. The Chatbot will be assisting the victims, especially the women and children, who have faced harassments or abuses from various sources (e.g. office harassment, domestic violence and cyberbullying etc.), by providing mental counseling and legal assistance.

The Chatbot will be open-source and pluggable with the following features:

1. The Chatbot will be multilingual in nature, supporting English and Hinglish (Code-mixed Hindi and English).

2. The Chatbot will provide the stakeholders (e.g. victim, parents, humanitarian agencies) the following information: counseling in an empathetic way so that the victims feel more relaxed and stress-free; legal assistance by making the victims aware about various steps of filing the complaints to the legal cell, about the various IPC sections pertaining to the particular case etc.
3. The conversational system will exploit background knowledge in the form of a knowledge base to generate more informed and accurate information.
4. The Chatbot will be empathetic in nature because of the sensitivity of the domain, and should be able to converse with the users by understanding various degrees of emotions.
5. We shall launch the Chatbot's service through Facebook messenger, WhatsApp, and Web based service.

Technology & Research:

Data Collection and Annotation: Domain specific conversation data is required to develop the Chatbot. We aim at collecting data from various case studies, reports, and different internet sources. We will also follow a semi-automatic strategy for conversational data preparation. We shall collect the FAQ, and design dialogue templates to create the conversational data. A basic bot would be created using an existing framework, and then this would be made available to create more conversations.

External Background Knowledge Acquisition: The Chatbot will be enriched with external knowledge. The external knowledge will be gathered from the various reports, government websites, sites of the humanitarian agencies etc. The external knowledge will be used as a background knowledge to the conversational bot. This will be utilized for more informed and accurate response generation from the conversational system.

Natural Language Understanding (NLU)/ Spoken Language Understanding (SLU): This is the very first step of every dialogue agent. There are three specific sub-modules, namely Dialogue Act Classification (DAC), Intent Detection (ID) and Slot Filling (SF). To this end, we will adopt our existing deep learning based end-to-end hierarchical multi-task model that can jointly perform both intent detection and slot filling tasks for the datasets of varying domains.

Dialogue Manager (DM): Once the pre-processing of the utterance/s is over it would be forwarded to the dialogue manager. The overall conversational loop will be under the supervision of the DM which controls the various phases (welcome, information seeking, feedback, bye).

Natural Language Generation (NLG): Natural language generation (NLG) is an important module in dialogue systems, question answering systems, and many other natural language interfaces. To help the users achieve their desired goals, response generation provides the medium through which a conversational agent is able to communicate with its user. We will adopt our already developed NLG system for our task.

Empathetic Dialogue Generation: Empathetic dialogue generation plays an important role in engaging the users in long conversation with the Chatbot. As our domain is related to assisting the victims through legal assistance and mental health counseling, incorporating empathy

into the conversational system will be an important feature. We shall adapt our prior works on sentiment and emotion in dialogues, generating courteous or polite natural language responses, personalized response generation frameworks and sentiment and emotion-controlled dialogue generation for our tasks.

Search engine for legal document retrieval

Context/Background: In Today's world if a person has to look for a legal document online, the websites which provide such services are Indian kanoon, SCC Online and Desikaanoon, allof these Indian services provide a Keyword Search engine even for their paid subscriptions their database is increased but the searching algorithm is the same.

Issues with these Solutions: With the present day searching system, if we were to find the cases which are exceptions to a specific law, i.e., cases that don't follow the norm. These exceptions cannot be searched just by typing "Exceptions to the mentioned Law", these cases need to be searched by case number or their header will have to be remembered or indexed somewhere for anyone to be able to retrieve it.

Aims & Objectives: Our Aim is to develop a search based system using Natural Language Processing which interprets the legal documents and connects relevant information to it so as to increase the scope of language that can be used to search a particular document and furthermore provide a brief description of that document. Hence, allowing for people outside the legal domain, able to pick up and understand relevant documents.

Possible Problems:

1. The Data provided to work upon is in one of the 3 stages; it is in the form of Scanned PDFs, Scanned PDFs converted to typed information by OCR (unreliable) or Typed Documents (Best source).
2. Need of Structuring and summarizing data for an ordinary citizen to understand, This data can be further used by the Searching Engine to possibly reduce the data to traverse through, allow it to also be able to interpret data on the basis of summary language and understand the likelihood of the requirement of data.
3. Need for a Model for legal data and building a repository.

Strategy: The way we are planning to deal with the above issues and substantially benefit from it is by:

1. For starters our main focus will be on acquiring a Sample Dataset which we will take from the Supreme Court Legal Document Repository. The reason for picking up this dataset is due to the fact that the Supreme Court has typed and maintained these documents themselves and have provided it online for public use. This Dataset is very reliable and is a perfect start to work upon.
2. Next our focus shifts to summarizing the documents. If we are able to make a machine which can read, learn and interpret Legal English language to Normal use English. This machine would immensely benefit us in providing the reader a very concise, to the point, document. Not only can we provide this document to the general public but also we can teach our AI Searching Engine to understand the language of these summarized documents and extract data quickly, which will increase the SearchEngine's competence by manifold.

3. While researching the potential of this project we came across free online Repositories which were well structured as per our needs and free to use. These repositories can also be a great starting point for our project and allow us to develop the NLP Search Engine.

Multilingual and Culturally Customized Knowledge Management for Legal Information Processing and Management

1. The technologies to be delivered include a globalized (multilingual and multicultural) platform for data collection, curation, management, and sharing. A globalized knowledge base will be prepared for the legal information processing incorporating the semantic data model.
2. Cognitive Data Governance: Multilingual and Multicultural Linked Data Store, Semantic Repositories, Knowledge Base. The legal data will be modeled as RDF. Existing legal information regarding the legal cases, law suits, various departments, and various tribunals will be converted from the existing heterogeneous (structure or unstructured) formats into interoperable RDF format. As the data will be put on the linked open data cloud, it will be easier for experts to annotate. Means semi-automatically annotating the data will be provided. Once annotated, it lies on the cloud itself, open and linked accessible to all the APIs and services.
3. Cognitive Services and APIs: Natural Language, Multi-modal data relationship discovery, real time events/social sensing, Real time data/text feeds. The API developed will be generic and can be utilized for any domain and specifically for legal information processing. The information in the RDF store will be accessed through a user-friendly interface. The semantic technologies thus utilized facilitate the interpretability and explainability of the decision making and the reasoning procedure applied.
4. Application Layer: Following applications are proposed:
    a. Legal Document Classifier: Given the legal document, this use case will classify the document into categories as provided by experts in the adjoining ontology. The ontology-based text classification will comprise all the qualities of the symbolic reasoning and the statistical approaches.
    b. Legal Document Retrieval: Given the abstract of a legal case, all the similar relevant legal cases will be retrieved based upon the semantic similarity. The multilingual knowledge graphs will facilitate semantic search on multilingual corpora while increasing the efficiency of search because of the exploitation of relationships between the attributes. The documents thus retrieved could be classified as a next step to retrieve keywords and textual description summarizing their contents.
    c. Question Answering: Once the legal documents are collected and converted into knowledge graphs by virtue of the semantic data models, SPARQL queries could be written for many known queries and a user friendly interface beprovided to the end-user for querying the legal knowledge base.

Multimodal Knowledge Graphs for Legal Question Answering and Reasoning

An important aspect in the field of cognitive computing is to build machines that can think, behave, and act in the way that humans do. In order to do this, machines would need

background knowledge and this can help in intelligent decision making. Knowledge Graphs (KG) are graph structures that capture knowledge in the form of entities, relationships between them, properties, and additional information including provenance. Along with the W3C recommended Semantic Web standards such as RDF, OWL, and SPARQL, advances in Machine Learning, Deep Learning, Natural Language Processing, and Information Retrieval has led to automated construction of large KGs such as DBpedia, YAGO, Wikidata,
and proprietary KGs such as those from Google, Yahoo, LinkedIn, Microsoft, Amazon, eBay, and Alibaba. But all these KGs have been built from only the text documents. There is important information captured by images and videos, especially in the legal domain and this has largely been ignored. Along with the textual data such as FIR, we plan to use the information that can be extracted from other modalities such as images, videos (CCTV footage) and audio to enrich the quality of the constructed KG. Doing so involves several challenges here, and they are listed below.

1. Extraction of structured data in the form of triples (subject, predicate or relation, and object) from the multiple modalities such as image, video, and audio. This is commonly referred to as the extraction of scene graphs from the visual source.
2. After the scene graph is extracted, identify whether some or all of the information in this graph falls into one of the following categories.
    a. Complement the information in the existing KG.
    b. Corroborate with the information in the existing KG.
    c. Contradict the information in the existing KG.

Depending on the category into which the information in the scene graph falls, it would be handled accordingly. To the best of our knowledge, none of the multimodal KG construction techniques consider these three aspects. Having such a KG facilitates deductive reasoning that can lead to inferring information which is implicitly hidden. Along with that, a structureddata source such as a KG would help in answering complex questions.

Technology & Research Methodology: The proposed project would involve the following steps:

1. Step-1: Extraction of structured information from the different modalities. Structured data in the form of triples (subject - predicate - object) needs to be extracted from images, videos, and audio related to the text using which the KG was built. A scene graph is built from these triples.
2. Step-2: Enriching the KG with the extracted information. Triples in each of the three categories have to be handled differently in terms of adding them to the KG. The triples that are complementary can be added to the KG, but the challenge here would be to determine where exactly in the KG these triples should be added to, i.e., who should be the neighbors of these triples in the graph. Corroboratory triples need not be added to the KG, but they reinforce the information that is already part of the KG. They can be considered as synonymous or alias information that can be useful for the downstream applications. For the third category, contradiction needs to be resolved. There are a few options that can be considered and this depends on the context of the triples and the contradiction. A human in the loop approach can also be taken here, where the human can go over the ranked options of resolving the contradictions and giving the feedback in each case.
3. Develop applications. The first one would be an application that can complement/corroborate/contradict the image/video with that of the text summary.

This will be very useful in law enforcement. FIR is filed and notes are taken to describe any crime that would have happened. Along with this short report, generally, there will also be corresponding images and videos taken at the crime scene. In order to cross check the text in the report with that of the visual data, the techniques developed in this project can be used. We can find the missing or wrong information in the report. The second application is a question answering system that makes use of the multi modal KG and the ontologies to understand the query intent and answer the query by making use of the context provided by the query.

Legal Text Simplification, Summarization and FAQ Retrieval

India has a vast system of laws, but most citizens are either unaware of their rights under the laws or do not understand how to enforce them. The lack of simple, authoritative and accurate information that provides guidance on how to solve their legal problems along with multiple factors like a user's identity, income, social status, geographical location and level of education also affect their experience of interacting with the law and lead to a situation where the law can often become a tool of exploitation, rather than empowerment. In trying to be exhaustive, the law gets filled with jargon and convoluted sentences. It is full of words, phrases, and terms like 'prima facie', 'proviso', 'notwithstanding', etc., only understood by lawyers therefore, making it inaccessible to many. The problem is further amplified because English is not the first language for most of the population. There are hundreds of laws and regulations at the Central and State level and they are modified often and the final interpretation of the law is based on judgements by the various courts across the country. The information is structured in a way that to get an accurate solution to a specific problem, laws have to be read together with amendments, judgements, rules and guidelines, creating a complex data set consisting of multiple documents. For example, the Indian Penal Code, which covers all the substantive aspects of criminal law in India and has a bearing on the lived realities of citizens, has 511 sections, multiple state and central amendments, and numerous judgements expanding on different aspects of the code. For this information to be accessible and useful to an average citizen, it has to be processed together, reviewed and structured in a simple, jargon free language. Therefore, with this proposal we plan to solve the following challenges to alleviate the problems faced by millions in India everyday:

1. Legal Language Simplification and Summarisation - Our team in collaboration with Nyaaya has created a framework where information is collected from legislations, judgements and rules to provide Simple, Actionable, Recallable and Authoritative Legal (SARAL) information. However, this is currently done manually by a team of editors. This is not scalable for the large number of Indian laws. We believe that an AI powered solution can help us to do this efficiently.
2. Translation of legal language - India has 22 scheduled languages and many more unscheduled ones. While a majority of the laws exist in the English language, 94% of the population does not speak English. Even amongst those who do speak English, their knowledge is not enough to understand the legal nitty-gritties. It is observed that even the legal experts find it difficult to understand legal documents, it is unfair to assume a thorough understanding from the common literate. At the same time, state of the art systems like Google translate do not work for Indian language and much less so for legal text. Therefore, we plan to train models on already available legal texts to cover most of the 22 scheduled Indian languages.

3. FAQ Retrieval - Taking the help of search engines for legal document information extraction is an error prone task. Therefore, most people have to rely on a human legal expert for the same. This process is difficult and expensive. With the able help from legal experts from Nyaaya and Vidhi Legal Center, we aim to annotate a set of legal documents with their similarity measures, and build models to provide the most relevant legal articles to a user as per their query, both in text using chatbot and speech using interactive voice response (IVR) system.

Technology & Research Methodology:

1. Collection and curation of datasets suitable for each of the above stated problems. This would involve searching, collection, parsing, cleaning and analyzing legal documents available in various public forums like the constitutional articles, schedules, sections, land deeds, state laws, court judgements, ministry notices and notifications.
2. Training models for Extractive and abstractive legal document summarisation and simplification and designing metrics suitable for both the tasks.
3. Dataset curation for similarity of legal documents by annotation from legal experts.
4. Training models for information extraction and recommendation of articles based on keywords, keyphrases, topical and other document similarity metrics.

Linked data model for legal information processing system

The legal system has key stakeholders such as judges, lawyers, plaintiffs, defendants, government departments and corporate/business entities. Legal information typically includes both structured data (e.g., title of case, date/time of trial hearings, names of parties, i.e., plaintiffs and defendants, names of impleaders, if any, amicus curiae, names of judges and lawyers, etc.) as well as unstructured data (e.g., evidence, information relevant to the case, witness testimonies, out-of-court depositions by witnesses, court transcripts, etc.). Observe that the unstructured data can be multi-modal, i.e., text, images, audio and video. Moreover, legal information essentially entails heterogeneous data sources linked to each other. For example, different government departments, law firms and corporate entities may maintain information in their own formats and use different database/data warehousing technologies since there is generally no de facto standard for legal data storage formats and processing technologies.

Given the sheer complexity of data in legal systems, legal information processing systems need to be organized in a way such that the queries of various stakeholders of the legal system can be answered both efficiently and seamlessly across such multiple and heterogeneous linked data sources. Now let us consider some typical queries to obtain a better understanding of the requirements of a legal information processing system. A lawyer, plaintiff or defendant may issue the following types of queries: “Retrieve all the judgments related to tribal land rights”, “Retrieve all the judgments given by the Supreme Court in the category of Medico-Legal matters”, “Retrieve all the judgments given by a specific Supreme Court judge in the category of Freedom of Speech during years 2010-2020”, “What kinds of arguments were made by the prevailing party in successful cases contesting land acquisition for highway development?” and so on.

Any legal information processing system needs to have the capability to cognitively understand, contextualize, discover, ingest, reason and infer about various kinds of data

(possibly multi-modal) across multiple disparate and heterogeneous data sources. Moreover, the legal information processing system also needs to be capable of acquiring, recording, indexing, cleaning, integrating, curating, visualizing and disseminating the data appropriately to ensure data reliability as well as consumability. Furthermore, the cognitive capabilities of the system should include cognitive querying based on query context, user context & domain context, as well as cognitive computation services for cognitive ingestion of the data. For example, there would be a significant difference in query context between cases involving corporate accounting frauds versus cases involving real property disputes or breach of contract or divorce. Such differences in context arise primarily because different domains work in fundamentally different ways. Furthermore, new laws may emerge and/or old laws may be rescinded or superseded by new laws and so on. This further exacerbates the problem of search in legal information processing systems because the search query needs to be contextualized based on the laws prevalent at the time of judgment. In essence, the legal data currently resides in traditionally passive data stores, and we need to convert such passive data stores into cognitive data stores and cognitive knowledge bases.

For effectively handling complex user queries in legal information systems, we propose a linked data model, which is capable of linking multiple disparate and heterogeneous data sources and contextually answer a wide variety of complex queries, whose results would be valuable to various stakeholders of the legal system. We will search through the Relational database (containing structured data) and then execute search on the unstructured data (stored in MongoDB) using “contains” for performing effective keyword search. In particular, we will use faceted search to enhance the search mechanism with a well-defined set of filters to facilitate search navigation. In effect, this would facilitate stakeholders in applying various filters to quickly zero in on the search results that would be most valuable to them. For example, let us consider the following query: “Retrieve all the judgments given by Judge X while on the bench in the High Court of State Y”. In this case, the filter categories for facets could relate to different types of cases e.g., lawsuits associated with land disputes, inheritance, and industrial wage claims. Facets could also relate to time, e.g., the judgments given by Judge X during a particular period of time or while at a particular court. Such faceted search and filter-based navigation can be extremely valuable to stakeholders in terms of obtaining targeted results.

Established legal research tools such as LexisNexis (https://www.lexisnexis.com/en-us/gateway.page) provide legal solutions & insights based on data analytics. In particular, Lexis DaaS (Data-as-a-Service) adds value to litigation analytics by providing APIs, which enables effective data-driven research into a wide variety of court cases. Similarly, a service such as Manupatra (https://www.manupatrafast.com/) provide a comprehensive database of court cases, and is one of the market leaders in the field of enabling law practitioners towards digital transformation. It also has data analytics and visualization capabilities to facilitate obtaining actionable insights. However, in order to use the full suite of tools/technologies in these existing systems, the subscription fees are generally too expensive for most solo legal practitioners, and generally out-of-reach for most ordinary litigants in India. What this means in practice is that established senior lawyers and wealthy litigants can access these systems, but ordinary lawyers and litigants, particularly those practicing in small towns and district courts are neither able to afford such subscription fees nor could they afford to hire IT specialists for sorting through masses of legal documents. In this regard, our vision is to develop an easy-to-use, cost-efficient (i.e., either free or with nominal charges) and open access legal information processing system, which serves as a one-stop platform for

catering to the requirements of various stakeholders in the Indian legal system. Furthermore, our proposed legal information processing system is differentiated from existing systems in that we will convert traditional passive data stores into cognitive data stores and cognitive knowledge bases.

A system for legal document search

The legal system can be a complicated maze of information given that different subsets of laws apply in different situations: election laws, personal laws, criminal laws, etc. Lawyers and judges – who are trained professionals – also have to refer carefully to the letter of the law before making their arguments and pronouncing judgments. When dealing with a complex case, other similar cases and precedents are important sources of information. With the digitization of many laws as well as court records, including the judgments of the supreme court from 1950 onwards, there is a big opportunity to make use of information retrieval methods to help in the research of past cases that are relevant to a case of interest.
In this project, our aim is to address two problems: i) given a summarized description of a case, identify other similar cases which have received judgments from the court, and, ii) expand the court's judgments with appropriate sections of the written laws and other documents. In addition, we aim to develop a suitable web interface that will allow users to navigate the results in an intuitive manner.

Retrieving judgments

When doing research on a case, a user may be interested in simple questions, such as: "How does the IPC define 'Robbery"'. A simple keyword search will easily retrieve the result for this question (specifically, Section 390 of the IPC defines the term "Robbery"). However, deciding whether an incident constitutes "Robbery" is more difficult. For example, the incident may be described as follows: "A person A entered the home of person B and stole jewellery after breaking the lock of a wooden cupboard. Person A was known to person B.". In order to interpret this incident as a "Robbery", there also has to be a physical threat to B, otherwise, the incident remains a "theft". These kinds of distinctions can be tricky to determine when conflicting evidence is given by witnesses.

For a lawyer who wants to prove or disprove that the incident was a Robbery, past case judgments on similar incidents will help to determine how to proceed.
Our aim is to provide an interface where a description or summary of the case is given and the system automatically retrieves a set of documents (typically, judgments) which are relevant to the case. Further, since these judgments can be very long, we aim to highlight exactly the sentences/passages that are relevant.

Expanding judgments

Retrieving judgments is sometimes a first step in trying to make sense of complex cases. Long judgments typically refer to different acts and laws by name which readers may not immediately
be aware of. Further, a set of relevant judgments for a particular case may even refer to other judgments as precursors to the current decision or as precedent. For example, the following as an excerpt from a judgment: "Feeling aggrieved and dissatisfied with the impugned judgment and order dated 27.11.2019 passed by the High Court of Judicature at Allahabad in

Criminal Miscellaneous Application No. 5213 of 2018,....". Clearly, the judgment refers to another previous judgment on the same matter, but not in a way that is easily accessible to the reader. Another reference from a judgment to precedents is also not easily accessible: "Heavy reliance is placed on the decision of this Court in the case of Prof. R.K. Vijayasarathy v. Sudha Seetharam (2019) 16 SCC 739 and Dr. Lakshman v. State of Karnataka (2019) 9 SCC 677."
The second part of this project aims to automatically connect these references to the actual documents. That is, provide navigable links for judgments and other documents relevant to them.

Grand Problems

Besides the broad area of Cognitive Computing & Social Sensing, we have decided to focus on two specific problems and provide cognitive computing and social sensing based technical solution to address the major challenges of our grand problems. The aim is conduct deep-tech, actionable, and translational research to provide solutions for the challenges that are faced by these grand problems. The grand problems are:

1. Grand Problem 1: **Designing Cognitive Computing Solutions for transforming the Education system in India working in collaboration with the Ministry of Education:** We are actively engaging with the Ministry of Education and discussing in different possibilities in which we can support technology for Education. Following are the ones suggested in our joint meetings:

Possible collaboration areas addressed by Ministry of Education:

1. **NDEAR:** VISION STATEMENT of NDEAR: A globally pioneering effort in education - A unifying national digital infrastructure to energise and catalyse the education ecosystem

   NDEAR is federated, unbundled, interoperable, inclusive, accessible, evolving aims to create and deliver diverse, relevant, contextual, innovative solutions that benefit students, teachers, parents, communities, administrators and result in timely implementation of policy goals

   We will work closely with the Ministry of Education to develop this unifying national digital infrastructure.

2. **DIKSHA** - Digital Infrastructure for School Education. This is an initiative of the National Council of Educational Research and Training (Ministry of Education, Govt of India).

   DIKSHA (Digital Infrastructure for Knowledge Sharing) is a national platform for school education, an initiative of National Council for Education Research and Training (NCERT), Ministry of Education. DIKSHA was developed based on the core principles of open architecture, open access, open licensing diversity, choice and autonomy as outlined in the Strategy and Approach Paper for the National Teacher Platform released by the former Hon' Minister for Human Resources Development Shri Prakash Javdekar.

   We will work on DIKSHA with the objective to make the policies and tools robust enabling the

education ecosystem (educationist, experts, organisations, institutions - government, autonomous institutions, non-govt and private organisations) to participate, contribute and leverage a common platform to achieve learning goals at scale for the country.

3. **UDISE+**: timely and accurate data is the basis of sound and effective planning and decision-making. Towards this end, the establishment of a well-functioning and sustainable educational management information system is of utmost importance today.

   Unified District Information System for Education (UDISE) initiated in 2012-13 integrating DISE for elementary and secondary education is one of the largest management information systems on school education covering more than 1.5 million schools, more than 9.6 million teachers and more than 264 million children.

   UDISE+ is an updated and improved version of UDISE. The entire system is now online and has been collecting data in real-time since 2018-19. However the data that is being captured is not in a form which can have a meaningful interpretation for any actionable.

   We will be working on this platform to (a) add on the various layers of data which is needed to gauge the progress of the various educational institutes and (b) working on the analytics to determine almost a real time performance of all the educational institutes on various parameters. This will also help in determining which interventions are leading to what all benefits.

In addition, iHub-Anubhuti will address following problems to advance the state of Indian Education System.

- Skill training: iHub-Anubhuti will make the vocational training feasible and attractive for the school children. It will employ Augmented Reality (AR) and Virtual Reality (VR) as a medium to train students with vocational skills at the K12 level to cater to the highly dynamic future of work and prepare quality skilled workers for the country.

- Efficient digital learning: There are numerous teaching / learning / MOOC platforms available for upskilling such as NPTEL Swayam etc. These platforms offer a variety of course content according to learner's ability, interest, technological innovation, demographic context of language, etc. However, it is difficult to find the right content using regular search techniques. iHub-Anubhuti will work on MOOC video summarization to enable the users from different backgrounds, with different languages/dialects, different stages of a skill/domain find it difficult to jump to the right content for learning. Via another project, it will also improve digital illustrations by focusing on using digital multi-media as a tool for enhanced learning experience by means of animation and interactivity. It will propose to design an algorithmic framework to enable creation of interactive and animated educational illustrations for eBooks (e.g. for NCERT). A number of software tools for sculpting and mesh editing in

  3D attempt to build better interfaces to work in 3D. Most of the tools mentioned above lack collaboration in creation and modification of 3D geometry. By developing collaborative training and design methods in 3D spaces iHub-Anubhuti will address limitations of existing systems and build a system for collaborative shape creation and editing framework usable in Mixed Reality environments. Our system will offer multiple users to share virtual space and collaborate in real-time, thus enabling a host

of applications including training, remote assistance, and collaborative creation. Consistent with the overall aims of improving socio-economic equality via computational means, the objective of this project is to build artificial intelligence (AI) enabled educational apps for mobile devices in an effort to digitize and democratize pre-primary education. Many instructional apps aim to teach pre-school children everyday concepts, including objects, shapes, colors, numbers, and alphabets. From our review, the existing apps cannot evaluate what the child has learned (except mathematics apps). Effective learning, however, requires a cycle of instruction-evaluation-instruction. We propose to develop a complete learning experience by expanding AI for evaluation beyond mathematics.

- Engaged and personalized learning: iHub-Anubhuti will work on developing personalized learning frameworks which will address the problem of deep analytics of learning contents, including automated meta-tagging of concepts, alignment to curriculum standards, and finding similar content. AI can make these learning contents smarter by associating metadata related to key concepts discovered. This includes learning standards and assessment alignment, complexity and comprehension burden, content suitability, learning perquisites, open references etc. This metadata can be discovered through learning content analytics to establish its linkages to the learning standards and other datasets. By discovery of metadata and its association with the original learning content, AI creates self- describing cognitive content that allows learners to more effectively grasp the underlying concepts and achieve their expected learning outcomes. As learning systems collect both structured (learner demographic, attendance, assessments, engagement etc.) and unstructured (observed behaviours, social learning data, feedback, learning standards, usage of content) data, it will become possible to construct learning models to estimate progress against expected learning outcomes, specific areas of interest, as well as particular areas where they struggle. These models, together with the learner's social context at the time of learning, will then be securely stored as records of the learner's life-long learning profile. Based on these learning models, AI-powered learning systems could aggregate and organize the learning contents more accurately to assess learning progress, identifying the learning or skills gap, design more individualized learning plan/pathways tailored to each learner, recommending the customized learning contents, provide timely guidance and explanations through augmented reality and/or virtual reality, and adapt the contents based on the observed learning behaviour and content complexity. By recommending how to read, understand, and approach concepts based on learning models and developing unique and personalized learning pathways, learners will receive a more effective and stimulating learning experience.
- Inclusive education: SEN children were struggling pre-COVID 19, and are in dire need of solutions to make their learning more effective, personalised and outcome oriented post COVID. There is still no single tool that is available to measure and monitor attention and thereby enhance online learning experience for SEN children. iHub-Anubhuti will enable online education for SEN children by creating a closed loop feedback system that monitors and measures attention; assists children in self regulation and enables teachers to take corrective measures continuously. Considering the learning aspects, cognitive ability and performance capability of the individuals with autism LD and intellectually challenged using computer, tablets and I-

pad as a medium some selected games can be played with them. Focus can be laid on the performance based skill development of the individuals with autism, LD and intellectually challenged using digital games. [OE1]. Most schools in India don't cater to children with learning disabilities, and teachers are not trained to deal with this glaring issue. The schools have no kind of special programs for these differently abled children. Few private schools that may provide special education often charge enormous prices which are not affordable for everyone. Educators and many analysts found out that as schools in India have become more and more competitive, the emphasis on textbook studies has increased and not enough attention is given to other skills. Schools might simply ignore children with such learning disabilities as misbehaved or hopeless. Unfortunately, India still thinks of dyslexia as a 'mental illness' rather than considering it a learning disability. The Indian government passed a detailed disabilities law that guaranteed job quota, rehabilitation and houses for people with visual, hearing, mental or physical disabilities. However, no learning disabilities were mentioned. Now is the time to change the attitude which the society in India has shown towards those with learning difficulties including dyslexia. And with the growing technology, we can tackle this issue without the need to invest exorbitant amount of money.

Within the Education vertical, there are four main themes to investigate that will provide competitive advantage in terms of enhanced education, building support systems for special educational needs, and creating software/hardware infrastructure for continuing education and training.

Within the purview of strengthening digital education and providing enhanced understanding of educational content we will utilize extended reality based immersive technologies. AR/VR is a promising technology for the future of education. Building interactive content creation tools will empower both content creators and educators to develop better educational material. Developing such enabling technologies will provide enhancement to illustrative education. Further AI-based automation will be pivotal in content creation, retrieval and organization. Systems that cater to learner specific needs will be able to provide precise content for focussed learning. A large amount of digital educational content is in the form of online MOOC courses that includes video content, however retrieval of specific content or topics could be a very time consuming process. Building systems that use AI to automate learner specific content retrieval and summarization will enhance access to the existing educational data. In addition to this, understanding the cognitive load in education is an important aspect in learning. Such an analysis will allow for better understanding of current education delivery systems and tools. Developing hardware and software for the same will provide understanding of current educational needs and pave the way for designing better learning systems. While indicates that there are 300+ AI related start-ups in India, and 11% of them focus on education, according to recent reports like 3 there is not a single finished product. There are many successful online education start-ups in India, such as Byju, Unacademy, Toppr etc.; however, they simply impart video lectures and conduct online exams. These companies are definitely built on cloud and IoT technologies but are not based on artificial intelligence. Moreover they target a much higher age range (middle school onwards).

Developing creative systems for children with special needs is a priority today. These children

require more attention and effort for their cognitive development. Presently the teachers and parents need to spend a lot of time with these children, however with current technology it is possible to develop hardware and software tools that can provide more engaging and customized interface to training and education. Further such tools need to be specific to the needs of the children. For instance, children with dyslexia can benefit a lot fromspeech tools built on state-of-the-art AI speech synthesis technology. Another important classis of children with Autism Spectrum Disorder (ASD) where the training itself becomes very difficult given the wide range of spectrum. Worldwide research on AR/VR technologies have shown promising results compared with classical training techniques. Developing digitalgames for training is a powerful technique to reinforce learning in such children. Similar is thecase with children exhibiting Attention Deficit Hyperactivity Disorder (ADHD). It is critical to create systems to advance understanding of behaviours of such children.

Training and continuing education is the third important area where technology can play a pivotal role. With current AR/VR technologies it is possible to create remote collaboration systems and telepresence capabilities. These powerful technologies are comparatively much better than existing video-based training techniques. 3D sensing and augmentation of ambient space can enable immersive remote learning. Such a virtual infrastructure is highly relevant in many aspects such remote education, collaborative learning and training, and remote assistance. In fact, the possibilities go beyond education and find applications in health, military simulations, disaster mitigation, etc. In addition to this, vocational skilltraining can benefit from AR/VR. Such systems will not only reduce the cost involved in settingup vocational training facilities in schools while letting students learn by doing in a virtual environment.

The fourth important area is the ethical implications of cognitive computing systems, especially in the field of education. Existing AI systems, very rarely, if at all, consider the ethical implications of their decisions. Therefore, it is very important to design, develop, and incorporate the ethical framework into the decision making of these AI systems. There are a few ad hoc solutions, but a methodical study and a mechanism to compute the ethical score and make decisions based on that are missing. Apart from that, the developed technology will be ethically compatible and will be compatible with societal and legal regulations. Since the technology will undertake the needs of society and sensitive to social needs it will add value to society.

Grand Problem 2: Designing Cognitive Computing & Social Sensing Solutions for Healthcare

The practices being followed in modern healthcare often need improvement with collection and analysis of data from humans and devices (social sensing). The collected data for healthcare not only needs novel algorithms or cognitive computing methods but an integrative approach for solving the problem in a holistic manner. Such a holistic cognitive computing approach is needed in multiple sectors of healthcare and most importantly in public health especially in the context of maternal and child health. Besides the public health, there are several areas, e.g., mental health, personalized therapy, non-communicative diseases, etc. that will benefit from cognitive computing and social sensing solutions.

Similarly, healthcare of elderly population at home needs infrastructure based on the IOT (internet over things) enabled social sensing approach as well as strong AI enabled cognitive computing systems to predict the risk of disorders. In addition, there are multiple possibilities of using a solution oriented cognitive computing approach on existing data in public as well as with collaborating health institutes. Such sources of data could be social media portals, schemes of government like eSanjeevani and national health stack and pre-existing genomic data for liquid biopsy. There is a need for a context specific and adaptive information processing approach, which utilizes previously known information and dependencies among variables in the dataset for better insights and disease prognostics and therapeutics.

In the Health domain, we aim to solve following challenges in the domain of public health:

Public Health Data Collection, Curation, Analysis and Modelling:

Currently most of the healthcare data is scattered and are in silos and every data-set has its own format and feature. It is assumed that information in healthcare data silos has the potential to reveal many important trends which could be used to save lives. However, we can not deny the fact that there is a need to collect more healthcare data directly from individuals and patients as there could be new questions and new opportunities. There are 3 possible solutions to handle current problem of healthcare data silos. The first solution is to collaborate with organisations keeping data in isolation from others, the second solution is to collect new data with help of IOT systems and instruments in clinics and hospitals. The third solution is to explore data with a limited number of parameters, made available due to a few initiatives by governments such as eSanjeevani and social media. However there could be multiple other kinds of healthcare data-sets which are not recorded yet. Hence, there is a need to add a new dimension in the data being collected for healthcare.
The non-availability of experts and well-equipped clinics in rural areas as well as scarcity of hospital beds during pandemics highlights the importance of home-healthcare. However to achieve sustainable and reliable home-healthcare in India, we need a multi-disciplinary approach such as: wearable technology and IOT system for data-collection, computational method and infrastructure to detect anomaly and risk of disorder, data security and search engines.

Efficient methods to collect, store and analyze on multitude of devices, e.g., mobile phones, tablet, laptop, are required. A good amount of data collection is manual and through paper forms. This leads to delay and human errors. Moreover, the data is then further needed to be stored in a digital format leading to more delays and errors. There is also no standardization of data collection and thus making it almost impossible for further analysis. New technologies to directly collect data in digital form, check for errors, verify the authenticity (e.g., with GPS triangulation), and storing in a standard format making it available for sharing, analysis, and report generation is essential. Analytical methods for public health data needs to be developed to generate insights into different public intervention, their effectiveness, and also for deciding future directions to further improve public health. Also, across India, different states implement different public health initiatives under the National Health Mission. As India has a large migrant population, there is a need to provide a cloud-based platform for making the services accessible even when people move from one state to another. Most of the migrant population is low-literate and does not have enough information about the support schemes available from the Government of India. Making this information available

and helping them in accessing the services while collecting vital health data would lead to improvement of health services across the country.

The data collection can be justified only when there is a system to analyse it for the purpose of making fundamental discoveries and prediction tools and solving problems of society. Especially there is a need for more comprehensive and holistic approach to process data-sets so that multiple hypotheses could be formulated and tested.

Intelligent Training and Learning Solutions for Community Health Workers (CHW):

One of the challenges in providing equitable access to proven life-saving interventions to end preventable deaths is the acute shortage of health workforce in high burden countries. Data for 2013–2018 showed that almost 40% of all countries have fewer than 10 medical doctors per 10 000 people, where 90% of low-income countries suffer from such shortages compared to only 5% of high-income countries. Further, according to the estimates of the WHO Global Strategy on Human Resources for Health: Workforce 2030, it is predicted that by 2030, there is going to be a global shortfall of 18 million health workers, primarily in low-income and lower-middle-income countries.

To address this challenge, a global strategy is to increase the capacity of the existing health workforce through "task-shifting" approach. This involves rational delegation and redistribution of tasks among health workforce teams. Specific tasks where appropriate are shifted from highly qualified professionals to less qualified workers to make efficient use of available human resources for health. In this context of task shifting that community health workers are gaining recognition. WHO defines CHWs as "Members of the community,selected by and answerable to the community they work for, and supported by the health system but with shorter training than professional health workers". Although the fundamental characteristics of the definition are universal, depending on program goals, variations may be found across and within countries in the roles, responsibilities, training, recruitment, and compensation.

Community Health Workers are the last mile actors of healthcare who serve communities often vulnerable populations towards accessing health services and adopting healthy behaviors. In particular, their role in reducing mortality rates for maternal and child health has been significant. Thus, increasingly more programs are aiming towards innovating methods to empower CHWs at the global level. Particularly, empowerment through technology integration in the ongoing endeavors has been an important direction.

Since CHWs often work in resource-constrained settings, whereby, infrastructure and technology penetration is limited, efforts have been towards making simple, low-cost and effective solutions. For instance, because of the high use of phones by CHWs, mostly feature phones, a rich body of work surrounds designing mobile-based solutions. By providing assistance through mobile applications, researchers focus on furthering CHWs performance which gets affected due to factors including limited supervision, education, and training. For example, mobile applications e.g. SMSs, IVR, e-checklists have been used in improving CHWs adherence to house visits and case management guidelines. In addition to supporting supervision, data collection by CHWs has been facilitated. Community education has been highlighted to be one of the most challenging tasks for CHWs. Studies have found a number

of causing factors. For example, in persuading families CHWs face resistance due to low literacy, compliance with old age traditions and practices. Due to minimum focus on the role of activists by existing training and supervision, CHWs remain unclear on how to execute the role of counsellor. Further, Limited, irregular payment of incentives and overburdened duties have been reported to negatively affect CHWs motivation towards counselling job. In this regard, use of mobile multimedia is increasingly being investigated in multiple ways because it offers an effective method of equipping CHWs with authentic health information which they can use anytime and anywhere in the fields.

Shortage of skilled trainers is a known challenge in upgrading skills of CHWs on a regular basis. A number of studies and projects have thus attempted to focus on using technology-based solutions to provide better learning opportunities for CHWs.

In India, Accredited Social Health Activists (ASHA) are the largest cadre of Community Health Workers and are the frontline health workers in India. The associated ASHA program was launched by National Health Mission in 2005 to achieve public health agendas in vulnerable communities such as slum and rural areas. An ASHA is a woman selected by the government to serve a population of 1000. The recruitment criteria include the age range of 25 to 45, minimum education of eight grade, leadership and communication skills. The responsibilities of an ASHA include acting as a promoter of preventive health practices, service provider, and healthcare facilitator. Under maternal and child health, ASHAs activities include visiting pregnant women, counseling on anemia prevention, escorting for ante-natal checkups and deliveries, providing post-natal care, monitoring danger signs of the newborn, immunizations, counselling on birth spacing, and family planning. The ASHA network is primarily made of women recruited form the community. While this provides easy access to the community, the poor training of ASHAs have been singled out as a major reason for inefficient execution of services on the field. The inadequate training is a result of lack of infrastructure, mobility, and lack of experts. WHO suggests that to enable CHWs function successfully, they must receive regular training and supervision. Computing solutions to provide distance-based training to community health workers are needed to support their growth and efficient execution on the ground. There is a need to develop personalized solutions that are available anywhere, anytime to provide the training. Use of Chatbots in native languages, or other AI-enabled solutions that can run on mobile devices may help in filling the current training gaps.

Intelligent solutions for Maternal and Child Health:

Maternal and child deaths have declined remarkably globally, however, they remain high in many low- and middle-income countries (LMICs), including India. Of the 5.1 million under-five deaths in 2019, 824,448 and of the 295,000 maternal deaths in 2017, 35,000 occurred in India. India is among five countries of the world that contributed to nearly half of the deaths. Poor maternal and child health continues to be a major public health concern.

To reduce deaths, women need good-quality reproductive health care, effective interventions, and skilled birth attendants – and the autonomy and resources to demand and secure access to these. However, there are high disparities in the availability and utilization of health services across different contexts. Under-resourced settings are the most affected. In India, broken links with the health system and shortage of human resources have been the

bottleneck problems. Women and children continue to die from leading preventable causes due to lack of access to simple and affordable life-saving interventions. Developing novel cognitive computing and social sensing solutions, using Chatbots to support Q&A solutions in native languages may help in providing vital information leading to health gains.

In recent year, HCI research exploring the role of digital technologies for women's health has increased, particularly covering different aspects of motherhood. Often new mothers face isolation in their initial phase of motherhood and suffer from depression, there is need to provide cognitive computing based solutions to support mental well-being of mothers to improve mother and child health.

Outside India, public health services are also showing interest in leveraging technology to support maternal and child health. For example, Public Health England has launched a chatbot application under the program—Start4life over Facebook Messenger and Alexa platforms to provide 24/7 support to mother for breastfeeding support. Tools are being developed to enhance communication between mothers and hospital staff. Also, commercially, there is an increase in applications targeting mother- hood needs. A simple google or app store search on motherhood leads to various options.
While the above-mentioned works are towards women who are digitally literate and belong to urban settings of the developed world, there are others from marginalized groups, for whom we have limited understanding. Recent affordances to smartphones in developing nations, have made women in under-served contexts to start owning smartphones and explore the digital world, which earlier was a remote possibility for them. These women represent an important class of users for Coginitive Computing and Social Sensing based technology solutions.

Intelligent Sensing Solutions for supporting Severe Mental Illness (SMI):

India suffers from a poor ratio of psychiatrist to people leading to severely impacting medical support for growing mental health problems, especially for severe mental illness like Bi-polar and Schizophrenia. The sensing platform provided by mobile devices and wearables combined with intelligent algorithms can help in detecting possible occurrence/re-lapse a SMI condition. We will be working with AIIMS, India to develop intelligent Sensing based Cognitive Computing solutions to support Severe Mental Illnesses.

Middleware solutions for efficient data distribution under intermittent network conditions:

The community health network usually serves in low connectivity regions and they usually possess low-resource computing devices such as feature phones. Therefore any data distribution faces challenges in efficient dissemination. There is a need to develop efficient Middleware solutions that can enable reliable data dissemination in such environments.

Intelligent public health and preventive medicine:

it involves use of emerging technologies such as Artificial Intelligence (AI), Augmented reality (AR), wearable sensors and telemedicine for prevention of infectious, non-infectious and mental health diseases in the community. Few project leaders in our hub are following the theme of social sensing and trying to handle the problem of collecting unique data-sets using sensors and wearable technology for different purposes such as Intelligent Physical Exercise Recognition and Evaluation and integrated health monitoring system. Currently there is rarely any comprehensive integrated dataset for health from wearable sensors for Indian population. In addition, technology of sensors evolves, hence there is very less chance of overlap of parameters with other international studies. Thus, the approach of social sensing to collect synchronized and integrated dataset wearable sensors for the Indian population would prove beneficial in long term as well as lead to product development for commercial and welfare purposes. Most of the teams proposed collection of health data using the IOT system while few teams are trying to make comprehensive data-set of cancer-cell images for solving the problem of developing intelligent and affordable diagnostic tools for hematopoietic cancers. Other team's proposals under the sub-category of "intelligent public health and preventive medicine", are related to using AI and cognitive computing for providing solutions to few unsolved problems such as anti- microbial resistance and tuberculosis and other respiratory diseases.

Intelligent diagnostics and prognostics for healthcare

It encompases diagnostic and prognostic techniques using genomic, radiological, molecular, biochemical and histopathological patterns and associated computational challenges of processing corresponding data-sets for better results. In the current version, two teams in our hub are trying to use two different approaches for diagnosis and prognosis of Autism. While another team is trying to build AI-enabled Smartphone based platforms for Maternal and Child Health in Rural India. Similarly, a few teams are developing and using AI and Deep learning approaches for detecting anomalies in lungs and olfactory systems. One of the teams has proposed to use a holistic cognitive computing approach for diagnostics using liquid biopsy, especially for early detection of cancer using blood plasma. Timely detection of cancer is still a challenge world-wide and the cost of such preventive diagnosis for cancer is high.

Computational approaches to therapeutics

it includes challenges associated search and discovery of drugs, vaccines, therapeutic approaches including food and nutritional interventions. Here we have teams working on 4 different kinds of problems associated with therapeutics; Assessment of drug efficacy in Indian population, Computational Gastronomy Framework for Achieving Better Nutrition and Health, deep-learning based approach for characterization of functional heterogeneity in cancer and using technology for therapeutics of children with Autism Spectrum Disorder(ASD) and chronic lung diseases to facilitate pulmonary rehabilitation. The task of using a computational approach for assessment of drug efficacy in India is crucial as India is the second largest market for internationally manufactured drugs mostly developed for European or American populations. Using holistic approach like cognitive computing toresolve issues of drug-efficacy in India has some unique features.

Health Data Analytics

It includes analysis of existing data-sets with new algorithms and technologies for analytics, cognitive computing, AI. It also includes use of virtual reality and data-visualisation so that hypothesis oriented projections can be made to understand any trend. Several teams have proposals under this category where they are trying to use data-sets which are either publicly available or being collected by their collaborators. Such as one of the teams have proposed to develop solutions for storing large healthcare data-sets collected by IOT systems or clinics in abstract (portable on laptop) and searchable format so that the physiological profile of a potential patient could be matched to millions of previous disease cases. While one of the teams is working on detecting mental disorders using social media posts, the other team is trying to use cognitive approach for counselling psychiatric patients on chatbot. One of the teams has proposed to use AI and integrative approach to contribute towards telemedicine service platform eSanjeevani of government of India by Tele-triaging of Rare Genetic Disorders. Given that eSanjeevani is poised to become one of mainstream free healthcare services of India, contribution towards it through computational work has a national importance.

AIMS AND OBJECTIVES

4. Aims and Objectives

Our aim is to support research studies which use a holistic approach to solve problems in the domain of Cognitive Computing and Social Sensing with a special focus on the two grand problems. As part of the TiH, the aims are to develop fundamental deep-tech research solutions and innovative translational technology that can lead to start-ups and other innovative activities. In order to achieve our goals of supporting different teams working on current relevant problems, we also aim to build computational infrastructure. Thus, our objectives are as follows.

Fundamental research

The TiH is committed to support project doing fundamental research in the Cognitive Computing and Social Sensing domain and with special focus on the challenges that are to be addressed in the context of the grand problems. The objective would be to develop research and technical solutions that address the fundamental problems in the target domain and towards solving the grand problems.

Platforms for Computing and data sharing: We aim to support teams for developing channels and pipelines for data and provide them a common platform to store and host the data so that it can be used by different researchers and industry for diagnosis and prognosis of diseases. This computing and data platform will be hosted at IIIT-Delhi and will provide a platform for researchers, national and international, to conduct research and development activities in the domain of Cognitive Computing and Social Sensing and towards targeting the grand problems.

Product development and Computational support

We would support teams working towards development of products through support for prototype development and technology transfer using the incubation hub and accelerator at IIITD

Knowledge exchange and training

We would support teams working on techniques for better training and dissemination of knowledge for healthcare, such as training to reduce child mortality rate and improvement of nutrition for children and older population.

Skill development for encouraging new start-ups: Our effort would be to provide support for exposing the young generation towards new technologies related to health-care and finding their applications. Exposure and training on new technologies to the young generation also open possibilities for innovations which could lead to new start-ups.

Establish Collaborations

One of the major aim of the TiH-Anubhuti would be to establish a network of collaborators that will help us in identifying ground-level problems, field-test our solutions, and eventually adopting our technology. We will be seeking active partnerships with NGOs, Govt., and Industry to create an eco-system of Cognitive Computing and Social Sensing.

STRATEGY

5. Strategy

iHUB Anubhuti has a 2- pronged strategy:

1) Top Down Approach and 2) a Bottoms Up Approach.

1) Top Down Approach – This includes working on the two Grand Challenges in the areas of Education and Healthcare. These are the big challenges where the impact is going to be huge. iHUB Anubhuti will be pursuing projects and actively developing various modules around education and healthcare. We are actively pursing collaborations with Ministry of Education and also partners in healthcare, e.g., AIIMS (New Delhi and Rishikesh) and PGIMER, Chandigarh, and more to develop technological solutions and doing necessary research to support them in addressing major health challenges.

2) Bottoms Up Approach – iHUB Anubhuti is working on various projects in the areas of cognitive computing and social sensing around its domain. We have invited the call for proposals and the projects were selected under the categories of (i) Predict mortality/ diagnosis/ diagnostic tools for patients suffering from cancer/ liver failure, (ii) Challenges in Public Health, (iii) Cognitive Computing and Social Sensing, (iv) Law Enforcement and Security and (v) Others.

   We are also going forward with other initiatives like Chanakya Fellowships, Call for start ups, etc. The overarching strategy of iHUB Anubhuti by way of this is to undertake unique activities that can catalyze data and intelligence transformation of the four key verticals, i.e., (i) Healthcare, (ii) Education, (iii) Environmental Sustainability (including smart cities), and (iv)Law enforcementand Security. The verticals chosen represent the major sectors of importance to India's visionto leapfrog into global challenges such as universal healthcare, effective education, safety in pervasive computing and smart cities. iHUB-Anubhuti has a strong emphasis on interdisciplinary collaboration, therefore the shortlisted projects will be encouraged to exchange technology and insights among each other to create a multiplier effect within the hub. A major strategic advantage of interdisciplinary emphasis is to avoid the creation of silos.The technologies developed in cognitive computing and social sensing verticals can be easilyexchanged and such cross-domain collaborations will be fostered in the hub. Many of the technologies that are being developed are domain independent, i.e. these technologies can be applied in different domains such as education, ecommerce, health, space exploration, to name a few. Individually, the shortlisted projects have adopted several strategies for achieving a competitive edge, and these can be summarised into the following principal components:

   1. Timeliness. The shortlisted verticals and projects under iHUB Anubhuti are timely and in alignment with the national strategy developments. By taking a focused lead in these directions, iHUB Anubhuti will create a competitive advantage. For example, the objectives and outcomes of shortlisted projects in healthcare are aligned with the National Digital Health Mission. Similarly, education vertical addresses the new-normal in education ushered in by the COVID-19 pandemic. Projects shortlisted under environment, law enforcement and security also offer unique computing and social sensing solutions for pertinent issues such as waste management and smart cities.

2. Data and Domain Expertise. Projects in each vertical have a strong focus contributing to making the respective verticals more data-rich through smart technologies and cognitive sensing. Many projects have proposed creation of novel datasets in public health, smart waste management, personalized education platforms, multimodal data fusion and knowledge graphs. The applicants of shortlisted proposals have a rich experience in data technologies.
3. Deep Tech. iHUB Anubhuti will create technologies that are beyond the state-of-the-art in the respective verticals. This is exemplified by the shortlisted proposals contributing to deep technological development beyond those currently available, such as energy efficient blockchain, and AI/ML architectures, novel algorithms for extracting structured information from raw data, scalable applications in healthcare, 3D Augmented Reality, Multi-object Tracking for vehicles etc. The focus on innovating in niche areas such as mobile computing and embedded computing for cognitive computing and social sensing is expected to be the first of its kind in India at such a scale.
4. Interdisciplinary collaborations. iHUB Anubhuti is encouraging cross-institutional and cross-domain collaborations. Nearly 80% of proposals are multi-institutional. This is critical to our strategy of creating a competitive edge by setting a higher bar of delivery within the verticals. For example, many of our healthcare projects are aligned to leverage and advance the delivery of primary care in India, a key component in the National Digital Health Mission. By developing interdisciplinary collaborative partnerships with universities from around the world, researchers will strive to leverage the expertise of faculties and scientists from different complementary disciplines.
5. Entrepreneurial edge and Industry partnerships. The projects undertaken by iHUB Anubhuti are encouraged to show a clear path to translation at scale in the real world. By forging strategic partnerships with the industry, iHUB Anubhuti aims to develop technologies that meet the industry needs and standards, and will provide them an opportunity to test these technologies in real world environments. Such partnerships will also open up entrepreneurial opportunities to monetize such technologies and create revenue flows.
6. Having a clear strategic management system. iHUB Anubhuti will take an outcome driven approach. The probability of successful outcomes will be increased by a clear strategic management plan that will include
    a. Selection and mentoring of projects to be aligned with the vision and mission of the hub
    b. Regular communication and appraisal of deliverables and goals.
    c. Planning for real-world deployment and scale-up of cognitive computing and social sensing solutions.
    d. Provide early feedback to projects through availability of mentors and experts.

TARGET BENEFICIARIES

6. Target Beneficiaries

The proposed TiH has identified few target domains besides the grand problems: 1) Healthcare 2) Legal System 3) Education 4) Law-enforcement & security, 5) Smart Cities. The target beneficiaries of each domain are discussed separately.

1. Healthcare:

In the domain of Health, there are project ideas related to mainly three categories, projects that analyse cell/gene data to detect different types of cancers, projects which propose to build a Chatbot to enable individuals get their queries related to various medical problems answered, projects that propose to collect individuals data through sensors and IOT technology to be analysed for various purposes. The various stakeholders identified are:

1. Patients and General public
2. Doctors
3. Medical Service Providing Organizations such as hospitals, pathology laboratories
4. Pharmaceutical Industry
5. IT Companies
6. Government organizations and NGOs
7. Biomedical Researchers

As of now, we have identified following Industry, academia and NGOs for collaborations as well as implementation of various proposed projects:

- Hospitals: All India Institute of Medical Science (AIIMS), Apollo Hospitals Enterprise Limited, Medanta Hospitals, Indian Spinal Injuries Centre, PGIMER-Chandigarh
- CDAC: e-Sanjeevani
- NGOs: WISH Foundation, Action for Autism (AFA), Genetico Research, SWACH Foundation, SWASTI
- Industry: CareOnco Biotech Pvt. Ltd., Circle of life, Google Research India, Elucidata, Persistent
- Diagnostics Companies: Lalpath, SRL, Dr. Reddy etc.

2. Legal System:

We will be engaging with Ministry of Justice to provide technologies for supporting the legal system. The technologies developed will help the different stakeholders engaged in the legal eco-system. Below is the list of beneficiaries in the Legal Domain:

1. Police Systems: With technologies to help police in analyzing data available in complaints, FIRs, records, judgements, etc., we believe that our technologies will directly benefit the Police systems.
2. Advocates: Today, the search among the judgements is manual and therefore requires significant time and effort, our technologies directly benefit advocates and other paralegal personnel.

3. Judicial System: Today, there are many pending cases because of paucity of judges at all levels. This often leads to justice being delayed for several years. The technologies developed by our TiH in close collaboration with the Ministry of Justice will directly benefit the judicial system.
4. Public: The awareness of law and judicial system is weak and fragmented in the public. Currently, there is no public system that helps an individual in searching and identifying relevant judgements that may benefit them. Our solutions will fill this gap.

3. Education:

The target beneficiaries of the projects in the Education domain are principally educational planners and educational policy makers who want to improve the equity and quality of their education system. Key beneficiaries are the poor or weaker section of society who cannot afford offline teaching aids. The Internet is now available everywhere with affordable cost. Using online teaching, it will cost very less as compared to offline teaching and also save the cost of transportation. The other main beneficiaries are learners whose chances of getting quality education and its benefits will be effectively enhanced. Other potential stakeholders of the projects will be AR and VR headset manufacturers like Oculus VR, HTC Vive, Microsoft who would like to step into the market of developing newer and cheaper hardware. Other beneficiaries are the students who need special assistance. There solutions like EEG devices can be used for stress monitoring for students who are facing stress due to their studies load, lifestyle or environment. These students can be provided interactive content created using AR and VR technologies. Below is a list of beneficiaries in the Education domain:

1. Students: who are in their schools or colleges learning towards a goal such as attaining better higher education, vocational skills or job. It also includes students that have special needs, like ADHD, Dyslexia.
2. Teachers and Content Creators: who are skilling themselves for teaching a new technology, doing research for Ph.D. or for funded projects. The solutions would enable a teacher to prepare multi-model and interactive content.
3. Regulators: Regulators need real feedback from the learners, so that they can improve their policies, platforms, and schemes accordingly. Due to the huge varieties of learners, it has been difficult to address those issues till now.
4. Institutions: Institutions need to understand the gaps and opportunities in their curriculum design, pedagogy, and delivery.

As of now, we have identified following Industry, academia and NGOs for collaborations as well as implementation of various proposed projects:

- Google
- Facebook
- Microsoft
- Walmart
- IBM
- Optum
- Asian Paints color academy
- Special Needs Technologies

- ITI arab Ki Sarai,
- ITI Malviya Nagar
- NSIC Okhla
- ShikshAI

4. Law Enforcement and Security:

The societal impact of the projects in the Security domain is quite straightforward. There are ideas proposed that will ensure that the benefit of social sensing technologies are more equitably distributed across various demographic groups within the population. The main target groups for uptake of the algorithmic advances would be government agencies and third sector groups (NGOs, think tanks etc.) who would be able to influence policy. Companies in the social sensing space may also be interested in embedding the proposed technologies to ensure that their products are morally neutral, or at least not evidently unfair. The work will benefit NGO/Govt. Organisations/start-ups working in Public Health, education, environment etc. Many of these areas are part of Sustainable Development Goals (SDG) and have a direct impact on society. The ideas related to countering fake news will help all responsible citizens, particularly those who want to get and disseminate true information. It will be helpful for journalists, bloggers, and social media influencers who can truth-check before writing an article.

As of now, we have identified following Industry, academia and NGOs for collaborations as well as implementation of various proposed projects:

- Daksh
- STMicroelectronics
- VVDN Technologies

5. Smart Cities:

In the domain of Smart Cities, there are project ideas related to mainly transportation, air/soil/weather sensing and smart city. For all these projects, citizens, policy makers and government agencies are the main target beneficiaries. Smart city projects have significant potential to enhance tourist experience thereby helping the tourism industry. In the same way, projects related to the environment such as soil, weather and water sensing can help the farmers to make informed decisions so as to improve the quality and yield of their crops. In fact, the policymakers who decide insurance premiums and other input subsidies as a function of soil quality can be one of the potential beneficiaries.
As of now, we have identified following Industry, academia and NGOs for collaborations as well as implementation of various proposed projects:

- Awidit Systems Pvt. Ltd.
- National Highways Authority of India (NHAI)
- Ministry of Road Transport and Highways, Government of India
- CSIR-Central Road Research Institute (CRRI)
- VVDN Technologies
- STMicroelectronics
- Varcas Automobiles Private Limited, Hyderabad, Telangana
- Centillion Networks Private Limited, Hyderabad (Solar Cells)

LEGAL FRAMEWORK

7. Legal Framework

For each project idea, we will perform the requirement analysis in collaboration with the different stakeholders. For example in the Education domain it will include parents/legal guardians of the children, child psychiatrists, psychologists, pediatricians, and the residential care facilities. The requirement analysis would help us to identify Ethical and legal issues that would be shared with the institutional review board (IRB) of IIIT-Delhi in collaboration with representatives of these stakeholder organizations and accordingly, each project will be designed, planned, and executed. Each of the projects will also address ethical issues regarding study design, inclusion/exclusion criteria of the participants, recruitment of the participants, participants' involvement, data protection, risk and benefits, compensation, conflicts of interest, and research ethics in terms of suitable methodology and use of gained data. Ethical conduct and guidelines will be written to be adhered to by the IRB as a subject of regular discussion and discourse within the consortium.

Regarding projects that would perform analytics on publicly available datasets, we may require access to specific datasets that are privileged and/or protected. Similarly, for projects where crowdsensing and social sensing is used to collect hyperlocal and user data, we may require to ensure user privacy. Depending on the traction developed with think tanks, potential lawmakers and law enforcement agencies, we will need relevant legal provisions for working with datasets/user data that are not publicly available. A suitable policy and oversight committee for data protection is planned to be implemented as part of the initiative.

For the projects dealing with medical history of patients, we will ensure that all the procedures and protocols of the study will be in compliance with the Declaration of Helsinki (World Medical Association, 2008) and "National Ethical Guidelines for Biomedical Research on Human Participants (2017)" provided by the Indian Council of Medical Research (ICMR), and approved by the IERB of IIIT-D, UoA and FMRI.

ENVIRONMENTAL IMPACT

8. Environmental impact

Most of the ideas to be taken up are related to soft computing in nature and thus, there will not be any environmental impact including in terms of

a) Land acquisitions
b) Environmental clearances
c) Forestry clearances
d) Wildlife clearances

The programme is structured to be operated at iHub Anubhuti where already built in building infrastructure is available. Thus, there are no new constructions, land acquisitions, dislocations and other environmental clearances are not required.
Moreover, the technologies developed will help tackle some of the important environmental issues.

TECHNOLOGY

9. Technology

For research and development supported by iHub-Anubhuti, a variety of technology will be used. For example, for social sensing we are relying on several technologies which can be grouped in different categories according to applications Internet of things (IOT), Smartphones, wearable technology: These are supposed to be widely used technologies in near future. However, they still need fine-tuning and testing, especially for our projects.

Augmented and Virtual Reality: Augmented reality is one of the most promising digital health technologies. The reason is the possibility of being more immersive and convincing as human eyes are trained for 3D visualization. Some possibilities with AR are: Patients describing their symptoms better through AR, Surgeons taking help of AR for surgery, physical training at home could be made more lively by AR and better visualization of molecules with augmented reality.

Video/Image processing: We will be using state-of-the art technologies and will also develop novel technologies and solution for different projects supported by the TiH.

AI, ML/Deep Learning methods: For developing cognitive computing solutions, existing state-of-the art algorithms and methods and new methods in the AI, Machine Learning, Deep Learning domains will be developed and used among different projects.

Middleware System: For storing, accessing, and retrieving data from multiple sources, a slew of middleware solution and technologies will be developed to enable data collection and accessing them.

Storing and managing structured and unstructured data: The volume and heterogeneity of data often creates a challenge of storing them so that they can be retrieved easily by APIs. There are some known frameworks for unstructured and semi-structured data-sets such as noSQL and markup language XML. However, they do not guarantee real-time access and search for different types of data collected through IOT systems or clinical devices. Hence our teams would also be exploring and investigating different ways for managing and storing data in a context specific manner.

Stream and Event computing: Streaming data coming from wearable devices or clinical setup pose a challenge of real-time analysis. There are a few data-independent Hashing Algorithms for Time Series such as SSH (Sketch, Shingle, & Hash). However, without contextualization and tuning by experts such methods also do not perform well. Hence, we would be exploring such methods as well as new ways to analyze streaming data which exploit graphs between the variables whose time-series signal is being streamed. One such example is sentiment flow analysis using streaming information on graphs of social networks to detect mental disorders. Another example is searching matching profiles for individuals using their physiological data collected through wearable devices for home-health care.

MANAGEMENT

10. Management

The “iHub Anubhuti-IIITDelhi Foundation” will be set-up with a committed funding of 100 Crore by DST over five years in the vertical of “Cognitive Computing and Social Sensing”. The management of day-to-day activities, monitoring of progress, and disbursal of the funds will be done by iHub Anubhuti. For the same a comprehensive monitoring system consisting of the following entities will be developed at IIIT-Delhi.

Board of Directors

Ranjan Bose
CHAIRMAN, IHUB ANUBHUTI
DIRECTOR, IIITD

Pushpendra Singh
PROJECT DIRECTOR

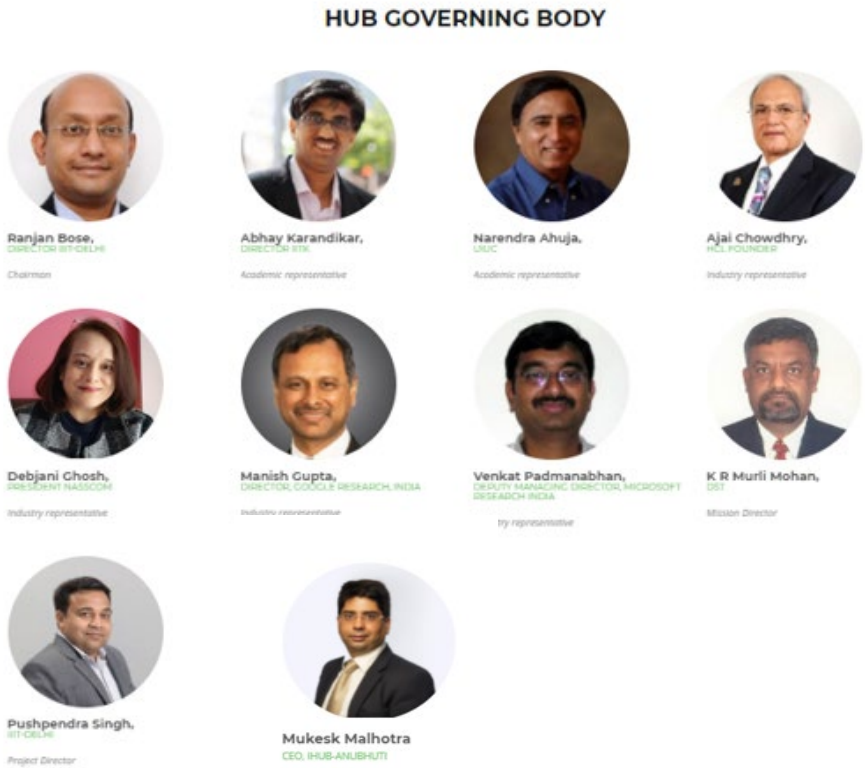
Tanmoy Chakraborty
PROJECT DIRECTOR

Vikram Goyal
PROJECT DIRECTOR

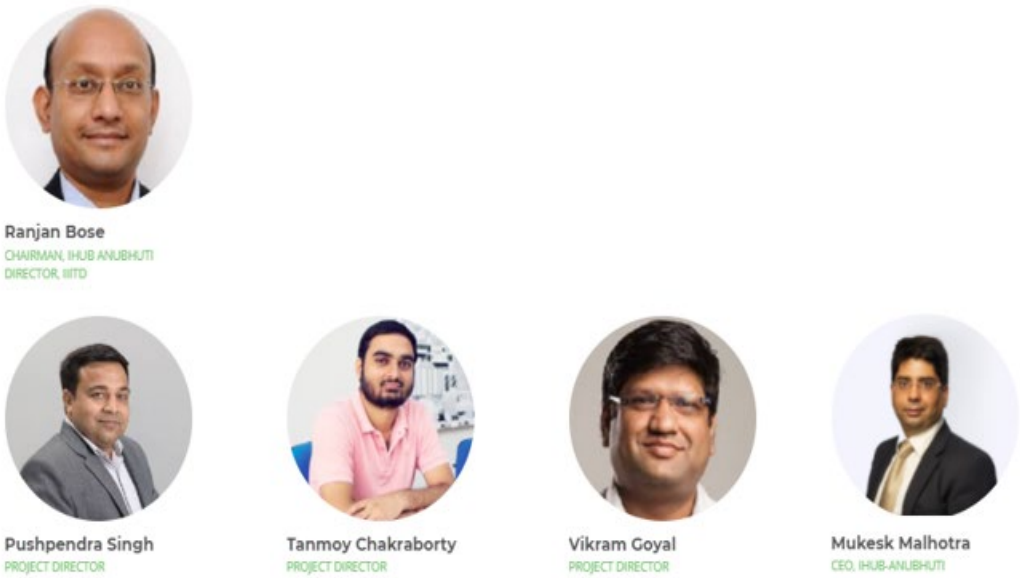
The Hub Governing Body (HGB)

The HGB, as advised by the DST, will be the apex governing body of the TiH and will follow the mandate of DST in all matters related to administration, legal, and financial. The Director of the IIIT-Delhi will chair the HGB and the composition of the HGB is given herewith:

Chairperson	Director IIITD (Ranjan Bose)
Academic Representatives	Abhay Karandikar, Director IITK Narendra Ahuja, Professor UIUC
Industry Representatives	Ajai Chowdhry, HCL Founder Debjani Ghosh, President NASSCOM Manish Gupta, Director Google Research Venkat Padmanabhan, Deputy MD, Microsoft Research India
Mission Director (or rep.), Mission Office, DST	Murali Mohan
Project Director	Pushpendra Singh
CEO	Mukesh Malhotra



iHUB Anubhuti-IIITD Foundation Management Structure



TiH Executive Committee for TiH (TEC)

The TiH Executive committee will be an internal committee, headed by Pushpendra Singh, external experts and IIITD Faculty members. The TEC will also have the CEO (Chief Executive Officer), Managers, and other staff members of iHub-Anubhuti once they are hired. As the

TiH progresses, the TEC may be expended with permission from the HGB. The TEC will be responsible for, including but not limited to, the following activities:

a. Reporting to HGB and DST: The ECE will be responsible for reporting the progress of any other information about the TiH to the HGB or DST, as and when, required.
b. Call for Proposals (CFP): The EC will come with regular CFPs during the tenure of the TiH to invite new proposals for the vertical. The selection of proposals will be done ina fair and transparent manner and on a meritorious basis. The decisions and all administrative actions related to projects will also be taken by the EC.
c. Call for Awards, fellowships, etc.: Apart from the research proposals the EC may also announce dedicated calls for fellowships, in addition to the projects, to fulfil the targets of the TiH.
d. Call for Startups/Incubation: To promote the entrepreneurial activities, the TEC may also announce dedicated calls for supporting start-ups in dedicated pre-defined areas,

to fulfill the targets of the TiH. The TiH will also work closely with the incubation hub of IIIT-Delhi to accomplish the above mentioned targets.

e. Industry liaison: Strengthening industry-academia relationship is another major activity for the TiH. With the projects working on the vertical of "Cognitive Computingand Social Sensing" under the four verticals, there is a strong need to engaging with relevant industry partners for knowledge exchange, transfer of know-how, and othercollaborative opportunities.
f. Administrative, financial, and legal issues: The TiH will also provide support for all administrative, financial, and legal aspects as and when they arise in executing the projects under the TiH.

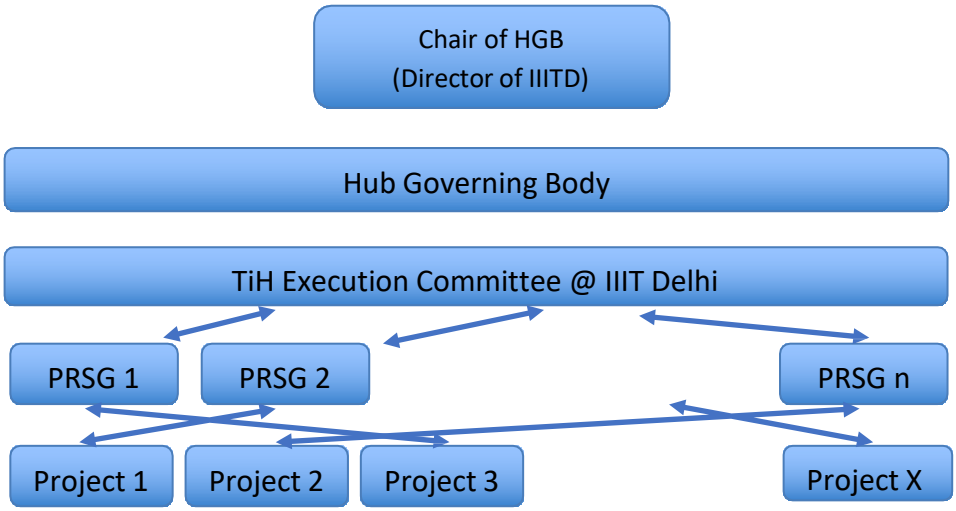
Project Review and Steering Groups (PRSG)

Each approved project will have an assigned PRSG committee for itself. The PRSG committee for each project will monitor the progress of each project every six month and will advise on the funds release depending on the project deliverables achieved as per the plan. The PRSG will consist of IIIT-Delhi as well as external members depending on the project and necessary expertise required to monitor the projects. The new funds release will only take place after the submission of a report by the corresponding PRSG of the project.

Project Teams

Each project team will be led by a Principal Investigator (PI). The PI will be the contact point and the nodal person for the project team. The PI will also be responsible for meeting the deliverables. The PI will also have to schedule the PRSG meetings and get regular reviews done in order to have a smooth release of funds etc.

A pictorial view of TiH Management is following:


Chair of HGB
(Director of IIITD)
Hub Governing Body
TiH Execution Committee @ IIIT Delhi
PRSG 1
PRSG 2
PRSG n
Project 1
Project 2
Project 3
Project X

FINANCIAL BUDGET

11. Finance

This section details the financial plan of the TiH complying with the Tri-partite agreement and the Budget allocated.
Budget release schedule as per the Tri-partite agreement:

Budget Head	Budget in Rs. Crore					
	1st Year	2nd Year	3rd Year	4th Year	5th Year	Total
Recurring	7.25	12.5	19.75	10.5	7	57.00
Non-Recurring	0.00	22	8.4	7.8	4.8	43.00
Total	**7.25**	**34.50**	**28.15**	**18.30**	**11.80**	**100.00**

Detailed Year wise Expenditure (in lakhs):

Sr. No.	TIH Areas		1st Year	2nd Year	3rd Year	4th Year	5th Year	Total
1	**Technology Development**	Recurring	50	300	300	220	100	970
		Non-Recurring	0	1840	555	420	250	3065
		Total	**50**	**2140**	**855**	**640**	**350**	**4035**
2	**Entrepreneurship Development**	Recurring	100	200	635	325	100	1360
		Non-Recurring	0	50	75	100	120	345
		Total	**100**	**250**	**710**	**425**	**220**	**1705**
3	**Human Resource Development**	Recurring	500	590	860	425	320	2695
		Non-Recurring	0	100	100	100	100	400
		Total	**500**	**690**	**960**	**525**	**420**	**3095**
4	**International Collaboration**	Recurring	15	30	50	50	50	195
		Non-Recurring	0	0	0	0	0	0
		Total	**15**	**30**	**50**	**50**	**50**	**195**
5	**TIH Management Unit**	Recurring	60	130	130	130	130	580
		Non-Recurring	0	210	110	60	10	390
		Total	**60**	**340**	**240**	**190**	**140**	**970**
	Gross Total		**725**	**3450**	**2815**	**1830**	**1180**	**10000**

Targets as per the Tripartite agreement:

S. No.	Target Area	Targets					
		1st Yr	2nd Yr	3rd Yr	4th Yr	5th Yr	Total
1	**Technology Development**						
(a)	No of Technologies (IP, Licensing, Patents etc.)	0	0	4	5	10	19
(b)	Technology Products	0	0	4	4	4	12
(c)	Publications, IPR and other Intellectual activities	3	5	8	10	10	36
(d)	Increase in CPS Research Base	3	5	10	19	25	62
2.	**Entrepreneurship Development**						
(a)	Technology Business Incubator (TBI)	0	0	0	1	0	1
(b)	Start-ups & Spin-off companies	1	2	5	10	14	32
(c)	GCC - Grand Challenges & Competitions	0	0	1	0	0	1
(d)	Promotion and Acceleration of Young and Aspiring technology entrepreneurs (PRAYAS)	0	0	1	0	0	1
(e)	CPS-Entrepreneur In Residence (EIR)	1	2	3	5	8	19
(f)	Dedicated Innovation Accelerator (DIAL)	0	0	1	0	0	1
(g)	CPS-Seed Support System (CPS- SSS)	0	0	1	0	0	1
(h)	Job Creation	50	500	1000	2500	3700	7750
3.	**Human Resource Development**						
(a)	Graduate Fellowships	10	50	50	50	50	210
(b)	Post Graduate Fellowships	5	10	12	12	0	39
(c)	Doctoral Fellowships	7	8	8	0	0	23
(d)	Faculty Fellowships	0	1	2	0	0	3
(e)	Chair Professors	0	1	2			3
(f)	Skill Development	10	50	100	100	120	380
4.	**International Collaboration**						
	International Collaboration	0	1	0	0	0	1

Detailed Year-wise Estimated Costs for Individual Key Areas

1. *Technology Development*

The "Technology Development" will be done primarily through research projects. The individual projects may be more than 40 in numbers and individual budget may vary from 20 lakhs to 5 crore depending upon the project content, duration, deliverables, etc. The manpower or international collaboration expenses can be supported from the budget of other relevant key areas, e.g., HRD. The computing requirements of the projects will be supported by the computing infrastructure of the TiH. Only small equipment may be allowed to be purchased from the project budget if felt appropriate by the evaluation committee. The budget may be asked under recurring and non-recurring categories.

TiH will ensure that there is a good balance of fundamental research projects aiming for knowledge creation and translational projects aiming for developing technology. At least 25% of the budget will be used for external projects. Each project will be encouraged to have industry collaboration or a field partner, e.g. NGO, govt. organizations, that will be ready to contribute either in cash or in kind (mentorship, use of facilities, technology deployment etc.). The additional contribution from industry or other agency will be over and above the budget approved by the TiH.

- The Technology Development is expected to deliver about 100 research publications and 20 patents over 5 years.
- Each project will be monitored by a dedicated PRSG which will recommend releasing money depending on the deliverables achieved and with overall progress of the project.

Establishment of Computing and Data Facilities: The establishment of the Data and Computing Facilities will be conducted under the Technology Development head. The computing infrastructure will also be used to generate revenue by renting out the computing infrastructure. The facility will provide all the computing facility for all the technical projects, start-ups, and external projects supported by the TiH. The manpower of the TiH will maintain the computing facilities and will help in providing infrastructure as service for revenue generation.

The following table gives the budget of manpower and equipment for the Technology Development.

Sr.No.	Budget Head	1st Year	2nd Year	3rd Year	4th Year	5th Year	Total (in lakhs)
Recurring							
1	Manpower	35	230	230	200	80	**775**
2	Contingencies	5	50	50	10	10	**125**
3	Consumables (Electricity, maintainence charges etc.	10	20	20	10	10	**70**
	Total (A)	**50**	**300**	**300**	**220**	**100**	**970**
Non-Recurring-Equipment							
4	GPU-Enabled Servers	0	1200	350	300	130	**1980**
5	Computing Servers, Storage,Other computing equipment, and lab infrastructure	0	500	150	90	90	**830**
6	IoT, Wearables, Mobiles, etc.	0	140	55	30	30	**255**
	Total (B)	**0**	**1840**	**555**	**420**	**250**	**3065**
	Gross Total Budget (A+B)	**50**	**2140**	**855**	**640**	**350**	**4035**

2. *Entrepreneurship Development*

The TiH will encourage and promote the entrepreneurial activities and start-ups emanating from TiH projects and in the vertical of "Cognitive Computing and Social Sensing". The IIIT-Delhi already has a Technology Business Incubator (TBI) and the TiH will work with the IIITD Incubation Center to support incubating activities related to start-ups. The TiH has its on 30,000 sq ft dedicated facility and the IIIT-Delhi Incubation Center has its own dedicated facility. Both of the facilities will be used to support the entrepreneurial activities.

a) The TiH will have a support for Technology Business incubation in addition to the existing TBI of the IIITD. The TiH will support manpower, travel, marketing and other related activities.
b) The TiH expects to support a minimum of 32 startups and spinoff companies during the tenure of 5 years. The selection of start-up for support will be done by the Executive Committee of the TiH. We assign a unit cost of 25 lakhs but the actual support may vary from 5 lakhs to 2 Crore as per the evaluation of the startup by the experts. The TiH will enter into a revenue sharing agreement with start-ups that it supports for revenue generation and sustainability.
c) The TiH will announce a conduct Grant Challenges and other nation-wide competitions to encourage entrepreneurship activities. The theme of the challenge will align with Sustainable Development Goals to encourage entrepreneurship activity for supporting a problem of social good in the vertical of "Cognitive Computing and Social Sensing".
d) The TiH will support entrepreneurs under the CPS-PRAYAS. While it is mentioned in the third year, it is only indicative and it could be anytime in the five years.
e) The TiH will also support at least 19 entrepreneurs in residence (CPS-EIR).
f) The TiH will support one dedicated Innovation Accelerator (CPS-DIAL). While it is mentioned in the third year, it is only indicative and it could be anytime in the five years.
g) The TiH will provide support for 1 initiative under seed support system (CPS-SSS). While it is mentioned in the third year, it is only indicative and it could be anytime in the five years.

Sr.No.	Budget Head	1st Year	2nd Year	3rd Year	4th Year	5th Year	Total (in lakhs)
Recurring							
1	Manpower	65	165	600	275	80	**1185**
2	Contingencies	25	25	25	25	10	**110**
3	Consumables	10	10	10	25	10	**65**
	Total (A)	**100**	**200**	**635**	**325**	**100**	**1360**
Non- Recurring-Equipment							
4	Specialized Equipment etc.	0	50	75	100	120	**345**
	Total (B)	**0**	**50**	**75**	**100**	**120**	**345**
	Gross Total Budget (A+B)	**100**	**250**	**710**	**425**	**220**	**1705**

3. *Human Resource Development*

The Human Resource Development in the vertical of "Cognitive Computing and Social Sensing" is a major goal of the TiH. Towards this, the TiH will support manpower at different levels – under-graduates, graduates, PhDs, Postdocs, engineers. The aim is to increase the research base of CPS and also support the technical projects that are run under TiH.

a) The TiH will support 210 graduate fellowships. The graduate fellowships will be announced in an open call and will support work in the vertical of "Cognitive Computing and Social Sensing". While we mention the equal number for all 5 years but actual numbers may vary each year. The graduate fellowship will be provided over the duration of the graduation of 4 years.

b) The TiH will support 39 post-graduate fellowships. The post-graduate fellowships will be announced in an open call and will support work in the vertical of "Cognitive Computing and Social Sensing". While we mention the equal number for all 5 years but actual numbers may vary each year. The post-graduate fellowship will be provided over the duration of the post-graduation of 2 years.

c) The TiH will support 23 doctoral fellowships. The doctoral fellowships will support PhD students under the technical projects supported by the TiH. While we mention the equal number for all 5 years but actual numbers may vary each year. The doctoral fellowship will be provided over the duration of the doctorate for 5 years.

d) The TiH will support 3 faculty fellowships. Each faculty fellowship will last for one year. The faculty fellowships will be announced in an open call and will support work in the vertical of "Cognitive Computing and Social Sensing". While we mention one in initial three years but it will depend on the applications.

e) The TiH will support 3 Chair Professor fellowships. Each fellowship will last for one year. The Chair Professor fellowships will be announced in an open call and will support work in the vertical of "Cognitive Computing and Social Sensing". While we mention one in initial three years but it will depend on the applications.

f) The TiH will support skill development in the vertical of "Cognitive Computing and Social Sensing". The skill development may happen through an ongoing project of the TiH. The project will be chosen carefully to support the aims of the TiH.

The numbers in each year are indicative.

Sr.No.	Budget Head	1st Year	2nd Year	3rd Year	4th Year	5th Year	Total (in lakhs)
Recurring							
1	Manpower	450	540	750	375	270	**2385**
2	Contingencies	25	25	50	25	25	**150**
3	Consumables (travel, publishing, etc.)	25	25	60	25	25	**160**
	Total (A)	**500**	**590**	**860**	**425**	**320**	**2695**
Non-Recurring-Equipment							
4	Specialized Equipment etc.	0	100	100	100	100	**400**
	Total (B)	**0**	**100**	**100**	**100**	**100**	**400**
	Gross Total Budget (A+B)	**500**	**690**	**960**	**525**	**420**	**3095**

4. International Collaboration

Supporting International collaboration is a major objective of the TiH. For this purpose, The TiH will encourage investigators in the project to strengthen the industrial collaboration. The support for international collaboration will be to support their travel and stay in India for the project purposes. The visit of the international collaborator in a project will be approved during the evaluation of the project. A PI would also be able to make a request for the visit of an international collaborator to the EC of the TiH and it may be approved. A total of 2 crore INR is allocated to support the international collaboration. While we indicate equal sum in all years, the actual numbers may change.

Sr.No.	Budget Head	1st Year	2nd Year	3rd Year	4th Year	5th Year	Total (in lakhs)
Recurring							
1	Travel , Honorarium, and DA/TA etc.	10	20	40	40	40	**150**
2	Contingencies	2	5	5	5	5	**22**
3	Consumables	3	5	5	5	5	**23**
	Total (A)	**15**	**30**	**50**	**50**	**50**	**195**
Non-Recurring - Equipment							
		0	**0**	**0**	**0**	**0**	**0**
	Total (B)	**0**	**0**	**0**	**0**	**0**	**0**
	Gross Total Budget (A+B)	**15**	**30**	**50**	**50**	**50**	**195**

5. *TIH Management Unit*

The TiH has been allocated a dedicated space of 30,000 sq ft to support the activities under the TiH. The development of this space, e.g. furniture etc., and running of the TiH with dedicated staff to support its functioning will be supported. For the same a total of 5 Crore is allocated. The first year will see the development and running cost and rest of the years, it will mainly be the running costs, i.e., salaries of the staff that is hired and expenses related to electricity etc. Chief Executing Officer has been hired to manage and run the TiH throughan open advertisement. The TiH will also have manpower of accountants, managers, webmasters etc. A total of around 1 Crore each year is dedicated for the salaries. An indicative list of manpower is given here:

Sr.No.	Budget Head	1st Year	2nd Year	3rd Year	4th Year	5th Year	Total (in lakhs)
Recurring							
1	Manpower	50	100	100	100	100	**450**
2	Contingencies	5	10	10	10	10	**45**
3	Consumables (travel, electricity, maintenance etc.)	5	20	20	20	20	**85**
	Total (A)	**60**	**130**	**130**	**130**	**130**	**580**
Non-Recurring - Equipment							
	Equipment	0	10	10	10	10	**40**
	Space Development	0	200	100	50	0	**350**
	Total (B)	**0**	**210**	**110**	**60**	**10**	**390**
	Gross Total Budget (A+B)	**60**	**340**	**240**	**190**	**140**	**970**

Commercialization and Revenue Generation:

The TiH will have to become self-sustainable after the five years. Moreover, the efforts towards the self-sustainability will be an important criterion to measure the success of the TiH during the five years. Therefore, a major activity of the TiH would be towards commercialization of the technology developed and efforts to raise funds to support activities of the TiH. The following are few directions that will be taken towards achieving self-sustainability:

a. Infrastructure as service: As part of the TiH, a large amount of infrastructure will be developed to support cutting-edge computing, e.g. DGX servers. Currently, there are multiple cloud service providers, e.g., Amazon, which let people hire the computing resources. We may follow a similar strategy for the computing infrastructure developed at IIIT-Delhi.

b. Software as service: A major outcome of projects under TiH would be different software to solve problems in the domain of Cognitive Computing and Social Sensing. The software can be converted to APIs which can then be commercialized as a service or even as stand-alone software.

c. Data Access and Analytics Services: The projects will also collect a good amount of data in different domain. While some of the data cannot be used for commercialization purposes but other data may be usefulfor such purposes. Therefore, the commercialization aspect of Data would also be explored.

TIME FRAME

12. Time frame

Major Activity	FY 20-21		FY 21-22				FY 22-23				FY 23-24				FY 24-25			
	Q1-Q3	Q4`	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
DPR Submission		✓																
TiH Set-up		✓																
HBG Constitution		✓																
Core Team onboarding		✓																
Establishment of Technology Platform			✓	✓	✓	✓												
Technology Platform Upgradtion							✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
!st CFP Projects Finalization			✓	✓														
PRSG Formation			✓															
2nd CFP and Project finalization				✓	✓	✓												
3rd CFP and Project finalization							✓	✓	✓									

Plan of action for accomplishment of targets in next 5 years

Technological Development

- Industry collaboration or a field partner, e.g. Pharma companies, Healthtech companies, hospitals like AIIMS, NGO, govt. organizations with each project- cash or in kind (mentorship, use of facilities, technology deployment etc.)
- Monitoring by dedicated PRSG which will recommend releasing money depending on the deliverables achieved
- Review of projects at regular quarterly intervals
- Establishment of Computing and Data Facilities to earn revenue
- Expected to deliver about 100 research publications and 20 patents over 5 years
- A spent of Rs.40.35 cr over a period of 5 years

Entrepreneurship Development

- Support a minimum of 32 startups and spinoff companies over 5 years- *by way of Equity participation or Revenue share*
- Work with IIITD Incubation Center to support incubating activities and develop our own 30,000 sq feet facility
- Conduct Grant Challenges and other nation-wide competitions to encourage entrepreneurship activités
- Support entrepreneurs under the CPS-PRAYAS and further 19 entrepreneurs in résidence (CPS-EIR)
- Will support one dedicated Innovation Accelerator (CPS-DIAL)
- Support for 1 initiative under seed support system (CPS-SSS)
- A spent of Rs.17.05 cr over a period of 5 years

Human Resource Development

- Support 210 graduate fellowships, 40 post-graduate fellowships, 45 doctoral fellowships, 40 post-doctoral fellowships, 3 faculty fellowships, 3 faculty fellowships
- Announcement through open call
- A spent of Rs.30.95 cr over a period of 5 years

International Collaboration

- Encourage investigators in the project to strengthen the industrial collaboration
- Visit of the international collaborator in a project will be approved during the evaluation of the project
- A spent of Rs.2 cr over a period of 5 years

CHANAKYA FELLOWSHIPS

Call for Chanakya fellowships to be started in September 2021.

CALL FOR CHANAKYA FELLOWSHIPS (UG & PG)
IHUB ANUBHUTI-IIITD FOUNDATION

Technology Innovation Hub (TiH) on Cognitive Computing and Social Sensing for Interdisciplinary Cyber-Physical Systems (ICPS), called iHub Anubhuti, has been established by Indraprastha Institute of Information Technology Delhi as a Section 8 Company under the aegis of the National Mission on Interdisciplinary Cyber-Physical Systems (NM-ICPS) of the Department of Science and Technology, Government of India. TiH will focus on new knowledge, technology solutions, skilled human resource and an enabling eco-system for entrepreneurship in the field of CPS. TiH aims at building a collaboration between industries, academia and government agencies on developing data driven cognitive computing solutions, mainly in verticals - Health, Law Enforcement & Security, Education and Environmental Sustainability. TiH will also encourage and promote the entrepreneurial activities and start-ups emanating from TiH projects and in the areas of "Machine Learning, Artificial Intelligence and Cognitive Computing & Social Sensing".

iHUB Anubhuti's Comprehensive and Holistic Advancement of National Knowledge Yield and Analytics (CHANAKYA) UG and PG fellowships are great opportunity for the undergraduate and postgraduate students to enhance their learning, explore entrepreneur skills and innovate across domains of **legal Information processing system**, **maternal and child public health, predict mortality/diagnosis/diagnostic tools for patients suffering from cancer/liver failure, cognitive computing & social sensing, education and environmental sustainability** (a non-exhaustive list).

Benefits to enhance your career growth:

- Opportunity to enhance the learning, explore entrepreneur skills and innovate.
- Opportunity for entrepreneurial activities and start-ups.
- Monthly fellowship grant.
- Get opportunity to interact with industries.

Proposals are invited from individual students or team of students. There can maximum 2 students in a team and the students can be from different disciplines/institutions. Each team will also be assigned a

faculty member who will help the team execute the project and look after the progress regularly. The students availing this fellowship can also have opportunity to work with Infosys Center for AI (CAI) or Center of Excellence in Healthcare (CoEHe) in IIITD.

Fellowship Name	**iHub Anubhuti CHANAKYA UNDER GRADUATE FELLOWSHIP**
Objective	To create a pool of trained professionals in CPS domain to be problem-solvers or turn entrepreneurs
Eligibility	B.Tech students from any Indian institution/discipline in their 3rd or 4th year. Applicants must be aspiring to work in the areas of "Machine Learning, Artificial Intelligence and Cognitive Computing & Social Sensing". CGPA for each team member should be greater than or equal to 7. If selected, the applicant cannot avail any other fellowship in parallel.
Duration	10 months
Fellowship Grant	Rs.10,000/- month per student *subjected to change as per DST guidelines and approval of TiH (iHub Anubhuti) management.
Selection Procedure	Each team needs to submit a proposal (max 2 pages) that should consist of the team details of the team members, the aim, the research gap, the methodology and the expected deliverables with timelines. Any crazy idea that has the potential to lead to a technology (startup) is highly encouraged. The proposals will be shortlisted based on the eligibility criteria, feasibility and originality of the proposal. Shortlisted applicants may be asked to give presentation on their proposal, to the evaluation committee.
Expected Outcomes	Novel method that may lead to translational technology (potential startup), source codes and datasets, research papers or patents, project report, trained professionals.
Performance Metrics	No. of IP generated, technology transferred. Contribution of UG students in solving issues of industries.
Number of fellowships	40 (maximum)

Terms & Conditions:

- Monthly monitoring of the selected proposals.

- iHub Anubhuti reserve the right to terminate the fellowship based on unsatisfactory progress.
- iHub Anubhuti reserve all rights to revise the eligibility criteria, reject any proposal without assigning reason at any time.
- Graduate fellowship is applicable during actual number of months (including summer and winter vacations) student has worked on the project with limit of maximum 10 months of academic year.

Fellowship Name	**iHub Anubhuti CHANAKYA POST GRADUATE FELLOWSHIP**
Objective	To provide a platform for PG students to work on a real problem of industry and solve it using CPS
Eligibility	M.Tech students from any Indian institution/discipline in their 1st and 2nd Year. Applicants must be aspiring to work in the areas of "Machine Learning, Artificial Intelligence and Cognitive Computing & Social Sensing". CGPA for each team member should be greater than or equal to 7. If selected, the applicant cannot avail any other fellowship in parallel.
Duration	24 months
Fellowship Grant	Rs. 12,400/- per month per student *subjected to change as per DST guidelines and approval of TiH (iHub Anubhuti) management.
Selection Procedure	Each team needs to submit a proposal (max 2 pages) that should consist of the team details of the team members, the aim, the research gap, the methodology and the expected deliverables with timelines. The proposals will be shortlisted based on the eligibility criteria, feasibility and originality of the proposal. Shortlisted applicants may be asked to give presentation on their proposal, to the evaluation committee. Any crazy idea that has the potential to lead to a technology (startup) is highly encouraged.
Expected Outcomes	Novel method that may lead to translational technology (potential startup), source codes and datasets, research papers or patents, project report, trained professionals.
Performance Metrics	No. of IP generated, technology transferred.

	Contribution of UG students in solving issues of industries.
Number of fellowships	8 (maximum)

Terms & Conditions:

- Monthly monitoring of the selected proposals.
- iHub Anubhuti reserve the right to terminate the fellowship based on unsatisfactory progress.
- iHub Anubhuti reserve all rights to revise the eligibility criteria, reject any proposal without assigning reason at any time.

Submission Procedure:
Deadline to submit the proposal: Oct 14, 2021 (11:59 pm IST)
Submission link:

PROPOSAL FORMAT FOR THE CHANAKYA UG & PG FELLOWSHIPS

1. Project Title:

2. Applicant's Name (Name, Department, Email id and Mobile Number):

3. Team Details (Name, Department, Email id and Mobile Number):

4. Faculty Mentor Details (Name, Department, Email id, if it has already been decided):

5. Domains (Please indicate one):

   - Legal Information Processing System
   - Healthcare
   - Cognitive Computing and Social Sensing
   - Education
   - Environmental Sustainability

6. State the problem (Max 300 words):

7. Objectives:

8. Executive Summary (with pictorial view of the expected prototype/design/Front-End, if possible) (Max 500 words, clearly write how your idea solves the real-world problem related to above mentioned verticals):

9. Methodology (should be technical):

10. Expected Deliverables/Outcomes and timelines:

Other open calls – Doctoral, Post-doctoral, Faculty, and Chair Professors, and more – will be announced during the course of TiH.

COST BENEFIT ANALYSIS

13. Cost-Benefit Analysis

Today, India is on a path to self-reliance and for the same, Prime Minister Shri Narendra Modi has recently espoused the dream of Atma Nirbhar Bharat–a self-reliant India. One of the five important pillars of this strategic vision is–leveraging technology for self reliance. The technologies that are being developed as part of the TiH-Anubhuti across the four verticals– Law and Security, Environmental Sustainability, Health, and Education, are aimed at accruing a wide range of benefits to the society and cater to the goal of India becoming self-reliant by developing technologies and technological solutions that are beyond state-of -the-art, and leveraging these technologies to address a number of problems at hand, and monetise these technologies through entrepreneurial ventures.

Technologies that are being developed as part of the Health Vertical include but not limited to algorithms and applications for India's National Health Stack. Applications include but are not limited to prediction and diagnostic AI models for life diseases and disorders-–cancer, genetic disorders, chronic liver failure, respiratory diseases, Autism Spectrum DIsorders among several others. In addition, projects also focus on the development of AI-based decision platforms for antimicrobial stewardship, maternal and child healthcare, and triage related decisions.

The recently released new National Education Policy, gave high quality contemporary education a necessary impetus. In this direction, technologies that are being proposed to be developed under the Educational Vertical would provide competitive advantage in terms of enhanced education, building support systems for special educational needs, and creating software/hardware infrastructure for digital education, continuing education and training.
In the area of Environmental Sustainability, the emphasis of research is on improving the quality of people's lives by addressing several environmental related issues that typically come with urbanization and industrialisation. Specifically, projects in this area are focusing on developing intelligent AI-based technological platforms and decision-making tools for addressing problems such as pollution and waste management, and the effective management of natural resources by preventing their over-exploitation. A second area of focus is on developing AI-based Intelligence Transportation Systems to manage transportation infrastructure and address some of the urban social problems such as traffic congestion. A third important focus is to develop AI-based surveillance and disaster response and decision systems that prove invaluable for saving lives and reducing human casualties during natural and man-made disasters.

In the area of Law and Security, the emphasis is on creating technologies and applications that focus on securing our nation's space assets, building systems that help us fight the problems of mis-information, and detecting and mitigating security threats from rumours and disinformation that pose a serious threat to law and order situation in the country. Lastly as e-governance continues to gain traction in the country, there is a significant shift towards making strategic policy decisions by leveraging AI-based decision support systems. In this context, projects under this vertical also focus on developing technologies that counter algorithmic injustice in AI-based decision support systems in order to accrue benefits of e-governance fairly to the public.

A second important objective of the iHUB Anubhuti initiative is that of exploring entrepreneurial opportunities by monetizing the technologies developed as part of these projects. From this perspective, many of the research projects are being executed in

collaboration with a wide variety of partners including but not limited to technology companies, research centers, government organizations, non-profit organizations, and hospitals to name a few. This collaborative strategy has many benefits. First, researchers would be able to draw upon the experience of industry experts and develop technologies that cater to the needs of emerging markets. Second, such partnerships will also pave the way for entrepreneurial opportunities to monetize such technologies and create revenue flows. Collaborative partnerships also provide opportunities for testing these technological applications in real world environments and obtaining feedback from end users. This ensures that researchers are able to ensure that their products meet industry standards and cater to the needs of their customers. This further increases the prospects of monetizing these technologies as they are able to better align their products to the needs of the market.

Third, the focus of all the projects is on developing innovative applications in niche areas such as mobile computing, edge computing, embedded computing, 3D imaging, and in creation of novel datasets and applications that have universal applications. Lastly, many of the technologies that are being developed are domain independent, i.e. these technologies can be applied in different domains such as education, ecommerce, health, space exploration, to name a few. These factors we strongly believe will ensure that the projects are successful and yield the anticipated benefits and are poised to bring in transformational changes in various sectors. Lastly, one of the most important outputs across all the verticals is the vast number of people who would be trained in high end skills, as part of these projects. In summary the various technological and human resource outputs from projects proposed as part of TiH-Anubhuti add tremendous value to the society and contribute to our country's vision of self-reliance.

RISK ANALYSIS

14. Risk Analysis

Large scale projects come with some inherent risks. However, strategic planning and implementation of the projects with due diligence, will help us to anticipate and minimize such risks and their adverse impacts. Projects under TiH-Anubhuti have been meticulously designed with clearly laid out timelines and plans for their execution. This has been complemented by a number of mechanisms built in to minimize any adverse impacts on the successful development of the proposed technologies and their eventual monetization. They are as follows.

First, a number of projects have been designed with collaborative relationships with leading research universities, renowned industrial partners, and research institutions. These collaborative partnerships provide access to diverse sources such as proprietary data, trained manpower, vast amounts of accumulated man hours of domain knowledge and expertise, and industrial know-how and expertise, which ensure that adequate resources are available for the successful completion of the projects.

Industrial collaborative partnerships are particularly noteworthy because they prove to be invaluable, as the ultimate goal of these projects is to develop innovative technologies and to explore entrepreneurial opportunities as a means to monetise these technologies and explore new markets for these technologies. To this end, industrial partnerships provide access to latest technological knowledge, and provide the right kind of environment for deploying and testing these technological products, and evaluate their viability and market value.

Furthermore, industrial partnerships also provide channels for valuable user feedback that can be incorporated into various stages of the product's development cycle. In the case of technologies aimed at the healthcare market, a number of projects have established collaborative partnerships with hospitals and or medical research institutions where these technologies can be tested in real world environments for obtaining feedback for product improvement.

Minor setbacks are inherent in any technology's development lifecycle. One common strategy that a number of projects in the TiH-Anubhuti have adopted is a module-based approach, where design, testing and feedback incorporation occur incrementally, and at an ideal pace. This allows for minimising losses from minor setbacks and any challenges that arise can be immediately recognised at initial stages, and allows researchers to go back to the drawing board and address the deficiencies in the technology design.

This strategy is also accompanied by well planned timelines, and a strong commitment to adhere to the deadlines. This is further bolstered by checks and balance mechanisms such as mandatory annual reviews and feedback by a committee of academic and industrial experts that helps project leaders to monitor the program closely, at every stage, from ideation to implementation and testing, and provides ample opportunities to assess and mitigate any risks. Annual reviews will also be opportunities to monitor not only the development of technologies, but also to evaluate the progress made in terms of other outcomes identified as part of the project.

Many cognitive computing and social sensing projects rely on corpus amounts of data that is drawn from various sources. Of particular importance is the use of social data which raises concerns about individuals' privacy and security. To this end, projects in TiH-Anubhuti, that involve the use of social data will ensure that such data is devoid of any personal identifiers, and data will be completely anonymised in order to protect the privacy of individuals. Data shall be stored in secure locations in digital format with adequate security measures in place. Lastly, some of the projects would also involve collection of data from vulnerable populations such as neurotypical children, and adults and children with cognitive impairments. Each of these projects would be scrutinized by the Institutional Review Board (IRBs) of the respective participating institutions to ensure that proper protocols are put in place to ensure that no harm comes to participants. Lastly, these projects will be carried out under supervision of trained health professionals, to ensure that no harm comes to the participants.

OUTCOMES

15. Outcomes

S No	Target Area	Targets					
		1st Yr	2nd Yr	3rd Yr	4th Yr	5th Yr	Total
1	**Technology Development**						
(a)	No of Technologies (IP, Licensing, Patents etc.)	0	0	4	5	10	19
(b)	Technology Products	0	0	4	4	4	12
(c)	Publications, IPR and other Intellectual activities	3	5	8	10	10	36
(d)	Increase in CPS Research Base	3	5	10	19	25	62
2.	**Entrepreneurship Development**						
(a)	Technology Business Incubator (TBI)	0	0	0	1	0	1
(b)	Start-ups & Spin-off companies	1	2	5	10	14	32
(c)	GCC - Grand Challenges & Competitions	0	0	1	0	0	1
(d)	Promotion and Accelerationof Young and Aspiring technology entrepreneurs (PRAYAS)	0	0	1	0	0	1
(e)	CPS-Entrepreneur In Residence (EIR)	1	2	3	5	8	19
(f)	Dedicated Innovation Accelerator (DIAL)	0	0	1	0	0	1
(g)	CPS-Seed Support System (CPS- SSS)	0	0	1	0	0	1
(h)	Job Creation	50	500	1000	2500	3700	7750
3.	**Human Resource Development**						
(a)	Graduate Fellowships	10	50	50	50	50	210
(b)	Post Graduate Fellowships	5	10	12	12	0	39
(c)	Doctoral Fellowships	7	8	8	0	0	23
(d)	Faculty Fellowships	0	1	2	0	0	3
(e)	Chair Professors	0	1	2			3
(f)	Skill Development	10	50	100	100	120	380
4.	**International Collaboration**						
	International Collaboration	0	1	0	0	0	1

EVALUATION

16. **Evaluation**

The projects under the TiH-Anubhuti will be evaluated annually by a committee consisting of senior scholars and industry experts. Every year, the Principal Investigators will submit a progress report in a prescribed format that clearly outlines the progress made in specific categories. Projects would be evaluated based on the progress made in terms of the following outputs: Academic outputs, technological development related outputs, outreach, and skill development outputs.

Under academic related outputs evaluation criteria include: Publications, case studies, courses or course modules developed as the result of the research. In terms of skill development evaluation criterion includes the number of students trained at the Graduate Level, Postgraduate Level, Doctoral Level, and at Post Doctoral Level, and training of software engineers and personnel in the effective deployment, testing and supporting of the application.

Technology Related Outputs will include the potential for patenting, intellectual property (IP), start-up potentials, prototypes, algorithms, tools and technologies, results from deployment and testing. Criteria for evaluating overall usability of the project may be evaluated with measures like user feedback, size of the user base in case of product usage, demonstrations of product prototype and functional effectiveness of technologies, quality and effectiveness of prototype chips, etc.

Outreach as an evaluation criterion will include workshops conducted, fruitful International collaborations established, and stakeholder engagement. At the end of three years wherever appropriate, projects will also be evaluated in terms of potential for monetization of the technologies or tools developed as outlined in the entrepreneurship and start-ups category.

COLLABORATION WITH OTHER TIHs

17. Collaboration with other TiHs

There is an opportunity to build long-lasting technologies by combining expertise of different TiHs. At IIIT-Delhi, we have had discussions with TiHs whose technology focus align with the TiH of IIIT-Delhi, We have identified initial avenues of collaborations and moving forward, we would be collaborating with them on the problems of mutual interest.

1. Collaboration with TiH at IIT-Delhi: The Project Coordinators at both of the TiHs have had multiple discussion and have planned to work on collaboration more closely. Following are the points on which the TiHs plan to collaborate:
    a. Forming joint project evaluation committees
    b. Launching Grand challenge together
    c. Outreach activities
    d. Supporting projects of mutual interest

2. Collaboration with TiH at IIT-Bombay: The TiH at IIT-Bombay is targeted towards providing technology solutions in the area of Technologies for the Internet of Things & the Internet of Everything. At the TiH at IIIT-Delhi, our focus is towards cognitive computing and social sensing, with emphasis on creating computer systems for different applications that behave, think, and interact the way humans do. The internet of things and the Internet of everything aligns naturally with social sensing which relies on collecting data through various means and with the proliferation of IoT/IoEs, data communication and sharing will primarily happen through IoT/IoE networks. Similarly, Cognitive Computing allows us to build information systems that can proactively understand the incoming data (from a diverse set of sources), situations arising in the systems, and reacting automatically by learning from patterns pertaining to these situations. There is a natural synergy between the TiHs of IIT-Bombay and IIIT-Delhi. The core team members at TiHs of both places have had a discussion and agree that moving forward, the two TiHs will work together actively on the problems of mutual interest.

3. Collaboration with TiH at IIT-Patna: The TIH of IITP aims to focus on developing NLP and Speech technologies for various applications. NLP is a core ability of cognitive computing systems and is often defined as helping computers process and understand human language. Since IIITD TIH will focus on cognitive computing, there is a natural synergy between these two domains. Both the teams had a discussion and reached a consensus that they will collaborate actively on the problems of mutual interest.

4. Collaboration with TiH at IIT-Kgp: The TIH of IITKGP aims to focus on developing AI and ML technologies for various applications. One of the applications could be cognitive computing and social sensing that the TIH of IIITD will target. Both the TIHs agreed that there is a natural synergy between two domains in terms of providing cognitive computing support in AI/ML-related applications.

5. Collaboration with TiH at IIT-Mandi: The TiH at IIT-Mandi is targeted towards providing technology solutions in the area of Human-Computer Interaction. The HCI research forms the basis of technologies developed to be used for the general public. The focus

application areas of IIT-Mandi and IIIT-Delhi, especially in the areas of Healthcare and environment, are of special interest and provides a natural avenue for collaboration. With the TiH at IIIT-Delhi, our focus is towards cognitive computing and social sensing, with emphasis on creating computer systems for different applications that behave, think, and interact the way humans do. Cognitive information systems can proactively understand the incoming data (from a diverse set of sources), situations arising in thesystems, and reacting automatically by learning from patterns pertaining to these situations. There is a natural synergy between the TiHs of IIT-Mandi and IIIT-Delhi. Thecollaboration between the TiHs will go towards providing cognitive computing and social sensing enabled solutions with foundations of Human-Computer Interaction research in the healthcare, education, environment domains among others. The coreteam members at TiHs of both places have had a discussion and agree that moving forward, the two TiHs will work together actively on the problems of mutual interest.

6. Collaboration with TiH at IIT-Bhilai: The TiH at IIT-Bhilai is targeted towards providing a customer-centric model for Fintech solutions. By providing customized, targeted products and services, financial technology companies are not just acknowledging, butalso catering to the evolving consumer demands by offering convenience, personalization, transparency, accessibility and ease of use factors that empower customers to a great extent. With the TiH at IIIT- Delhi, our focus is towards cognitive computing and social sensing, with emphasis on creating computer systems for different applications that behave, think, and interact the way humans do. Cognitive information systems can proactively understand the incoming data (from a diverse set of sources), situations arising in the systems, and reacting automatically by learning from patterns pertaining to these situations. To elaborate, cognitive computing understands the psychology of end-users and how they are going to consume the information thereby facilitating immense amounts of personalization, precise profile mapping, accurate trend analysis, and futureprediction. From the perspective of end products, cognitive computing plays a key rolein human-centric computing driven personalized visualization tools, recommended interactions, adapt the user behavior and interaction patterns to improve context, etc.There is a natural synergy between the TiHs of IIT-Bhilai and IIIT-Delhi. The collaboration between the TiHs will go towards providing cognitive computing and social sensing enabled solutions for the FinTech area. The core team members at TiHsof both places have had a discussion and agree that moving forward, the two TiHs will work together actively on the problems of mutual interest

As the TiH progresses, formal collaborations with other TiHs will also be explored and formalized. We will enter into specific MoUs with TiHs based on areas of mutual interest.

GRANTING OF RESEARCH PROJECTS

1. EngageMe: Multimodal Analysis of Attention among Children with Attention Deficit Hyperactivity Disorder for Digital Learning

PI: Jainendra Shukla, IIITD

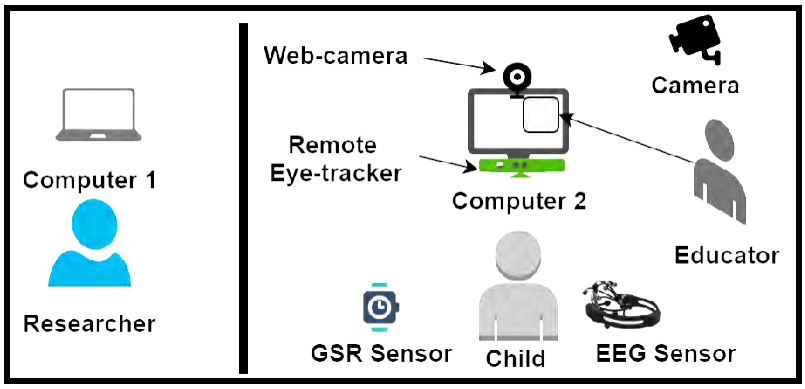
- **Aim & Objectives:** The aim of this project is to develop an Intelligent Education Platform, **EngageMe** that will offer personalized, monitored, and evidence-based identification of attention level among children with Attention Deficit Hyperactivity Disorder (ADHD) to enable adaptive learning in digital settings. EngageMe will help the special educators and pedagogues in reaching an objective and reliable assessment of the child's attention level during online learning. We plan to leverage machine learning and artificial intelligence techniques to tackle the specialized manpower issue and reduce the burden of the educators/parents. To achieve thisaim, the project has set forth the following objectives:

1. To investigate the effectiveness of physiological measures in the evaluation of the level of attention among children with ADHD during online learning settings. The investigated physiological signals will include Galvanic Skin Response (GSR), Heart-Rate Variability (HRV), Electroencephalogram (EEG), and pupil-dilation (Eye-tracker).
2. To investigate the effectiveness of audio-visual signals in the evaluation of the level of attention among children with ADHD during online lecture settings.
3. To develop an ICT system integrating the findings of objectives 1 and 2 by employing machine or deep learning algorithms for real-time continuous assessment of attention level among children with ADHD during online lecture settings.

- **Competitive Advantages:**

1. **Multimodal Interaction:** The use of audio-visual and physiological sensing-based interaction will not only enrich the interaction quality but will also improve the system's reliability in several tasks such as in approximating the user's level of attention.
2. **AI supported monitoring and learning:** The proposed solution will employ assorted artificial intelligence, machine learning & deep learning algorithms to perform automated detection and/or prediction of a child's level of attention and for the development of an adaptive learning system. It will enable intelligent training and learning thereby reducing the learning time for the child and improving the overall efficiency.
3. **Objective Assessment:** To date, monitoring of attention among children with ADHD by special educators is done based on their expertise, and hence, it introduces subjectivity in the assessment. Our proposed solution will provide an objective assessment thereby making it standard.
4. **Increased Availability in terms of Location and Time:** The availability of special educators is limited even in tier-1 Indian schools and cities. Our proposed solution will facilitate the extension of the services to even the remote areas of India. Our proposed solution can perform the tasks repeatedly without wear leading to better confidence in the monitoring of the level of attention.

- **Methodology:** The project will collect the audio-visual and physiological data among 30 children with ADHD (clinically diagnosed with ADHD) aged 8-12 years and 30 typically developing children of the same group. Each child will undergo two experiments over a week. Hence, a total of (60x2=) 120 data collection sessions will be performed, resulting in a ~40 hours of data collection time, [considering 20 minutes for each experiment] altogether over an estimated period of 24 days as we plan to conduct five 20 minutes sessions per day. During the experiment, offline sessions will be conducted in which the participant will perform a series of computerized experiments. The experiment will consist of standard psychological tasks which are proven to stimulate particular types of attention. During the experiment, the child will be wearing 2 wearable physiological sensors, one wristband-shaped sensor for GSR and another headset-shaped sensor for EEG signals. In addition, one remote desktop eye-tracker and a web-camera



will be attached to the computer device. The duration of the experiment will be roughly 20 minutes depending upon the child interaction. Each session will be recorded using a high-resolution video camera. Data from 1 web camera, 1 eye-tracking camera, and wireless physiological sensors will be recorded during the experiment. Figure 1 below shows the setup for the data collection.

Figure 1: Setup for the data collection EEG

A. Recurring		**Year I**	**Year II**	**Year III**	**Total**
1.Salaries/wages					
Designation & number of persons	Monthly Emoluments				
Junior Research Fellow (JRF), 1	38,440	461,280.00	461,280.00	-	922,560.00
Senior Research Fellow (SRF), 1	45,500	-	-	546,000.00	546,000.00
Research Assistant, 1	25,000	300,000.00	-	-	300,000.00
2. Consumables		20,000.00	15,000.00	15,000.00	50,000.00
3. Travel		50,000.00	75,000.00	75,000.00	200,000.00
4. Contingency		50,000.00	50,000.00	50,000.00	150,000.00

5. Other costs [Licensing + Participation Costs + Consultation fees]	30,000.00	25,000.00	25,000.00	80,000.00
Total (A)	**911,280.00**	**626,280.00**	**711,000.00**	**2,248,560.00**
B. Non-Recurring Equipment				
1 Eye-tracking device	600,000.00	-	-	600,000.00
2 Laptop	150,000.00	-	-	150,000.00
Total (B)	**750,000.00**	**-**	**-**	**750,000.00**
Total (A+B)	**1,661,280.00**	**626,280.00**	**711,000.00**	**2,998,560.00**

2. **Proposal Title:** Explainable AI Methods for interpretable transcriptomic cancer data analysis.

   PI: Dr. Ranjitha Prasad; **Co PI**: Dr. Anubha Gupta, Dr. Ritu Gupta

   a. **Perceived TIH Application under Cognitive Computing and Social sensing:** Intelligent diagnostics and prognostics for healthcare

   b. **Application Accelerators:** Addressing specific industry functions in healthcare

   c. **Proposal Details:** 1. **Context/Background**: Cognitive computing powered by artificial intelligence is revolutionizing the area of healthcare globally. Researchers are making efforts to build automated analysisand decision-making tools on enormous volumes of healthcare data using advanced AI-based cognitive computing. Cancer is the one of the leading causes of death globally. Cancer arises through a combinationof genetic changes and mutations that are unique to an individual patient. These changes can be studied from the genomics or transcriptomics data. While the single cell resolution provided by sc-RNA data provides a unique opportunity to make observations, it also offers unique challenges such as obtaining algorithms to handle high volume of data in an interpretable and scalable manner. In the proposed study, we will address the problem of scalable and interpretable machine learning methods for the analysis of single-cell transcriptomic data (sc-RNA) using tools such as Bayesian causal inference.

   - **Problems to be addressed**: Single single-cell RNA sequencing (scRNA-seq) promises to provide higher resolution of cell-level variations as compared to bulk-RNA sequencing, providing medical and statistical researchers a plethora of opportunities for data analysis. Although scRNA-seq data provides immense opportunities for novel data analysis, it also entails several issues such as being noisy with severe missing data problems in high-dimensions. The following are the primary issues with scRNA-seq data:

   - **Batch Effect**: High-throughput single-cell RNA-seq data are collected in multiple batches under different conditions, leading to difference in gene expression values, which may be construed as the biological variations due to inherent inter-cell heterogeneity. This bias due to measurement is often referred to as the batch effect.
   - **Dropout Events**: In the case of lowly expressed genes, the stochastic transcriptional process could lead to zero entries in scRNA-seq expression data, leading to dropout events.
   - **Technical and biological noise**: In addition to batch effect and dropout events, some other technical factors could also cause biases in the scRNA- seq data, especially for lowly expressed genes. Such biases are called technical noise and biological noise.
   - **Curse of Dimensionality and scalability**: An scRNA-seq dataset consists of the expression profiles of a large number of genes.. In such a high-dimensional space, as the data points become sparse, the distance measures become ineffective. This is called the curse of dimensionality. Further, the sheer amount of data available is huge and hence, this calls for efficient and scalable algorithms for modeling and data analysis.

We address the interpretability issues that arise in sc-RNA data analysis that are listed above, and propose a study on local interpretable models which provide explanations for a single test instance. The problems addressed are:

- Interpretability of joint approaches for batch effect removal and clustering to obtain cluster-specific gene expression representation and cluster assignments, while including cells from all batches in the analysis. The proposed study will employ Bayesian generative nonparametric methods. Interpretable imputation methods to address the increased sparsity in the scRNA data. The proposed study will be based on causal Bayesian techniques which uncover confounding factors that lead to technical and biological zeros, especially in the context of lowly expressed genes.
- Transcriptomic sc-RNA data is high-dimensional, where the number of features are much larger compared to the number of samples. The importance of raw input features are not individually meaningful, but the

structured patterns that emerge in an explanation is important. In this work, we develop methods for explaining predictions in terms of a semantically meaningful concept which we propose to obtain using hierarchical Bayesian methods.

- **Strategy**: In our study, we propose causal Bayesian methods for interpretable data analysis of single-cell RNA data, where we propose local interpretable algorithms to facilitate per-sample explanation of a prediction. We will design algorithms based on publicly available datasets such as METABRIC, SEER, TCGA (bulk RNA sequencing data) and GSE72056 and GSE118828 sc-RNA dataset from gene expression omnibus . The proposed steps are:

- In the first phase, we design explainable methods for bulk-RNA data based on publicly available datasets such as METABRIC, SEER and TCGA. This will give us a handle on high-dimensionality of data, without having to look at other issues such as batch effects and dropouts.
- In the next phase, we consider publicly available single-cell RNA from the gene expression omnibus database. We will extend our high-dimensional methods towards handling dropout, sparsity and noise aspects specific to sc-RNA data.

- **Target beneficiaries and Management**: Since the cancer lab of AIIMS is the collaborator in this work, we deploy and test the proposed tool to obtain feedback from the doctors and clinicians of the cancer hospital. The project will consist of two teams, one at IIIT Delhi and the other at AIIMS, Delhi. The team at IIIT Delhi willconsist of Dr. Ranjitha Prasad and Dr. Anubha Gupta, with one project associate I. Further, the team at AIIMS consists of Dr. Ritu Gupta.

- **Finance**:

A. Recurring		**1st Year**	**2nd Year**	**3rd Year**	**Total (in Lakhs)**
Salaries/wages	Monthly Emoluments				
Project Associate I (IIITD)	31,000/-+(24% HRA)(Monthly) for first and second year, and 35,000/-+(24% HRA) for the third year onwards.	461,280.00	461,280.00	520,800.00	1,443,360.00
Contingency @ IIITD		100,000.00	100,000.00	100,000.00	300,000.00
Travel		100,000.00	100,000.00	100,000.00	300,000.00
Consumables		50,000.00	50,000.00	50,000.00	150,000.00
Total		**711,280.00**	**711,280.00**	**770,800.00**	**2,193,360.00**
B. Non-recurring: Equipment					
Equipment @ IIITD (year 1)	1 Laptop for the project (10th Generation IntelCore, i7-10510U Processor-Year 1	250,000.00	-	-	250,000.00
	1 NAS for storage	350,000.00	-	-	350,000.00
	1 Server -Year 1	200,000.00	-	-	200,000.00
Total		**800,000.00**	-	-	**800,000.00**
Gross Total		**1,511,280.00**	**711,280.00**	**770,800.00**	**2,993,360.00**

➢ **Outcomes**

The tangible outcomes of this project are as follows:

- Technology Development: Development of a software tool (beta version) for explainable methods forclustering and differential analysis of bulk and single-cell transcriptomic cancer data.
- HRD and Skill Development
- Publications and Patents: In Core A conferences, computational methods journals for single-cell genomics.

3. Multimodal Knowledge Graphs for Legal Question Answering and Reasoning

PI and Team

- Dr. V. Raghava Mutharaju (PI), IIIT-Delhi. raghava.mutharaju@iiitd.ac.in
- Dr. Rajiv Ratn Shah (Co-PI), IIIT-Delhi. rajivratn@iiitd.ac.in

➢ Problem Description

An important aspect in the field of cognitive computing is to build machines that can think, behave, and act in the way that humans do. In order to do this, machines would need background knowledge and this can help in intelligent decision making. Knowledge Graphs (KG) are graph structures that capture knowledge in the form of entities, relationships between them, properties, and additional information including provenance. Along with the W3C recommended Semantic Web standards such as RDF, OWL, and SPARQL, advances in Machine Learning, Deep Learning, Natural Language Processing, and Information Retrieval has led to automated construction of large KGs such as DBpedia, YAGO, Wikidata, and proprietary KGs such as those from Google, Yahoo, LinkedIn, Microsoft, Amazon, eBay, and Alibaba. But all these KGs have been built from only the text documents. There is important information captured by images and videos, especially in the legal domain and this has largely been ignored. Along with the textual data such as FIR, we plan to use the information that can be extracted from other modalities such as images, videos (CCTV footage) and audio to enrich the quality of the constructed KG. Doing so involves several challenges here, and they are listed below.

1. Extraction of structured data in the form of triples (subject, predicate or relation, and object) from the multiple modalities such as image, video, and audio. This is commonly referred to as the extraction of scene graphs from the visual source.
2. After the scene graph is extracted, identify whether some or all of the information in this graph falls into one of the following categories.
    a. **Complement** the information in the existing KG.
    b. **Corroborate** with the information in the existing KG.
    c. **Contradict** the information in the existing KG.

Depending on the category into which the information in the scene graph falls, it would be handled accordingly. To the best of our knowledge, none of the multimodal KG construction techniques consider these three aspects. Having such a KG facilitates deductive reasoning that can lead to inferring information which is implicitly hidden. Along with that, a structured data source such as a KG would help in answering complex questions.

➢ Technology & Research Methodology

The proposed project would involve the following steps.

Step-1: Extraction of structured information from the different modalities. Structured data in the form of triples (subject - predicate - object) needs to be extracted from images, videos, and audio related to the text using which the KG was built. A scene graph is built from these triples.
Step-2: Enriching the KG with the extracted information. Triples in each of the three categories have to be handled differently in terms of adding them to the KG. The triples that are complementary can be added to the KG, but the challenge here would be to determine where exactly in the KG these triples should be added to, i.e., who should be the neighbors of these triples in the graph. Corroboratory triples

need not be added to the KG, but they reinforce the information that is already part of the KG. They can be considered as synonymous or alias information that can be useful for the downstream applications. Forthe third category, contradiction needs to be resolved. There are a few options that can be considered and this depends on the context of the triples and the contradiction. A human in the loop approach can also be taken here, where the human can go over the ranked options of resolving the contradictions and giving thefeedback in each case.

Step-3: Develop two applications. The first one would be an application that can complement/corroborate/contradict the image/video with that of the text summary. This will be very useful in law enforcement. FIR is filed and notes are taken to describe any crime that would have happened. Along with this short report, generally, there will also be corresponding images and videos taken at the crime scene. In order to cross check the text in the report with that of the visual data, the techniques developed in this project can be used. We can find the missing or wrong information in the report. The second application is a question answering system that makes use of the multi modal KG and the ontologies to understand the query intent and answer the query by making use of the context provided by the query.

- **Contribution to the grand problem.** The proposed project directly contributes to the grand problem of legal information processing and management. In particular, we address the topics of legal data curation and modeling, legal reasoning and legal question answering.

- **Finance**

Item			**Budget**		
A. Recurring		**1st Year**	**2nd Year**	**3rd Year**	**Total (in Rupees)**
1.Salaries/wages					
Designation & number of persons	Monthly Emoluments				
1 Junior Research Fellow	40,950	491,400.00	491,400.00	-	982,800.00
1 Senior Research Fellow	43,400	-	-	520,800.00	520,800.00
1 Research Assistant	35,000	420,000.00	-	-	420,000.00
Consumables		20,000.00	20,000.00	20,000.00	60,000.00
Travel		50,000.00	53,000.00	53,000.00	156,000.00
Contingency		20,000.00	20,000.00	20,000.00	60,000.00
Total (A)		**1,001,400.00**	**584,400.00**	**613,800.00**	**2,199,600.00**
B. Non-Recurring (Equipment)					**-**
Server	150000	800,000.00	-	-	**800,000.00**
GPU Cards*3@190000/-	570000	-	-	-	**-**
Laptop	80000	-	-	-	**-**
Total (A+B)		**1,801,400.00**	**584,400.00**	**613,800.00**	**2,999,600.00**

4. Visual Traffic Analytics

PI: Dr. A V Subramanyam

- **Overview** Intelligent transportation system (ITS) is one of the major requirements in a smart city setup. ITS encompasses traffic data analytics using various sensors including visual sensors and form the major segment as part of ITS surveillance systems. These visual sensors capture plethora of data and if analysed properly, can lead to better monitoring, planning and managing traffic related issues. While several solutions exist under the umbrella of visual surveillance, we target different and important issues of vehicle reidentification (VRID), and vehicle counting.

In terms of visual sensors which form the major segment of surveillance systems, traditionally, Automated Number Plate Recognition (ANPR) has been one of the prominent services [1]. However, ANPR cannot be used in the cases where number plate is occluded or blurred. Thus, there are limitations in the problems addressed via ANPR technique. Our VRID system overcomes this problem as it matches the complete vehicle across the cameras.

Smart city requirements also include many other services such as – classification of vehicles into a predefined number of classes, simultaneous classification and counting under extreme conditions of congestion as well as stop-and-go. While having all services is a challenging task, we would like to emphasize that these components can be incorporated at a suitable later stage.

- **Objectives** Our proposal aims at traffic data analytics that can be further used for city wide traffic planning which has become a pinching necessity in urban India. While there are a good number of players in this domain, the technological innovations that have happened in the recent years have not yet been adopted. We are targeting to bridge this gap and collaborating with a startup which has an emerging presence in some smart cities. In the following, we briefly discussthese problems.

  - **Vehicle reidentification (VRID)** is one of the major components in traffic surveillance and is also an important requirement for law enforcement agencies. VRID involves re-identifying a given object across multiple cameras.
  - **Vehicle counting** between a source and a destination is one of the essential objectives to effectively manage the traffic. The model will be able to provide an estimate of the count of the vehicles travelling between a source camera to destination camera.

- **Research Methodology** Our system comprises of multiple modules – object detection, re-identification, and counting. Given an input video frame, object detection is first applied to determine the bounding boxes of objects in the frame.

  - **Object Detection** Given an input frame, we first detect or label an object to be tracked. Models such as Yolov4 [3] are deployed to this end. The following figure shows CCTV

images from Chandigarh. The two cameras are non-overlapping. The left two images show a vehicle and the detected vehicle from the frame. The right two show the footageof the same vehicle and detected vehicle in another camera.

- **Vehicle Re-identification** VRID presents severe challenges in terms of huge intra class dis-similarity and significant inter class similarity. The illumination variation, low resolution and lack of sufficiently large annotated dataset further adds to the complexityof the problem. Towards this, we first investigate training of deep learning models such as ResNet using ranking losses like triplet loss which has shown tremendous performance in person reidentification. Further, since the global vehicle image only hassmooth texture and may have high similarity for similar make of the model, prior literature extracts orientation based keypoints. We leverage both deep features and keypoints and perform a late fusion to obtain a unified representation. This representation can then be used to match a given pair of vehicles and further ranked based on the similarity. Some sample images are shown below. The leftmost image enclosed in red box is the query and the images in right are 5 best retrieved matches indecreasing order.

- **Counting** This is achieved by reidentifying the vehicles between the pair of given cameras.

➢ **Finance** (Duration – 2 years)

A. Recurring		**Budget**			
1. Salaries/wages		**1st Year**	**2nd Year**	**3rd Year**	**Total**
Designation & number of persons	Monthly Emoluments				
Research Fellow	32000, 36000	384,000.00	432,000.00	432,000.00	1,248,000.00
Research Assistant	20000	240,000.00	240,000.00	240,000.00	720,000.00
2. Consumables		30,000.00	30,000.00	-	60,000.00
3. Travel		-	25,000.00	-	25,000.00
4. Contingency		25,000.00	25,000.00	-	50,000.00
Total (A)		**679,000.00**	**752,000.00**	**672,000.00**	**2,103,000.00**
B. Non-Recurring-Equipments					
High End Server with GPU capability for executing complex image and video processing		800,000.00	-	-	800,000.00
Laptop		72,000.00	-	-	72,000.00
Printer		25,000.00	-	-	25,000.00
Total (B)		**897,000.00**	**-**	**-**	897,000.00
Gross Total (A+B)		**1,576,000.00**	**752,000.00**	**672,000.00**	**3,000,000.00**

References

[1] https://smartnet.niua.org/sites/default/files/TENP2.pdf, last accessed on 26th February, 2021.

[2] Bochkovskiy, Alexey, Chien-Yao Wang, and Hong-Yuan Mark Liao. "YOLOv4: Optimal Speed and Accuracy of Object Detection." arXiv preprint arXiv:2004.10934 (2020)

5. A pretrained model based approach to characterisation of functional heterogeneity in cancer

PI: Debarka Sengupta **| Collaborators:** Prof. Colleen Nelson (QUT, Brisbane), Dr. Gaurav Ahuja (IIIT-D), Prof. Jayadeva (IIT-D), Mr. Pallab Mitra (CareOnco Biotech Pvt. Ltd.)

Thematic connections: Cognitive Computing and Social Sensing (cognitive computing applied to cancers); Maternal and Child Public Health (Solutions described here entail all cancer types including pediatric and female cancers)

Budget (30 Lac for 2 years): A. Recurring:- (i) Postdoc or equivalent: 12 Lac (6 Lac each year); (ii) Consumables: 13 Lac (8 Lac and 5 Lac in first year and second year respectively); (iii) Contingency: 3 Lac (1.5 Lac each year). **B. Non-recurring:-** 2 Lac in the first year.

Key publications from the Sengupta Laboratory: Gupta et al., Genome Research (IF: 11), 2021; Jindal etal., Chawla et al., Nucleic Acids Research (IF: 11.5), 2020; Jindal et al., Nature Communications (12), 2018; Srivastava et al., Nucleic Acids Research (IF: 11.5), 2018; Sinha et al., Nucleic Acids Research (IF: 11.5), 2018.

- **Summary**

Human DNA harbors roughly ~3 billion, leaving an explosive number of possibilities of point mutations and their combinations. Both nationally and internationally, there is a dearth of systemic studies exploring the possibility of universal representation learning from genomic datasets. DeepSEA (Zhou and Troyanskaya 2015) and DanQ (Quang and Xie 2016) are two prominent deep learning architectures which allow representation learning and supervised classification based on DNA sequence information. DeepSEA leverages chromatin profiling data to predict chromatin effects due to genomic alterations, thereby putting forth a strategy to prioritise functional variants, including the non-coding ones. DanQ implements a unique strategy to encode sequence data as a binary image, which is then subjected to convolutional layers, followed by long short-term memory networks (LSTMs). Both these methods can be generalised for diverse prediction tasks involving sequence data. For coding mutations, we developed an alternative strategy i.e. Aminoacid Switch Sequence Model (ASSM), wherein a mutation, along with the context nucleotides can be represented as numeric vectors learned from large scale exome sequencing data (Gupta et al. 2020). Such numeric representations were learned using “Skipgram”, a shallow neural network, typically used for word-embedding tasks as part of Natural Language Processing (NLP) (Guthrie et al. 2006). Further we used LSTMs for classifying coding variants into two categories “related to cancers”, and “not related to cancers”. Coding mutations constitute a minor fraction of mutations in human cancers. As such, there is an urgent need to accommodate non-coding mutations under a similar framework. Some recent reports have suggested transfer learning workflow for single cell expression data, where deep neural networks and matrix factorisation based methods have been used for single cell data denoising, and clustering respectively (Stein-O’Brien et al. 2019; Wang et al. 2019). However the potential of word embedding techniques have not been exploited. Moreover, there has been no effort in inter-linking genomic alterations and expression based cellular phenotypes.

- **Aims and Objectives**

As discussed above, there is a clear gap in terms of availability of computational frameworks that enable transfer-learning from large scale genomic data. For this project our focus will be on evaluating multiple cognitive computing and deep learning based methods for constructing pre trained neural

networks/embeddings, which can be repurposed for precise delineation of functional heterogeneity in cancer. To this end, below we enumerate the main aims and objectives of the project.

a. **Embedding genomic alterations in cancer**
    i. Ingesting a large number of coding and non-coding variants to learn universal representation of mutations.
    ii. Using these embeddings under supervisory learning settings, to detect functional variants, with their potential influence on gene expression.

b. Transfer learning from single cell gene expression profiles
    i. Ingesting large volumes of expression data to train reconfigurable neural networks drawing inspiration from word embedding technologies.

c. Characterisation of functional heterogeneity in gallbladder tumors
    i. Cross linking embeddings of genomic alterations and gene expression for precise delineation of cellular heterogeneity in cancer. As a model we will use Gallbladder cancer, a demographic challenge with highest concentration in India and under-representations in global cohort studies such as TCGA.

- **Methodological advancement.** The very concept of pre-trained model stems from machine learning and vision research, and highly underutilised in the genomics field. Generating sequence data of cancers is a resource intensive process. Most of this data is publicly available and centrally archived at a handful of repositories (such as NCBI GEO, EMBL-EBI ENA). As such there is a both an urgent need and opportunity of harnessing the power of the existing sequence data. This project is targeted to evaluate and exploit the various neural network architectures, which allow learning from sequencing and the expression data. As a proof of concept study, we have developed a machine learning workflow for learning continuous representation of coding variants leveraging large scale public databases of exome sequencing studies. For this we developed Aminoacid Switch Sequence Model (ASSM) to represent mutations as a sequence of aminoacid switches, factoring in the context nucleotides. Continuous representation (300 sized numeric vectors) of 640 possible aminoacid switches were learned by ingesting variant data from ~60,000 exomes, using Skipgram (a commonly used single layer neural network in Natural Language Processing) (Guthrie et al. 2006). As an application, using this, we for the first time could recognise somatic mutations without any matched normal samples. We made use of a deep neural network architecture involving LSTM for the same. Notably the unavailability of matched normal samples hinders the characterisation of unknown variants in clinical settings (Gupta et al., bioRxiv, 2020). This work has drawn significant scientific attention from the cancer genomic community, and opened up multiple doors for international collaborations. This inspired us to generalize our approach to accommodate non-coding mutations, which will be taken up as partof the current project.

- **Impact:** The proposed work is of applied nature with direct potential impact on the public health. Below we enumerate the major research outputs and potential industry partners who could be interested in commercialisation of the same. India has the highest cases of gallbladder cancers. Due to the lack of its global footprint, this cancer is under/not represented in international cohorts such as TCGA. Single cell RNA sequencing of gallbladder cancer will generate highly valuable resources that can provide numerousleads to targeted therapies. 5 companies, who can take interest in the software, as well as translational outcomes of the research are as follows. A. MedGenome, B. Strand Life Sciences, C. CareOnco Biotech Pvt Ltd., D. BioCon, E. Dr. Reddy's

Laboratory. Notably, PI Dr. Sengupta's laboratory has recently transferred a cancer liquid biopsy technology to CareOnco Biotech Pvt. Ltd., where IIIT-D retained a significant equity.

- **Finance:**

A. Recurring		**1st Year**	**2nd Year**	**3rd Year**	**Total**
1.Salaries/wages	Monthly Emoluments				
1 Ph.D. student @35,000 pm	35000	420,000	420,000	420,000	1,260,000
Total (A)		**420,000**	**420,000**	**420,000**	**1,260,000**
B. Non-recurring (Workstation)					
2. Consumables (Next Generation Sequencing, and single cell)		700,000	300,000	0	1,000,000
3. Travel		45,000	45,000	0	90,000
4. Contingency + Publication cost		300,000	150,000	100,000	550,000
5. Equipment		100,000	0	0	100,000
GPU-Enabled workstation		0	0	0	0
Storage + RAM expansion		0	0	0	0
Total (B)		**1,145,000**	**495,000**	**100,000**	**1,740,000**
Gross Total (A+B)		**1,565,000**	**915,000**	**520,000**	**3,000,000**

6. **Title: Assessment of drug efficacy in Indian maternity hereditary lineage: A pharmaco-genomicstudy**

Technology Innovation Hub (TIH) for Maternal and Child Public Health, IIIT Delhi, India

PI and Team (Add rows if necessary. List foreign academic/industry partners as well.)
Name: Arjun Ray
Designation: Assistant Prof.
Affiliation Address: IIIT Delhi CB department, IIIT Delhi
Email & Mobile No. : arjun@iiitd.ac.in , 9582948808

- **Context/Background**

India is the second largest market for internationally manufactured drugs. The human trials of these drugs are mostly done on European or American populations. The current healthcare system prescribes the same drug therapy regardless of the genetic diversity of the Indian population's maternal healthcareconcerns. Keeping the role of epigenetic in causing genetic variants in mind, these datasets are biased toward European populations and failed to address the Indian populations in these databases. Most ofthe research outcomes are based on these databases only and are most likely to be biased towards US and UK based populations.
We propose to do an exhaustive and comparative study of India-specific variants (using IndiGen data)on kinase proteins on maternal and child pair to decipher the efficacy of the drug-protein interaction, owing due to hereditary genetic variations. The project shall involve studying the structural effect of these variants in protein function, stability, protein-ligand binding and disease. Our study shall elucidatecritical insights into the structural and functional changes caused by single nucleotide variants includingchange in structure stability, binding and pathway enrichment analysis. Finally, drug-design accompanied with experimental validations shall have the potential to propose novel drug molecules for personalized medicine purpose.

- **Aims and Objectives**

  - Using publicly available dataset, to identify the critical variations that can destabilize kinase-drug interaction for genetic disorders for Indians and create a database for thesame.
  - Using machine learning techniques, to propose alternative drug modifications to minimize the possible adverse drug reactions in the offspring.

- **Strategy**

As protein kinases are the second most targeted group of drug targets after the G-protein coupled receptors (Bhullar et al., 2018), for this study our approach includes druggable genes of Kinase's family. There are 538 protein Kinases a human genome encodes. Many of these Kinases are associated with maternal and pre-natal health disorders. Our project strategy can be summarized into two phases:
Phase I : In-silico Assessment of drug efficacy in Indian population.
Phase II: Customized drug optimization and discovery in-lieu of Indian variations

Salient steps:
Phase I
● Assembling a comprehensive druggable kinase gene dataset for the maternal and child pair.
● Preprocessing of Annovar output from Indigen and other genomic repositories.
● Processing master data for sequence- and structure-level analysis
● Structural modeling of the variants using advanced homology/ ab-initio modelling techniques.
● Virtual screening of the proteins against its associated drug partners to ascertain any difference in binding due to the variations.

Phase II

- Structure-activity relationship (SAR) and quantitative structure-activity relationship (QSAR) modeling
- Experimental validation of the structures of variant proteins.
- Kinase activity validity with experimental validation
- High-throughput experimental screening using multiple assay.

- **Risk analysis**

In this proposed project, we aim to assess the implications of India specific variations in the context of drug efficacy. Being a collaborator of the IndiGen consortium, the lab has the rare access to all the raw read files of the genomic variation data, acquired from more than a thousand people from a heterogenous population sample of the country. The potential insights shall have immense implications in the pharmaceutical industry. We aim to commercialize both the datasets, as well as the software workflows which we will generate during the funding period. Moreover, targeting of the kinase proteins, the second most druggable gene, makes any potential discovery highly desirable.

Last, depending on the novelty, and the immense amount of resources and creativity involved, we areconfident that the results obtained from this project will generate more revenues than its mere investment.

- **Budget**

A. Recurring					
1.Salaries/wages	Monthly Emoluments	**1st Year**	**2nd Year**	**3rd Year**	**Total**
Designation and number of persons					
Phd Student (1) @ 35000/-	35,000.00	-	420,000.00	420,000.00	840,000.00
Total (A)		**-**	**420,000.00**	**420,000.00**	**840,000.00**
B. Non-recurring					2,160,000.00
2. Consumables:-					
Primary Antibodies, HRP-conjugated Secondary Antibodies, Kinase Activity Assay Kit, Kinase Inhibitors, Experimental Validation cost		400,000.00	100,000.00	-	500,000.00
3. Travel		10,000.00	50,000.00	-	60,000.00
4. Contingency + Publication cost		300,000.00	300,000.00		600,000.00
5. Equipment					
GPU-Enabled workstation		900,000.00	-		900,000.00
Storage + RAM expansion		50,000.00	-	50,000.00	100,000.00
Total (B)		**1,660,000.00**	**450,000.00**	**50,000.00**	**2,160,000.00**
Gross Total (A+B)		**1,660,000.00**	**870,000.00**	**470,000.00**	**3,000,000.00**

7. **Title: Cognitive computing approach based analysis of genomic information in blood plasma for diagnosis and prognostic for cancer**

PI and Team

Project investigators

Name	Designation	Affiliation	Address	Email & MobileNo.
Dr Vibhor Kumar	Assistant professor	**Department of computational Biology, IIIT Delhi**	A-304 (R&D Block), IIIT Delhi, Okhla Ph. 3, New Delhi	**vibhor@iiitd.ac.in mobile: 9205424866**
Dr Sunil Kumar	Additional Professor	**Surigical Oncology section, AIIMS Delhi**	Institute Rotary Cancer Hospital All India Instituteof Medical Sciences, New Delhi 110029	**sksunilkr1976@gmail.com mobile: 9968300241**

- **Context and background**

Intelligent diagnostics and prognostics for healthcare need cognitive computing and social sensing. This proposal is about developing smart diagnostics and prognostics for diseases especially cancer. Instead of relying of regular biopsy and staining based tests for detection of cancer and follow up aftersurgery, we aim to use intelligent methods which can predict tissue of origin and presence of cancer using molecular signature in cell free DNA and RNA in blood. There is need for non-invasive technique to detect early or growing tumor stage. Patterns in DNA, RNA and Proteins in cell-free components has the potential to reveal cancer stage, however they also offer challenge in terms of computational analysis.

- **Problem to be addressed:**

We aim to use intelligent methods which can predict tissue of origin and presence of cancer using molecular signature in cell free DNA and RNA in blood. However unlike traditional we do not rely on single pattern to make decision. However Combining multilple types of patterns from cell free DNA and training on a data-set of patient does not warranty correct detection of disease. The reason is the batch effect in data-sets from different labs***. Hence we would be using adaptive transfer learning which would also involve self-learning (a type of cognitive approach) to make a prediction system which can predict disorders from cf DNA of patients.***

- **Aim**

Hence the objective related with our aims of our proposal is to

a) Develop and validate sensitive computational methods for detection of cancer.

b) Develop cognitive computation based models to test efficacy of drug on cancer patient and for theirfollow up using cell-free genomic components in blood plasma

c) Extend our approach of exploiting signals in cell-free DNA and RNA for detecting other disorders

also.

d) Collaborate with diagnostic agencies/companies to implement our methods for practical use.

Social sensing components

Curating existing published data : Our project would involve a lot of data curation and collection of published cfDNA profile. First we will develop our computational methods on available data. For this purpose we will download DNA methylation profile of cfDNA from different studies. We would also be curating data-sets of fragmentation pattern of cfDNA

Collecting data from patients

It would be important to test our computational approach on data-sets generated by us for cancer patients in India. At Aiims (under supervision of Dr Sunil Kumar) the blood and biopsy samples would be collected at AIIMs and for some sample cfDNA or cfRNA would also be extracted at AIIMs. The library preparation for sequencing cfDNA would be done at IIIT Delhi, in the laboratory of computational biology department. The cost of sequencing would be also be paid by group at IIITDelhi.

Computational Approach:

First we would apply reference-free deconvolution methods. However such methods are also affected by batch effect in data. The predictive system trained on a data-set might not work efficiently on data-set generated from data from other labs. Here data is cell-free DNA methylation profile, however the concept remains similar. We would developed self-learning systems which could combine multiple types of pattern in cfDNA with adapative approach to achieve accurate prediction of disorders. Hence combining deconvolutional method with adaptive transfer learning will be done to find non-hematopoietic source of cfDNA. However adaptive transfer learning itself may need small amount of data from target lab. Here we would use self-learning based on multiple pattern to make an initial guess of strong positive and negative for adaptive transfer learning.

- ***Outcome***

1. Tools to predict risk and occurrence of cancer and other diseases using cfDNA and cfRNA profile
2. Knowledge bout markers and techniques tom improve sensitivity of diagnosis using liquid biopsy
3. Training of PhD students
4. Cheaper non-invasive technique for detecting cancer

Overall: Budget

Item		Budget			
A.Recurring		**Ist Year**	**2nd Year**	**3rd Year**	**Total (In Rupees)**
1. Salaries/wages					
Designation & number of persons	Monthly Emoluments				
SRF-1 **(IIITD)**	35000	420,000.00	420,000.00	210,000.00	1,050,000.00
Laboratory Assistant/Technician **(AIIMS)**	25000	300,000.00	300,000.00	150,000.00	750,000.00
2.Consumables **IIITD**		600,000.00	100,000.00	50,000.00	750,000.00
Consumables **AIIMS**		100,000.00	100,000.00	50,000.00	250,000.00
3. Contingency **IIITD**		40,000.00	30,000.00	-	70,000.00
Contingency **AIIMS**		50,000.00	50,000.00	30,000.00	130,000.00
B.Non-Recurring-Equipment					
Total (A+B) IIITD		**1,060,000.00**	**550,000.00**	**260,000.00**	**1,870,000.00**
Total (A+B) AIIMS		**450,000.00**	**450,000.00**	**230,000.00**	**1,130,000.00**
Gross Total		**1,510,000.00**	**1,000,000.00**	**490,000.00**	**3,000,000.00**

8. Building an Intelligent Platform for Childhood Pneumonia and TuberculosisSurveillance

PI and Team

Principal Investigators:

A. Tavpritesh Sethi (Assistant Professor, IIIT-Delhi, TavLab, Dr. Tavpritesh Sethi)
B. Shama Karkal (Chief Executive Officer, Swasti | Swasti)

Alignment with grand challenges: (i) SDG 3.2: By 2030, End preventable deaths of newborns and children under 5 years of age, and (ii) SDG 3: By 2030, EndTB

Summary of our proposal: Respiratory illnesses are the leading cause of mortality in India, especially in children. India accounts for 20% of global deaths due to childhood pneumonia. Inaddition, India also leads in childhood mortality due to Tuberculosis (TB) with more than 22% of global burden and 2,50,000 children died of TB in 2017[1]. Our proposal takes a proactive approach to identify, track and predict pneumonia and TB starting with neonates and young children. Our solution is to build an AI enabled digital platform for primary health care and is aligned with the maternal and child health and the social sensing verticals of TiH-Anubhuti. Since 2002, Swasti has mobilized, coordinated and trained community & State actors in TB andHIV responses and is currently working with 20,000 families in 4 States on prevention, treatment and social determinants of health, in the primary healthcare context. However, currently there are no platforms in India that include causal temporal analytics for proactive identification and prediction of respiratory infections. Therefore, this is not only a health challenge, but also a social computing challenge as an interplay of various social and geographical factors. The partnership between Swasti and IIIT-Delhi is complementary. Swastihas created a Digitally Inclusive Community Engagement Flow, DICEFlow, an extensible platform for creating primary care workflows, using visual coding technologies that is being deployed across various healthcare interventions in 6 States. IIIT-Delhi has developed multipleAI/ML based solutions and platforms for public health challenges including COVID-19. The overarching goal of this project is to build an active detection and prediction platform for childhood respiratory infections such as pneumonias and TB, working with 5000 low income households who are currently reached by Swasti through a primary health care program in Gurgaon, Haryana.

- **Objectives:** (1) Onboarding of families and monitoring primary health care visits on DICEFlow. The target beneficiaries are approximately 2500 mothers and children(2) Longitudinal modeling, Causal modeling and Contextual AI models for capturing and predicting spikes in respiratory illnesses including MDR-TB in the pediatric age group. (3) Deployable analytics platform for childhood pneumonia and TB surveillance

- **Technology Deliverables:** (1) Prediction models for respiratory illness outbreaks including viral pneumonias and community increase in TB incidence or drug resistance. Changepoint detection using validated methods[2],[3] and novel deep learning based approaches will be

[1] "India Tb Report 2019 - Central TB Division." 7 Feb. 2019, https://tbcindia.gov.in/WriteReadData/India%20TB%20Report%202019.pdf. Accessed 27 Feb. 2021.
[2] "Package 'changepoint' - CRAN." 4 Oct. 2016, https://cran.r-project.org/web/packages/changepoint/changepoint.pdf. Accessed 8 Sep. 2020.
[3] "Comparison of change point detection methods." 28 Sep. 2019, https://www.marinedatascience.co/blog/2019/09/28/comparison-of-change-point-detection- methods/. Accessed 8 Sep. 2020.

carried out in Python and R. (2) NDHM compliant plugins for integration of these models intoplatforms including DICEFlow (3) Novel algorithms for addressing missingness and integrationof heterogeneous public health datasets will be implemented. (4) Custom libraries for causal analytics and policy learning will be developed in Python and R along with plugins for deployment in the DICEFlow platform and will be made available as APIs to the DICEFlow platform for integration at no cost.

- **Milestones and Timelines:**

0 - 6 Mo: Onboarding 2500 mothers and children on the DICEFlow platform

6 - 12 Mo: Data Processing and exploratory data analytics (IIITD)

12 - 30 Mo: Design of plugins, applications, user analysis, predictive analytics (IIITD, Swasti)

30 - 36 Mo: Feedback for deployment and implementation (Swasti)

- **Budget**

Item			**Budget**		**Total Rs.**
		Ist Year	2nd Year	3rd Year	
A. Recurring					
1. Salaries/wages	Monthly Emoluments				
JRF (IIITD)	Rs 38440/ month	461,280.00	461,280.00	461,280.00	1,383,840.00
Field Study (Personnel)	Rs 50000/ month	600,000.00	660,000.00	-	1,260,000.00
Total A		**1,061,280.00**	**1,121,280.00**	**461,280.00**	**2,643,840.00**
B. Non-Recurring					
1. Consumables		50,000.00	50,000.00	50,000.00	150,000.00
2. Travel		30,000.00	-	30,000.00	60,000.00
3. Contingency		50,000.00	50,000.00	45,000.00	145,000.00
Total B		**130,000.00**	**100,000.00**	**125,000.00**	**355,000.00**
Gross Total (A+B)		**1,191,280.00**	**1,221,280.00**	**586,280.00**	**2,998,840.00**

9. SAHAYAK: An Automated Complaint to First Information Report (FIR) Drafting

PI: Md Shad Akhtar, IIIT Delhi

Category: Legal Information Processing System

Local police stations are the gateways to initiate a legal investigation proceedings by reporting any crime or a misfortune event. The first information report (FIR) is a legal document containing essential information about the event. It is a fairly technical document and the use of legal jargons are common. However, a majority of the public are not used to such language. Therefore, while reporting an event they use the most commonly used terms to explain the events, and an expert (a police officer) translates their narratives into an FIR.

In recent years, many state police departments (e.g., Delhi Police[1], Bihar Police[2], etc.) have started the service of registering online complaints, where the complainants are asked to fill-in some predefined fields, such as name, contact details, etc., along withthe description of events. These platforms have two fundamental issues -

1) Similar to the offline setup, the manual intervention is required to prepare an FIR from the complainant-provided description. This often limits the throughput to handlea long list of online complaints.
2) Another critical issue is the absence of essential information in the description, whichrequires a series of back-and-forth communication between the complainant and theofficer to finalize the FIR, thus a delay in processing the complaint is undeniable.

To this end, we propose, SAHAYAK, an automated system that analyzes the complaint description, highlights the essential-but-missing information to the complainant, and generates the corresponding FIR once all essential information is acquired. More specifically, the system will operate in two phases:

- **Analysis:**
  - **Event identification:** The first task in the analysis phase is the identification of the event/crime reported by the complainant. This phase is crucial as different types of events require different essential information. For example, a report for theft of an automobile may require the vehicle identification number, make and model of the vehicle, etc.
  - **Entity extraction and Relationship identification between entities:** This module will be responsible for extracting all the entities from the description and establishing relationships among the identified entities. Moreover, it will also extract the anaphoric and cataphoric relationship between the entities.
  - **Semantic-role labeling with the event:** Once the entities and relationships are identified, the semantic-role labelling will map the entities to the event as per their role in the event. For example, a report for a violence and associated injury needs the information of any equipment that was used; therefore, relating guns, sticks, etc. to the event as equipment is an important piece of information. The resulting output (a graph) will be matched with the predefined template of the identified event and any missing information will be reported back to the complainant and request for revising the description.

[1] http://www.delhipolice.nic.in/citizen_services.html

[2] http://biharpolice.in/Registration/OnLineRegisterComplaint.aspx

- **Generation Phase:**
  - **FIR Draft Generation:** After collecting all essential information, the generation module will generate a cohesive FIR, which will be used for further legal proceedings.

Languages: In Indian context, the description can belong to different regional languages. Initially, we propose to develop the above modules to handle three languagesetups -- English, Hindi, and code-mixed Hinglish.

Duration: 2 years

- **Deliverables and Timeline:** At the end of the project, SAHAYAK will be developed as amobile/web-based application. It will be developed as per the following timeline:
  - [1-6 months]: Data Collection, Analysis, and Annotation.
  - [7-9 months]: Event or Type of Crime identification.
  - [10-13 months]: Entity Extraction and Relationship Identification.
  - [14-17 months]: Semantic-Role Labelling and Template Matching.
  - [18-22 months]: FIR Draft Generation.
  - [23-24 months]: System Integration and Mobile/web application development

- **Budget for 2 years:**

Item		**Budget**			
Recurring-A		**Ist Year**	**2nd Year**	**3rd Year**	**Total (In Rupees)**
1. Salaries/wages					
Designation & number of persons	Monthly Emoluments				
Phd-1	35000	420,000.00	420,000.00	420,000.00	1,260,000.00
RA-1	25000	300,000.00	300,000.00	150,000.00	750,000.00
Total (A)		**720,000.00**	**720,000.00**	**570,000.00**	**2,010,000.00**
Non-Recurring-Equipment-B					-
Server		650,000.00	-	-	650,000.00
NVIDIA GPU Card		200,000.00	-	-	200,000.00
Laptop		140,000.00			140,000.00
Total (B)		**990,000.00**	**-**	**-**	**990,000.00**
Gross Total		**1,710,000.00**	**720,000.00**	**570,000.00**	**3,000,000.00**

10. **Title of the Proposal**: An Empathetic Knowledge Grounded Conversational System for Mental Health Counseling and Legal Assistance

PI: Dr. Asif Ekbal, IIT Patna Collaborator: Dr. Shad Akhtar, IIIT Delhi

- **Aims, Objectives and Deliverables**

The proposal aims at developing an Empathetic Knowledge grounded Chatbot for Mental Health Counseling and Legal Assistance. The Chatbot will be multilingual in nature (English and Code-mixed Hinglish), and will exploit background knowledge present in the knowledge graph and/or knowledge base, and conditioned on user-specific information, domain-specific information and empathy. This is the very first attempt to make a conversational agent in this particular domain.

The Chatbot will be assisting the victims, especially the women and children, who have faced harassments or abuses from various sources (e.g. office harassment, domestic violence and cyberbullying etc.), by providing mental counseling and legal assistance.

The Chatbot will be open-source and pluggable with the following features:

i. The Chatbot will be multilingual in nature, supporting English and Hinglish (Code-mixed Hindi and English).
ii. The Chatbot will provide the stakeholders (e.g. victim, parents, humanitarian agencies) the following information: counseling in an empathetic way so that the victims feel more relaxed and stress-free; legal assistance by making the victims aware about various steps of filing the complaints to the legal cell, about the various IPC sections pertaining to the particular case etc.
iii. The conversational system will exploit background knowledge in the form of knowledge base to generate more informed and accurate information.
iv. The Chatbot will be empathetic in nature because of the sensitivity of the domain, and should be able to converse with the users by understanding various degrees of emotions.
v. We shall launch the Chatbot's service through Facebook messenger, WhatsApp, and Web based service.

- **Technology & Research**

- **Data Collection and Annotation:** Domain specific conversation data is required to develop the Chatbot. We aim at collecting data from various case studies, reports, and different internet sources. We will also followa semi-automatic strategy for conversational data preparation. We shall collect the FAQ, and design dialogue templates to create the conversational data. A basic bot would be created using an existing framework, and thenthis would be made available to create more conversations.
- **External Background Knowledge Acquisition:** The Chatbot will be enriched with external knowledge. The external knowledge will be gathered from the various reports, government websites, sites of the humanitarian agencies etc. The external knowledge will be used as a background knowledge to the conversational bot. This will be utilized for more informed and accurate response generation from the conversational system.
- **Natural Language Understanding (NLU)/ Spoken Language Understanding (SLU):** This is the very first step of every dialogue agent. There are three specific sub-modules, namely Dialogue Act Classification (DAC), Intent Detection (ID) and Slot Filling (SF). To this end, we will adopt our existing deep learning basedend-to-end hierarchical multi-task model that can jointly perform both intent detection and slot filling tasks for the datasets of varying domains [6-7].
- **Dialogue Manager (DM):** Once the pre-processing of the utterance/s is over it would be forwarded to the dialogue manager. The overall conversational loop will be under the supervision of the DM which controls the various phases (welcome, information seeking, feedback, bye).
- **Natural Language Generation (NLG):** Natural language generation (NLG) is an important module in dialogue systems, question answering systems, and many other natural language interfaces. To help the users achieve their desired goals, response generation provides the medium through which a

conversational agent is able to communicate with its user. We will adopt our already developed NLG system [2,7] for our task.

- **Empathetic Dialogue Generation:** Empathetic dialogue generation plays an important role in engaging theusers in long conversation with the Chatbot. As our domain is related to assisting the victims through legal assistance and mental health counseling, incorporating empathy into the conversational system will be an important feature. We shall adapt our prior works on sentiment and emotion in dialogues [8], generating courteous or polite natural language responses [7], personalized response generation frameworks [10] and sentiment and emotion-controlled dialogue generation [11] for our tasks.

➢ The project is planned for 24 months.

[M1] Months 1-3: Data collection and Annotation
[M2] Months 4-6: Data collection and Annotation to continue; FAQ template creation; Template generation for multi-turn dialogues
[M3] Months 7-9: Data collection and Annotation to continue; Background Knowledge Acquisition; A Prototype Chabot (English) with limited functionalities; Integration with WhatsApp, Facebook Messenger Integration and Web interfacing.
[M4] Months 10-12: Conversational data creation (English) through the Prototype Chatbot; Background Knowledge Acquisition to continue; Chabot with knowledge base encoding; Emotion and Intensity annotation
[M5] Months 13-15: Updated version of English Chatbot; Chatbot with external knowledge augmentation; Emotion annotation to continue; Emotion incorporation in the model
[M6] Months 16-18: Machine Translation of English conversational data to Hindi; Post-editing; Code-mixed English-Hindi conversational data creation; Refinement of English chatbot
[M8] Months 19-21: Code-mixed English-Hindi conversational data creation; Prototype Chatbot for Hinglish; Feedbacks on English Chatbot and improving its quality
[M9] Months 22-24: Updated version of Chatbot in Hinglish; Integration with WhatsApp, Facebook Messenger Integration and Web interfacing

➢ **Finance**

Item		**Budget**			
A. Recurring		**Ist Year**	**2nd Year**	**3rd Year**	**Total (In Rupees)**
1. Salaries/wages					
Designation & number of persons	Monthly Emoluments				
Senior Research Fellow -1	35000	420,000.00	420,000.00	420,000.00	1,260,000.00
Post Gradutate Fellowship /Annotator/JRFs-1	28000	336,000.00	336,000.00	336,000.00	1,008,000.00
2. Contingency		100,000.00	100,000.00	100,000.00	300,000.00
3. Consumables		50,000.00	100,000.00	100,000.00	250,000.00
4. Travel		-	-		-
Total		**906,000.00**	**956,000.00**	**956,000.00**	**2,818,000.00**

B.Non-Recurring-Equipment (Laptops)		182,000.00	-	-	182,000.00
Gross Total		**1,088,000.00**	**956,000.00**	**956,000.00**	**3,000,000.00**

Relevant Publications of PI:

1. M. Firdaus, A. Ekbal and P. Bhattacharyya (2021). More the Merrier: Towards Multi-Emotion and Intensity Controllable Response Generation. In AAAI 2021 (accepted)
2. D. Vershney, A. Ekbal and P. Bhattacharyya (2021). Modelling Context Emotions using Multi-task Learning for Emotion Controlled Dialog Generation. In EACL 2021 (accepted)
3. M Firdaus, A Pratap Shandeelya, A Ekbal (2020). More to diverse: Generating diversified responses ina task oriented multimodal dialog system. PloS one 15 (11), e0241271
4. Z. Ahmad, A. Ekbal, S. Sengupta, A. Mitra, R. Rammani, P. Bhattacharyya (2020). Active Learning based Relation Classification for Knowledge Graph Construction from Conversation Data. In Proceedings of the 27th International Conference on Neural Information Processing (ICONIP), pp. 617- 625, Nov 18-22, Bangkok.
5. M. Firdaus, A. Kumar, A. Ekbal and P. Bhattacharyya (2019). A Multi-task Hierarchical Approach for Intent Detection and Slot Filling. *Knowledge based Systems*, Elsevier, https://doi.org/10.1016/j.knosys.2019.07.017.
6. M Firdaus, H Golchha, A Ekbal, P Bhattacharyya (2020). A deep multi-task model for dialogue act classification, intent detection and slot filling, *Cognitive Computation*, Springer, 1-20.
7. Hitesh Golchha, Mauajama Firdaus, Asif Ekbal, Pushpak Bhattacharyya (2019). Courteously Yours: Inducing courteous behaviour in Customer Care responses using Reinforced Pointer Generator Network. In *Proceedings of the HLT NAACL 2019*, pp. 851-860. 2019.
8. Dushyant Singh Chauhan, Md Shad Akhtar, Asif Ekbal and Pushpak Bhattacharyya, *Context-aware Interactive Attention for Multi-modal Sentiment and Emotion Analysis*, In *Proceedings of EMNLP-IJNLP 2019*, Hong-Kong, China, Nov 3-7, 2019
9. Hardik Chauhan, Mauzama Firdaus, Asif Ekbal and Pushpak Bhattacharyya (2019). Ordinal andAttribute Aware Response Generation in a Multimodal Dialogue System. In *Proceedings of Association for Computational Linguistics (ACL), pp. 5437-5447*
10. Mauajama Firdaus, Naveen Thangavelu, Asif Ekbal and Pushpak Bhattacharyya (2020). Persona aware Response Generation with Emotions. In Proceedings of the 2020 International Joint Conference on Neural Networks, pp. 1-8.
11. M. Firdaus, H. Chauhan, A. Ekbal, P. Bhattacharyya (2020). EmoSen: Generating Sentiment and Emotion Controlled Responses in a Multimodal Dialogue System (2020). IEEE Transactions on Affective Computing, IEEE
12. M. Firdaus, N. Thakur and A. Ekbal (2021). Aspect-Aware Response Generation for Multimodal Dialogue System. ACM Transactions on Intelligent Systems and Technology (TIST) 12 (2), 1-33

11.Proposal Title: A Tool to Detect Privacy Leaks in Deep Learning Models

- **PI and Team** (Add rows if necessary. List foreign academic/industry partners as well.)

Technology Innovation Hub (TIH) for Cognitive Computing and Social Sensing, IIIT Delhi, India

[A Section-8 Company: iHub-Anubhuti Foundation]

[Please read the executive summary carefully before filling up this form]

Format of the Detailed Project Report (DPR)

Name	Designation	Affiliation	Address	Email & Mobile No.	Website Link
Dr. Vikram Goyal Principle Investigator	Associate Professor &HoD	IIIT Delhi	Indraprastha Institute of Information Technology(IIIT), Delhi, Okhla Phase-III, Near Govindpuri Metro Station, New Delhi-20	vikram@iiitd.ac.in	www.iiitd.ac.in/vikram
Dr. Chetan Arora Principal Investigator	Associate Professor	IIT Delhi	IIT Delhi Hauz Khas	chetan@cse.iitd.ac.in	http://www.cse.iitd.ernet.in/

C. Perceived Applications Theme

Cognitive computing and Social Sensing

D. Activity Focus:

- Knowledge Generation **√**
- Technology Product Development **√**
- Skill Development **√**
- Collaboration (National/International) **√ (National)**

Proposal details as per the DPR Requirement

- **Context/Background**

In the last decade, Machine learning has progressed rapidly. The driving factors that enabled it are availability of fast computing technologies and unprecedented growth of data. The performance of the ML techniques has made their use ubiquitous almost across all the domains. Specifically, the Deep learning (DL) models have shown their great potential due to their high accuracy. On the other hand, these models work like a black box and it is difficult to interpret them. Hence, it raises questions regarding the integrity, bias and fairness of these models. To address these questions, we aim to study different attacks on Deep learning models and develop a tool that would score ML models in different aspects. The tool

would help an end user agency to take an informative decision regarding the deployment of models in real settings.

The tool fits perfectly in the cognitive computing settings as DL models are one of the main building blocks for Cognitive computing and we aim to investigate privacy issues in DL models.

1. **Problems to be Addressed** (1500 words) [In view of the international and nationalstate-of-the-art]

Problem: Our problem is to design an effective tool that can help to investigate different deep learning techniques with respect to their robustness for adversarial attacks and privacy.

Recently due to the great success of deep learning techniques, they are being used in numerous applications including image classification, language translation, voice synthesis. Many applications being life crucial and related to the nation security, the recent studies find that DL is vulnerable to various attacks, e.g. a minor input perturbation may lead to wrong output, membership inference attack where one can identify an object used to train a DL model, model inversion attack where the objective is to infer missing attributes of an input object, model stealing attack where an adversary tries to learn a DL model parameters or knowing what features of an input object contributed to a specific decision.

We aim to study and develop privacy and security attacks on different types of DL models in thisproject. The discovered attacks would help users like state defense agencies or crucial service provider to weigh pros and cons before deploying a particular DL model. The work would also help the DL model developers to modify their models so as to increase their robustness and performance with respect to the discovered attacks. The knowledge acquired for performing various attacks on DL models would be subsequently used to create a tool to assist end users.

We plan to uncover attack surfaces in different settings. Some examples are as follows:

a) Is it possible to infer input records used in training a DL model when we know only the model input and output domains? The answer to this question leads to a privacy attack. There can be scenarios where a user had shared her information for a purpose other than allowing one to use it for a training model. Such inference attack allows a user agency to be careful about future lawsuits and avoid sabotage of its reputation. The same question can also be asked in scenarios where the adversary gets access to a model's parameters. Itmakes the adversary more powerful and the probability of inference increases.
b) Is it possible to fool a DL model to classify an object wrongly by minor feature perturbation? This is one of the most studied problems in literature. We want to focus on universal perturbation methods that once devised would apply on each input objectinstead of devising perturbation differently for each object.
c) Can different outputs at two different times due to online learning effect leak information? Deep learning models deployed in real settings are retrained continuously inonline fashion when new data become available, as training from scratch is prohibitive due to high training cost. This attack surface leads to a privacy attack for the new updating set where the adversary can identify objects from the new updating set.

Since the current practice of using existing datasets mainly developed for training the modelsmay not give a true picture of the attacks, we also aim to devise new benchmarks specific to the problem of privacy and security attacks on the DL models.

- **Aims and Objectives** (500 words)

  - Investigate and develop privacy/security attacks on DL models for the scenarios when DL technique is available as a Black box. The adversary knows only the input and output domain of the model and is not provided with any model parameters.
  - Investigate and develop privacy/security attacks on DL models for the scenarios when the adversary is powerful and knows model parameters.
  - Investigate and develop privacy and security attacks on DL models when the adversaryhas access to the model during its training. It is the most powerful adversary scenario.
  - Develop different benchmarks to evaluate privacy and security vulnerabilities as wellaccess efficacies of newly proposed attacks by other researchers.
  - Design and develop a tool to detect privacy and security vulnerabilities.

- **Strategy** [What is the competitive advantage of the proposed technology over other alternatives?]

  A simple alternative strategy could be to simply ignore the privacy and security aspects of DL models used in various application systems. A serious drawback of this strategy could be catastrophic depending on the application scenarios. The performance of the system could degrade as a result of data poisoning attack and the system may be discarded by the users. In the case of a privacy leak, an end user may sue the organization which can result in huge monetary loss and trust degradation. Furthermore, it would become difficult to identify limitations of the system.

  The proposed strategy would not be able to completely eliminate all the above and similarproblems. But it would help to minimize these concerns.

- **Target Beneficiaries** (500 words) [Elaborate on your proposed research objectives societal impact, national level consequences. Name 5 companies who may be interested.Are you in contact with any/all of them?]

  The proposal is forward looking, and would be useful for all the states' as well as private institutions which will be using AI and specifically DL based models for collecting or analyzing the public data. The penetration is low at the moment but is widely expected to grow exponentially in future. The auditing of such systems in terms of their vulnerability to privacy/security attacks will be an important step for the production release of AI systems.

- **Technology & Research Methodology**

  The project is focused on investigating Deep Learning models with respect to their privacy guarantees. Asthe work would be mainly using Machine Learning platforms as the technology, the first task is to performliterature surveys to understand the working of Deep learning models. The Second step would be to generate hypotheses regarding performing attacks over a DL model. As an example, one hypothesis can be that the output of a DL model is highly correlated with the input data instances used for learning and not correlated with instances that are not used during learning. For each such hypothesis formed, a systematic study would be carried out across multiple DL models to accept or reject the Hypothesis. It would involve performing experiments with different datasets and statistically analysis of observations obtained through experiments. Accepting a hypothesis makes the possibility of performing an attack asfeasible.

Finally, all the knowledge thus collected would be used to implement APIs and a tool for investigating DL models.

- **Management**

There are two principal investigators working at two different places: a) IIIT Delhi and c) IIT Delhi. It is also proposed to recruit two numbers of Research Associates/PhDs (RA). Thearrangement of postings of research personnel shall be subjected to adjustments on need basisand can be finalized with mutual consent after the project implementation starts.

We also plan to publish research papers in reputed conferences/journals during the course of the project. We expect that students from IIIT Delhi, IIT Delhi and other Institutes will be workingon this project as interns.

- **Finance (in Lakh Rupees)**

Budget Head		**1st Year**	**2nd Year**	**3rd Year**	**Total**
A. Recurring		**IIIT-D**	**IIIT-D**	**IIIT-D**	
1. Salaries/wages					
Designation	Monthly Emoluments				
PhD student@1	35000	420,000.00	420,000.00	420,000.00	1,260,000.00
	10% HRA	42,000.00	42,000.00	42,000.00	126,000.00
2. Intern	10000	120,000.00	120,000.00	120,000.00	360,000.00
Total		**582,000.00**	**582,000.00**	**582,000.00**	**1,746,000.00**
3. Travel		10,000.00	60,000.00	64,000.00	134,000.00
4. Contingency		-	10,000.00	10,000.00	20,000.00
Subtotal -A (1-4)		**592,000.00**	**652,000.00**	**656,000.00**	**1,900,000.00**
B. Non-Recurring					
1. Laptops/Notebooks@1		100,000.00	-		100,000.00
2. GPU Cards		1,000,000.00	-		1,000,000.00
Total		**1,100,000.00**	**-**		**1,100,000.00**
Gross Total (A+B)		**1,692,000.00**	**652,000.00**	**656,000.00**	**3,000,000.00**

- **Time frame [Please include a PERT chart to show technology development phase andtranslational**

phase.]

	Task	Quarter											
		1	2	3	4	5	6	7	8	9	10	11	12
1	**Developing algorithm to detect privacy attacks on DL models in a white box/black boxsetting.**												
a	Recruitment of manpower												
b	Literature survey												
c	Implementation												
d	Validation and comparison with state of the art												
e	Manuscript communication and publication												
2	**Developing algorithm to detect privacy attacks on DL models in life-long learningsetting.**												
a	Recruitment of manpower												
b	Literature survey												
c	Implementation												
d	Validation and comparison with state of the art												
e	Manuscript communication and publication												

- **Cost Benefit Analysis [Mention about how your technology will benefit the applicationdomain/users with respect to what exists at present.]**

There does not exist any technology that scans DL models for various privacy and security vulnerabilities. However, there has been some recent research on the topic where researchers uncovered some attacks. We aim to exploit the existing research and uncover new vulnerabilities to finally develop a tool for the same. The tool can be associated with a set of APIs which can be provided to users and researchers on chargeable basis.

- **Risk analysis [Financial and economic cost-benefit analysis of the project should beundertaken wherever such returns are quantifiable and some outcomes/measurables should be suitable defined] (500 words)**

As such, we do not foresee any risk factors w.r.t. this project. We will closely monitor the progress of

the project both in terms of ideation and implementation to ensure that all project specifications are satisfied within deadlines and within the budget. Both the PIs (Dr. Vikram Goyal and Dr. Chetan Arora) have between them an adequate level of experience and expertisein managing and executing industry projects. Hence, we will use our industry experience for mitigating project management risks. Another important point that we have to mention is relatedto our previous large experience in working together. Both the PIs together successfully managed a Meity sponsored project and a DST SERB sponsored project.

- **Outcomes [Success criteria to assess whether the development objectives have beenachieved should be spelt out in measurable terms. Base-line data should be available against which success of the project will be assessed at the end of the project (impact assessment)): Please provide details as per Item 10 of Annexure II in the table below] (500 words)**

Components	Activity	Target
Technology Development	No. of Technologies (IP, Licensing, Patentsetc.)	1
	Technology Product	1
	Publications, IPR and other Intellectual Activities	2+
HRD & Skill Development	High End Skill Development	4
	Graduate Fellowships	0
	Post graduate Fellowships	0
	Doctoral Fellowships	2
	Post-Doctoral Fellowships	0
	Faculty Fellowships	0
	Chair Professors	0
Entrepreneurship & Startups	CPS-GCC - Grand Challenges and Competitions	0
	CPS-Promotion and Acceleration of Youngand Aspiring technology entrepreneurs (CPS-PRAYAS)	0
	CPS-Entrepreneur In Residence (CPS-EIR)	0
	CPS-Start-ups & Spin-off companies	0

	CPS-Technology Business Incubator (TBI)	0
	CPS-Dedicated Innovation Accelerator (CPS-DIAL)	0
	CPS-Seed Support System (CPS- SSS)	0
International Collaboration	International Collaborations	2

- **Evaluation [How you would like to be evaluated? What are the measures?]**

We will constitute a committee of experts from both academia and industry to periodically reviewthe progress of the project. We will also maintain communication with the Technology Innovation Hub and submit project progress reports, as required.

12.Proposal Title: Legal Text Simplification, Summarization and FAQ Retrieval

PI and Team: (Add rows if necessary. List foreign academic/industry partners as well.)

1. Rajiv Ratn Shah (PI), Assistant Professor, IIIT Delhi, rajivratn@iiitd.ac.in, 8800629998
2. Ponnurangam Kumaraguru (Co-PI), Professor, IIIT Delhi, pk@iiitd.ac.in, 9650122772
3. Junyi Jessy Li (Collaborator), Assistant Professor, UT Austin, jessy@austin.utexas.edu
4. Anisha Gopi (TL) and Sumeysh Srivastava (Sr. Resident Fellow), Nyaaya Foundation and Vidhi Center for Legal Policy, anisha@nyaaya.in, sumeysh@nyaaya.in (Collaborators)

- **Problem Description:**

India has a vast system of laws, but most citizens are either unaware of their rights under the laws or do not understand how to enforce them. The lack of simple, authoritative and accurate information that provides guidance on how to solve their legal problems along with multiple factors like a user's identity,income, social status, geographical location and level of education also affect their experience of interacting with the law and lead to a situation where the law can often become a tool of exploitation, rather than empowerment.

In trying to be exhaustive, the law gets filled with jargon and convoluted sentences. It is full of words, phrases, and terms like 'prima facie', 'proviso', 'notwithstanding', etc., only understood by lawyers therefore, making it inaccessible to many. The problem is further amplified because English is not the firstlanguage for most of the population. There are hundreds of laws and regulations at the Central and State level and they are modified often and the final interpretation of the law is based on judgements by the various courts across the country. The information is structured in a way that to get an accurate solution toa specific problem, laws have to be read together with amendments, judgements, rules and guidelines, creating a complex data set comprising of multiple documents. For example, the Indian Penal Code, which covers all the substantive aspects of criminal law in India and has a bearing on the lived realities of citizens, has 511 sections, multiple state and central amendments, and numerous judgements expanding on different aspects of the code.

For this information to be accessible and useful to an average citizen, it has to be processed together, reviewed and structured in a simple, jargon free language. Therefore, with this proposal we plan to solve the following challenges to alleviate the problems faced by millions in India everyday:

1. Legal Language Simplification and Summarisation - Our team in collaboration with Nyaaya has created a framework where information is collected from legislations, judgements and rules to provide Simple, Actionable, Recallable and Authoritative Legal (SARAL) information. However, thisis currently done manually by a team of editors. This is not scalable for the large number of Indian laws. We believe that an AI powered solution can help us to do this efficiently.
2. Translation of legal language - India has 22 scheduled languages and many more unscheduled ones. While a majority of the laws exist in the English language, 94% of the population does not speak English. Even amongst those who do speak English, their knowledge is not enough to understand thelegal nitty-gritties. It is observed that even the legal experts find it difficult to understand legal documents, it is unfair to assume a thorough understanding from the common literate. At the same time, state of the art systems like Google translate do not work for Indian language and much

less sofor legal text. Therefore, we plan to train models on already available legal texts to cover most of the22 scheduled Indian languages.

3. FAQ Retrieval - Taking the help of search engines for legal document information extraction is an error prone task. Therefore, most people have to rely on a human legal expert for the same. This process is difficult and expensive. With the able help from legal experts from Nyaaya and Vidhi Legal Center, we aim to annotate a set of legal documents with their similarity measures, and build models to provide the most relevant legal articles to a user as per their query, both in text using chatbot and speech using interactive voice response (IVR) system.

- **Technology & Research Methodology:**

1. Collection and curation of datasets suitable for each of the above stated problems. This would involvesearching, collection, parsing, cleaning and analyzing legal documents available in various public forums like the constitutional articles, schedules, sections, land deeds, state laws, court judgements, ministry notices and notifications.
2. Training models for Extractive and abstractive legal document summarisation and simplification anddesigning metrics suitable for both the tasks.
3. Dataset curation for similarity of legal documents by annotation from legal experts.
4. Training models for information extraction and recommendation of articles based on keywords,keyphrases, topical and other document similarity metrics.

- **Contribution to the grand problem:** The proposed project directly contributes to the grand problem of legal information processing and management. In particular, we address the topics of legal document simplification, summarization of legal judgement, legal data curation and modeling (from translation of legal documents in (most of the) 22 Schedule-8 Languages point of view), and chatbot for legal conversation. We plan to build a dashboard/ website/ applications to provide these services to users.

- **Budget**

Item		**Budget**			
Recurring-A		**Ist Year**	**2nd Year**	**3rd Year**	**Total (In Rupees)**
1. Salaries/wages					
Designation & number of persons	Monthly Emoluments				
Phd Student 1	35,000.00	420,000.00	420,000.00	420,000.00	1,260,000.00
Research Assistant 1	30,000.00	360,000.00	360,000.00	-	720,000.00
2. Consumables, Travel & Contingency		50,000.00	70,000.00	50,000.00	170,000.00

Total (A)		**830,000.00**	**850,000.00**	**470,000.00**	**2,150,000.00**
Non-Recurring-Equipment-B					
Server		660,000.00	-	-	660,000.00
1*3090 GPU		190,000.00	-	-	190,000.00
Total (B)		**850,000.00**	**-**	**-**	850,000.00
Total (A+B)		**1,680,000.00**	**850,000.00**	**470,000.00**	**3,000,000.00**

13. Linked data model for legal information processing system

PI and Team:

PI: P. Krishna Reddy, Professor, IIIT Hyderabad, pkreddy@iiit.ac.in
Co-PI: Mukesh Mohania, Professor & Dean, IIIT Delhi, mukesh@iiitd.ac.in
Co-PI: Anirban Mondal, Associate Professor, Ashoka University, anirban.mondal@ashoka.edu.in
Co-PI: Prof. Girish Agrawal, Professor, Jindal University, gagrawal@jgu.edu.in

- **Executive Summary:** The legal system has key stakeholders such as judges, lawyers, plaintiffs, defendants, government departments and corporate/business entities. Legal information typically includes both structured data (e.g., title of case, date/time of trial hearings, names of parties, i.e., plaintiffs and defendants, names of impleaders, if any, amicus curiae, names of judges and lawyers, etc.) as well as unstructured data (e.g., evidence, information relevant to the case, witness testimonies,out-of-court depositions by witnesses, court transcripts, etc.). Observe that the unstructured data can be multi-modal, i.e., text, images, audio and video. Moreover, legal information essentially entails heterogeneous data sources linked to each other. For example, different government departments, law firms and corporate entities may maintain information in their own formats and using different database/data warehousing technologies since there is generally no de facto standard for legal data storage formats and processing technologies.

Given the sheer complexity of data in legal systems, legal information processing systems need to be organized in a way such that the queries of various stakeholders of the legal system can be answered both efficiently and seamlessly across such multiple and heterogeneous linked data sources. Now let us consider some typical queries to obtain a better understanding of the requirements of a legal information processing system. A lawyer, plaintiff or defendant may issue the following types of queries: *"Retrieve all the judgments related to tribal land rights"*, *"Retrieve all the judgments given by the Supreme Court in the category of Medico-Legal matters"*, *"Retrieve all the judgments given by a specific Supreme Court judge in the category of Freedom of Speech during years 2010-2020"*, *"What kinds of arguments were made by the prevailing party in successful cases contesting land acquisition for highwaydevelopment?"* and so on.

Any legal information processing system needs to have the capability to cognitively understand, contextualize, discover, ingest, reason and infer about various kinds of data (possibly multi-modal) across multiple disparate and heterogeneous data sources. Moreover, the legal information processing system also needs to be capable of acquiring, recording, indexing, cleaning, integrating, curating, visualizing and disseminating the data appropriately to ensure data reliability as well as consumability. Furthermore, the cognitive capabilities of the system should include cognitive querying based on query context, user context & domain context, as well as cognitive computation services for cognitive ingestion of the data. For example, there would be a significant difference in query context between cases involving corporate accounting frauds versus cases involving real property disputes or breach ofcontract or divorce. Such differences in context arise primarily because different domains work in fundamentally different ways. Furthermore, new laws may emerge and/or old laws may be rescinded or superseded by new laws and so on. This further exacerbates the problem of search in legal information processing systems because the search query needs to be contextualized based on the lawsprevalent at the time of judgment. In essence, the legal data currently resides in traditionally

passive data stores, and we need to convert such passive data stores into cognitive data stores and cognitive knowledge bases.

For effectively handling complex user queries in legal information systems, we propose a linked data model, which is capable of linking multiple disparate and heterogeneous data sources and contextually answer a wide variety of complex queries, whose results would be valuable to various stakeholders of the legal system. We will search through the Relational database (containing structured data) and then execute search on the unstructured data (stored in MongoDB) using "contains" for performing effective keyword search. In particular, we will use faceted search to enhance the search mechanism with a well-defined set of filters to facilitate search navigation. In effect, this would facilitate stakeholders in applying various filters to quickly zero in on the search results that would be most valuable to them. For example, let us consider the following query: *"Retrieve all the judgments given by Judge X while on the bench in the High Court of State Y"*. In this case, the filter categories for facets could relate to different types of cases e.g., lawsuits associated with land disputes, inheritance, and industrial wage claims. Facets could also relate to time, e.g., the judgments given by Judge X during a particular period of time or while at a particular court. Such faceted search and filter-based navigation can be extremely valuable to stakeholders in terms of obtaining targeted results.

Established legal research tools such as LexisNexis (https://www.lexisnexis.com/en-us/gateway.page) provide legal solutions & insights based on data analytics. In particular, Lexis DaaS (Data-as-a-Service) adds value to litigation analytics by providing APIs, which enables effective data-driven research into a wide variety of court cases. Similarly, a service such as Manupatra (https://www.manupatrafast.com/) provide a comprehensive databases of court cases, and is one of the market leaders in the field of enabling law practitioners towards digital transformation. It also has data analytics and visualization capabilities to facilitate obtaining actionable insights. However, in order to use the full suite of tools/technologies in these existing systems, the subscription fees are generally too expensive for mostsolo legal practitioners, and generally out-of-reach for most ordinary litigants in India. What this means in practice is that established senior lawyers and wealthy litigants can access these systems, but ordinary lawyers and litigants, particularly those practicing in small towns and district courts are neither able toafford such subscription fees nor could they afford to hire IT specialists for sorting through masses of legal documents. In this regard, *our vision is to develop an* ***easy-to-use***, ***cost-efficient*** *(i.e., either free or with nominal charges) and* ***open access legal information processing system****, which serves as a* ***one- stop platform*** *for catering to the requirements of various stakeholders in the Indian legal system.* Furthermore, our proposed legal information processing system is differentiated from existing systemsin that we will convert traditional passive data stores into cognitive data stores and cognitive knowledge bases.

- **Deliverables:**

- Discuss with key stakeholders to understand pain points & requirements
- Convert above requirements to technology requirements
- Identification of entities and the relationship between them and all constraints
- Develop a linked data model on top of MySQL and MongoDB; we will use both SQL and NoSQL databases.
- Deliver a platform where entities can upload data and stakeholders can search

We plan to publish research papers in reputed conferences/journals. We expect that students from IIITDelhi, Ashoka University, IIIT Hyderabad and other Institutes will be working on this project as

interns. Prof. Mukesh Mohania, Prof. P. Krishna Reddy and Dr. Anirban Mondal will be providing the necessary data management expertise in our team. Prof. Girish Agrawal (who is a licensed Attorney atLaw with the degree of Juris Doctor from the UC Berkeley School of Law) will be providing domain expertise in the field of law in our team.

- **Budget – Initial Phase**

Item		**Budget**			
A. Recurring		**Ist Year**	**2nd Year**	**3rd Year**	**Total (In Rupees)**
1. Salaries/wages					
Designation & number of persons	Monthly Emoluments				
JRF*2	30000	720,000.00	720,000.00	720,000.00	2,160,000.00
2. Consumables		60,000.00	60,000.00	60,000.00	180,000.00
3. Travel		100,000.00	100,000.00	100,000.00	300,000.00
4. Contingency		50,000.00	50,000.00	50,000.00	150,000.00
Total (A)		**930,000.00**	**930,000.00**	**930,000.00**	**2,790,000.00**
B. Non-Recurring-Equipment					
Laptops*3 @ 50000/-		150,000.00	-	-	150,000.00
Total (B)		**150,000.00**	**-**	**-**	**150,000.00**
Gross Total (A+B)		**1,080,000.00**	**930,000.00**	**930,000.00**	**2,940,000.00**

14. Development of efficient homomorphic secret sharing schemes for applications inpublic health

PI: Dr. Anuradha Sharma (IIIT-Delhi)and

Co-PI: Dr. Vaneet Aggarwal (Purdue University, USA)

There is an increasing need to support privacy and security in today's data driven world, while inferring meaningful information from the data obtained from different domains, e.g., health, education, transportation, energy, etc.

- **Homomorphic Encryption:** Recently, use of encryption techniques, such as homomorphic encryption, has been proposed to allow necessary operations on data while maintaining the data secrecy. Craig Gentry (2009) provided the first plausible construction for a fully homomorphic encryption scheme using lattice-based cryptography. Later, Dijk et al. (2010) provided the second fully homomorphic encryption scheme that uses most of the tools developed in Gentry's work but does not require ideal lattices. It was noted that Gentry's original construction is of little practical significance. Later, several new constructions of fully homomorphic encryption schemes and implementation techniques to improve the efficiency of these schemes have been proposed. Such encryption schemes make it possible to develop programs that can run on encrypted inputs. By the virtue of working on encrypted input, i.e., such a program need not decrypt the input, such programs can be run by an untrusted party while maintaining privacy. Fully homomorphic cryptosystems have great practical implications in the outsourcing of private computations, for instance, in the context of cloud computing, mobile systems, etc. One example of such a scheme is that on a health data: one may want to find if a medical condition is covered under health insurance or not, without providing the details of the health condition to the hospital and insurance agency. With increasing use of data in public and private life, e.g. using biometric data to authenticate without revealing it, several such challenges will emerge.

- **Secret Sharing Schemes:** Similarly, the other challenge is to ensure that all parties agree for sharing common data, i.e., data cannot be accessed without the consent of the concerned party. A simple example is that whenever the biometric data of an individual is used, the person's consent is taken. For the same, secret sharing schemes have been developed using several approaches. In the secret sharing scheme, a secret is divided among a group of participants and the reconstruction of the secret is only possible when enough shares are combined. Therefore, in a secret sharing scheme, the individual shares are of no use on their own. Towards this, several probabilistic secret sharing schemes for compartmented access structures have been proposed and studied. In another direction, several constructions of secret sharing schemes have been proposed using linear error-correcting codes. In the era of cloud computing, new secret sharing schemes are being developed to allow data sharing while providing higher privacy and security for data. Homomorphic secret sharing schemes are secret sharing analogues of homomorphic encryption and have recently attracted a lot of attention. A major challenge that the secret sharing schemes face is the computational ability of devices as a good number of today's secret

sharing schemes cannot be applied to computationally weak devices. However, today a good amount of data is collected from mobile devices and IoT devices, which are computationally limited. Developing secret sharing schemes that can work on multitude of devices, including mobile and IoT, is a much-needed direction of research for enabling privacy and security solutions in a data-driven world.

- **Outcomes**: Our plan is to develop efficient homomorphic secret sharing schemes to provide data privacy and security solutions.

- **Budget:**

A. Recurring	**Year 1**	**Year 2**	**Year 3**	**Total**
*Postdoc (1)	600,000.00	600,000.00	600,000.00	1,800,000.00
Travel (Indian PI) Including TA/DA for stay up to 2 month per year	300,000.00	300,000.00	300,000.00	900,000.00
Contingency & Consumables	50,000.00	50,000.00	50,000.00	150,000.00
Total	**950,000.00**	**950,000.00**	**950,000.00**	**2,850,000.00**
B. Non-Recurring-Laptop (1)	150,000.00	-	-	150,000.00
Total	**1,100,000.00**	**950,000.00**	**950,000.00**	**3,000,000.00**

*Postdoc will be hired at IIIT-Delhi only.

15. Project Title: Ethics of Cognitive Computing and Social Sensing

PI and Team

Name	Designation	Affiliation	Address	Email & Mobile No.	Website Link
Manohar Kumar	Assistant Professor	IIIT Delhi	Okhla Phase 3, IIIT Delhi, Delhi-110020	manohar.kumar@iiitd.ac.in	https://iiitd.academia.edu/ManoharKumar
Jainendra Shukla	Assistant Professor	IIIT Delhi	Okhla Phase 3, IIIT Delhi, Delhi-110020	jainendra@iiitd.ac.in	https://jainendra.in/
V. Raghava Mutharaju	Assistant Professor	IIIT Delhi	Okhla Phase 3, IIIT Delhi, Delhi-110020	raghava.mutharaju@iiitd.ac.in	http://raghavam.github.io/
Daniele Santoro	Researcher	University of Minho, Portugal	Centre for Ethics, Politics, and Society, ILCH - University of Minho, Braga, Portugal	dsantoro@ilch.uminho. pt daxanto@gmail.com	https://uminho.academia.edu/DanieleSantoro

The current proposal seeks to study the ethical implications of Cognitive Computing and Social sensing. Cognitive computing systems that behave and act like humans and that are modelled on human beings pose significant opportunities and challenges. Super-intelligent systems can overcome the limitations of human imagination and thought and achieve outcomes hitherto inconceivable. The development alsoposes significant challenges for society: would these systems enable or control human beings? Would the decisions taken by such systems be fair and not reflect bias and prejudicial behaviour? Would they threaten human freedom or enable it? Would the decision procedures and the algorithms that enable them be accountable to human beings? Would they be predictable to ensure legal scrutiny? Would they respect the user's privacy? Who would be responsible in case of harm? Similarly, social sensing raises questions of profiling, manipulation, harm, privacy, ownership and control over data. The proposal will seek to study the ethical, societal, and legal challenges that emerge in the development of cognitive computing and social sensing, and seeks to inbuild ethics into the design and implementation. In addition, the project will demonstrate the mechanism of embedding ethics into the design of artificially intelligent systems by choosing a few use cases. This proposal will explore (but not limited to) the following questions:

(i) How do cognitive computing and social sensing affect the autonomy of individuals?

(ii) To what purposes should cognitive computing and social sensing be put to?

(iii) What implications on privacy do cognitive computing and social sensing have?

(iv) What concerns of fairness and justice arise in cognitive computing and social sensing?

(v) What concerns of trust, transparency, and accountability arise in the application of cognitive

computing and social sensing?
(vi) How can we address ethical concerns in the design of cognitive computing and social sensing? Howdo different designs affect privacy?
(vii) What regulatory and societal challenges emerge through the application of cognitive computing andsocial sensing?
(viii) To what extent should governments and public institutions be involved in the development andimplementations of cognitive computing and social sensing?

- **Technology and Research Methodology**

The first part of the project involves understanding the key ethical problems in Cognitive Computing and Social Sensing. To achieve this, first, a literature survey of the state-of-the-art will be conducted. Key ethical problems will be identified through an application of the major theories: consequentialism, deontology, virtue ethics, feminism, distributive justice, race theory etc. Ethical problems will be identified at the level of thinking and design and for its impact on society. Following this, in the second part, the team will work together to identify the level at which ethics can be integrated.

Finally, we will develop algorithms to let AI systems make ethical decisions, i.e., they should be able to evaluate all the possible ethical choices and take the best decision (ethically) based on the context. This would require knowledge of the domain and the possible ethical issues that emerge.The domain knowledge from the perspective of ethics can be captured in the form of an ontology that describes the concepts and their relationships. Since the number of examples involving ethical choices and their implications is pretty small, we plan to build a learning model by using inductive logic programming. This model learns from a set of positive and negative examples of ethical choices related to a particular domain under various contexts. The AI system should be able to compute an ethical score for the various ethical choices that it has at any given point of time and resolve the ethical dilemma while making a decision. An added advantage of using inductive logic programming is that it makes the decision and hence the system explainable, which is very important in critical decision making AI systems. We will demonstrate these algorithms by considering a specific use case related to healthcare and social robotics.

- **Relevance for iHUB Anubhuti**

This proposal is imagined at the intersection of ethics and AI and combines researchers from both disciplines. The results of the research will also be beneficial for the entire thematic area as findings can be used to develop user interfaces that are sensitive to the ethical concerns of the end-user and the regulatory challenges. In specific, the proposal by identifying use cases in healthcare (a core area identified for the TiH) and social robotics will demonstrate how ethics can be integrated in the design thinking stage itself. The proposal through policy papers, white papers, and workshops with industry will enlighten practitioners, government, and industry about the ethical and societal challenges of Cognitive Computing and Social Sensing. We believe that building ethics into the design will make the machines compliant with regulatory and societal guidelines. In the long run, it will differentiate between firms who integrate ethics and who don't, and possibly also between those who are profitable and those are not.

Recent studies suggest that the main barriers to ethical integration in cognitive computing and social sensing are not only technical but are related to different issues concerning stakeholders, such as user acceptability, lack of knowledge exchange, lack of norms etc. This project has identified the following key stakeholders and will work on identifying and addressing their requirements to reduce the market

barrier and building a stakeholders community around the project consortium:

1. Academic/Research Institutions
2. Regulatory/Standard Agencies
3. Government Agencies
4. Private Industries working in the area of AI and/ML

Following are the names of the six companies who may be interested in our work.

1. Google
2. Facebook
3. Microsoft
4. Walmart
5. IBM
6. Optum

We (PI and Co-PIs) have ongoing collaborations with some of these companies. We plan to leverage our contacts to discuss the proposed project and work with them to develop an ethical decision making framework that can be incorporated into their AI systems. By raising questions of accountability, transparency, and fairness the project will enable the design solutions to be just and fair. Finally, the project will add to the scant literature on the ethics of cognitive computing and social sensing.

- **Budget**

A. Recurring					
1.Salaries/wages	Monthly Emoluments	**Year I**	**Year II**	**Year III**	**Total**
Designation & number of persons					
Junior Research Fellow (JRF), 2	76,880.00	922,560.00	922,560.00	520,800.00	2,365,920.00
2. Consumables		10,000.00	20,000.00	10,000.00	40,000.00
3. Travel		40,000.00	60,000.00	60,000.00	160,000.00
4. Contingency		24,000.00	30,000.00	30,000.00	84,000.00
5. Other costs [Workshop]		-	250,000.00	-	250,000.00
Total (A)		**996,560.00**	**1,282,560.00**	**620,800.00**	**2,899,920.00**
B.Non-Recurring-Equipment		-	-	-	-
2 Laptops		100,000.00	-	-	100,000.00
Total (A+B), without overhead		**1,096,560.00**	**1,282,560.00**	**620,800.00**	**2,999,920.00**

16. COLOURS: A COgnition-enabLed information system for personalizedtOURistic experienceS in India

PI and Team

PI: Anirban Mondal, Associate Professor, Ashoka University, anirban.mondal@ashoka.edu.inCo-PI:
Mukesh Mohania, Professor & Dean, IIIT Delhi, mukesh@iiitd.ac.in
Co-PI: Ladjel Bellatreche, Professor, ENSMA, Poitiers, France, Ladjel.bellatreche@ensma.fr

Executive Summary: The tourism sector is of paramount importance to India's economic growth. Tourism contributed INR16,907.9 Billion (9.2% of the total GDP) and 42.673 million jobs (8.1% of the total employment) to the Indian economy in 2018. India is a potential gold- mine for tourism with all kinds of climate, a wealth of historical places, a rich and diverse cultural & spiritual heritage and wildlife. It can cater to tourists with a diverse range of interests. Additionally, the Government of India has been investing heavily on improving tourisminfrastructure and towards promoting tourism e.g., Incredible India. However, a huge gap stillremains between India's actual extent of tourism and its potential to become tourism powerhouse. Among other reasons, a key reason for this gap is lack of information (especially hyperlocal information) about touristic attractions. Hence, there is a huge scope for improving tourism by exposing tourists to hyperlocal information about touristic attractions.

We propose COLOURS, a personalized ***cognitive computing*** and ***social sensing based tourist assistance information system*** for creating a highly enhanced and immersive tourism experience in India. COLOURS collects hyperlocal information from local people, experts and historical/cultural documents to build a cognitive knowledge base. Notably, hyperlocal information concerns information, which only a local person living in a particular area would know e.g., knowledge of where the autorickshaws congregate in a given local area,the location of the best vegetarian restaurants in that area and so on. COLOURS has two kinds of users, namely ***tourists (both domestic and foreign)*** and ***tourism authorities***.

For tourists, it augments touristic planning by cognitively capturing user preferences and context via social sensing and crowdsourcing, and then cognitively delivering contextually relevant snippets of hyperlocal Point-Of-Interest (POI)/event spatio-temporal information. Thiscreates a positive feedback ripple effect due to in-person or social media based word-of-mouthrecommendations from tourists to their friends/neighbors, thereby further boosting tourist arrivals in India. Notably, tourists typically have preferences about the type of touristic attractions and also about factors such as safety, convenience and cost. Suppose Alice, whois visiting Hyderabad from New York, is interested in Indian classical music. The system couldrecommend her to visit relatively "safe" places that have classical Indian music performances at say 7 pm. Observe how the system should not only present the hyperlocal information to users, but should also render it cognitively to match user preferences and context.

For tourism authorities, COLOURS contextualizes and delivers statistical information, actionable insights and *what-if analytics capabilities* for improving touristic foot traffic. Tourismauthorities can use the what-if capabilities of the system towards effectively shaping policy decisions (e.g., investing funds on improving the road and transportation infrastructure in theroutes leading to specific touristic attractions, reducing admission ticket prices, providing discounts of say 20% for nearby shops/restaurants for tourists with an admission ticket vicinityetc.) to improve tourism. Observe how this can create value for small businesses (e.g., shopsselling memorabilia and artefacts, and restaurants) nearby touristic locations. Additionally, this can also generate employment for the locals, thereby improving the local economy as well.

Popular apps for tourism include Kayak, Booking.com, Skyscanner and Airbnb. Several apps(e.g., Google Maps, Ola, Uber etc.) can also help tourists. However, to our knowledge, there currently exists

no system/app, which has the social sensing and cognitive capabilities of COLOURS for enabling immersive touristic experiences by providing hyperlocal information based on user preferences and context. While some of the hyperlocal information may exist in bits and pieces across a wide gamut of websites, it becomes extremely tedious for users tomanually gather such information. Moreover, no existing system/app provides a one-stop platform for tourism authorities to gather actionable insights and to perform what-if analysis. In essence, COLOURS is ***easy-to-use*** and provides significant differentiators in terms of its ***social sensing*** and ***cognitive capabilities*** as well as ***what-if analytics capabilities***.

Research challenges include the design of novel cognitive computing algorithms/techniques for handling spatial queries with constraints (user preferences/context, batteries of devices, real-time recommendations, etc) concerning POIs/events and abstract directives such as "safety", which are non-trivial to cognitively quantify. Inferring and cognitively computing the actual contextual preferences of users from the users' stated preferences also poses challenges. Challenges also concern the investigation of methods for caching/prefetching of the POI/event data for optimizing the search communication cost. This also requires researchinto the reuse of the intermediate results of POI/event queries (with constraints) across the journeys of multiple users. Important research challenges also include data mining, analyticsand creation of what-if capabilities to contextualize touristic data and deliver actionable insights in a cognitive manner. Other significant research challenges include the evaluation ofthe reliability of the data obtained from social sensing and crowdsourcing, and rewarding userswith incentives based on the extent of reliability of the data they provide to our system. Another issue concerns the choice of the relevant representation of this data that strongly impacts the performance of proposed algorithms: Matrix-based, Tensor-based, Graph-based and Hybrid.

Our proposal is most closely aligned with the vertical of **Environmental sustainability (Smartcities) – Intelligent transportation system** of the TIH in **Cognitive Computing and SocialSensing.** However, the cognitive & social sensing technologies that we will develop would also be applicable to the other verticals of the TIH. For example, our technologies will be designed to cognitively understand, contextualize, discover, ingest, reason and infer about various kinds of data such as multi-modal data (e.g., text, image & video data) and real-time feeds about events that are collected by using social sensing and crowdsourcing. Our systemwill also clean, integrate and curate the data appropriately to ensure data reliability and consumability. Furthermore, the cognitive capabilities of our system include cognitive querying based on query context, user context & domain context as well as cognitive computation services for cognitive ingestion of the data. *In turn, this facilitates in converting traditionally passive data stores into cognitive data stores and cognitive knowledge bases.*

- **Project deliverables:**

1. Investigate and develop a social sensing and crowdsourcing based framework to collect Point-of-Interest (POI) & event data and hyperlocal information by considering a sample of coordinators from across a diverse set of pilot regions in a given city
2. Develop a cognitive computing based spatial data management framework for conceptually modelling POI & event data and hyperlocal information such that the framework is amenable to targeted extraction of knowledge for answering various userqueries by considering user preferences and context and also for responding to cognitive queries from tourism authorities
3. Develop effective (user) context-aware & cognitive indexing and query processing algorithms with pruning heuristics for facilitating quick retrieval of personalized (query)results from the spatial database in response to constraint-based user queries and cognitive queries from tourism authorities
4. Design, implement, pilot & deploy a web-based information system and a mobile app with user-friendly interface for seamlessly providing cognitively computed, contextuallyrelevant & personalized results in response to queries by users and tourism authorities

We plan to publish research papers in reputed conferences/journals. We expect that studentsfrom IIIT

Delhi, Ashoka University and other Institutes will be working on this project as interns.International collaborator, Prof. Ladjel Bellatreche, will also be involved in co-supervising the students & interns. with Dr. Anirban Mondal and Prof. Mukesh Mohania.

- **Budget – Initial Phase**

A. Recurring		**Ist Year**	**2nd Year**	**3rd Year**	**Total (In Rupees)**
1. Salaries/Wages					
Designation & number of persons	Monthly Emoluments				
Ph.D Student 2	32000.00	768,000.00	768,000.00	768,000.00	2,304,000.00
2. Consumables		15,000.00	15,000.00	15,000.00	45,000.00
3. Travel		-	-	176,000.00	176,000.00
4. Contingency		12,000.00	12,000.00	12,000.00	36,000.00
5. Workshop/ Conference Organization		-	-	79,000.00	79,000.00
Total		**795,000.00**	**795,000.00**	**1,050,000.00**	**2,640,000.00**
B. Non-Recurring					**-**
1. Go Pro Cameras (5)		125,000.00	-	-	125,000.00
2. Mobile phones with GPS capability (5)		75,000.00	-	-	75,000.00
3. Laptops (2)		160,000.00	-	-	160,000.00
Total		**360,000.00**	-	-	360,000.00
Total (A+B)		**1,155,000.00**	**795,000.00**	**1,050,000.00**	**3,000,000.00**

17. **Project Title**: A Computational Gastronomy Framework for Achieving Better Nutrition and Public Health

TiH proposal for creation of a Computational Gastronomy framework

Project Investigator: Ganesh Bagler

- **Summary**: Cooking forms the core of our cultural identity other than being the basis of nutrition and health. The increasing availability of culinary data and the advent of computational methods for their scrutiny is dramatically changing the artistic outlook towards gastronomy. Starting with a seemingly simple question, 'Why do we eat what we eat?' data-driven research conducted in our lab has led to interesting explorations of traditional recipes, their flavor composition, and health associations. Our investigations have revealed 'culinary fingerprints' of regional cuisines across the world, starting with the case study of Indian cuisine. Application of data-driven strategies for investigating the gastronomic data has opened up exciting avenues giving rise to an all-new field of 'Computational Gastronomy'. This emerging interdisciplinary science asks questions of culinary origin to seek their answersvia the compilation of culinary data and their analysis using methods of statistics, computer science, and artificial intelligence. Along with complementary experimental studies, these endeavors have the potential to transform the food landscape by effectively leveraging data-driven food innovations for better nutrition and public health. The proposal of creating 'A Computational Gastronomy Framework for Achieving Better Health and Nutrition' aims at capturing the culinary creativity into computational strategies thereby contributing to the larger goals of 'Maternal and Child Public Health' and SDG3 (Good Health and Well-being).

- **Aims and Objectives:**

- **To compile and collate structured data of recipes, ingredients and nutritional values.**One of the key datasets required towards building the computational gastronomy framework is that of a structured repository of recipes, their ingredients and nutritionalvalues of ingredients. While these details are available in a fragmented manner, no existing repertoire provides them in an integrated fashion. We intend to build computational strategies for compiling this information and creating a structured library of recipes.

- **To create an exhaustive repository of flavor compounds found in natural ingredients**Beyond the details of recipe structure and nutritional value of the ingredient, the otheraspect of food that needs to be encoded towards creating the framework is that of itsflavor (taste + odor). We intend to create an exhaustive repository of flavor compoundsin natural ingredients building on the existing state of the art. This repertoire is expected to help in achieving the desired food pairing pattern in the recipes.

- **To design algorithms for estimating nutritional content of a recipe**
  Having built the above-mentioned datasets, we intend to design a strategy for estimating the nutritional value of recipe. As a first step, this requires creation of Natural Language Processing (NLP) strategies for correct identification of ingredients, quantities and units in the ingredient section of a recipe using Named Entity Recognition (NER). Subsequently, the nutritional content of the recipe can be estimated by bridging this information with nutritional details such as those available
  from USDA (United States Department of Agriculture).

- **To establish an integrative Computational Gastronomy framework**
  With the databases and algorithms thus built, we intend to blend the knowledge extracted

from traditional recipes, food pairings and nutritional values of recipes to create a Computational Gastronomy framework. Such framework integrates differentaspects of food facilitating creation of healthy and nutritious recipes that are palatable.

- **To design algorithms for generating novel recipes**
  All the above objectives are aimed at knowledge creation aspect of the project. Goingforward, we intend to generate novel recipes that are palatable as well as nutritious. Towards this, we intend to create novel recipe generation algorithms using text generation strategies. This project attempts to create a 'Turing Test for Chefs' and would be a major step in Artificial Intelligence towards automating the culinary creativity.

➢ **Strategy:**

The existing state-of-the-art for achieving better public health via the nutritional interventionis heavily driven by heuristics rather than data-driven inferences. Using an intensely data- driven strategy, the proposed Computational Gastronomy framework intends to fill this gap.The proposed technology compiled various layers of data associated with food to make it computable. By generating novel recipes that are not only palatable but also nutritional, it provides data-driven measures for achieving better public health.

➢ **Technology and Research Methodology:**

- **Recipe Compilation and Nutrition Estimation**
  We intend to compile recipes from across the cultural zones from around the world representing the culinary diversity in a computable format. Further, we intend to integrate nutritional data from the United States Department of Agriculture so as to automate the estimation of the nutritional content of recipes.
- **Flavor Database Creation**
  Going beyond the database we have already created in the lab, we intend to create adatabase of flavor compounds accounting for all the major taste and odor attributes.
- **Computational Gastronomy Framework**
  All the above-described elements form the jigsaw pieces of the food puzzle: recipes, flavors, nutrition, and health. By seamlessly putting them together in the form of a Computational Gastronomy framework, we intend to create an integrative resource for data-driven investigations of food. Such a resource will not only facilitate queryingthe food data from its taste and nutrition perspective but will also form the corpus for data analytics and mining.
- **Novel Recipe Generation**
  We expect to build Novel Recipe Generation algorithms to pass the 'Turing Test for the Chefs.' We expect to tweak ingredient combinations to generate recipes with desired nutritional profile.

➢ **Budget**:

Budget Head		1st Year	2nd Year	3rd Year	Total
A. Recurring					
1. Salaries/wages					
Designation	Monthly Emoluments				
PhD student@2 (Research Scholar)	31000	744,000.00	744,000.00	840,000.00	2,328,000.00
	24% HRA	178,560.00	178,560.00	201,600.00	558,720.00

Total		**922,560.00**	**922,560.00**	**1,041,600.00**	**2,886,720.00**
2.Contingency		50,000.00	50,000.00	13,280.00	113,280.00
Total		**972,560.00**	**972,560.00**	**1,054,880.00**	**3,000,000.00**

18. Proposal Title: **AI based Interactive and Multimodal 2D Illustrations in Mixed Reality**

PI: **Ojaswa Sharma, Associate Professor, IIIT Delhi**

- **Context/Background**

Our proposal on enriching traditional educational illustrations using AI and Mixed Reality is a technological solution to making digital education more engaging and interesting. This very well fits into the overarching theme of Cognitive computing and Social sensing by enhancing educational presentation and interactivity with Mixed Reality and by enabling AI based algorithms for delivery of visual content.

Traditional education illustrations are static and non-interactive. An understanding of the content is gained by comprehension of the accompanying text. With the new display technologies offered by various Mixed Reality devices, this could drastically change. We propose to build a framework for authoring interactive and animated 2D illustrations for educational use. Our system will make use of immersive technologies for greater interactivity and engagement that will aid in better and faster learning. The authoring tool will provide APIs and workflows to enable faster creation of the animated content.

The underlying algorithms for semi-automated content authoring will be based on computational geometryand artificial intelligence. We intend to build our framework on context inference using AI based methods.This will enable anyone to create content without much difficulty. Overall, we believe that our system willaid in e-learning and will be a great addition to remote education. This will benefit content creators, educators and students.

- **Aims and Objectives**

The proposed project aims to introduce interactivity and animation in educational illustrations that can be embedded in electronic textbooks. Examples of such educational illustrations include depiction of physical,biological, and mechanical systems. Many such systems and natural processes depict phenomena that varyin space and time, for which animation is a great tool to visualize. Also since the primary goal of an illustration is to induce learning, interactivity and feedback play important roles here. In building such a proposed framework, our primary objectives are:

1. Designing an algorithmic framework to enable creation of interactive and animated educational illustrations in Mixed Reality environments, and
2. Develop AI based algorithms for creating contextual workflows for content authors. Such algorithms will not only enable automation in content creation, but will also lead to smarter workflows.

We would like to provide an open-source software toolset for creating underlying simulations, animations, and interactivity. The community supported open-source framework will allow anyone to develop (with ease) curriculum specific illustration development tools.

- **Technology & Research Methodology**

Our project is about developing next generation algorithms for interactive content creation for education. Enabling technologies in the field of Mixed Reality will allow content authors to provide more engaging and immersive experiences. We will make use of the Hololens AR headset for live and interactive

animations that could be visualized in the real space of the user. The Hololens is not just a new type of display technology, but it also allows amalgamating the virtual world with the real. This will also open uppossibilities for the user to interact with real world objects in different ways.

There will be emphasis on using Web based interfaces for wide adaptability. For the same, initial prototypes will be experimented on HTML5. Such a web based platform will allow ease of access for both the content creators, teachers and students, thus benefiting all stakeholders. Using lightweight web technologies and distributed platforms will also make our system accessible over the internet.

One of the pillars of our proposed framework is an AI based backend for analysis of content and workflow automation. We plan to make the system smart and easy to use so that the content creators do not struggle to deal with complicated scenarios and illustrations. In other words, with AI based assistance, our system will become usable even for naive content creators. In fact our system will be usable even for teachers to create the content (eliminating the need for specialized content creators).

- **Finance**

1. Salaries/wages					
Designation & number of persons	Monthly Emoluments				
JRF/Phd *1	31,000-Yrs 1, 35,000-2,3rd Yrs+3HRA	461,280.00	520,800.00	520,800.00	1,502,880.00
RA*1	25000	300,000.00	-	-	300,000.00
2. Consumables		-	-	-	-
3. Contingency		-	50,000.00	50,000.00	100,000.00
Total (A)		**761,280.00**	**570,800.00**	**570,800.00**	**1,902,880.00**
B. Non-Recurring-Equipment					
GPU Workstation*1		500,000.00	-	-	500,000.00
Tablet*2@1,15,000/-		230,000.00	-	-	230,000.00
Microsoft Hololens*1		-	312,000.00	-	312,000.00
Storage*1		55,000.00	-	-	55,000.00
Total (B)		**785,000.00**	**312,000.00**	**-**	**1,097,000.00**
Gross Total (A+B)		**1,546,280.00**	**882,800.00**	**570,800.00**	**2,999,880.00**

19. Development of efficient block chain protocols for applications in public health

PI: Dr. Vaneet Aggarwal (Purdue University, USA) and

Co-PI: Dr. Anuradha Sharma(IIIT-Delhi)

- The number of businesses that use the IoT technologies has increased from 13 percent in 2014 to about 25 percent in 2019. Further, the worldwide number of IoT-connected devices is projected to increase to 43 billion by 2023, an almost threefold increase from 2018. The key concerns that have hindered the large-scale deployment of distributed IoT are security vulnerabilities and scalability. It has been widely studied that the blockchain can help alleviate the security and scalability concerns associated with IoT. This is in part due to (i) The distributed ledger in a blockchain system is tamper-proof and this removes the need for trust among the involved parties, (ii) Using blockchain to store IoT data adds another layer of security that hackers would need to bypass in order to get access to the network, (iii) Blockchain provides transparency, by allowing anyone who is authorized to access the network to track the transactions that happened in the past, and (iv) Blockchain can enable fast processing of transactions and coordination among billions of connected devices. Due to the blockchain providing improved trust between parties and devices, cost reduction by removing overhead associated with middlemen and intermediaries, and reduced settlement times, IBM launched an IoT Blockchain Service that makes it possible for groups of participating parties to use information from IoT devices, such asgeo-location, with smart contracts running on the IBM Blockchain networks or fabric.

In this proposal, we wish to innovate on efficient blockchain protocols that optimize throughput,latency, storage, ease of new miners, and energy efficiency. In order to account for addition of new miners, efficient coding-theory based protocols will be investigated. To reduce the computation burden to solve the proof-of-work (PoW) puzzles, an approach called sharding has been proposed. Sharding has been used widely in IoT blockchains. This work will consider sharded block chains and apply coding theoretic techniques to reduce the storage cost as well as the bootstrap cost. We will investigate secure regenerating codes that are not only storage efficient but also bandwidth efficient while repairing single node failures. We aim to draw the equivalence between the process of bootstrapping a node and repairing a failed node. This equivalence can make the bootstrap cost low as compared to uncoded sharding.

- The key research objectives include:

    1. Efficient capability for new miners: We aim to innovate on efficient blockchain protocols that optimize throughput, latency, storage, ease of new miners, and energy efficiency. In order to account for addition of new miners, efficient coding-theory based protocols will be investigated.

    2. Optimized decision of who verifies: This will be based on stochastic calculations of transactional throughput and latency. The different control knobs will be

investigated, and the metrics will be optimized for efficient queueing and decision statistics.

3. A joint decision making: All decisions like consensus, mining, queueing, etc. will beexposed for a decision making of the users. Such decision making will use multi-agent reinforcement learning.

Thus, overall tools from algebraic coding theory, stochastics and queueing theory, Markov Decision Process, and Multi-agent reinforcement learning will be used to obtain an optimized blockchain architecture.

- **Budget:**

A. Recurring	**Year 1**	**Year 2**	**Year 3**	**Total**
*Postdoc (1)	600,000.00	600,000.00	600,000.00	1,800,000.00
Contingency & Consumables/Conference Travel/Registration/Journal Page Charges	100,000.00	200,000.00	200,000.00	500,000.00
Travel (US PI) including TA/DA for stay upto 2 months per year	-	150,000.00	250,000.00	400,000.00
Total (Recurring)	**700,000.00**	**950,000.00**	**1,050,000.00**	**2,700,000.00**
B. Non-Recurring-Laptop (2)	150,000.00	150,000.00	-	300,000.00
Total	**850,000.00**	**1,100,000.00**	**1,050,000.00**	**3,000,000.00**

*Postdoc will be hired at IIIT-Delhi only.

20. **Proposal Title:** Design and Development of Interpretable AI methods for Flow Cytometric Immunophenotyping Data from Blood Cancers

    PI and Team:

    PI: Prof. Anubha Gupta, Collaboration: AIIMS

    Perceived TIH Application under Cognitive Computing and Social sensing: Intelligent diagnostics and prognostics for healthcare

    Application Accelerators: Addressing specific industry functions in healthcare

    Proposal Details:

- **Context/Background:** Hematopoietic (blood) cancers are prevalent in India with more than 70000 cases reported every year. India ranks 3rd highest in terms of reported cases after the US and China. As of now, multiparametric flowcytometry data analysis is the gold standard for the diagnosis and monitoring of the course and treatment of blood cancers. These are specialized tests that help in the evaluation of peripheral blood and bone marrow aspirate in blood cancers both for diagnosis and for detection of residual disease. The technology is also used in crossmatching organs for transplantation, research involving stem cells, vaccine development, apoptosis, phagocytosis, identifying immunological parameters in disease, studying the humoral and cellular response to vaccines, besides providing the counts of helper-T lymphocytes needed to monitor the course and treatment of HIV infection, in the diagnosis and monitoring of leukaemia and lymphoma patients, the evaluation of peripheral blood hematopoietic stem cell grafts, and so on.

  FCM is an expensive and specialized test. FCM data analysis is typically time consuming and prone to errors because the data of millions of cells is being produced that needs to be visualized and analysed by the doctors to infer the next course of treatment. Inferences are often subjective to the judgement of the experts rather than on the standardized statistical inference. In fact, the potential of this tests are undermined owing to the lack of reliable statistical methods and accompanying software implementations along with interpretability for the analysis of FCM data. With the advances in cognitive computing domain, it is now possible to develop advanced data analysis tools that can aid the specialists in the decision making. In the proposed research study, we will develop interpretable methods for the analysis of flow cytometry data for blood cancers. The digital interfacing will promote reliable and fast diagnosis, improve the survival and cure of cancer patients.

- **Problems to be addressed:** Multiparametric flow cytometry (FCM) can measure from a single to upto 40 cell characteristics, for up to millions of individual cells per sample. It works on the principle of scatter properties of light in the forward and side directions that differs for various cell types, besides the type and the density of the expression of antigen on the cell surface that helps in the specific identification of cell types. Typical FCM data analysis involves unsupervised clustering for the cells based on their cell types, gating for the identification of homogenous cell populations that share a particular function, interpretation (i.e., finding (or using) correlations between some characteristics of the identified cell populations (e.g., percentages of cells in a cell population, median fluorescent intensity of a cell population for different markers) and the visualization for the pathologists/clinicians to understand the results of the FCM data analysis. All the above tasks have their own challenges.

  A) Unsupervised clustering: The number of abnormal cells can be large and small in numbers. Likewise, different blood cell types are present in varied proportion in the peripheral blood or bone marrow. Accordingly, the clusters would be big/small in size, may be near each other or overlapping as well. The number of clusters

is also not known apriori. Since the data is high dimensional as well as large, this poses a problem in extracting the correct number of clusters that are also meaningful for cancer diagnosis and treatment.

B) Gating: Gating is a highly subjective process, wherein the experts analyze the regions in multiparametric space based on their knowledge and experience. This is a tedious, timeconsuming, and often inaccurate task, typically accomplished using proprietary software provided by instrument manufacturers. These software provide limited capability to analyse the data. Intersections or unions of polygonal regions in hyperspace are used to filter data and define a subset or subpopulation of events for further analysis. A relatively minor differences in gating can produce different quantitative results.

C) Interpretation: As stated earlier, FCM data analysis is typically time consuming and prone to errors because inferences are often subjective to the judgement of the experts rather than on the standardized statistical inference. In fact, the potential of this tests are undermined owing to the lack of reliable statistical methods and accompanying software implementations along with interpretability for the analysis of FCM data.

D) Visualization: The clusters extracted in the high dimensional data are required to be presented in the 2D visualization space for the doctors that is again fraught with challenges. Since visualization is again critical when gating needs to be decided for the detection of clusters of abnormal cells.

- **Aims and Objectives:**

In our study, we will work on the Acute Lymphoblastic Leukemia of B-cell type (B-ALL). AIIMS will collect the flow cytometry data of 10 bone marrow aspirate from 10 control subjects and 100 leukemic bone marrows. We will design algorithms for the unsupervised clustering with the additional prior information on the biological data. We will propose newer clustering methods that can

a) find the optimal number of clusters wherein some clusters will be very large, while some will be relatively smaller in size;

b) identify smaller versus bigger clusters in the vicinity wherein some might be overlapping as well to infer the normal versus abnormal cell population;

c) since FCM data is a big data in multidimensional space (10-20-dimensional space) capturing millions of cells, the algorithms need to perform efficiently with reference to time complexity for the ease of deployment;

d) provide interpretability on the methodology developed along with the statistical validation to help the doctors with reliable inference;

e) provide efficient automated gating mechanism along with the handle to the doctors to change/edit gates according to their inference. Hence, the tool would be interactive in nature.

e) provide efficient visualization as well as generate an automated report based on the analysis.

- Strategy: As stated earlier, multiparametric flowcytometry data analysis is the gold standard for the diagnosis and monitoring of the course and treatment of blood cancers. In this project, we would like to exploit the potential of FCM data by the development of advanced interpretable AI based solutions with big data clustering, visualization, and interactive interface. A significant amount of knowledge is available big data processing and AI, but a concerted effort by the diagnostic pathologists and technical experts on data processing and AI is required to harness the translational potential of these technologies in the domain of health. The proposed project is aimed to open new avenues with the goal of using the current gold standard

of cancer diagnostics as the basis of developing a novel technology.

- **Target beneficiaries and Management:**

  The ultimate user for such a work is a cancer hospital. Since the cancer lab of AIIMS is the collaborator in this work, it is one of the best places to deploy and test the proposed tool. Cancer patients from all over the country approach AIIMS, New Delhi. Also, it is the ultimate referral hospital for handling critical cases in India. Hence, the deployment of technology at AIIMS and the feedback from the doctors and clinicians of the cancer hospital is most valuable in this project. The target beneficiaries of this project are:
  a. Cancer hospitals
  b. Pathology laboratories
  c. Patients suffering with blood cancer

- **Legal Framework** (500 words) [If applicable, kindly elaborate, what kind of legal framework will be required for achieving your proposed research objectives?] The patients will be diagnosed and recruited at AIIMS, New Delhi and data will be collected at that end. The project will go through the appropriate procedures of ethical approval through the Institutional Ethics Committee (IEC) of AIIMS, New Delhi.

- **Environmental impact** (500 words) [If any, kindly elaborate.] Not applicable

- **Finance:**

Item		Ist Year	2nd Year	3rd Year	Total
Recurring-A					
Salaries/wages	Monthly Emoluments				
Project Associate-1@31,000/-+HRA 24% for Ist 2 Years and Rs. 35000/-+HRA 24% in the 3rd Year	38440	461,000.00	461,000.00	521,000.00	1,443,000.00
Project Associate-1@31,000/-+HRA 24% for 2 years	38440	-	461,000.00	461,000.00	922,000.00
Contingency		100,000.00	125,000.00	110,000.00	335,000.00
Total Recurring (A)		**561,000.00**	**1,047,000.00**	**1,092,000.00**	**2,700,000.00**
Non-Recurring-B					
Network Access Storage (12TB*5) **@ IIIT-D**		**300,000.00**	**-**	-	300,000.00
Gross Total (A+B)		**861,000.00**	**1,047,000.00**	**1,092,000.00**	**3,000,000.00**

- **Outcomes**

The tangible outcomes of this project are as follows:

● Technology Development: Development of a software tool (beta version) for the analysis of FCM data for B-ALL blood cancer.
● HRD and Skill Development
● Publications and Patents: Publication in high impact factor journals and conferences.

21. Proposal Title: Computational-assisted and data-driven delineation of olfactory cognitive abilities.

PI and Team:

Principal Investigator: Dr. Gaurav Ahuja; **Co-Principal Investigator:** Dr. Debarka Sengupta

- **Aims and Objectives**

1. Deorphanization of the orphan olfactory receptor using Deep Learning-based models.
2. Establishment of EvOlf database, the largest database representing odor-receptor pairsacross the animal kingdom.
3. Defining the molecular clues imparting distinct tuning properties in the olfactory receptors.
4. Development of novel visualization techniques to co-embedded ligand-receptor pairs across vertebrate evolution.
5. Upgradation of the Machine-Olf-Action computational workflow.
6. Linking metabolic intermediates with tumor-associated-olfactory receptors.
7. Large-scale high-throughput experimental validation of the ligand-receptor pairs using human cell lines-based heterologous systems.

- **Problems to be addressed**: Olfaction, the sense of smell, is phylogenetically one of the most ancient senses and is primarily carried out by the olfactory receptors (ORs), a special class of G-Protein Coupled Receptors, residing within the olfactory sensory neurons of the nose (Fleischer etal.). Activation of OR by agonist in the cell periphery activates a cascade of biochemical reactions leading to depolarization of the neuronal membrane and creates action potential (Fleischer et al.). Despite the human genome encoding ~400 functional olfactory receptors (Trimmer et al.), recent reports suggest that humans can detect and discriminate trillions of odorant molecules (Bushdid et al.). This is feasible due to the olfactory system's combinatorial coding strategies, meaning that eachOR can detect more than 1 odorant and vice versa (Su et al.).

- **Strategy:**

Part 1: Unlike other senses such as vision and auditory perception, where the triggering stimulus can be quantified in wavelength (λ) or frequency (Hz), respectively, the sense of smell largely lacks similar quantifiable metrics. However, in the case of olfaction, the prime drivers, the olfactory receptors are largely classified as broadly or narrowly tuned. Given the aforementioned lack of an accepted odorant similarity metric, historically, this question was addressed by quantifying the number of known agonists for a particular receptor. However, such an approach is prone to bias and largely obscures the accurate assessment of the tuning breadth of olfactory receptors, since it completely negates the information about **(a)** the total number of tested odorants, **(b)** similarity/dissimilarity among the agonists/non-agonists, and **(c)** deployment of stringent statistical validations. All these collectively reinforce the urgent requirement to devise computational methods to identify the molecular clues that regulate the tuning properties of an olfactory receptor. *We aim to address this apparent gap of knowledge by devising and utilizing a state-of-the-art Machine Learning/Deep Learning-based statistical framework. Moreover, to ensure the availability of adequate data points for the execution of the aforementioned methods, we aim to manually compile the (non)agonist-olfactory receptor information from literature across the animal kingdom. In summary, it will*

enable us to gain molecular insights into the receptor tuning parameters and would also shed light on the evolutionary basis of receptor tuning.

Part 2: Although olfactory receptors are also known to express in pathological states, such as tumors, neurodegenerative diseases, etc, their functional contribution to their pathogenesis is far from complete. We have recently identified a substantial number of tumor-associated olfactory receptors using the high-dimensional single-cell transcriptomics approach. Mechanistic insights revealed that ORs are linked to the aggressive proliferative behavior of the malignant cells, suggesting an opportunity to extrinsically modulate their activation status using odorant analogs, potentially the endogenous metabolic intermediates. A recent study established a potential link between the activation/deactivation of a gut-expressed olfactory receptor by metabolic intermediates secreted by the gut microbiomes.

This further suggests that, although in the nasal epithelium, the olfactory receptors are activated by odorants, but within the ectopic tissue, they could be potentially activated by the oncometabolite and therefore, might be assisting the tumor growth. The functional feasibility of this hypothesis is highlighted in a recent publication, where the authors have identified the activation of OR51E2 (olfactory receptor) by testosterone metabolite in the malignant prostate cells. Functional insights revealed that the activation leads to transdifferentiation of the cancer cells, therefore transforming the tumor into an aggressive state. We have recently developed a cross-platform, Machine Learning based Chemoinformatics tool (Machine-Olf-Action) with inbuilt explainable Artificial Intelligence (AI) modules.

Machine-Olf-Action enables the synthesis of prediction models that can identify new potential agonists and non-agonists for the olfactory receptors, and also highlights the key features of the odorant chemistry which reinforce this interaction. In the second part of the proposal, we aim to extend the features of the Machine-Olf-Action framework, by introducing more classifiers, improvement towards the integration of new databases including the manually curated, yet high dimensional EvOlf database (our lab work, in preparation), and more layers of recently developed explainable AI algorithms.

Such a highly productive, Graphical User Interface-enabled chemoinformatics tool will be one of its kind in the cutting edge field of drug-development, olfaction, alchemy, and all other fields of science that includes chemical features. We will use the updated version of Machine-Olf-Action in establishingthe relationship between the oncometabolite, and the tumor-associated olfactory receptors. Lastly, we will experimentally validate the predictive interactions using heterologous system-based assays.

- **Budget**

A. Recurring		**1st Year**	**2nd Year**	**3rd Year**	**Total**
1.Salaries/wages	Monthly Emoluments				
1 Ph.D. student @31,000 pm +16% HRA	35960	-	431,520.00	431,520.00	863,040.00
Contingency		300,000.00	200,000.00	100,000.00	600,000.00
Total (A)		**300,000.00**	**631,520.00**	**531,520.00**	**1,463,040.00**

B. Non-recurring (Workstation)					
Cell-Culture facility (Equipments)		986,960.00	-	-	986,960.00
Cell Culture consumables		550,000.00	-	-	550,000.00
Total (B)		**1,536,960.00**	**-**	**-**	**1,536,960.00**
Grand Project Total (A+B)		**1,836,960.00**	**631,520.00**	**531,520.00**	**3,000,000.00**

- **Outcomes**

The success of each of the detailed objectives can be measured by:

(1). Its commercial interest among the industrial folks, especially the perfume, aroma-therapy,healthcare industry, and repellant industries.

(1) the number of high-impact publications emanating from this project.

(2) The number of academic users will be utilizing our generated resources.

22. **Proposal Title:** Multilingual and Culturally Customized Knowledge Management for LegalInformation Processing and Management

PI: Dr. Sarika Jain, Collaborator: Semantic Web

As it is very limited time to revise the complete proposal along with the thorough literature and gap analysis, we provide here an addendum to the already submitted proposal aligning to the problem ofLegal Information Processing and Management. It is important to mention that the crux of the proposal supports the "Cognitive Computing" at its heart and is therefore suitable for any problem at the application layer based upon the domain to be applied for.

- We propose to change the domain, rest proposal remaining intact. Instead of business analytics, we will pursue Legal Information Processing and Management Solutions. Here are the notable points:

1. **The technologies** to be delivered include a globalized (multilingual and multicultural) platform for data collection, curation, management, and sharing. A globalized knowledge base will be prepared for the legal information processing incorporating the semantic data model.

2. **Cognitive Data Governance:** Multilingual and Multicultural Linked Data Store, Semantic Repositories, Knowledge Base
   The legal data will be modeled as RDF. Existing legal information regarding the legal cases, law suits, various departments, and various tribunals will be converted from the existing heterogeneous (structure or unstructured) formats into interoperable RDF format. As the data will be put on the linked open data cloud, it will be easier for experts to annotate. Means to semi-automatically annotating the data will be provided. Once annotated, it lies on the cloud itself, open and linked accessible to all the APIs and services.

3. **Cognitive Services and APIs:** Natural Language, Multi-modal data relationship discovery, real time events/social sensing, Real time data/text feeds.
   The API developed will be generic and can be utilized for any domain and specifically for legal information processing.
   The information in the RDF store will be accessed through a user-friendly interface. The semantic technologies thus utilized facilitate the interpretability and explainability of the decision making and the reasoning procedure applied.

4. **Application Layer:** Following applications are proposed:
   (a) Legal Document Classifier: Given the legal document, this use case will classify the document into categories as provided by experts in the adjoining ontology. The ontology-based text classification will comprise all the qualities of the symbolic reasoning and the statistical approaches.
   (b) Legal Document Retrieval: Given the abstract of a legal case, all the similar relevant legal cases will be retrieved based upon the semantic similarity. The multilingual knowledge graphs will facilitate semantic search on multilingual corpora while increasing the efficiency of search because of the exploitation of relationships between the attributes. The documents thus retrieved could be classified as a next step to retrieve keywords and textual description summarizing their contents.
   (c) Question Answering: Once the legal documents are collected and converted into knowledge graphs by virtue of the semantic data models, SPARQL queries could be written for many known queries and a user friendly interface be provided to the end-user for querying the legal knowledge

base.

- **Budget**

Item		Budget			
A. Recurring		**Ist Year**	**2nd Year**	**3rd Year**	**Total (In Rupees)**
1. Salaries/wages					
Designation & number of persons	Monthly Emoluments				
Post Doc/JRF/RA (2)		496,000.00	744,000.00	434,000.00	1,674,000.00
2.Students Interns		-	10,000.00	10,000.00	20,000.00
3. Travel		20,000.00	300,000.00	20,000.00	340,000.00
4. Consultancy/ Outsourcing		200,000.00	150,000.00	186,000.00	536,000.00
5. Contingency		30,000.00	30,000.00	20,000.00	80,000.00
Total (A)		**746,000.00**	**1,234,000.00**	**670,000.00**	2,650,000.00
B. Non-Recurring-Equipment					
High End Laptops/Desktops (3)		350,000.00	-	-	350,000.00
Total (B)		**350,000.00**	**-**	**-**	350,000.00
Grand Total (A+B)		**1,096,000.00**	**1,234,000.00**	**670,000.00**	**3,000,000.00**

23. **Title:** Development, validation and deployment of a novel prediction model utilizing artificial intelligence on clinical and proteomic features to predict mortality among patients with acute-on-chronic liver failure

PI: Dr. Nipun Verma

Perceived Applications Scenarios: Intelligent diagnostics and prognostics for healthcare ***

Technologies to be delivered − Prediction models − AI-based portable platforms − Visual Dashboards

- **Activity Focus: −**

Knowledge Generation **
− Technology Product Development **

− Collaboration (National) ** Contributions: The project will contribute to predictive analytics, application development, AI, and MLdriven cognitive services. The project will provide a user-friendly web-based application that can be used at the bedside of a patient in hospital or outpatient, which will be used for the prediction of death in patients with liver cirrhosis

- Summary

Liver cirrhosis is the most common cause of death among gastrointestinal diseases, with a global burden of 1.5 billion persons being affected annually. It is the 4th and 10th most common cause of disease in males and females, respectively. Acute-on-chronic liver failure (ACLF) development is the most common cause of death in these patients. The mortality of this dynamic syndrome ranges from 15-100% within 30-days and 90-100% within one year of presentation. Multiple life-threatening events like sepsis, organ failures, and gastrointestinal bleeding may occur during its course. This proposal is aimed to develop, validate, and deploy a novel model to predict mortality among patients with acute-on-chronic liver failure (ACLF). The model will be derived from clinical, biochemical, radiological and plasma proteomics data obtained from patients with ACLF. Principles of AI, machine learning, and bioinformatics will be utilized to build such a model. The model aims to incorporate the systems approach to medicine with clinical sciences and information technology sector. The proposal includes multiple innovations. It will develop a multimedia processing engine (converting the existing data in a ready format for ML and AI modeling), development, and deployment of AI-derived model for realtime predictive analytics. The models' deployment will contribute to the development of user-friendly data analytic apps. The models may be used to assess the futility of care and allocating significant resources like ICU and mechanical ventilators to deserving patients. Importantly, patients' lives can be saved with a timely decision for definitive treatment (liver transplant) using these models. This proposal aligns with the mandate of TiH in terms of knowledge generation (development of clinical and proteomic database), technology product development (portable AI model), skill development (integration of skills such as clinical, biochemical, bioinformatics and information technology), collaboration (between clinicians, basic scientist, computational biologist, and information technologist).

- Aims: We aim to develop, validate, and deploy a novel AI-derived model on clinical and proteomic features to predict death in patients with ACLF.

The objectives and brief methods will include

a. To develop a clinical database of patients with acute-on-chronic liver failure

   Relevance: This will help create a knowledge database and involve the utilization of a multimedia processing engine where the data as the text will be translated into a statistically sound format for computation. The platform could be used for real-time multi-centric data collection, curation, and sharing.
   Plan: we will develop a web-based application for patients' data capture in a structured format.

Patients will be recruited prospectively over 1.5 years (approx. N=100) and retrospectively (approx. N=1000) over the last five years. Clinical, biochemical, radiological, and outcomes will be noted.

b. Develop a proteomic database of patients with acute-on-chronic liver failure stratified by their mortality and validate the mortality-specific proteins/pathways.
Relevance: This phase will involve multimodal data discovery using systems unbiased approach to medicine, bioinformatics, and generation of data required for cognitive services. It will enhance the knowledge database in ACLF patients.

   Plan: We will conduct proteomic experiments in the plasma of prospectively collected patients. The samples will be analyzed as per the patient's mortality. Top three or four altered proteins identified on bioinformatic and ML analysis between survivors and non-survivors will be validated with ELISA.

c. To develop and internally validate a prediction model using supervised machine-learning and Deep learning models for death from the collected data (from objectives a and b)

   Relevance: This phase will involve cognitive services-predictive analytics, AI and ML-driven computing of models and testing and comparing the performance of models, and preparing for deployment.

   Plan: We will apply multiple ML/DL algorithms to predict mortality in ACLF patients using the data collected. The model will be selected based on accuracy and area under the roc curve, and cross-validation and hyperparameter tuning will be performed to choose one model. The model explanation will be sought, and top features may derive a simpler ML model.

d. To validate the derived model in a separate prospective cohort of ACLF patients.

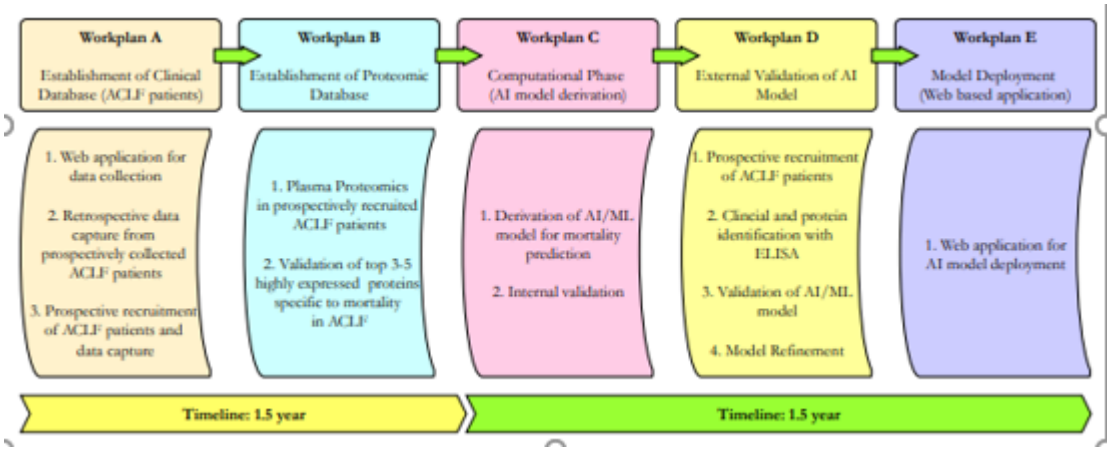
   Relevance: This phase will involve cognitive services-validation of AI-derived predictive model

   Plan: A prospective cohort of ACLF patients (N=100) will be recruited, and an AI-derived model will be validated.

e. To develop a user-friendly mobile or computer-based web application for real-time prediction of death in patients with ACLF

   Relevance: Development of AI-based portable platforms

   Plan: we will collaborate with an IT expert to develop a web-based AI model for the real-time prediction of events in ACLF.



- **Budget**

A. Recurring Costs	Year 1	Year 2	Year 3	Total
i) Proteomics related chemicals, reagents, experiments	2,300,000.00	-	-	2,300,000.00
ii) ELISA Kits for validation of proteins	-	400,000.00	-	400,000.00
iii) Data capture application, cloud storage, troubleshooting, training, web hosting	20,000.00	15,000.00	15,000.00	50,000.00
iv) Data analysis and model development	-	100,000.00	-	100,000.00
v) Contingency	50,000.00	50,000.00	50,000.00	150,000.00
Total (A)	**2,370,000.00**	**565,000.00**	**65,000.00**	**3,000,000.00**

24. **Title: Leveraging Multi-modality in Understanding and Summarizing Task-Oriented Dialogue Systems**

PI: Tanmoy Chakraborty (http://faculty.iiitd.ac.in/~tanmoy/)

Category: Cognitive Computing and Social Sensing

- **Introduction:** In recent times, a seamless effort has been made on dialogue understanding and generation. Often, while working with the dialogue system, people refrain from analyzing the emotions, and conversation sensitivity to better understand the dialog. The need to study the impact of emotions and conversation sensitivity increases significantly as we talk about dialogue systems in the domain, such as a therapeutic conversation between doctor and patient. Since not just one or two but more than 90% of the responses generated in such settings are emotion-driven. This connection is called emotional intelligence. Also, task-oriented dialogue systems revolve around a certain number of topics and while generating summaries of such conversation using current deep learning-based approaches, a very common problem of 'topic drift' occurs often. Sticking to a topic and covering all relevant utterances is crucial while generating summaries of conversations. After a careful investigation, it can be clearly observed that the work done is on a very limited scale and needs massive exploration. Our state-of-the-art chatbots are able to generate grammatically correct and more human-like summaries but are not able to generate a topic-aware language.

  Considering the scope of research in both dialogue understanding as well as summarization, we will summarize our project into two tracks –

  ● Track 1: Dialogue understanding in task-oriented dialogue system: ○ Emotion-aware dialogue-act and intent understanding to better analyze the co-dependency of emotions and other features in the dialogue system. ○ Incorporating multi-modality to better understand emotions and dialogues. We will focus on the textual, visual, and speech features to design multi-modal models.

  ● Track 2: Summarization of conversations: ○ Topic-aware summarization of conversations in a dialogue system and analyze the comparison of unsupervised and supervised methodologies for topic modeling. ○ Detection of important utterances in the conversation and utilize them in the generation of meaningful summaries.

- **Related Study:** Earlier work by [Huang et al.2018] explicitly shows how lack of emotion labeling is a problem in solving emotion-aware dialogue generation. They used a sequence to sequence model to generate utterances containing emotional context. Later, Song et al. [2019] showed the utility of incorporating emotional words within the utterance and generating responses accordingly. Their model, EmoDS, a sequence to sequence architecture with lexicon-based attention, was able to generate emotion-plugged utterances. We are planning to leverage the multi-modal nature of dialogue interactions as discussed by Kurt et al. [2020] to improve the engagement metrics with the users. Given a user request, we may have multiple options -- visual representation will give a good glimpse of understanding as well as summary generation. Even in the real world, doctor-patient interactions highly rely on the visual features for most of the examples. Therefore, considering the multi-modality of images as well as speech will be a vital point with the users.

- **Proposed Methodology:** Currently available datasets in conversation understanding and summarization are very limited in several domains like health. The ones that closely resemble this domain are either limited in the varying input length or do not combine other modalities such as visuals and audio to amend the quality further. We plan to study co-dependency of both emotion features and intent features, and hence we intend to implement attention-awareness which could leverage the joint training. Moreover, the systems proposed for abstractive summarization tasks provide benchmarks on the standard corpora, which often belongs to a particular family of genres. Hence, the problems that arise in processing the task-oriented topic-aware dialogue summarization

are often ignored by the NLP community, leading to minimal advancements. Given that, we strategize our plan of action first to collect the relevant data to the domain. Also, to use the data, we will dive deep into sophisticated annotation guidelines to prepare the data for the dialogue understanding and summarization tasks.

Moving ahead, once we have our own data in hand, our plan of action will open up several branches to work on modules. To initiate, we plan to work with emotion-aware intent/dialog-act joint classification task as well as dialogue summarization in parallel for the corpus in extractive summarization track to extract important utterances from the dialogue. Once we know the data specifics and how we function on our data textual modality, we plan to move our research to incorporating different modalities. We intend to work on multimodal domains to generate summaries via important utterances. In succession, we plan to question the existing state-of-the-art language models in comparison to topic-guided attention-based models. This investigation could further interpret some deep insights on topic-guided extractive summary generation and improve upon existing conversational methodologies.

Duration: 2 years

- **Deliverables and Timeline**

● [1-2 Months]: Our primary focus in the first two months would be towards data collection for each sub-task, refine it using various data preprocessing techniques and annotate the data with emotion labels as well as reference summaries. By the end of two months, we expect to deliver a novel dataset of conversations.

● [3-5] Months: In the next three months, we aim to design a novel emotion-aware dialogue-act joint attention-aware classification model which would be ready to be deployed in chatbots for emotion and intent understanding.

● [6-10] Months: In the span of five months, our aim would be to devise a deep-learning based architecture for dialogue summarization. We would focus on two modules -- important utterance detection, and dialogue topic modeling. We would aim to finalize these two modules and in parallel work on the information extraction part with an analysis of co-dependence of key utterances and topics in the healthcare domain.

● [11-17] Months: In the next seven months, we aim to incorporate multi-modality in the dialogue with text, image and speech understanding and deliver a complete emotion-aware dialogue understanding module.

● [18 - 24] Months: In the last seven months, our aim would be to develop a language generation model for generating abstractive summaries using the important utterance detection and topic detection modules delivered earlier. In the end, we intend to deliver a complete extractive to abstractive summary generation deployment-ready module capable of generating healthcare conversation summaries with important utterances, topics, and generated summaries.

- **Budget (in INR)**

A.Recurring	**Ist Year**	**2nd Year**	**3rd Year**	**Total (In Rupees)**
2 Phd/RA @35000/-	840,000.00	840,000.00	840,000.00	2,520,000.00
B.Non-Recurring- 1 GPU Card	480,000.00	-	-	480,000.00

Total	**1,320,000.00**	**840,000.00**	**840,000.00**	**3,000,000.00**

References:

[Huang et al.2018] Chenyang Huang, Osmar Za ̈ıane, Amine Trabelsi, and Nouha Dziri. 2018. Automatic dialogue generation with expressed emotions. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers), pages 49–54, New Orleans, Louisiana, June. Association for Computational Linguistics.

[Song et al.2019] Zhenqiao Song, Xiaoqing Zheng, Lu Liu, Mu Xu, and Xuanjing Huang. 2019. Generating responses with a specific emotion in dialog. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 3685–3695, Florence, Italy, July. Association for Computational Linguistic

[Kurt et al.2020] Kurt Shuster, Eric Michael Smith, Da Ju, and Jason Weston. 2020. Multi-ModalOpen-Domain Dialogue.arXiv:2010.01082 [cs] (Oct. 2020). http://arxiv.org/abs/2010.01082 arXiv: 2010.01082

25. Title: Neev: An AI-Enabled Cloud- and Smartphone-Based Learning Platform for Community Health

- Rationale for the proposed work

Accredited Social Health Activists (or ASHAs) have played a crucial role in bringing down the under-5 child mortality rate in India by delivering public health services to mothers residing in low-resource settings. Though ASHAs play an essential role, they are inadequately trained, leaving them ill-equipped to handle many situations in the field. Currently, ASHAs travel to the Primary Healthcare Center (PHC) to receive training through face-to-face sessions, which can be time and cost-intensive. The lack of trained instructors and poor infrastructure for organizing such training sessions often leads to inadequate learning opportunities. COVID-19 further highlights the need to deliver remote and timely training.

We plan to develop an AI-enabled, cloud and smartphone-based learning platform for ASHAs and mothers residing in low-resource settings. Our platform will enable interactive learning sessions with experts as well as asynchronous means to reach experts with queries. ASHAs will have flexibility in attending learning sessions and be able to personalize their learning using self-assessment and self-learning. Additionally, the system will leverage community support, AI-enabled chatbots, and automatic curation of data for building knowledge repositories.

Our platform will provide a low-cost, extensible, learning framework for ASHAs and mothers to learn and access authenticated information without having to travel or coming in contact with outsiders. The platform's extensible nature will make it future-ready to allow new functionalities that can support future requirements. Data collected will provide essential insights into building large-scale learning platforms for the masses.

- Work plan (maximum 3 pages)

Our proposed learning platform, Neev, would provide the foundation to build knowledge through learning sessions with instructors, accessing learning materials, peer engagement, and self-assessments to find and fill the knowledge gaps. Unlike phase I, where only the host had the smartphone app, in Neev, the ASHAs with the smartphones will also have an app to manage and execute their learning. With Neev, we aim to implement an extensible learning platform that can be updated with newer learning echnologies with time. The initial phase of the project will focus on developing the following components:

• An AI-enabled cloud-based Middleware Platform: The middleware will form the core of our system, managing the learning material, connecting with ASHAs and mothers in a low-resource context, and enabling the learning app to deliver effective personalized learning. We will handle the learning material, as well as content created during the learning sessions on the cloud which can then be made accessible to learners on demand using the app developed to deliver learning. The cloud platform would implement AI algorithms to support the feature described above for automatic annotation of learning material and content created during the sessions to curate them for the knowledge resource data set. We will explore Edge Computing to enable AI solutions in low-resource settings.

• Mobile-based applications for ASHAs, mothers, Medical Experts, and Administrators: A series of mobile applications will be developed to achieve the mentioned objectives. The apps for ASHAs and mothers will allow them to access learning material, contact experts, connect with peers, access assessment and self-learning tools, etc. Additionally, the ASHA app will empower them to schedule learning sessions as per their needs. For the app for mothers, we would also have a feedback system for reporting on the services rendered by ASHAs and sharing success stories. The problem of assessing ASHAs' work in the field is challenging as usually direct access to mothers is

difficult, and therefore previous studies have evaluated ASHAs' performance through self-reporting. Limitations in assessing ASHAs' work in the field make it difficult to assess the effectiveness of training or to identify gaps that still need to be covered. Previous studies have evaluated the impact of training through ASHAs' self-reporting. However, cross-validating this information is necessary to understand the effectiveness of training on the ground. The mobile app for mothers will enable them to receive their feedback about the ASHAs' work and understand what impact the training is bringing to their work. This will give us important inputs to tailor our training to provide maximum impact on the ground. Moreover, the app for mothers will allow them to share success stories directly instead of through ASHAs, thus providing further validation of the training. The app for Medical experts will enable them to conduct synchronous training sessions as well as reply asynchronously to queries by ASHAs and mothers. The app for administrators will help them in managing learning sessions, understanding learning needs, consumption of learning material, etc.

The development phase will follow a human-centered design process for developing the app. Deployment and Experiments: With our NGO partner SWACH, we aim to deploy our learning platform to about 1000 ASHAs and 100 mothers. We will provide ASHAs and mothers with smartphones loaded with our application and learning materials or install the application on their phones.

- **Budget Head**

Item		Ist Year Amount(Rs.)	2nd Year Amount(Rs.)	3rd Year Amount(Rs.)	Total (In Rupees)
A. Recurring					
Designation and number of persons	Monthly Emoluments				
Research Associate/Phd Student		-	420,000.00	420,000.00	840,000.00
Travel		300,000.00	300,000.00	300,000.00	900,000.00
Contingency		100,000.00	100,000.00	100,000.00	300,000.00
Consumable		35,000.00	35,000.00	40,000.00	110,000.00
Total (A)		**435,000.00**	**855,000.00**	**860,000.00**	**2,150,000.00**
Non-Recurring-Equipment*	850000				
High-end Workstation to act as a server		500,000.00	-	-	500,000.00
Two laptops		300,000.00	-	-	300,000.00
Printer (one b/w and one color)		50,000.00	-	-	50,000.00
Total (B)		**850,000.00**	-	-	**850,000.00**
Gross Total (A+B)		1,285,000.00	855,000.00	860,000.00	3,000,000.00

Equipment costs are tentative and actual costs of printers and laptops may be different than indicated; however, the overall limit of 8,50,000 will be maintained.

26. Title: Enabling Smart-sensors via Novel Edge-AI and In-memory Compute Paradigms: From Design, Prototype to Fabrication

- **PI and Team**

Name	Designation	Affiliation	Email & Mobile No.
Dr. Anuj Grover	Associate Professor	IIIT Delhi	anuj@iiitd.ac.in 9910013387
Dr. Sumit Darak	Associate Professor	IIIT Delhi	sumit@iiitd.a c.in 8800890561
Nitin Chawla	ST Fellow, Strategy and Innovation, Technology and Design Platforms	STMicroelectronics (I) Pvt Ltd	nitin.chawla @st.com 9560641115

- **Introduction and Problem Definition**

  With the proliferation of Industry 4.0 and large scale cyber-physical systems, more than 30 billion sensors are expected to be deployed by 2022. Existing battery powered devices such as Alexa, Google home, Siri, have limited computation capability and hence, communicate the sensed information to cloud servers. Augmenting such sensors with built-in cognitive computing capabilities (referred to as EDGE-AI devices) can minimize the communication overhead. Also, privacy and security threats demand data processing at edge devices.

In this project, we aim to design and efficiently map novel AI/ML algorithms on low-power and lowarea hardware chips. In collaboration with our industry-partner, STMicroelectronics, we aim to develop FPGA prototype of the proposed work followed by on-chip fabrication on exclusive state-ofthe-art technology.

To bring these applications to reality, a new research domain with focus on the dedicated hardware chips for the acceleration of AI/ML algorithms has gained significant importance and is the main focus of this proposal.

• Efficient and synthesizable architectures for mapping the state-of-the-art AI/ML algorithms is the first problem to be addressed in this project.

• The IMC circuits design, analysis and realization is the second problem to be addressed in this project along with an aim to reach the goal of 2000 TOPS/W.

- In summary, the proposal targets ultra low power envelope cognitive computing for edge devices, where mainstream research targets AI in the cloud. We follow a two pronged approach of reducing the resource footprint of AI algorithms and also increasing energy efficiency by doing IMC. While the AI algorithms will be validated and verified on FPGAs, IMC will be verified by fabricating and validating on Silicon.

- Research Methodology

First stage is to build a clear specification for the design and formulate a verification and optimization strategy followed by design and implementation. The next stage is the hardware and software partitioning as the design proposed in this project would function as a separate subsystem within the chip. Efficiency of AI algorithms will then be enhanced to efficiently operate in resource constrained battery operated edge devices. For instance, in addition to conventional parameters such as number of layers, filter size, appropriate quantization level is also an important factor in hardware architecture.

Improving energy efficiency of the system requires minimizing this power consumption. This can be done by moving the computation close to the memories (Near Memory Compute) or inside the memories (In-Memory Compute). We aim to design practical approaches for In-Memory Computations involving innovation not only in memory circuits, but also in interface, instructions, and computation management. This has the potential to bring in large parallelism in a very small energy envelope, thereby improving the energy efficiency of hardware accelerators.

Integrating IMC SRAM on a SoC additionally requires design of associated controllers. Once the design is complete, the next step is the complete functional verification on FPGAs, where Xilinx MPSoC boards such as ZCU102, ZCU104, ZC706 will be utilized for exercising real time scenarios.

A full RTL to GDS flow will be carried out on the design over multiple iterations. Flows for different technologies at the industry partner - STMicroelectronics - will be utilized for design implementation. The RTL will be designed using Verilog and SystemVerilog. Synthesis, Formal verification, Design for Test (DFT), Power and Timing Analysis is to be performed on the design as it moves forward through the flow. Floorplan, Clock Tree Synthesis and Routing follows, which is a part of the backend flow. Once the backend flow is complete, timing checks are performed on the design, power analysis will be done on the generated netlist for the design, and functionality is verified again through GLS (Gate Level Simulation).

- Budget

Item			**Budget**		
Recurring		**Ist Year**	**2nd Year**	**3rd Year**	**Total (In Rupees)**
1. Salaries/wages					
Designation & number of persons	Monthly Emoluments				
JRF (2-hired mid-first year)/RA	38400	460,800.00	921,600.00	460,800.00	1,843,200.00
SRF (2-half-year)/RA	43400	-	-	520,800.00	520,800.00
3. Travel (Industry Visits)		-	30,000.00	50,000.00	80,000.00
4. Contingency (Conference Reg. and Page Length Charges)		-	30,000.00	50,000.00	80,000.00
Total		**460,800.00**	**981,600.00**	**1,081,600.00**	**2,524,000.00**
Non-Recurring			-	-	-
ZYNC FPGA Boards		476,000.00	-	-	476,000.00
Laptop		-	-	-	-
Chip Fabrication Cost		(Paid by STMicroelectronics-ranging between 30 to 50 lacs per interation)			
Total (A+B)		**476,000.00**	**-**	**-**	**476,000.00**
Grand Total (A+B+C)		**936,800.00**	**981,600.00**	**1,081,600.00**	**3,000,000.00**

Pert Chart

Phases of the project	Quarters											
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
1) Design and Implementation of AI algorithms												
2) Design of circuits and architectures for IMC												
3) Integration of AI algorithms with IMC and optimization												
4) Design and implementation of security algorithms												
5) FPGA Prototypes												
6) On Chip fabrication and validation												
7) Demonstration via suitable application												

27. Title: Dataset Curation of Wearable Health Monitoring Data for Indian Population

PI : Sujay Deb, IIIT Delhi

1. **Context/Background:** Wearable sensors have the potential to enable the monitoring of individuals' health data in a reliable and continuous manner. A range of scientifically validated wearable sensors is already available commercially (e.g., Fitbit1 , Apple Watch2 , Embrace3) that can measure different physiological parameters of the human body including heart rate, oxygen saturation, Electrocardiogram, Photoplethysmography, Galvanic Skin Response, etc. These data can not only be used to monitor one's physical health but when combined together it can also be indicative of one's emotional health. However, to date, no integrated dataset of health data from wearable sensors is available for the Indian population. Furthermore, there is a lot of variation in the wearable sensors technology and hence the data is of different types. Enabling a national level synchronized and integrated dataset of health data from wearable sensors for the Indian population will add significant value to the researchers and other stakeholders in the e-Health and m-Health domain.

2. **Problems to be addressed**

Remote healthcare monitoring allows people to continue to stay at home rather than in expensive healthcare facilities such as hospitals or nursing homes. It thus provides an efficient and cost-effective alternative to on-site clinical monitoring.

Such systems equipped with non-invasive and unobtrusive wearable sensors can be viable diagnostic tools to the healthcare personnel for monitoring important physiological signs and activities of the patients in real-time, from a distant facility. Therefore, it is understandable that wearable sensors play a critical role in such monitoring systems that attracted the attention of many researchers, entrepreneurs, and tech giants in recent years. A variety of application specific wearable sensors, physiological and activity monitoring systems were proposed in the literature.

Wearable devices can monitor and record real-time information about one's physiological condition and motion activities. Wearable sensor-based health monitoring systems may comprise different types of flexible sensors that can be integrated into textile fiber, clothes, and elastic bands or directly attached to the human body. The sensors are capable of measuring physiological signs such as electrocardiogram (ECG), electromyogram (EMG), heart rate (HR), body temperature, electrodermal activity (EDA), arterial oxygen saturation (SpO2), blood pressure (BP) and respiration rate (RR). In addition, micro-electro-mechanical system (MEMS) based miniature motion sensors such as accelerometers, gyroscopes, and magnetic field sensors are widely used for measuring activity related signals. Continuous monitoring of physiological signals could help to detect and diagnose several cardiovascular, neurological and pulmonary diseases at their early onset. Also, realtime monitoring of an individual's motion activities could be useful in fall detection, gait pattern and posture analysis, or in sleep assessment. The wearable health monitoring systems are usually equipped with a variety of electronic and MEMS sensors, actuators, wireless communication modules and signal processing units. The wearables have the potential to lead to an overall increase in the efficiency and effectiveness of under-resourced health infrastructures, ultimately translating into benefits for patients.

In general, however, the scalability of the wearable-based mHealth projects from pilot projects to large-scale implementation has been low. This can change if an ecosystem of data collection and sharing can be developed that will work as a catalyst for further development and

1. https://www.fitbit.com/in/home
2. https://www.apple.com/in/apple-watch-series-5/
3. 3https://www.empatica.com/embrace2/

scaling of the solutions. Towards that we propose to develop an integrated, synchronized, and annotated dataset of wearable health data, consisting of the different physical and physiological human body parameters. In consultation with a team of medical doctors, the project will start with the identification of the different types of physical and physiological data that can be of significant potential in the monitoring of human health. Further, we will identify/ develop commercially viable wearable sensors that can collect these parameters at a mass scale and share the data with a cloud based service for reporting and analysis. Then, we will perform a careful selection of wearable devices on different parameters, such as ease of use, minimal or no training requirement for the user, calibration-free, compatibility with existing compute and communication infrastructure, etc. As an example, for continuous blood pressure monitoring, we will use a cuffless noninvasive sensor developed in-house. Further, a significantly representative population of India considering different gender, age-groups, etc will be chosen for data collection. The wearable health data will be collected among the chosen sample population while performing different activities of daily life. The project will also develop an intelligent middleware consisting of built-in artificial intelligence and machine learning techniques. These techniques can be invoked through a set of application programming interfaces (APIs) to process different types of smart queries (e.g. prediction, classification, recommendation, etc.) produced by different end-users such as the researchers, clinicians, and patients. The end-users will be able to develop patient-centric applications in a faster manner by making use of these APIs. The developed dataset and the intelligent middleware platform will be made in agreement with India's Privacy and Data Protection (PDP) law, Health Insurance Portability and Accountability Act (HIPAA), and General Data Protection Regulation (GDPR) guidelines to ensure data security and data privacy.

In this project we propose to work on the following aspects:

We will develop an integrated, synchronized, and annotated health dataset collected from multiple wearable body sensors. The dataset will correspond to different health conditions, collected from healthy as well as diseased patients of different age groups, say patients suffering with heart abnormality, old patients needing day care, patients with neurodevelopmental disorders, patients with spinal injury, patients in ICU, and so on. The dataset would capture physical and physiological human body parameters collected over longer durations. This dataset will be made available as a national repository, easily accessible by researchers countrywide, for carrying out research in the domain of affordable low-cost AI enabled healthcare solutions.

We will evaluate existing sensor technologies for wearable health monitoring systems along with the possible suggestions on improvement on hardware technology and software components. We will also explore different use cases where we will demonstrate the end-to-end solutions, e.g., a low-cost integrated augmentative and alternative communication tool for individuals with special needs for bed-ridden individuals would be developed as part of the project using a part of the data. We do expect that the collected data will be able to drive many more diverse solutions of related healthcare needs including maternal and child public healthcare.

We will also provide an intelligent middleware consisting of built-in artificial intelligence and machine learning techniques. These techniques can be invoked through a set of application programming interfaces (APIs) to process different types of smart queries (e.g., prediction, classification, recommendation, etc.) produced by different end-users such as the researchers, clinicians, and patients.

3. Aims and Objectives

1. To develop an integrated, synchronized, and annotated dataset of health-data from wearable sensors to facilitate the research and development in the e-Health and m-Health domain for the Indian population.

2. Development of the smart application programming interface (API) tools for the aggregated health data based on artificial intelligence and machine learning techniques enabling intelligent queries.

4. **Strategy**

To date, no integrated dataset of health data from wearable sensors is available for the Indian population. Furthermore, there is a lot of variation in the wearable sensors technology and hence the data is of different types. Enabling a national level synchronized and integrated dataset of health data from wearable sensors for the Indian population will add significant value to the researchers and other stakeholders in the e-Health and m-Health domain. The developed dataset and intelligent middleware will enable different stakeholders to advance the state of the art in e-Health as well as in m-Health for the Indian population, improving their physical and mental health thereby creating a healthy India.

5. **Target beneficiaries**

The project will develop an integrated, synchronized, and annotated health dataset collected from multiple wearable body sensors. The dataset would correspond to different health conditions, collected from healthy as well as diseased patients of different age groups, say patients suffering with heart abnormality, expecting mothers, new born babies, old patients needing day care, patients with neurodevelopmental disorders, patients with spinal injury, patients in ICU, and so on. The dataset would capture physical and physiological human body parameters collected over longer durations. This dataset will be made available as a national repository, easily accessible by researchers countrywide, for carrying out research in the domain of affordable low-cost AI enabled healthcare solutions. This kind of ecosystem will accelerate the research and development in this domain and will pave the way for effective m-Health and e-Health solutions for masses. Companies: Microsoft Research, IBM, TCS Research, Google, Samsung etc.

6. Budget

Item		**Budget**			
A. Recurring		**Ist Year**	**2nd Year**	**3rd Year**	**Total (In Rupees)**
1. Salaries/wages					
Designation & number of persons	Monthly Emoluments				
PHD/RA-01	40,950.00	491,400.00	491,400.00	491,400.00	1,474,200.00
2. Consumable		100,000.00	100,000.00	100,000.00	300,000.00
3. Travel		10,000.00	10,000.00	10,000.00	30,000.00
4. Contingency		115,000.00	115,000.00	115,000.00	345,000.00
Total (A)		**716,400.00**	**716,400.00**	**716,400.00**	**2,149,200.00**

B. Non-Recurring-Equipment					
Storage Server		400,000.00	-	-	400,000.00
Wearable Sensing Solutions		300,000.00	-	-	300,000.00
Mobile Data Collection Platform		150,000.00	-	-	150,000.00
Total (B)		**850,000.00**	**-**	**-**	**850,000.00**
Grand Total (A+B)		**1,566,400.00**	**716,400.00**	**716,400.00**	**2,999,200.00**

HUB POLICIES

18. Hub Policies

IHUB Anubhuti-IIITD Foundation Purchase Policy and Financial Delegation of Authority (DoA)

IHUB ANUBHUTI-IIITD FOUNDATION

(A Section 8 Company Promoted by IIIT-Delhi)

IHUB/E/2021/2 Date: 27/7/2021

IHUB ANUBHUTI-IIITD FOUNDATION PURCHASE POLICY

1. PREAMBLE

This policy has been framed, based on series of meetings held by a committee consisting of members from IHUB Anubhuti, Faculty members of IIITD, and by the consent of Executive Committee of IHUB Anubhuti, for purchase from IHUB's planned and non-planned funds generated from sponsors and self-efforts and other funds allocated to IHUB Anubhuti.

This policy is also in consultation with the Purchase Policy and DOA of IHFC Purchase Policy (IIT Delhi Technology Innovation Hub).

These rules for purchase and accounting of various expenses of setting-up and running the organization and thereby provide a conducive working environment for employees/staff, partners and associated entities to promote excellence expected from IHUB and as such should be interpreted in that context, so that the procurement of the needed equipment/stores/services is done in time and without procedural wrangles, which permits the desired task to be pursued with greater vigor.

Every authority delegated with the financial powers of procuring goods/services in public interest shall have the responsibility and accountability to bring efficiency, economy, and transparency in matters relating to procurement and for fair equitable treatment of suppliers and promotion of competition in procurement. The procedure to be followed in making procurement must confirm to the following yardsticks.

a) The specification in the terms of quality, type, duration etc. and quantity of goods to be procured, should be clearly spelt out keeping in view the specific needs of the procuring entity without including non-essential features, which may result in unwarranted expenditure.
b) Offers should be invited following a fair, transparent and reasonable procedure.
c) The procuring authority should be satisfied that the equipment(s)/item(s) / material/ service partner(s)/vendor(s) selected offer(s)/adequately meet(s) the requirement in all respects.
d) The procuring authority should satisfy itself that the price of the selected offer is reasonable and consistent with the quality required.
e) Purchase should not be split to avoid obtaining approval of appropriate competent financial authority.
f) The procedure to be followed for a purchase would follow the Financial DOA guidelines.

2. CLASSIFICATION OF STORES

Non- Consumable Stores (NC)		Consumable Stores
Permanent Assets (PA)	Limited Time Assets (LTA)	
Stores satisfying anyone of the following conditions shall be classified as permanent Assets 1. Stores which are intended to be used over prolonged period (more than 5 years) before becoming unusable or obsolete Eg. Office furniture and fixtures	Stores satisfying anyone of the following condition shall be classified as Limited Time Assets • Store consisting any amount and having useful life of less than five years, which rapidly lose their value/ relevance with the lapse of time or have very little or negligible disposal value • Stores which can be upgraded either by replacing components/part or which can be rendered obsolete by the release of new versions or editions. • Stores which can be used over any period of time but costing less than Rs.75,000/ excluding office furniture and fixtures. Examples of such items PCs, Laptop, Tablet Cameras, Mobiles, Software Projector, LFT monitors	Stores satisfying anyone of the following condition shall be classified as Consumable stores. • Stores which exhaust rapidly with the lapse of time (normally within one year) • Stores which are rapidly rendered unusable due to normal wear and tear. • Stores which have negligible disposal value • Spares of equipment etc. • Examples, Chemicals, stationery items, printer toner/cartridge, electronic storage media like pen drive. Computer parts requiring replacement such as the mother board, RAM hard disc etc. Batteries of any kind (like UPS, laptops/mobiles portables instruments etc.) electronic components like capacitors, connectors, electrical, wires, plugs switches, tool bits and hand tool etc.

a. **Any item with value of greater than INR 2500 (including taxes) will automatically get classified as Non-Consumables**

b. In case of any ambiguity with respect to classification of stores, the same may be resolved by the CEO, IHUB Anubhuti and one of the Directors of IHUB Anubhuti.

Version 1: 2021

2. SALIENT FEATURES OF THE POLICY

i. Upgradation and Repairs: In case of upgradation of assets, the old asset which has been upgraded will be treated as written-off i.e., its value will be written-off from the records. The upgraded new asset will be entered in the records. Sum of cost of old asset and the additional cost paid, will be considered as cost of the upgraded asset. For any repairs of imported assets, in case the items is to be sent back to the OEM in foreign country, re-import process should be initiated.

ii. Purchase of Proprietary Items: IHUB Anubhuti Single Tender Enquiry form may kindly be used.

iii. Earnest Money: To safeguard against a bidder's withdrawing or altering its bids during the bid validity period, the Purchase Committee may obtain bid Security (also known as Earnest Money) (2.5% of the estimated cost).

iv. Performance Guarantee: In order to ensure due performance of the contract, Purchase Committee may obtain performance security from the successful bidder awarded the contract (5-10% of the order value). Performance security may be obtained in the form of demand draft or Bank Guarantee from a scheduled bank in an acceptable form safeguarding the purchaser's interest in all respect.

v. For all purchase, Purchase Requisition Form needs to be used.

vi. Role of Store Purchase Section
   1. Processing and clearance of Indian/ International purchase
   2. E-Procurement, E-Publishing, E-Auction, Onboarding/Registration of Vendors
   3. Maintenance of Central Asset Registers for all items/ Vendor Details
   4. Conducting physical annual stock/contract verification
   5. Centralized GIS
   6. Maintenance of Purchase Requisition Forms against all purchases

vii. Payment: Payment should be made to vendors through RTGS/Cheque/NEFT.

viii. The value of a service contract would be considered basis its annual value (from the date of initiation of contract), irrespective of the tenure of the contract, for deciding the authority to finalize/terminate the vendor for that service.

ix. Employee recruitment would not be considered a part of service contracts and would be as per the HR and Recruitment Policy.

x. Write-off/Disposal/Re-use/abnormal loss/theft of Stores: IHUB Anubhuti will initiate the process of write-off/disposal/Re-use of stores after the completion of project life-time is over. It would be governed by Financial DOA.

a) Any loss or shortage of public money, departmental revenue, receipt, stamps, stores or other property held by or on behalf of the IHUB. Irrespective of the cause of loss and manner of detection, shall be immediately reported to the next higher authority. Cases involving serious irregularities shall be brought to the notice of CEO/Project Director.

b) Report of loss contemplated in (a) above shall be made as follows - An initial report with the security officer should be made as soon as a suspicion arises that a loss has taken place. FIR be made immediately, and "Non-Traceable Report" issued by police station be submitted for the purpose of write –off of the item/s.

xi. Any rule/purchase procedure not covered here shall be governed through requisite approval from Board of Directors/Hub Governing Body of IHUB Anubhuti

Initiated by:	Recommended by:	Approved by:
Mukesh Malhotra	Prof. Pushpendra Singh	Prof. Ranjan Bose
CEO - IHUB Anubhuti	Project Director - IHUB Anubhuti	Director - IIIT Delhi Chairman(HGB)-IHUB Anubhuti

Version 2021

IHUB ANUBHUTI – IIITD FOUNDATION — Financial Delegation of Authority (DoA)

1. Approval authority for store purchase/service contract

Amount (INR)	Process/Guideline	Initiator	Recommendation	Approver
Upto 25k	Need to accompany actual invoices	Employee/PI	None	CEO
25k - 2.5 Lac	Single vendor allowed	Employee/PI	Two member purchase committee from the EC	CEO
2.5Lac - 10 Lac	Minimum 3 vendors	Employee/PI	Two member purchase committee from the EC	CEO
10 Lac - 25 Lac	Minimum 3 vendors	Employee/PI	Three member purchase committee from EC (CEO mandatory member	Project Director
Above 25 Lac	Minimum 3 vendors	Employee/PI	Three member purchase committee from the EC (Project Director and CEO mandatory members)	Director

2. Approval authority for travel expenses

Amount (INR)	Process/Guideline	Initiator	Approver
Upto 1 Lac	Need to accompany actual invoices for travel and requisite approval for travel from CEO/ Project Director	Employee, IHUB	CEO

Version 2021

1 Lac and above	Need to accompany actual invoices for travel and requisite approval for travel from CEO/ Project Director	Employee, IHUB	Project Director

3. Approval authority for salary expenses

Amount INR	Process/Guideline/ Initiator	Recommendation	Approver
	HR person to initiate the process basis employee records and contractual obligations*	CEO, IHUB	Project Director

*Authority to issue an offer letter or appointment letter would be with CEO or above

4. Financial power to write-off the assets

Amount (INR)	Process/Guideline	Initiator	Recommendation	Approver
Upto 25k	As per Purchase Policy	Employee/PI/Vendor	None	CEO
25k - 2.5 Lac	As per Purchase Policy	Employee/PI/Vendor	Two member committee from the EC	CEO
2.5Lac - 10 Lac	As per purchase policy	Employee/PI/Vendor	Two member committee from the EC	CEO
10 Lac - 25 Lac	As per purchase policy	Employee/PI/Vendor	Three member purchase committee (CEO mandatory member	Project Director

Version 2021

Above 25 Lac	As per purchase policy	Employee/PI/Vendor	Three member purchase committee (Project Director and CEO mandatory members)	Director

1. For any service contracts, financial authority would be defined based upon the annual value of contract, irrespective of the tenure of the contract

2. Any exception to the above DOA requires approval from the CEO and Project Director, IHUB Anubhuti.

Initiated by:

Mukesh Malhotra
CEO - IHUB Anubhuti

Recommended by:

Prof. Pushpendra Singh
Project Director - IHUB Anubhuti

Approved by:

Prof. Ranjan Bose
Director – IIIT Delhi
Chairman (HGB) – IHUB Anubhuti

Version 2021

IHUB ANUBHUTI-IIITD FOUNDATION
(A Section 8 Company Promoted by IIIT-Delhi)
Website: https://ihub-anubhuti-iiitd.org/

IHUB/E/2021/1 Date: 27/7/2021

Guidelines for Honorarium by IHUB Anubhuti

The Board of Directors in their 2nd Board Meeting held on 27th July'2021 have proposed following guidelines for payment of honorarium to IIITD Office Staff (includes regular staff and staff hired through outsourced agency) engaged in the services for IHUB Anubhuti-IIITD Foundation (IHUB). This also includes the Faculty members who are serving as the Project Directors.

- Work by IIITD Faculty and Staff (Administrative Staff/Administrative Staff/ Support Staff) for IHUB is treated as "extra work" and would be applicable for Honorarium by IHUB.

- Services include work done by the IIITD Faculty and Staff during the set-up of IHUB, such as: consultation, hiring, staffing, advertisements, operational activities, banking facilities, IT Services, procurement, and other activities involving conferences/ workshops, etc.

- Services also include work done by the Project Directors on the various projects which IHUB is undertaking for implementing the objectives of National Mission on Interdisciplinary Cyber Physical Systems.

- Futures services by IIITD Faculty and Staff may also include consultation, hiring, staffing, advertisements, operational activities, IT Services, procurement and other activities involving conferences/workshops, etc. as per the requirement by IHUB.

- This honorarium will be approved by the IHUB CEO and the Project Directors for all the IIITD Faculty and Staff other than Project Directors.

- For the Project Directors, this will be approved by the CEO and the Director.

- Budget provision should be available for compensating staff. Number of hours/days for each individual will be mutually decided between IIITD concerned approver and the IHUB CEO/ Project Directors depending upon the quantum of assigned responsibility.

Regd.Office: GB Pant Polytechnic Extension, Okhla Phase-III, Delhi-110020, South Delhi, India
Tel: +911126907400 CIN: 73100DL2020NPL374793

Honorarium Structure is as follows:

- For services rendered by Senior Faculty (Project Directors/Professors) and Senior Administrative Staff(Registrar, Controller of Finance) = Rs. 5,000/- per hour
- For services rendered by other Faculty Members (Associate Professors, Assistant Professor) and Mid- level Administrative Staff(Senior Manager and above) = Rs. 3,000/- per hour
- For services rendered by Research Associates/Students and Junior Administrative Staff(Managers and below) = Rs. 1,000/- per hour

This is being issued with the approval of the competent authority and is applicable retrospectively right from the time of start of this company. It includes services rendered by IIITD Faculty and Staff at the time of pre-incorporation of the company also.

(Mukesh Malhotra)
CEO

Copy for information / necessary action to: (through email)

1. IHUB Chairman/ IIITD Director
2. IHUB Board of Directors
3. IHUB Project Directors
4. IHUB Staff
5. IIITD Faculty & Office Staff
6. IIITD CoF office
7. IIITD Registrar Office
8. IIITD – IRD Office

Addendum to the Guidelines for Honorarium by IHUB Anubhuti

IHUB ANUBHUTI-IIITD FOUNDATION

(A Section 8 Company Promoted by IIIT-Delhi)
Website: https://ihub-anubhuti-iiitd.org/

IHUB/E/2021/3 Date: 16/8/2021

Addendum to the Guidelines for Honorarium by IHUB Anubhuti

The following section is modified and replaced and will overwrite any earlier text mentioned in the original Guideline for Honorarium IHUB/E/2021/1 with respect to the Honorarium Structure.

Honorarium Structure is as follows:

- For services rendered by Senior Faculty (Project Directors/Professors) and Senior Administrative Staff (Registrar, Controller of Finance) = Rs. 5,000/- per hour
- For services rendered by other Faculty Members (Associate Professors, Assistant Professor) and Mid-level Administrative Staff (Manager and above) = Rs. 3,000/- per hour
- For services rendered by Research Associates/Students and Junior Administrative Staff = Rs. 1,000/- per hour

This is being issued with the approval of the competent authority.

(Mukesh Malhotra)
CEO

Copy for information / necessary action to: (through email)

1. IHUB Chairman/ IIITD Director
2. IHUB Board of Directors
3. IHUB Project Directors
4. IHUB Staff
5. IIITD Faculty & Office Staff
6. IIITD CoF office
7. IIITD Registrar Office
8. IIITD – IRD Office

Regd. Office: GB Pant Polytechnic Extension, Okhla Phase-III, Delhi-110020, South Delhi, India
Tel: +911126907400 CIN: U73100DL2020NPL374793

Memorandum of Association

MEMORANDUM OF ASSOCIATION

OF

IHUB ANUBHUTI-IIITD FOUNDATION

Form No. INC-13

[Pursuant to rule 19(2) the Companies (Incorporation) Rules, 2014]

I. The name of the company is "**IHUB ANUBHUTI- IIITD FOUNDATION**"

II. The Registered office of the company will be situated in the **"NCT OF DELHI".**

III. The objects for which the company is established are:

1. To establish a Technology Innovation Hub (TIH) at IIIT- Delhi for implementation and realising the part of objectives of National Mission on interdisciplinary Cyber-Physical Systems (NM-ICPS) on the theme of Cognitive Computing and Social Sensing to integrate computation and physical processes in a dynamic environment in order to spur growth and technology led economic development.

2. To carry on expert-driven research, Consortium based Research through Cluster-Based Network Programmes, directed research for the specific requirements of Industry, other Govt. verticals and International Collaborative Research Programmes.

3. To carry on research collaboration, drawing faculty from multiple departments, centres andother relevant entities/units in the institute.

4. To carry on HRD & Skill Development through Fellowship Based UG/ PG, Ph.D., Post-Doctoral and Short Term Training for Faculty in Innovation, Entrepreneurship and Start-upEcosystem, to enhance competencies, capacity building and training to nurture innovationand Start-up ecosystem and to establish and strengthen the international collaborativeresearch/ Technology Development in the field of development of intelligent machines forhuman assistance.

5. To introduce inter-disciplinary academic programs so that researchers from two or more disciplines pool their approaches and modify them so that they are better suited to the problem at hand, including the case of the team-taught course where students are required to understand a given subject in terms of multiple traditional disciplines.

1

6. Tofacilitate intra mural and extra mural research and to establish research labs with the intention of performingquality research, technology transfer and commercialization, and creating future generationhuman resources

7. To foster linkages with other National and International institutes to carry out programmes/projects in the field of development of intelligent machines on global, regional and national level.

8. To organize awareness campaigns on Cognitive Computing and Social Sensing and collaborate with neighbourhood Universities for student training/internshipprogrammes and to participate in Government formed platforms.

9. To provide education, consultancy and a unique set of role based competencies byTechnology Skill Certification and to offer a gap filling option and a brand-agnostic Independent certification and to carry on all other activities in line with the above captionedObjects the Company and to facilitate the dissemination of information by organizingconferences, seminars including future-vision seminars on forecasting of trends in technologydevelopment in identified generic areas defined above.

The Company will do all such other lawful things as considered necessary for the furtherance ofthe above objectives, provided that the company shall not support with its funds, or endeavour to impose on, or procure to be observed by its members or others, any regulation or restrictionwhich, as an object of the company, would make it a trade union.

IV. The objects of the company extend to India

V. (i) The profits, if any, or other income and property of the company, when soever derived, shall be applied, solely for the promotion of its objects as set forth in this memorandum.

(ii) No portion of the profits, other income or property aforesaidshall be paid or transferred, directly or indirectly, by way of dividend, bonus or otherwise by way of profit, to persons who, at any time are, or have been, members of the company or to any one or more of them or to any persons claiming through any one or more of them.

(iii) No remuneration or other benefit in money or money's worth shall be given by the company to any of its members, whether officers or members of the company or not, except payment of out-of-pocket expenses, reasonable and proper interest on money lent, or reasonable and proper rent on premises let to the company.

(iv) Nothing in this clause shall prevent the payment by the company in good faith of prudent remuneration to any of its officers or servants (not being members) or to any other person (not being member), in return for any services actually rendered to the company.

(v) Nothing in clauses (iii) and (iv) shall prevent the payment by the company in good faith of prudent remuneration to any of its members in return for any services (not being services of a kind which are required to be rendered by a member), actually rendered to the company;

VI. No alteration shall be made to this memorandum of association or to the articles of association of the company which are for the time being in force, unless the alteration has been previously submitted to and approved by the Registrar or any other relevant Authority.

VII. The liability of the members is limited.

VIII. The Authorized Share Capital of the Company is Rs. 10,00,000 (Rupees Ten Lakhs Only) divided into 1,00,000 (One Lakh) Equity Shares of Rs. 10 (Rupees Ten Only) each.

IX. True accounts shall be kept of all sums of money received and expended by the company and the matters in respect of which such receipts and expenditure take place, and of the property, credits and liabilities of the company; and, subject to any reasonable restrictions as to the time and manner of inspecting the same that may be imposed in accordance with the regulations of the company for the time being in force, the accounts shall be open to the inspection of the members.Once at least in every year, the accounts of the company shall be examined and the correctness of the balance-sheet and the income and expenditure account ascertained by one or more properly qualified auditor or auditors.

X. If upon a winding up or dissolution of the company, there remains, after the satisfaction of all the debts and liabilities, any property whatsoever, the same shall not be distributed amongst the members of the company but shall be given or transferred to such other company/ trust/ society having objects similar to the objects of this company, subject to such conditions as the Tribunal may impose, or may be sold and proceeds thereof credited to the Rehabilitation and Insolvency Fund formed under section 269 of the Act.

XI. The Company can be amalgamated only with another company registered under section 8 of the Act and having similar objects.

XII. We, the several persons, whose names and addresses are subscribed, are desirous of being formed into a company in pursuance of this Memorandum of Association, and we respectively agree to take the number of shares in the capital of the company set against our respective names:

S. No.	Name, addresses, Description and Occupation of each subscribers	Number and type of shares equity	Signature of subscribers	Name, addresses, Description and Signatures of witnesses
01	Indraprastha Institute of Information Technology-Delhi (IIIT-Delhi) (Registered Office: GB Pant extended polytechnic (Near Govindpuri Metro Station, Okhla Phase-III, New Delhi-110020) Through its authorized representative Prof. Ranjan Bose S/o Mr. Sudip K Bose R/o 46,New Campus, I.I.T Delhi, Hauz Khas,Delhi-110016 Occupation: Service	9,999(Nine Thousand Nine Hundred Ninety Nine) equity shares	Ranjan Bose 7/12/20	I witness to subscriber who has subscribed and signed in my presence at New Delhi Further I have verified his Identification Details(ID) for his identification and satisfied myself for his identification particulars as filled in Kapil Chawla S/o Late Sh. B.S. Chawla E/No.-102 E&H Apt. Pocket 1&2 D[illegible]
02	Indraprastha Institute of Information Technology-Delhi (IIIT-Delhi) (Registered Office: GB Pant extended polytechnic (Near Govindpuri Metro Station, Okhla Phase-III, New Delhi-110020) Through its Nominee Prof. Pushpendra Singh S/o Mr. Kanchhi Singh R/o 802, Faculty Residence, IIIT Delhi, Near Govindpuri Metro Station, Okhla Phase – 3, Okhla Industrial Area , Delhi – 110020 Occupation: Service	1(One) equity shares	P. Singh 7-12-20	I witness to subscriber who has subscribed and signed in my presence at New Delhi Further I have verified his Identification Details(ID) for his identification and satisfied myself for his identification particulars as filled in Kapil Chawla S/o Late Sh. B.S. Chawla E/No 102 E&H Apt. Pocket 1&2 D[illegible]

Total Equity Shares 10,000(Ten Thousand)

Place: New Delhi

Dated: 07.12.2020

Articles of Association

(THE COMPANIES ACT, 2013)

(COMPANY LIMITED BY SHARES)

UNDER SECTION 8 OF THE COMPANIES ACT, 2013

ARTICLES OF ASSOCIATION

OF

IHUB ANUBHUTI-IITD FOUNDATION

PRELIMINARY

1. Subject as hereinafter provided the Regulations contained in Table 'F' in the First Schedule to Companies Act, 2013 shall apply to the Company to the extent they are consistent with the notified Sections of the Companies Act, 2013 and Rules made there-under from time to time (including any statutory modification(s) or re-enactment thereof for the time being in force).

The provisions of the Companies Act, 2013 and rules made there under relating to Private Company shall be applicable to the company as and when notified by the Central Government.

INTERPRETATION

2.In these regulations:

(a) "Act" means the Companies Act, 2013, or any statutory modification or re-enactment thereof from time to time.

(b) "Board of Directors" or "Board", in relation to a company, means the collective body of the Directors of the Company.

(c) "Chief Executive Officer" means an officer of a company, who has been designated as such by it.

(d) "Chief Financial Officer" means a person appointed as the Chief Financial Officer of a company.

(e) "Company" means "IHUB ANUBHUTI-IITD FOUNDATION".

(f) "Director" means a Director appointed to the Board of a Company.

(g) "Financial statement" has been defined to include:

(i) A balance sheet as at the end of the financial year;
(ii) A profit and loss account, or in the case of a company carrying on any activity not for profit, an income and expenditure account for the financial year;

(iii) cash flow statement for the financial year;

(iv.) A statement of changes in equity, if applicable; and

(v.) Any explanatory note annexed to, or forming part of, any document referred to in sub-(i) to sub clause (iv)

1

(h) "Key Managerial Personnel (KMP)", means
(i) The Chief Executive Officer or the Managing Director or the Manager,
(ii) The Company Secretary;
(iii) The whole-time director;
(iv) The Chief Financial Officer;
(v) such other officer, not more than one level below the directors who is in whole-time employment, designated as key managerial personnel by the Board; and
(vi) Such other officer as may be prescribed by Companies Act, 2013.

(i) "Member", means

(i) The subscriber to the memorandum of the company, who shall be deemed to have agreed to become member of the company, and on its registration, shall be entered as member in its register of members;

(ii) Every other person who agrees in writing to become a member of the company and whose name is entered in the register of members of the company;

(iii) Every person holding shares of the company and whose name is entered as a beneficial owner in the records of a depository

(j) "Office" means the Registered Office of the Company.

(k) "Officer" includes any director, manager or key managerial personnel or any person in accordance with whose directions or instructions the Board of Directors or any one or more of the directors is or are accustomed to act.

Unless the context otherwise requires words or expressions contained in these regulations shall bear the same meaning as in the Act, or any statutory modification thereof in force at the date at which these regulations become binding on the Company.

PRIVATE COMPANY

3. The Company is a Private Company within the meaning of Section 2(68) of the Companies Act, 2013 and accordingly :-

a) The right to transfer shares in the Company is restricted in the manner and to the extent hereinafter appearing.

b) The number of members of the Company (exclusive of persons who are in the employment of the Company, and persons who having been formerly in the employment of the Company, were members of the Company while in the employment and have continued to be members after the employment ceased) shall be limited to two hundred; provided that for the purpose of this definition where two or more persons jointly hold one or more shares in the Company, the same shall, be treated as a single member, and

c) No invitation shall be issued to the public or subscribe for any securities of the Company.

SHARE CAPITAL AND VARIATION OF RIGHTS

4. Subject to the provisions of the Act and these Articles, the shares in the capital of the company shall be under the control of the Directors who may issue, allot or otherwise dispose of the same or any of them to such persons, in such proportion and on such terms and conditions and either at a premium or at par and at such time as they may from time to time think fit, may also allot and issue shares in capital of the Company in payment or part payment for any property sold or transferred to or for services rendered to the Company in or about the conduct of its business and the shares which may be so allotted may be issued as fully paid up shares and if so issued deemed to be fully paid up shares.

5. (i) Every person whose name is entered as a member in the register of members shall be entitled to receive within two months after incorporation, in case of subscribers to the memorandum or after allotment or within one month after the application for the registration of transfer or transmission or within such other period as the conditions of issue shall be provided,

(a) one certificate for all his shares without payment of any charges; or

(b) Several certificates, each for one or more of his shares, upon payment of twenty rupees for each certificate after the first.

(ii)Every certificate shall specify the shares to which it relates and the amount paid-up thereon and shall be signed by two directors or by a director and the company secretary, wherever the company has appointed a company secretary:

Provided that in case the company has a common seal it shall be affixed in the presence of the persons required to sign the certificate.

(iii) In respect of any share or shares held jointly by several persons, the company shall not be bound to issue more than one certificate, and delivery of a certificate for a share to one of several joint holders shall be sufficient delivery to all such holders.

6 (i) If any share certificate be worn out, defaced, mutilated or torn or if there be no further space on the back for endorsement of transfer, then upon production and surrender thereof to the company, a new certificate may be issued in lieu thereof, and if any certificate is lost or destroyed

then upon proof thereof to the satisfaction of the company and on execution of such indemnity as the company deem adequate, a new certificate in lieu thereof shall be given. Every certificate under this Article shall be issued on payment of twenty rupees for each certificate.

(ii) The provisions of Articles (5) and (6) shall mutatis mutandis apply to debentures of the company.

TRANSFER OF SHARES

7. (i) The company shall not register a transfer of securities of the company, unless a proper instrument of transfer, duly stamped, dated and executed by or on behalf of the transferor and the transferee and specifying the name, address and occupation, if any, of the transferee has been delivered to the company by the transferor or the transferee within a

period of sixty days from the date of execution, along with the certificate relating to the securities, or if no such certificate is in existence, along with the letter of allotment of securities.

(ii) Each Shareholder agrees that it shall not, without the prior written consent of the other Shareholder other than inter se Transfers between the Parties, Transfer any Equity Shares held by or on behalf of such Shareholder. The Shareholders shall procure, to the extent of the irrespective rights to vote as shareholders and/or as directors of the Company, that no person is registered as the holder of any Equity Shares except in accordance with the Articles and the Agreement.

(iii) The instrument of transfer of any share in the company shall be executed by or on behalf of both the transferor and transferee.

(iv) The transferor shall be deemed to remain a holder of the share until the name of the transferee is entered in the register of members in respect thereof.

8. The Board may, subject to the right of appeal conferred by section 58 decline to register-

(a) The transfer of a share, not being a fully paid share, to a person of whom they do not approve; or

(b) Any transfer of shares on which the company has a lien.

9. The Board may decline to recognize any instrument of transfer unless-
(a) The instrument of transfer is in the form as prescribed in rules made under sub-section (1) of section 56;
(b) the instrument of transfer is accompanied by the certificate of the shares to which it relates, and such other evidence as the Board may reasonably require to show the right of the transferor to make the transfer; and

(c) The instrument of transfer is in respect of only one class of shares.

TRANSMISSION OF SHARES

10. (i) On the death of a member, the survivor or survivors where the member was a joint holder,
and his nominee or nominees or legal representatives where he was a sole holder, shall be the only persons recognized by the company as having any title to his interest in the shares.

(ii) Nothing in clause (i) shall release the estate of a deceased joint holder from any liability in respect of any share which had been jointly held by him with other persons.

11. (i) Any person becoming entitled to a share in consequence of the death or insolvency of a member may, upon such evidence being produced as may from time to time properly be required by the Board and subject as hereinafter provided, elect, either-

(a) to be registered himself as holder of the share; or
(b) to make such transfer of the share as the deceased or insolvent member could have made.
(ii) The Board shall, in either case, have the same right to decline or suspend registration as it would have had, if the deceased or insolvent member had transferred the share before his death or insolvency.

ALTERATION OF CAPITAL

12. The Authorized Share capital of the Company shall be such amount and be divided into such shares as may, from time to time, be provided in Clause VIII of Memorandum of Association payable in the manner as may be determined by the Directors, from time to time and the company may, from time to time, by ordinary resolution increase the share capital by such sum to be divided into shares of such amount, as may be specified in the resolution.

13. Subject to the provisions of section 61, the company may, by ordinary resolution-

(a) Consolidate and divide all or any of its share capital into shares of larger amount than its existing shares;
(b) Convert all or any of its fully paid-up shares into stock, and reconvert that stock into fully paid-up shares of any denomination;

(c) Sub divide its existing shares or any of them into shares of smaller amount than is fixed by the memorandum;

(d) Cancel any shares which, at the date of the passing of the resolution, have not been taken or agreed to be taken by any person.

14. The company may, by special resolution, reduce in any manner and with, and subject to, any incident authorized and consent required by law,-
(a) Its share capital;
(b) Any capital redemption reserve account; or
(c) Any share premium account.

BUY BACK OF SHARES

15. Notwithstanding anything contained in these articles but subject to the provisions of sections 68 to 70 and any other applicable provision of the Act or any other law for the time being in force, the company may purchase its own shares or other specified securities.

GENERAL MEETINGS

16. All general meetings other than annual general meeting shall be called extraordinary general meeting.

17. (a) Subject to the provisions of the Act, the Board may convene a General Meeting of the Company. Not less than clear Fourteen day's notice) written notice of every General

Meeting shall be given to each Shareholder, at its usual address whether in India or abroad, and to the Auditor of the Company, provided that a General Meeting may be convened at a shorter notice of less than Fourteen days (14 days) in accordance with the provisions of the Act.

(ii) The Board may, whenever it thinks fit, call an extraordinary general meeting.

(iii) If at any time directors capable of acting who are sufficient in number to form a quorum are not within India, any director or any two members of the company may call an extra ordinary general meeting in the same manner, as nearly as possible, as that in which such a meeting maybe called by the Board.

PROCEEDINGS AT GENERAL MEETINGS

18. (i) No business shall be transacted at any general meeting unless a quorum of members is present at the time when the meeting proceeds to business.

(ii) Subject to the Provisions of the section 103 of the Companies Act, 2013, the quorum shall be two members personally present.

CHAIRPERSON

19. The Chairperson of the Board shall preside as Chairperson at every general meeting of the Company.

20. The Chairperson of the Board meeting of the Company shall be any Director as may be appointed and nominated by IIIT-Delhi.

21. The Chairperson of the General Meeting shall have the casting vote right.

ADJOURNMENT OF MEETING

22. (i) Subject to the provisions of the Articles, if a quorum as specified in above) is not present for a proposed General Meeting, the General Meeting shall be adjourned and reconvened on the same day in the next week at the same time and place, or that day is not a Business Day, on the next succeeding Business Day, at the same time or on such other day, not being National Holiday, or at such other time and place as may be determined by the Board.

(ii) The Chairperson may, with the consent of any meeting at which a quorum is present, and shall, if so directed by the meeting, adjourn the meeting from time to time and from place to place.

(iii) No business shall be transacted at any adjourned meeting other than the business left unfinished at the meeting from which the adjournment took place.

(iv) When a meeting is adjourned for thirty days or more, notice of the adjourned meeting shall be given as in the case of an original meeting.

(iv) Save as aforesaid, and as provided in section 103 of the Act, it shall not be necessary to give any notice of an adjournment or of the business to be transacted at an adjourned meeting.

VOTING RIGHTS

23. Subject to any rights or restrictions for the time being attached to any class or classes of shares,-

(a) On a show of hands, every member present in person shall have one vote; and

(b) On a poll, the voting rights of members shall be in proportion to his share in the paid-up equity share capital of the company.

24. A member may exercise his vote at a meeting by electronic means in accordance with section 108 and shall vote only once.

25. (i) In the case of joint holders, the vote of the senior who tenders a vote, whether in person or by proxy, shall be accepted to the exclusion of the votes of the other joint holders.

(ii) For this purpose, seniority shall be determined by the order in which the names stand in the register of members.

26. A member of unsound mind, or in respect of whom an order has been made by any court having jurisdiction in lunacy, may vote, whether on a show of hands or on a poll, by his committee or other legal guardian, and any such committee or guardian may, on a poll, vote by proxy.

27. Any business other than that upon which a poll has been demanded may be proceeded with, pending the taking of the poll.

28. No member shall be entitled to vote at any general meeting unless all calls or other sums presently payable by him in respect of shares in the company have been paid.

29. (i) No objection shall be raised to the qualification of any voter except at the meeting or adjourned meeting at which the vote objected to is given or tendered, and every vote not disallowed at such meeting shall be valid for all purposes.

(ii) Any such objection made in due time shall be referred to the Chairperson of the meeting, whose decision shall be final and conclusive.

PROXY

30. The instrument appointing a proxy and the power-of-attorney or other authority, if any, under which it is signed or a notarized copy of that power or authority, shall be deposited at the registered office of the company not less than 48 hours before the time for holding the meeting or adjourned meeting at which the person named in the instrument proposes to vote, or, in the case of a poll, not less than 24 hours before the time appointed for the taking of the poll; and in default the instrument of proxy shall not be treated as valid.

31. An instrument appointing a proxy shall be in the form as prescribed in the rules made under section 105.

32. A vote given in accordance with the terms of an instrument of proxy shall be valid, notwithstanding the previous death or insanity of the principal or the revocation of the proxy or of the authority under which the proxy was executed, or the transfer of the shares in respect of which the proxy is given. Provided that no intimation in writing of such death, insanity, revocation or transfer shall have been received by the company at its office before the commencement of the meeting or adjourned meeting at which the proxy is used.

BOARD OF DIRECTORS

33. The number of the directors and the names of the first directors shall be determined in writing by the subscribers of the memorandum or a majority of them.

The following shall be the First Directors of the Company:

1. Mr. Ranjan Bose
2. Mr. Pushpendra Singh
3. Mr. Vikram Goyal
4. Mr. Tanmoy Chakraborty

34. The Board may pay all expenses incurred in getting up and registering the company.

35. The company may exercise the powers conferred on it by section 88 with regard to the keeping of a foreign register; and the Board may (subject to the provisions of that section) make and vary such regulations as it may thinks fit respecting the keeping of any such register.

36. All cheques, promissory notes, drafts, hundis, bills of exchange and other negotiable instruments, and all receipts for monies paid to the company, shall be signed, drawn, accepted, endorsed, or otherwise executed, as the case may be, by such person and in such manner as the Board shall from time to time by resolution determine.

37. Every director present at any meeting of the Board or of a committee thereof shall sign his name in a book to be kept for that purpose.

38. (i) Subject to the provisions of section 149, the Board shall have power at any time, and from time to time, to appoint a person as an additional director, provided the number of the directors and additional directors together shall not at any time exceed the maximum strength fixed for the Board by the articles.

(ii) Such person shall hold office only up to the date of the next annual general meeting of the company but shall be eligible for appointment by the company as a director at that meeting subject to the provisions of the Act.

PROCEEDINGS OF THE BOARD

39. (i) The Board of Directors may meet for the conduct of business, adjourn and otherwise regulate its meetings, as it thinks fit. A meeting of the Board of Directors shall be held at least once in six calendar months.

(ii) Not less than seven day's notice of every meeting of the Board shall be given to every Director and to any alternate Director appointed in his place along with the proposed agenda.

(ii) A director may, and the manager or secretary on the requisition of a director shall, at any time, summon a meeting of the Board.

(iii) The quorum for a meeting of the Board shall either four Directors or 25per cent of its total strength whichever is less provided that the quorum shall not be less than Two (2)Directors.

(iv) If a quorum is not present for a proposed meeting of the Board, the meeting shall be adjourned and reconvened on the same day in the next week at the same time and place, or if that day is not a Business Day, on the next succeeding Business Day, at the same time and place or on such other day, not being a National Holiday, or at such other time and place as may be determined by the Board.

(v) In case the meeting held through audio-video conference mode, the recordings of such meetings will be stored on digital media

40. (i) Save as otherwise expressly provided in the Act, questions arising at any meeting of the Board shall be decided by a majority of votes.

(ii) The Chairperson of the Board shall have casting vote right.

41. The continuing directors may act notwithstanding any vacancy in the Board; but, if and so long as their number is reduced below the quorum fixed by the Act for a meeting of the Board, the continuing directors or director may act for the purpose of increasing the number of directors to that fixed for the quorum, or of summoning a general meeting of the company, but for no other purpose.

42. (i) The Chairperson of the board meeting of the Company shall be any Director as may be appointed and nominated by IIIT-Delhi from time to time.

(ii) If at any meeting the Chairperson is not present within five minutes after the time appointed for holding the meeting, the directors present may choose one of their members to be chairperson of the Meeting.

43. (i) The Board may, subject to the provisions of the Act, delegate any of its powers to committees consisting of such member or members of its body as it thinks fit.

(ii) Any committee so formed shall, in the exercise of the powers so delegated, conform to any regulations that may be imposed on it by the Board.

CHIEF EXECUTIVE OFFICER, MANAGER, COMPANY SECRETARY OR CHIEFFINANCIAL OFFICER

44. Subject to the provisions of the Act-

(i) A chief executive officer, manager, company secretary or chief financial officer may be appointed by the Board for such term, at such conditions as it may thinks fit; and any chief executive officer, manager, company secretary or chief financial officer so appointed may be removed by means of a resolution of the Board;

(ii) A director may be appointed as chief executive officer, manager, company secretary or chief Financial officer.

45. A provision of the Act or these regulations requiring or authorizing a thing to be done by or to a director and chief executive officer, manager, company secretary or chief financial officer shall not be satisfied by its being done by or to the same person acting both as director and as, or in place of, chief executive officer, manager, company secretary or chief financial officer.

GOVERNING BODY

46. (i) The Ministry of Science and Technology, Government of India has shortlisted Indraprastha Institute of Technology, Delhi as one of the Host Institute for establishing Technology Innovation Hubs(TIHs) having domain areas of Cognitive Computing and Social Censing, pursuant to National Mission on Interdisciplinary Cyber-Physical Systems (NM-ICPS). The Company is a Technology Innovation Hub (TIH) established by IIIT- Delhi (Host Institute). The affairs of the Company shall be managed by the Governing Body in the manner provided hereinafter and who shall exercise administrative, technical and financial powers, which may be necessary for such Governing except those, which are, by this Article or by statute, expressly directed to be done by the Board of Directors.

(ii) The Governing Body shall be formed with the following structure:

1. Head of the Host Institute (IIIT- DELHI): Chairman
2. Academic representatives (not less than 2): Members
3. Industry Representatives (not less than 2): Members
4. Mission Director (or representative), Mission Office, DST: Member
5. Project Director/CEO, TIH (Technology Innovation Hub): Member-Secretary

(iii) The Governing Body could co-opt eminent people (India/ abroad) as members.

The following will be eligible to be a member of a Governing Body:

1. Authorized Representative of the Shareholder
2. Individual Shareholder itself
3. Any other individual as may be approved by Governing Body.

(iv) The Governing Body shall have authority to invite persons in the interest of the Company to attend any Governing Body meeting, but such persons shall not have any right to vote.

(v) The Governing Body may constitute expert committees from time to time and special committees and task Force(s). Such Task Force(s) may be need based and for a specified period and preferable to be formed by the related expert committee(s). The Governing Body may also appoint sub committees from time to time and assign and/or mandate

them appropriate technical streams or assign tasks that fall within the scope of such Committees for efficient implementation of Hubs at Host Institutes.

(vi) The Governing Body shall be the Apex body for overall supervision, control, directions and mid-course correction in the implementation of Hubs at Host Institutes and will approve key guidelines for implementation of the Hub.

(vii) The Governing Body will be the final authority to provide guidelines for implementation and operating the Company (HUB) and all other matters related to them.

(viii) The Governing Body will have full financial and administrative powers, including approvals to, re-appropriation of the budget within the ceiling of sanctioned budget, hire the appropriate manpower as per industry standards, sign Memorandum of Understanding (MoU)with International institutions and approve Collaboration foreign visits, partner with industry, receive/ support for projects in their domain areas to academic, R&D institutions, Industry, other funding agencies and linkages with existing TBIs (Technology Business Incubators) or create a new TBI if there is no TBI in IIIT Delhi.

(ix) The Governing Body would meet as often as required and at least once in six months

(x) Notwithstanding anything to the contrary contained herewith, board of Directors and shareholders shall not take, approve or otherwise ratify at any of their Meeting any of the actions, deeds, matters or things set out in Reserved Matters as provided hereunder, without the prior approval of the Governing Body.

THE SEAL.

47. (i) The Board shall provide for the safe custody of the seal.

(ii) The seal of the company shall not be affixed to any instrument except by the authority of a resolution of the Board or of a committee of the Board authorized by it in that behalf, or except with the authority of the person as the Board may appoint for the purpose.

ACCOUNTS

48. (i) The Board shall from time to time determine whether and to what extent and at what times and places and under what conditions or regulations, the accounts and books of the company, or any of them, shall be open to the inspection of members not being directors.

(ii) No member (not being a director) shall have any right of inspecting any account or book or document of the company except as conferred by law or authorized by the Board or by the company in general meeting.

WINDING UP

49. Subject to the applicable provisions of Chapter XX the Act read with Insolvency and Bankruptcy Code, 2016 and rules made there under-

(i) In the event of any liquidation or bankruptcy proceedings or any threatened distress action against the Company or any of its assets; machineries and equipment procured for the purpose of the Project out of or with the support of grant-in- aid shall be outside such proceedings and the Government of India may assume the control and Governing of the Company and appoint any of its officer or authorized representative to run the Project handled by the Company.

(ii) If on the winding up or dissolution of a company registered under this section, there remains, after the satisfaction of its debts and liabilities, any asset, they may be transferred to another Company registered under this section and having similar objects, subject to such conditions as the Tribunal may impose, or may be sold and proceeds thereof credited to "Insolvency and Bankruptcy Fund formed under section 224 of the Insolvency and Bankruptcy Code, 2016".

INDEMNITY

50. Every officer of the company shall be indemnified out of the assets of the company against any liability incurred by him in defending any proceedings, whether civil or criminal, in which judgment is given in his favour or in which he is acquitted or in which relief is granted to him by the court or the Tribunal.

We, the several persons, whose names and addresses are subscribed, hereto are desirous of formed into a Company in pursuance of these Articles of Association.

S. No	Name, addresses, Description and Occupation of each subscribers	Signature of subscribers	Name, addresses, Description and Signatures of witnesses
01	Indraprastha Institute of Information Technology-Delhi (IIIT-Delhi) (Registered Office: GB Pant extended polytechnic (Near Govindpuri Metro Station, Okhla Phase-III, New Delhi-110020) Through its authorized representative Prof. Ranjan Bose S/o Mr. Sudip K Bose R/o 46,New Campus, I.I.T Delhi, Hauz Khas,Delhi-110016 Occupation: Service	Ranjan Bose 7/12/20	I witness to subscriber who has subscribed and signed in my presence at New Delhi Further I have verified his Identification Details(ID) for his identification and satisfied myself for his identification particulars as filled in Name : Kapil Chawla S/O : Late Sh. B.S. Chawla Add : F/NO. 102, ERA APPN. Pocket 1&2 Dwarka. OCC : Service.
02	Indraprastha Institute of Information Technology-Delhi (IIIT-Delhi) (Registered Office: GB Pant extended polytechnic (Near Govindpuri Metro Station, Okhla Phase-III, New Delhi-110020) Through its Nominee Prof. Pushpendra Singh S/o Mr. Kanchhi Singh R/o 802, Faculty Residence, IIIT Delhi, Near Govindpuri Metro Station, Okhla Phase – 3, Okhla Industrial Area , Delhi – 110020 Occupation: Service	P. Singh 7-12-20	I witness to subscriber who has subscribed and signed in my presence at New Delhi Further I have verified his Identification Details(ID) for his identification and satisfied myself for his identification particulars as filled in Name : Kapil Chawla S/O : Late Sh. B.S. Chawla Add : F/NO 102 ERA APPN. Pocket 1&2 Dwarka. OCC. : Service.

Place: New Delhi

Dated: 07.12.2020

13