Revision to the DPR: IIT Patna

Sl No	SAC Suggestion	Explanation	Revision details
1	Executive summary	Each DPR should contain a max of 3 page Executive Summary, a gist of the entire DPR.. Each DPR should contain a certificate from Project Director.	**Please refer to Section-1; Page no: 3-5**
2	NM-ICPS DPR	NM-ICPS Mission DPR was prepared by subject experts, is approved by the Govt of India and is wellorganised in terms of mission philosophy, structure, detailing, models, financial calculations, targets,deliverables etc. NM-ICPS DPR is for the entire Mission and the DPR by Hub is for the Technology Vertical. The Hub DPR should be in sync with Mission DPR.	The DPR has been prepared in sync with Mission DPR
3	Context and Background	The DPRs should bring out a brief description of the Technology Vertical and how it connects with theoverall objectives of the Mission	**Please refer to Section 2; Page no: 6-7**
4	Aims and Objectives	Please refer to the NM-ICPS DPR and elucidate the TIH vision and focus, and specific objectives which contain both foundational and applied aspects of the Tech Vertical. During the meeting, the SAC members asserted that they did not want tinkering with well known ideas and techniques, or theassembly of oft repeated prototypes in the name of indigenisation. Every TIH’s activities need to be research driven so that contributions have a place in the international CPS activities, or in addressing national needs. The contributions should range from foundational to truly novel prototype development. The TIHs must not focus haphazardly distributed over small, incremental projects. On the other hand each TIH must define its focus in the chosen technology vertical, and develop expertise in their vertical, expand the frontiers of knowledge in their focus area (disseminated via high quality peer-reviewed publications in top venues), conceive and develop technology that has internationally and nationally recognised impact, and become the “go-to” group in their technology focus area, nationally and internationally..	As suggested, both fundamental/core research and applied research have been included. ***Please refer to Section 4; Page no: 74-79***

5	Comprehensive analysis of the existing knowledge and practice in the area, ofthe gaps that need to be filled, and possible areas that can be opened up.	Each TIH needs to have understood the knowledge and applications landscape (national and international) of the area they are representing, and should write a strong technical section on the status of the area, an analysis of the gaps, and their own plans to close these gaps and/or open up and extend the area. For a mission of this size, it is expected that the TIHs will be able to open up and drive novel research intiatives in their chosen area, so that they become the international leaders in those topics.	As suggested, comprehensive analysis of the state-of-the-art have been done, gaps have been identified, and the focus have been defined. Prominent technologies to be developed have been earmarked. ***Please refer to Sections 3.1.1 and 3.1.2; Page no: 7-24***
6	Statements of challenging problems which are hard, and whose solution is elusive, but the solution of which reaps high rewards	The DPR must identify, in the chosen technology vertical, a few challenging open problems that the TIH will aim to solve. The problems should be such, that if solved, it will make an international impact, or will address major unmet national needs. It is important for a mission of this magnitude that the TIHs identify and aim to solve such hard problems; failure is acceptable as long as there is evidence of significant effort in tackling the problems.	As suggested, we have included the challenging problems with respect to theoretical or core algorithmic aspects; and also highlighted the technologies of national importance to develop. ***Please refer to Sections 3.1.3; Page no: 24-44.***
7	Target Beneficiaries	The technologies developed must have target beneficiaries. Such a technology could be (i) a contribution to a larger CPS technology being developed by a group of TIHs, or (ii) developed as a research prototype to demonstrate a truly novel concept, or (iii) a challenging requirement of the GoI,or the industry (private, or public, established or start-up). In this list, (i) is essential since CPSs are inherently interdisciplinary, involving many concepts and technologies. It is expected that NMICPS has a few cross-TIH technologies and demonstrations. For (iii), in the above list, each TIH will need to engage with the user agencies to determine their requirements. Again, it is emphasised that, mereindigenisation cannot be an aim for NMICPS developed technologies. Fundamental and theoretical research must be disseminated in the top publishing venues, so that the knowledge generated under the NMICPS program gets a wide audience.	The problems have been identified after consultation with various stakeholders, such as industries, start-ups as well various govt agencies. ***Please refer to Section 6; Page no: 85-92***
8	Management	The DPR should bring out the management structure of the Hub, roles of HGB, BoD and how variousissues are to be sorted out by the	Details have been included. ***Please refer to Section 10; Page no: 144-150***

		established procedures etc. Responsibilities of different agencies for Hub management & implementation should be elaborated. The organisation structure at various levels, Human resource requirements and as well as monitoring arrangements should be clearly speltout.	
9	Finance	Detailed cost estimates, component-wise-year-wise and detailed Bifurcation of Budget under differentheads within the ceiling of the TIH budget keeping in view targets as mentioned in Tripartite Agreement (Refer Chapter 4 of Mission DPR and prepare all tables accordingly going to the leaf level). Means of financing and phasing of expenditure also needs to be captured. Options for cost recovery and cost sharing, private partnerships should be explored. Issues relating to Hub sustainability, including stakeholder commitment, operation-maintenance of assets after 5 years andother related issues should also be addressed in this section.	As suggested, we have provided cost estimates, component-wise-year-wise and detailed bifurcation under different heads. Options of cost recovery and self-sustainability have been covered. ***Please refer to Annexure-1***
10	Strategy	The DPRs should clearly bring out in the strategy section about the implementation models in Tech. Development, HRD, Entrepreneurship development, international co-operation etc. NM-ICPS DPR clearly elucidates the units of cost, models, process and measurable outcomes. Now, every Hub has been assigned a set of targets and budget. The strategy should cover means by which the set targets are achieved within that cost & timelines.	As suggested, strategy section is included with the details of technology development, HRD, Entrepreneurship develop, International co-operation etc. ***Please refer to Section 5; Page no: 79-84***
11	Time Frame	Each Hub was given a time frame of 5 years. This section should indicate proposed time lines to achieve the objectives, targets, deliverables etc and also provide a PERT/CPM chart wherever applicable.	As suggested, time frame is divided into 5 years. Necessary details with have been provided. ***Please refer to Annexure 2***
12	Outcomes	The DPRs should clearly define expected outcomes from the Hub.	Outcomes have been stated. ***Please refer to Section 15, Page no 163-167.***
13	Evaluation	SAC & MGB will be at higher level for evaluation. But, each Hub should have internal mechanisms apart from HGB/ BoD etc for day to day requirements. It may be noted that	As suggested, project proposal submission, project evaluation, project selection, project progress monitoring has been included.

		continuation of the Hub from one year to another will not be permissible without a proper evaluation, it could be SAC or any other Committee constituted by MGB.. There will be a midterm technical review by the SAC, or committees set up by the SAC, which will decide on the continuation of each TIH to its full term, or early termination, and the elevation of some TIHs to TTRPs.	***Please refer to Section 17; Page no: 174-175.***

Detailed Project Report (DPR)

National Mission
on
Interdisciplinary

Cyber-Physical Systems (NM-ICPS)

Indian Institute of Technology Patna

Patna, Bihar, India

Pin- 801106

Contents

1. Section-1: Executive Summary

We propose to establish a Technology Innovation Hub (TIH) named as "***IIT Patna Vishlesan i-Hub Foundation***" in the area of "Speech, Video and Text Analytics" which aims to create a strong and seamless ecosystem for leveraging the potential and exponential growth of Interdisciplinary Cyber Physical Systems (ICPSs). The proposed HUB will mark an impact, both at the national and international level, by carrying out fundamental and translation research in the broad areas of speech, video and text analytics. This will facilitate the creation of national competence in essential technologies of the future and catalyze the translation of that technology into useable applications for greater welfare of the society. The CPS expertise available at the Indian Institute of Technology Patna (IITP) as well as in the country will drive the nationwide efforts of make-in India program for knowledge generation, innovation, product development, and commercialization. Developing technologies in Speech, Text, and Video Analytics has been in the forefront of many multinational R&D companies, such as Google, Amazon, Microsoft, Facebook, IBM, Uber etc. Several academic institutions all over in the world such as Stanford, MIT, UC Berkley, USA; DFKI, Darmstadt University, University of Aachen, Edinburgh University, Cambridge University in Europe and UK; NUS, NTNU, Kyoto University, Tokyo University in Asia; and IIT Bombay, IISC Bangalore, IIIT Hyderabad, IIT Madras, IIT Kanpur, IIT Delhi, IIT Kharagpur etc. in India have been involved in active research and development in the broad areas of Speech, Video and Text Analytics. While several Government funded R&D institutions such as CDAC Pune, CDAC Noida, CDAC Kolkata, CILL Mysore have been exploring research and developments in Speech, Video and Text Analytics, Technology Development in Indian Languages (TDIL), MeITY, Govt. of India has been promoting Research and Development with the vision of *digital unite and knowledge for all*, especially in Indian languages.

IIT Patna has been pursuing extensive research and development in the areas of "Speech, Video and Text Analytics", and showcasing its presence at both national and international levels. It has established itself as a leading institution of Research and Development in "Speech, Video and Text Analytics" by publishing its research in the well-acclaimed journals and conferences, undertaking important R&D projects duly sponsored by the various Government agencies such as MeITY, SERB, DRDO, MHRD; and Industries such as Elsevier, Accenture, LG Soft, Samsung Research, Honeywell, Wipro, Bosch etc. As per the CS ranking, it ranks 14th in Asia and 1st in India for Natural Language Processing research in terms of publications during the last 5 years (2015-2020).

Under the umbrella of "Centre of Excellence for Artificial Intelligence", several research groups, *viz.* Artificial Intelligence-Natural Language Processing-Machine Learning, Data Analytics and Network Science Lab, Centre for Endangered Language Studies and groups working in the areas of Computer Vision, Image Processing, IoT, Big Data Analytics, Robotics etc. in the Department of Computer Science and Engineering, Electrical Engineering, Mechanical Engineering, Mathematics, Civil and Environmental Engineering, Humanities and Social Sciences have undertaken research and development in the cutting edge areas of Artificial Intelligence, Natural Language Processing, Machine Learning, Data Science, Speech Processing, and Video Analytics. The proposed facility along with the existing facilities at IIT Patna will create a technology or knowledge or innovation-based start-up ecosystem in alignment with the National Mission on Interdisciplinary Cyber Physical Systems (NM-ICPS) with special emphasis on "Speech, Video and Text Analytics" for the holistic development of the nation.

Novelty: The proposed ***"IIT Patna Vishlesan i-Hub Foundation"*** on "Speech, Video and Text Analytics" has wide-spread activities ranging from the fundamental basic research, to translational research and development of cutting-edge technologies for creating solutions to serve the technological requirements of common man through the development of appropriate skills and technologies. The proposed mission aims at supporting research and development activities through a large number of schemes, programmes and missions.

Some of the distinguishing aspects of our TIH hub activities include**:** foundational research on speech, video and text analytics, especially efficient multimodal (text, video and/or speech) representation, multi-lingual and multi-dimensional embedding, multitasking models, meta learning and few-shot learning, transfer learning, knowledge infused machine learning models, techniques to solve a variety of problems in low-resource scenario etc. Based on these foundational concepts, proof-of-concepts, and technologies in the areas of "Speech, Video and Text Analytics" will be developed for societal applications, such as health, judiciary, education, national security, environment etc. Products will be created for commercial use, nurture startups and increase the job market. Some of the important applications to be developed include multilingual translation, chatbot, summarization system, multimodal and multilingual sentiment analysis and emotion recognition, edge computing-based health care platform, 5G-enbaled smart healthcare system, blockchain technology embedded with heterogeneous swarm systems for real time video analytics etc.

Objective: Under the proposed TIH, IIT Patna aims to promote foundational and translational research in CPS technologies, especially in "Speech, Video and Text Analytics". Novel algorithms for multimodal information analysis, multilingual information processing, handling low-resource scenario etc. will be developed using the latest techniques of natural language processing, computer vision and speech processing technologies in order to solve various real-life problems. Under the umbrella of text, video and speech analytics, various problems relevant to the national sustainable goals may be taken up in the following application domains, such as education, health, tourism, judiciary, railways, agriculture, national security etc.

The second objective of the proposed TIH is to develop technologies, prototypes and demonstrate associated applications pertaining to national priorities and competence in "Speech, Video and Text Analytics" by carefully selecting the impactful and innovative technologies for the future to work on; crafting harmonious collaboration within and outside IIT Patna for knowledge creation and dissemination; catalyzing the conversion of such knowledge into tools, platforms, products for wider use; creating and sustaining commercial viability for the long run.

The third objective of the proposed TIH is to nurture and scale up high-end researchers' base, Human Resource Development (HRD) and skill-sets in the emerging areas of “Speech, Video and Text Analytics”. TIH will aim to enhance core competencies, capacity building and training to nurture innovation and start-up ecosystems; to create the world-class multi-disciplinary Technology Innovation Hub in “Speech, Video and Text Analytics”, which will serve as the focal point for technology inputs for the industry and policy advice for the government in the allied disciplines. Government and Industry R&D labs will be engaged as partners in the proposed TIH. Private participation to encourage professional execution and management of pilot scale research projects will be incentivized. Under the umbrella of TIH, a CPS-TBI with a focus on “Speech, Video and Text Analytics” will also be set-up. Existing facilities at IIT Patna (Incubation Centre, IIT Patna and/or TBI, IIT Patna) will be utilized to foster close collaboration with the entrepreneurship ecosystem.

Another major objective of the proposed TIH is to enhance core competencies, capacity building and training to nurture innovation and Start-up ecosystem in the areas of “Speech, Video and Text Analytics”. IIT Patna TIH aims to equip the incubate entity with all the world class facility, equipment and services that are essential to convert the idea/ concept to successful business. The incubates will be provided with techno business mentorship to prune and refine the idea from the concept board level to an organizational setup. They will be encouraged fail-fast to ensure efficient utilization of high-tech resources made available. IIT Patna TIH aims to create a holistic ecosystem for encouraging R&D, innovation, and Entrepreneurship in the domain of speech, video and text analytics. It will enable creation of IPR within the country for maximizing the domestic value add and diminishing the external dependency in CPS domain providing assistance during prototyping, development and commercialization for the products produced through the scheme for India and other growth markets. Employments at various levels will be created under TIH IIT Patna. Long term partnerships with strategic sectors will be established focusing on the theme of “Speech, Video and Text Analytics”. The major emphasis will be on IP creation and product development to result in increased domestic value addition in the field of “Speech, Video and Text Analytics”. IIT Patna TIH will demonstrate unique integration of academia, industry, government and Incubation eco systems on the theme of “Speech, Video and Text Analytics”.

Grand Challenges: TIH will run the grand challenges in the broad areas of speech, video and text analytics. Some of the important problems identified include the process, technology and/or product development for multilingual machine translation, multilingual conversational agents, empathetic conversational agents, smart health care, swarm robotics for security, summarization, misinformation detection, sentiment and emotion analysis, human activity recognition etc.

2. Section-2: Context/Background

Developing technologies in Speech, Text, and Video Analytics has been in the forefront of many multinational R&D companies, such as Google, Amazon, Microsoft, Facebook, IBM, Uber etc. Several academic institutions all over in the world such as Stanford, MIT, UC Berkley, USA; DFKI, Darmstadt University, University of Aachen, Edinburgh University, Cambridge University in Europe and UK; NUS, NTNU, Kyoto University, Tokyo University in Asia; and IIT Bombay, IISC Bangalore, IIIT Hyderabad, IIT Madras, IIT Kanpur, IIT Delhi, IIT Kharagpur etc in India have been involved in active research and development in the broad areas of Speech, Video and Text Analytics. While several Government funded R&D institutions such as CDAC Pune, CDAC Noida, CDAC Kolkata, CILL Mysore have been exploring research and developments in Speech, Video and Text Analytics, Technology Development in Indian Languages (TDIL), MeITY, Govt. of India has been promoting Research and Development with the vision of *digital unite and knowledge for all*, especially in Indian languages.

IIT Patna has been pursuing extensive research and development in the areas of "Speech, Video and Text Analytics", and showcasing its presence at both national and international levels. It has established itself as a leading institution of Research and Development in "Speech, Video and Text Analytics" by publishing its research in the well-acclaimed journals and conferences, undertaking important R&D projects duly sponsored by the various Government agencies such as MeITY, SERB, DRDO, MHRD; and Industries such as Elsevier, Accenture, LG Soft, Samsung Research, Honeywell, Wipro, Bosch etc; and represented several national and international bodies such as "Chairman, AI Standardization Committee, Bureau of Indian Standard (BIS), Ministry of Consumer Affairs, Govt. of India", "Member, Advisory Committee on Artificial Intelligence, NITI Ayog", "Member-at-Large (MAL), Asian Federation of Natural Language Processing", "President, Association for Computational Linguistics" etc. As per the CS ranking[1], it ranks 14th in Asia and 1st in India for Natural Language Processing research in terms of publications during the last 5 years (2015-2021).

Under the umbrella of "Centre of Excellence for Artificial Intelligence", several research groups, *viz.* Artificial Intelligence-Natural Language Processing-Machine Learning[2], Data Analytics and Network Science Lab[3], Centre for Endangered Language Studies[4] and groups working in the areas of Computer Vision, Image Processing, IoT, Big Data Analytics, Robotics etc. in the Department of Computer Science and Engineering, Electrical Engineering,

[1] http://csrankings.org/#/fromyear/2015/toyear/2020/index?nlp&asia

[2] http://www.iitp.ac.in/~ai-nlp-ml/

[3] (https://www.iitp.ac.in/~danes/

[4] https://www1.iitp.ac.in/index.php?option=com_content&view=article&id=3145&Itemid=552

Mechanical Engineering, Mathematics, Civil and Environmental Engineering, Humanities and Social Sciences have undertaken research and development in the cutting edge areas of Artificial Intelligence, Natural Language Processing, Machine Learning, Data Science, Speech Processing, and Video Analytics.

The proposed facility, named *Vishlesan I-Hub Foundation* at IIT Patna will create a technology or knowledge or innovation-based ecosystem in alignment with the National Mission on Interdisciplinary Cyber Physical Systems (NM-ICPS) with special emphasis on "Speech, Video and Text Analytics" for the holistic development of the nation. While on one hand, the hub will take up some fundamental research problems on speech, video and text analytics, such as developing robust techniques for multilingual representation learning, multimodal representation learning, scalable multimodal learning, multitasking models, meta learning and few-shot learning for domain adaption, unsupervised and semi-supervised learning, knowledge infused machine learning models, investigating methods for low-resource scenario, efficient techniques for handling noisy data etc.; the algorithms will be applied to develop prototypes, and technologies. The prominent technologies to focus on Multilingual Machine Translation in Education, Health, Judiciary and Noisy data, TIH will run the grand challenges in the broad areas of speech, video and text analytics. Some of the important problems identified include the technology and product development for multilingual machine translation; multilingual and multimodal chatbots for agriculture, health, judiciary; empathetic conversational agents; smart health care, summarization, sentiment and emotion analysis; swarm robotics for security; human activity recognition etc.Products will be developed for commercial use, nurture startups and increase the job market.

3. Section-3: Problems to be Addressed

As part of this i-HUB we will take up a variety of problems related to the broad areas of "Speech, Video and Text Analytics". The Hub will focus on investigating novel algorithms and techniques that would solve a variety of problems; develop proof-concepts, prototypes, and products; and nurturing incubation activities.

3.1. **Comprehensive analysis of the existing knowledge and practice in the area, of the gaps that need to be filled, and possible areas that can be opened up**

Natural Language Processing (NLP), Computer Vision and Speech Processing are the three important research dimensions nowadays with widespread applications in different sectors of our lives. During the last decade, there has been tremendous growth in the fundamental as well as applied research in these areas. While on the one hand, there has been phenomenal growth in the amount of data being generated daily in the various forms (text, audio, video) from Twitter, Facebook, Uber, and different e-commerce platforms like Amazon, Flipkart etc; it has also posed various challenges, such as handling noisy data, efficient storing mechanism for such a large volume of data, retrieving relevant data, extracting relevant information, and use of the relevant data for various predictive modeling and higher level tasks. Deep learning has taken a giant leap with discoveries of theories of Convolutional Neural Networks, Recurrent Neural Networks, Long-Short Term Memory, Gated Recurrent Units, Recursive Neural Network, Memory Networks and Transformer networks. There has been tremendous growth in natural language processing and natural language understanding techniques for better context understanding, semantic analysis and discourse processing. Computer vision has made significant progress for image segmentation, object tracking, object detection, object level feature extraction, scene recognition etc. Speech processing has also made progress in terms of techniques related to speaker recognition, speech synthesis, speech analysis and coding.

3.1.1. Comprehensive Analysis:

The field of natural language processing (NLP), also known as computational linguistics, has two views, *viz.* science view (i.e. understanding the languages) and the engineering view (i.e. implementation of techniques and methods). The overall discipline can be divided into two broad sub-areas: foundational core areas and applications. However, it is often difficult to draw a boundary between these two views. The core areas address fundamental problems such as language modeling, morphological processing, syntactic processing, semantic processing, discourse and pragmatic modelling. The application areas involve topics such as extraction of useful information (e.g. named entities and relations), machine translation to translate between the languages, automatic summarization, question-answering, conversational systems, etc..

The progress of Natural Language Processing (NLP) greatly depends on the data-driven approaches which facilitates building more robust and powerful models [1]–[3]. Recent advances in computational power, as well as the tremendous growth of the variety of information, has enabled deep learning, one of the most prominent approaches in the NLP domain [1], [2], [4]. This is in line with given the fact that deep learning has been already demonstrating state-of-the-art performance in related fields like Computer Vision [5]–[9] and Speech Recognition [10]–[12]. These advancements have led to a complete paradigm shift from traditional to novel data-driven models aimed at advancing the field of NLP. One of the reasons behind this contribute to the facts that the new approaches are more promising, and are easier to engineer.

The core issues in NLP domains correspond to those which are inherently present in any computational linguistic system. To perform machine translation, summarization, image captioning, question-answering, conversational system modelling, or any other linguistic task, there must be some understanding of the language phenomenon. This understanding can be broken down into at least four main areas, *viz.* language model, morphology processing, shallow and deep parsing, and semantics.

Language Models: Arguably, the most crucial task in NLP is the language modeling. Language model (LM) is an essential piece of almost all the applications of NLP. LM denotes the process of creating a model to predict words or simple linguistic units given the previous words [13]. This is generally useful for applications in which a user types input, provides predictive ability for fast text entry. Its power and versatality, although, emanate from the fact that it can implicitly capture information at the various levels, e.g. syntactic and/or semantic relationships among words or components in a linear neighborhood, and making it useful for several downstream tasks such as machine translation, sentiment analysis, summarization etc. Statistical language models suffer are not robust to handle synonyms or out-of-vocabulary (OOV) words. Recent progress of neural language model [14] has made it possible to deal with such issues. The LM community immediately took advantage of them, and continued to develop sophisticated models, many of which were summarized by DeMulder et al. [15]. Daniluk et al. [16] explored several networks using variations of attention mechanisms. The very model had a simple attention mechanism, with a window length of five. Another recent study carried out focused on the usage of residual memory networks (RMNs) for LM [17]. This showed that residual connections skipping two layers were most effective, followed closely by those skipping a single layer. A CNN used recently in LM replaced the pooling layers with fully-connected layers [18]. These layers allowed the feature maps to be reduced to lower dimensional spaces just like the pooling layers.

Language modelling has been evolving very fast, with the phenomenal works by Radford et al. [19] and Peters et al. [20]. Generative Pre-Training (GPT) was introduced in [21]. This is a pre-trained language model based on the Transformer networks [22], that learns dependencies of words in sentences and longer segments of text, rather than just the immediate surrounding words. In [21], Peters et al. incorporated bi-directional features to capture the backward context in addition to the forward context, in the Embeddings from Language Models (ELMo). Here, vectorizations from at multiple level, rather than just the final layer, were captured. This eventually allowed for multiple encodings of the same information to be captured, which was empirically shown to boost the performance significantly. In [22], Devlin et al., introduced an additional unsupervised training tasks of random masked neighbor word prediction, and next sentence-prediction (NSP). Here, for a given a sentence or a continuous segment of a text, another sentence was predicted to either be the next sentence or not. These Bidirectional Encoder Representations from Transformers (BERT) were further built upon by Liu et al. [23] to create Multi-Task Deep Neural Network (MT-DNN) representations, which are the current state of the art in LM.

Morphology Learning: In morphology processing, one of the recent line of works is related to the universal morphology. The ultimate goal of universal morphology is to study the relationships between the morphologies of different languages and how they relate to each other. There has been a recent study [24] that made use of deep learning techniques for this specific problem.

In addition to this universal morphology, creating morphological embeddings could help in multilingual information processing. These could be possibly be used across the cognate languages, which would be valuable when some languages are more resourceful than others. Morphological structures, on the other hand, may be important in handling specialized languages such as those used in the biomedical literature.

Parsing: Parsing is one of the very crucial components for NLP. There are at least two distinct forms of parsing, *viz.* constituency parsing and dependency parsing [25]. In constituency parsing, phrasal constituents are extracted from a sentence in a hierarchical fashion. The dependency parsing, on the other hand, looks at the relationships between the pairs of individual words. Graph-based parsing constructs a number of parse trees that are then searched to find the correct one. Most graph-based approaches are generative models, in which a formal grammar, based on the natural language, is used to construct the trees [26]. More popular in recent years than graph-based approaches have been transition-based approaches that usually construct only one parse tree.

One early works of Socher et al. [27-28] included the use of Recurrent Neural Networks (RNNs) with probabilistic context-free grammars (PCFGs) [29-30]. Dyer et al. [31] proposed a model that used RNNs for parsing and language modeling. While majority of the approaches take a bottom-up approach to parsing, this took a top-down approach, taking as input the full sentence in addition to the current parse tree. Choe and Charniak [32] formulated as problem of language modeling, and used an LSTM to assign probabilities to the parse trees, achieving state-of-the art. A model created by Dozat and Manning [33] used a graph-based approach with a self-attentive network.

Universal dependency parsing is a relatively new task of parsing language using a standardized set of tags and relationships across all languages. While parsing varies greatly from language to language, this focuses on building a uniform model between them. Nivre [34] discussed the recent development of universal grammar and presented the challenges that are needed to solved in future, mainly the development of tree banks in more languages and the consistency of labeling between tree banks in different languages.

Semantic Processing: Semantic processing concerns with the understanding of the meaning of words, phrases, sentences, or documents at some level. Prominent word embedding techniques, such as Word2Vec [35-36] and GloVe [37], claim to capture the meanings of words, following the Distributional Hypothesis of Meaning [38]. When vectors corresponding to phrases, sentences, or other components of text are processed using a neural network, a representation that can be loosely thought to be semantically representative is computed compositionally.

Literature shows that deep learning generally performs very well, achieving state-of-the-art performance in many down-stream NLP applications. However, it is clear that natural language processing and natural language understanding are the enigmatically complex topics, with myriad core or basic tasks, of which deep learning has only grazed the surface.

The core concepts of NLP are applied to solve various higher level NLP tasks.

Information Extraction: Information Extraction (IE) identifies structured information from unstructured, semi-structured or structured data. Some of the important tasks of IE correspond are Named Entity Recognition, Relation Extraction, Coreference Resolution, and Event Extraction.

Named Entity Recognition (NER) aims to locate and categorize named entities in context into pre-defined categories such as the names of people, places, organizations, dates, time expressions, monetary expressions etc. Deep neural networks have been applied for NER, for example, CNN [39] and RNN architectures [40], as well as hybrid bidirectional LSTM and CNN architectures.

Coreference resolution includes identification of the mentions in a context that refer to the same entity. One of the very recent works makes use of Reinforcement Learning (RL) [41] for coreference resolution. One of the widely used techniques is to leverage an attention mechanism [42]. In [43], authors adopted a reinforcement learning policy gradient approach to coreference resolution and provides state-of-the art performance on the English OntoNotes v5.0 benchmark task. Authors in [44] have reformulated coreference resolution as a span prediction task as in question answering and provide superior performance on the CoNLL- 2012 benchmark task.

Event Extraction involves recognizing trigger words related to an event and assigning labels to entity mentions that represent event triggers. Convolutional neural networks have been utilized for event detection [45]. Nguyen and Grishman [46] applied graph-CNN (GCCN) where the convolutional operations are applied to syntactically dependent words as well as consecutive words. Re-inforcement learning [47] based on generative adversarial networks (imitation learning) has been tried to tackle joint entity and event extraction. Document level event extraction using a combined dependency based GCN and a hypergraph [48] has been exploited.

Sentiment Analysis: The primary goal in sentiment analysis is to uncover the subjective information from text by contextual mining. Sentiment analysis is sometimes called opinion mining, as its primary goal is to analyze human opinion, sentiments, and even emotions regarding products, problems, and varied subjects. Seminal works on sentiment analysis or opinion mining include [49], [50]. In 51], a comprehensive review of sentiment analysis tasks based on deep learning methods have been presented. Sentiment analysis can be carried at the various levels: document level sentiment analysis using GRU, LSTM and domain adaptive techniques [52-55]; sentence level sentiment analysis that focuses on determining sentiment class for each sentence [56-60]; and aspect level sentiment analysis that attempts towards more fine-grained sentiment analysis [61-67].

Machine Translation (MT) is one of the most fascinating areas of NLP. The very first demonstration of MT system took place in 1954 [68], where the authors tried to translate from

Russian to English. It was not until the 1990s that successful statistical implementations of machine translation emerged as more bilingual corpora became available.

Unlike traditional statistical machine translation, NMT is based on an end-to-end neural network [69]. This reduces the complexity greatly as there is no need for extensive pre-processing and word alignments. Instead, the focus mainly shifted towards the network structure. Improvement in NMT architectures have been happening due to the growth in people's interest and need to understand other languages. The work in [69] tries to addresses with the problem of rare words. The LSTM network consists of encoder and decoder layers using residual layers along with the attention mechanism. More recently, a single model to perform massive multilingual NMT has been provided in [70]. Adversarial networks [71] have been recently used for building a robust NMT that could handle the noisy inputs, and reported performance gains over the Transformer model. In [72], authors have presented an insightful recent work where the authors sampled context words from the predicted sequence as well as the ground truth to try to reconcile the training and inference processes.

Question answering (QA): Question-answering (QA) is a fundamental task of NLP. In QA specific answers are sought, typically ones that can be inferred from available documents. Reading comprehension and dialogue systems, the two other areas of NLP, intersect with QA. Smartphones (Siri, Ok Google, Alexa, etc.) and virtual personal assistants are common examples of QA systems with which many interact on a daily basis. Although earlier systems employed rule-based algorithms, nowadays most of the algorithms are based on deep learning. Dynamic memory network [73] has been widely used for the QA tasks. Given an input image, Visual Question Answering (VQA) tries to answer a natural language question about the image [74].

Document summarization: It refers to a system that, for a given document, generates the summary. Typically, there are two types of summarization methods, *viz.* extractive and abstractive methods. One of the very early work that made use of neural networks for extractive summarization is proposed in [75] that used a ranking technique to extract the most salient sentences in the input. The model was further improved in [76] that made use of a document-level encoder to represent sentences, and a classifier to rank these sentences. The pointer generator network is widely used in summarization [77] that focuses on selecting important segment from the input document and copies in the final summary. A copy mechanism was also adopted by [78] for the similar tasks. But their analysis reveals a key problem with attention-based encoder-decoder models: they often generate unusual summaries consisting of repeated phrases. In recent times, the state-of-the-art performance on abstractive summarization is achieved by [79].

Dialogue Systems: Dialogue systems have attracted the attention of the researchers and practitioners due to its widespread applications in a variety of tasks on human-machine conversation, due in part to their promising potential and commercial value [79]. Due to the high cost of knowledgeable human resources, companies frequently turn to intelligent conversational machines. Dialogue systems are usually task-based or non-task based. Recent task-oriented dialogue systems have been designed based on deep reinforcement learning, which provided promising results [80], domain adaptation [81], and dialogue generation [82]. This was due to a shift towards end-to-end trainable frameworks to design and deploy task-oriented dialogue systems. A non-task based system, on the other hand, has the capability to empower a machine with the ability to have a natural conversation with humans [83].

Multimodal Leaning for Computer Vision: During the last few years, there has been considerable growth in computer vision with respect to both theoretical research and practical oriented applications. The computer vision community, in recent years, has paid more attention to deep learning algorithms due to their exceptional capabilities compared to traditional handcrafted methods. Some of the following literature gives an idea about the progress made in these disciplines: survey of deep learning algorithms in the computer vision community [84], a comprehensive survey that focuses directly on the problem of deep object detection and its recent advances [85], and a collection of deep learning models including generative adversarial network, its related challenges and applications [86]. However, in modern machine learning applications, information aggregation from more than one modality (e.g., visual and textual modalities) is the priority. Some of these applications include question answering, vision-and-language navigation, queue detection, emotion recognition, summarization etc. Therefore, it is very vital to learn more complex and cross-modal information from different sources, types, and data distributions. From the very early research on speech processing to the recent advances in language and vision tasks, deep multimodal learning techniques have shown tremendous success in improving cognitive performance and interoperability of prediction models in a variety of ways.

Some of the core techniques in computer vision, especially with respect to multimodal machine learning include the following:

- Multimodal Representation: Learning multimodal representation from heterogeneous signals introduces a lot of challenges to the community. Typically, inter- and intra-modal learning involves investigating efficient way to represent an object of interest from different perspectives, that often offers complementary and semantic context where multimodal information is fed into the network. Another advantage of these interacting networks is the discriminating power of the perceptual model for multisensory stimuli by exploiting the potential synergies between modalities and their intrinsic representations [87].
- Fusion algorithms: One of the critical aspects of combining multimodal data is the flexibility to represent it at different levels of abstraction. Attention mechanism has been one of the most popular techniques used to combine the various heterogenous networks. Attention based methods have been exploited in the computer vision community in a variety of applications such as video description [88,89], salient object detection [90], etc.

- Multimodal alignment: Alignment of multimodal information consists of linearly linking the features of two or more different modalities. Some of its application areas include medical image registration [91], machine translation [92], etc. Multimodal image alignment, specifically provides a spatial mapping capability between the images taken by sensors of different modalities, which may be categorized into feature-based [93,94] and patch-based [95,96] methods.
- Multimodal transfer learning: Training a deep neural network model from scratch requires a large amount of labeled data to achieve an acceptable level of performance. One interesting solution to this problem is to find an efficient method that transfers knowledge already derived from another trained model onto a huge dataset (e.g., 1000kImageNet) [97].
- Zero-shot learning: In practical situations, the amount of labeled data samples for effective model training is often insufficient to recognize all possible object categories in an image (i.e., seen and unseen classes). Zero-shot learning [98] is a remedy under such situation. This supervised method has opened up many valuable applications, such as object detection [99], object classification and retrieval of videos [100] etc. This, in turn, solves a multi-class classification problem, when some of the classes do not have good representative samples to learn from. However, during the learning process, additional visual and semantic features such as word embeddings [101], visual attributes [102], or descriptions [103] can be assigned to both seen and unseen classes.
- Some of the interesting problems where multimodal machine learning can be applied include: Human recognition (recognizing same target from different time space); Face recognition (used in biometric systems for control and monitoring purpose); Gesture recognition (human-computer interaction to detect motion in real time); Image captioning (automatic writing of description of a given image); Video question-answering (answering questions for a question asked); Vision and language navigation (navigating in crowded locations and visual cues to perceive the surroundings); Style transfer (visual cues for art and painting); Medical data analysis (improving clinical accuracy and detecting abnormalities in medical images); Autonomous systems.

[1] X. Zhang, J. Zhao, and Y. LeCun, “Character-level convolutional networks for text classification,” in Advances in neural information processing systems, pp. 649–657, 2015.

[2] K. Cho, B. Van Merri¨enboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, “Learning phrase representations using RNN encoder-decoder for statistical machine translation,” arXiv preprint arXiv:1406.1078, 2014.

[3] S. Wu, K. Roberts, S. Datta, J. Du, Z. Ji, Y. Si, S. Soni, Q. Wang, Q. Wei, Y. Xiang, B. Zhao, and H. Xu, “Deep learning in clinical natural language processing: a methodical review,” Journal of the American Medical Informatics Association, vol. 27, pp. 457–470, 2020.

[4] R. Collobert and J. Weston, "A unified architecture for natural language processing: Deep neural networks with multitask learning," in Proceedings of the 25th international conference on Machine learning, pp. 160–167, ACM, 2008.
[5] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei, "Large-scale video classification with convolutional neural networks," in Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, pp. 1725–1732, 2014.
[6] M. Oquab, L. Bottou, I. Laptev, and J. Sivic, "Learning and transferring mid-level image representations using convolutional neural networks," in Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, pp. 1717–1724, 2014.
[7] A. Shrivastava, T. Pfister, O. Tuzel, J. Susskind, W. Wang, and R.Webb, "Learning from simulated and unsupervised images through adversarial training," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2107–2116, 2017.
[8] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, "Deep Learning for Computer Vision: A Brief Review," Computational Intelligence and Neuroscience, Feb 2018.
[9] N. O′Mahony, S. Campbell, A. Carvalho, S. Harapanahalli, G. V. Hernandez, L. Krpalkova, D. Riordan, and J. Walsh, "Deep learning vs. traditional computer vision," in Advances in Computer Vision (K. Arai and S. Kapoor, eds.) pp. 128–144, Springer International Publishing, 2020.
[10] A. Graves and N. Jaitly, "Towards end-to-end speech recognition with recurrent neural networks," in International Conference on Machine Learning, pp. 1764–1772, 2014.
[11] D. Amodei, S. Ananthanarayanan, R. Anubhai, J. Bai, E. Battenberg, C. Case, J. Casper, B. Catanzaro, Q. Cheng, G. Chen, et al., "Deep speech 2: End-to-end speech recognition in English and Mandarin," in ICML, pp. 173–182, 2016.
[12] U. Kamath, J. Liu, and J. Whitaker, Deep learning for NLP and speech recognition, vol. 84. Springer, 2019.
[13] D. Jurafsky and J. Martin, Speech & language processing. Pearson Education, 2000.
[14] Y. Bengio, R. Ducharme, P. Vincent, and C. Jauvin, "A neural probabilistic language model," J. of Machine Learning Research, vol. 3, 2003.
[15] W. De Mulder, S. Bethard, and M.-F. Moens, "A survey on the application of recurrent neural networks to statistical language modeling," Computer Speech & Language, vol. 30, no. 1, pp. 61–98, 2015.
[16] M. Daniluk, T. Rockt¨aschel, J. Welbl, and S. Riedel, "Frustratingly short attention spans in neural language modeling," arXiv preprint arXiv:1702.04521, 2017.
[17] K. Beneˇs, M. K. Baskar, and L. Burget, "Residual memory networks in language modeling: Improving the reputation of feed-forward networks," Interspeech 2017, pp. 284–288, 2017.
[18] N.-Q. Pham, G. Kruszewski, and G. Boleda, "Convolutional neural network language models," in EMNLP, 2016, pp. 1153–1162.
[19] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, "Improving language understanding by generative pre-training
[20] M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer, "Deep contextualized word representations," arXiv preprint arXiv:1802.05365, 2018.

[21] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in NIPS, 2017, pp. 6000–6010.

[22] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," arXiv preprint arXiv:1810.04805, 2018.

[23] X. Liu, P. He, W. Chen, and J. Gao, "Multi-task deep neural networks for natural language understanding," arXiv preprint arXiv:1901.11504, 2019.

[24]. M. Kawato, K. Furukawa, and R. Suzuki, "A hierarchical neuralnetwork model for control and learning of voluntary movement," Biological Cybernetics, vol. 57, no. 3, pp. 169–185, 1987.

[25] D. Jurafsky and J. Martin, Speech & language processing. Pearson Education, 2000.

[26] R. Socher, A. Perelygin, J. Wu, J. Chuang, C. D. Manning, A. Ng, and C. Potts, "Recursive deep models for semantic compositionality over a sentiment treebank," in EMNLP, 2013, pp. 1631–1642.

[27] R. Socher, J. Bauer, C. Manning et al., "Parsing with compositional vector grammars," in ACL, vol. 1, 2013, pp. 455–465.

[28] T. Fujisaki, F. Jelinek, J. Cocke, E. Black, and T. Nishino, "A probabilistic parsing method for sentence disambiguation," in Current issues in Parsing Technology, 1991, pp. 139–152.

[29] F. Jelinek, J. Lafferty, and R. Mercer, "Basic methods of probabilistic context free grammars," in Speech Recognition and Understanding. Springer, 1992, pp. 345–360.

[30] C. Dyer, A. Kuncoro, M. Ballesteros, and N. A. Smith, "Recurrent neural network grammars," arXiv preprint arXiv:1602.07776, 2016.

[31] D. K. Choe and E. Charniak, "Parsing as language modeling," inMNLP, 2016, pp. 2331–2336.

[32] T. Dozat and C. D. Manning, "Simpler but more accurate semantic dependency parsing," arXiv preprint arXiv:1807.01396, 2018.

[33] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa, "Natural language processing (almost) from scratch," Journal of Machine Learning Research, vol. 12, no. Aug., pp. 2493–2537, 2011.

[34] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," arXiv preprint arXiv:1301.3781, 2013.

[35] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in NIPS, 2013, pp. 3111–3119.

[36] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pretraining of deep bidirectional transformers for language understanding," arXiv preprint arXiv:1810.04805, 2018.

[37] R. Socher, B. Huval, C. D. Manning, and A. Y. Ng, "Semantic compositionality through recursive matrix-vector spaces," in Proceedings of the 2012 joint conference on empirical methods in natural language processing and computational natural language learning, pp. 1201–1211, Association for Computational Linguistics, 2012.

[38] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa, "Natural language processing (almost) from scratch," Journal of Machine Learning Research, vol. 12, no. Aug., pp. 2493–2537, 2011.

[39] G. Mesnil, X. He, L. Deng, and Y. Bengio, "Investigation of recurrent neural network architectures and learning methods for spoken language understanding.," in Interspeech, pp. 3771–3775, 2013.

[40] K. Clark and C. D. Manning, "Deep reinforcement learning for mention-ranking coreference models," arXiv preprint arXiv:1609.08667, 2016.

[41] K. Lee, L. He, and L. Zettlemoyer, "Higher-order coreference resolution with coarse-to-fine inference," arXiv preprint arXiv:1804.05392, 2018.

[42] H. Fei, X. Li, D. Li, and P. Li, "End-to-end deep reinforcement learning based coreference resolution," in Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pp. 660– 665, 2019.

[43] W. Wu, F. Wang, A. Yuan, F. Wu, and J. Li, "Corefqa: Coreference resolution as query-based span prediction," in Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp. 6953–6963, 2020.

[44] Y. Chen, L. Xu, K. Liu, D. Zeng, and J. Zhao, "Event extraction via dynamic multi-pooling convolutional neural networks," in Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), vol. 1, pp. 167–176, 2015.

[45] T. H. Nguyen and R. Grishman, "Graph convolutional networks with argument-aware pooling for event detection," in Thirty-Second AAAI Conference on Artificial Intelligence, 2018.

[46] T. Zhang, H. Ji, and A. Sil, "Joint entity and event extraction with generative adversarial imitation learning," Data Intelligence, vol. 1, no. 2, pp. 99–120, 2019.

[47] W. Zhao, J. Zhang, J. Yang, T. He, H. Ma, and Z. Li, "A novel joint biomedical event extraction framework via two-level modelling of documents," Information Sciences, vol. 550, pp. 27–40, 2021.

[48] T. Nasukawa and J. Yi, "Sentiment analysis: Capturing favorability using natural language processing," in Proceedings of the 2nd International Conference on Knowledge Capture, pp. 70–77, ACM, 2003.

[49] K. Dave, S. Lawrence, and D. M. Pennock, "Mining the peanut gallery: Opinion extraction and semantic classification of product reviews," in Proceedings of the 12th international conference on World Wide Web, pp. 519–528, ACM, 2003.

[50] A. Yadav and D. K. Vishwakarma, "Sentiment analysis using deep learning architectures: a review," Artificial Intelligence Review, vol. 53, no. 6, pp. 4335–4385, 2020.

[51] D. Tang, B. Qin, and T. Liu, "Document modeling with gated recurrent neural network for sentiment classification," in Proceedings of the 2015 conference on empirical methods in natural language processing, pp. 1422–1432, 2015.

[52] X. Glorot, A. Bordes, and Y. Bengio, "Domain adaptation for largescale sentiment classification: A deep learning approach," in Proceedings of the 28th international conference on machine learning (ICML- 11), pp. 513–520, 2011.

[53] G. Rao, W. Huang, Z. Feng, and Q. Cong, "Lstm with sentence representations for document-level sentiment classification," Neurocomputing, vol. 308, pp. 49–57, 2018.

[54] M. Rhanoui, M. Mikram, S. Yousfi, and S. Barzali, "A cnn-bilstm model for document-level sentiment analysis," Machine Learning and Knowledge Extraction, vol. 1, no. 3, pp. 832–847, 2019.

[55] R. Socher, J. Pennington, E. H. Huang, A. Y. Ng, and C. D. Manning, "Semi-supervised recursive autoencoders for predicting sentiment distributions," in Proceedings of the conference on empirical methods in natural language processing, pp. 151–161, Association for Computational Linguistics, 2011.

[56] X. Wang, Y. Liu, S. Chengjie, B. Wang, and X. Wang, "Predicting polarities of tweets by composing word embeddings with long short term memory," in Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), vol. 1, pp. 1343–1353, 2015.

[57] R. Socher, A. Perelygin, J. Wu, J. Chuang, C. D. Manning, A. Ng, and C. Potts, "Recursive deep models for semantic compositionality over a sentiment treebank," in Proceedings of the 2013 conference on empirical methods in natural language processing, pp. 1631–1642, 2013.

[58] R. Arulmurugan, K. Sabarmathi, and H. Anandakumar, "Classification of sentence level sentiment analysis using cloud machine learning techniques," Cluster Computing, vol. 22, no. 1, pp. 1199–1209, 2019.

[59] D. Meˇskel˙e and F. Frasincar, "Aldonar: A hybrid solution for sentencelevel aspect-based sentiment analysis using a lexicalized domain ontology and a regularized neural attention model," Information Processing & Management, vol. 57, no. 3, p. 102211, 2020.

[60] Y. Wang, M. Huang, L. Zhao, et al., "Attention-based LSTM for aspect level sentiment classification," in Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, pp. 606–615, 2016.

[61] Y. Ma, H. Peng, T. Khan, E. Cambria, and A. Hussain, "Sentic lstm: a hybrid network for targeted aspect-based sentiment analysis," Cognitive Computation, vol. 10, no. 4, pp. 639–650, 2018.

[62] H. Xu, B. Liu, L. Shu, and P. S. Yu, "BERT post-training for review reading comprehension and aspect-based sentiment analysis," arXiv preprint arXiv:1904.02232, 2019.

[63] H. Xu, B. Liu, L. Shu, and P. S. Yu, "Double embeddings and CNN-based sequence labeling for aspect extraction," arXiv preprint marXiv:1805.04601, 2018.

[64] H. H. Do, P. Prasad, A. Maag, and A. Alsadoon, "Deep learning for aspect-based sentiment analysis: a comparative review," Expert Systems with Applications, vol. 118, pp. 272–299, 2019.

[65] S. Rida-E-Fatima, A. Javed, A. Banjar, A. Irtaza, H. Dawood, H. Dawood, and A. Alamri, "A multi-layer dual attention deep learning model with refined word embeddings for aspect-based sentiment analysis," IEEE Access, vol. 7, pp. 114795–114807, 2019.

[66] Y. Liang, F. Meng, J. Zhang, J. Xu, Y. Chen, and J. Zhou, "A novel aspect-guided deep transition model for aspect-based sentiment analysis," arXiv preprint arXiv:1909.00324, 2019.

[67] L. E. Dostert, "The Georgetown-IBM experiment," 1955). Machine translation of languages. John Wiley & Sons, New York, pp. 124–135, 1955.

[68] D. Bahdanau, K. Cho, and Y. Bengio, “Neural machine translation by jointly learning to align and translate,” arXiv preprint arXiv:1409.0473,2014.

[69] Y. Wu, M. Schuster, Z. Chen, Q. V. Le, M. Norouzi, W. Macherey, M. Krikun, Y. Cao, Q. Gao, K. Macherey, et al., “Google’s neural machine translation system: Bridging the gap between human and machine translation,” arXiv preprint arXiv:1609.08144, 2016.

[70] R. Aharoni, M. Johnson, and O. Firat, “Massively multilingual neural machine translation,” 2019.

[71] Y. Cheng, L. Jiang, and W. Macherey, “Robust neural machine translation with doubly adversarial inputs,” arXiv preprint arXiv:1906.02443, 2019.

[72] W. Zhang, Y. Feng, F. Meng, D. You, and Q. Liu, “Bridging the gap between training and inference for neural machine translation,” 2019.

[73]. A. Kumar, O. Irsoy, P. Ondruska, M. Iyyer, J. Bradbury, I. Gulrajani, V. Zhong, R. Paulus, and R. Socher, “Ask me anything: Dynamic memory networks for natural language processing,” in International Conference on Machine Learning, pp. 1378–1387, 2016.

[74] S. Antol, A. Agrawal, J. Lu, M. Mitchell, D. Batra, C. Lawrence Zitnick, and D. Parikh, “VQA: Visual question answering,” in Proceedings of the IEEE international conference on computer vision, pp. 2425– 2433, 2015.

[75] R. Nallapati, F. Zhai, and B. Zhou, “SummaRuNNer: A recurrent neural network based sequence model for extractive summarization of documents.,” in AAAI, pp. 3075–3081, 2017.

[76] S. Narayan, S. B. Cohen, and M. Lapata, “Ranking sentences for extractive summarization with reinforcement learning,” in NAACL:HLT, vol. 1, pp. 1747–1759, 2018.

[77] R. Nallapati, B. Zhou, C. dos Santos, C. Gulcehre, and B. Xiang, “Abstractive text summarization using sequence-to-sequence RNNs and beyond,” in Proceedings of The 20th SIGNLL Conference on Computational Natural Language Learning, pp. 280–290, 2016.

[78] J. Gu, Z. Lu, H. Li, and V. O. Li, “Incorporating copying mechanism in sequence-to-sequence learning,” in ACL, vol. 1, pp. 1631–1640, 2016.

[79] A. See, P. J. Liu, and C. D. Manning, “Get to the point: Summarization with pointer-generator networks,” in ACL, vol. 1, pp. 1073–1083, 2017.

[80] E. Merdivan, D. Singh, S. Hanke, and A. Holzinger, “Dialogue systems for intelligent human computer interactions,” Electronic Notes in Theoretical Computer Science, vol. 343, pp. 57–71, 2019.

[81] C. Toxtli, J. Cranshaw, et al., “Understanding chatbot-mediated task management,” in Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, p. 58, ACM, 2018.

[82] V. Ilievski, C. Musat, A. Hossmann, and M. Baeriswyl, “Goal-oriented chatbot dialog management bootstrapping with transfer learning,” arXiv preprint arXiv:1802.00500, 2018.

[83] A. Ritter, C. Cherry, and W. B. Dolan, “Data-driven response generation in social media,” in Proceedings of the conference on empirical methods in natural language processing, pp. 583–593, Association for Computational Linguistics, 2011.

[84] Guo, Y., et al.: Deep learning for visual understanding: a review. Neurocomputing 187, 27–48 (2016)

[85] Wu, X., Sahoo, D. Hoi, S.C.H.: Recent Advances in Deep Learning for Object Detection. arXiv:1908.03673 (2019)

[86] Creswell, A., et al.: Generative adversarial networks: an overview. IEEE Signal Process. Mag. 35, 53–65 (2018)

[87] Peng, Y., et al.: CCL: cross-modal correlation learning with multigrained fusion by hierarchical network. IEEE Trans. Multimed. 20(2), 405–420 (2017)

[88] Hori, C., et al.: Attention-based multimodal fusion for video description. In: IEEE International Conference on Computer Vision (ICCV), pp. 4203–4212 (2017)

[89] Huang, X., Wang, M., Gong, M.: Fine-grained talking face generation with video reinterpretation. Vis. Comput. 37, 95–105 (2021)

[90] Liu, Z., et al.: Multi-level progressive parallel attention guided salient object detection for RGB-D images. Vis. Comput. (2020). https://doi.org/10.1007/s00371-020-01821-9

[91] Guan, S.-Y., et al.: A review of point feature based medical image registration. Chin. J. Mech. Eng. 31, 76 (2018)

[92] Bahdanau, D., Cho, K., Bengio, Y.: Neural Machine Translation by Jointly Learning to Align and Translate. arXiv:1409.0473 (2016)

[93] Chen, C., et al. Progressive Feature Alignment for Unsupervised Domain Adaptation. arXiv:1811.08585 (2019)

[94] Jin, X., et al.: Feature Alignment and Restoration for Domain Generalization and Adaptation. arXiv:2006.12009 (2020)

[95] Šˇcupáková, K., et al.: A patch-based super resolution algorithm for improving image resolution in clinical mass spectrometry. Sci. Rep. 9, 2915 (2019)

[96] . Bashiri, F.S., et al.: Multi-modal medical image registration with full or partial data: a manifold learning approach. J. Imag. 5, 5 (2019)

[97] Shamwell, E.J., et al.: Unsupervised deep visual-inertial odometry with online error correction for RGB-D imagery. IEEE Trans. Pattern Anal. Mach. Intell. (2019). https://doi.org/10.1109/TPAMI. 2019.2909895

[98] Wang,W., et al.: A survey of zero-shot learning: settings, methods, and applications. ACM Trans. Intell. Syst. Technol. 10, 13:1– 13:37 (2019)

[99] Wei, L., et al.: A single-shot multi-level feature reused neural network for object detection. Vis. Comput. (2020). https://doi. org/10.1007/s00371-019-01787-3

[100] Parida, K., et al.: Coordinated joint multimodal embeddings for generalized audio-visual zero-shot classification and retrieval of videos. In: CVPR, pp. 3251–3260 (2020)

[101]. Hascoet, T., et al.: Semantic embeddings of generic objects for zero-shot learning. J. Image Video Proc. 2019, 13 (2019)

[102] Liu, Y., et al.: Attribute attention for semantic disambiguation in zero-shot learning. In: ICCV, pp. 6697–6706 (2019)

[103] Li, K., et al.: Rethinking zero-shot learning: a conditional visual classification perspective. In: ICCV, pp. 3582–3591 (2019)

3.1.2. Gaps that need to be filled and the possible areas to explore:

Natural Language Processing, Computer Vision and Speech processing are the three prominent areas where researchers and developers have been focusing on to solve different problems, having immense practical potentials in day-to-day lives. The research can be carried out in two dimensions: foundational research to investigate algorithms and techniques; and the application-oriented research that will particularly focus on using these cores (or, foundational) concepts and techniques for solving a practical problem.

Pre-trained Language Model: Most of the language models are available for high-resource language like English. These language models do not perform well when applied to a low-resource language (like the most of the languages in India). Code-mixed language models will be an important area to explore.

Knowledge-infused Neural Network: Any deep learning technique is data hungry that needs good amount of annotated data for solving any problem. For many applications and languages, this is difficult to achieve. Investigating robust techniques to incorporate knowledge in the forms of structured and unstructured forms would play an important role to build efficient models.

Multimodal Representation: Multimodal Artificial Intelligence provides an efficient way to combine the knowledge emanating from the various sources such as images, texts, videos and audios. Some of the challenges include:
how to select important and relevant information from these sources?
how to remove noise?
how to fuse this information in a unified representation?
how to represent the information not only coming from the different modalities, but also from the different languages?

Multitask Models: One prominent research is to build an end-to-end model for solving a set of tasks simultaneously. This provides the flexibility to train the whole network in one-shot. Some of the benefits of such approaches include: improved efficiency, reduced overfitting through shared representations, and fast learning by leveraging auxiliary information. However, the learning multiple tasks introduce the new design and optimization challenges that include: balancing between hard and soft information sharing; appropriate gating mechanism to choose the most relevant information to share; data sampling strategy in case learning is carried out from more than one datasets; choice of appropriate loss functions; establishing intra-modal and inter-modal relationships while more than one modality is involved etc.

Transfer Learning and Domain Adaptation for Low-resource Environment: Transfer learning and domain adaptation are two key issues for dealing with the low-resource scenarios. Meta learning and few shot learning are the two recent techniques that help adapting to a new application or problem quickly.

Multimodal Learning: In recent times, many heterogeneous networks have been successfully deployed for both lower-level as well as higher level applications [1-3]. With the availability of such networks, unprecedent amount of information are generated daily from a variety of sources, such as Twitter, Facebook, Uber, Amazon review site, Flipkart review portals and a large no of blogs and sites. These data, often referred to as big data, hold such characteristics as high volume, high variety, high velocity, and high veracity [4-6]. These huge data contain structured, semi-structured, and unstructured data, which are coming from multiple-modality, i.e. text, image, audio and video. And each modality of different source, type, and distribution contains modality-specific information [7-8]. In order for Artificial Intelligence to make advancement to understand the real world around us, it is necessary to be able to interpret and reason about multimodal messages. In multimodal, machine learning we aim at building models that can process and relate information from multiple modalities. From early research on audio-visual speech recognition to the recent explosion of interest in language and vision models, multimodal machine learning is a vibrant multi-disciplinary field of increasing importance and with extraordinary potential [9]. In multimodal learning the challenges lie in the following issues:

- **Representation**: The very first fundamental challenge of multimodal learning is to represent and summarize multimodal data in such a way that exploits the complementarity and redundancy of multiple modalities. The diversity in the multimodal data makes it challenging to construct such representations. For example, language is generally symbolic, but audio and visual modalities correspond to the signals.
- **Translation**: The second challenge is to address how to translate (or, map) data from one to the other modality. The data is not only heterogenous, but the relationships between modalities is often open-ended or subjective. As an example, there exists a number of possible ways to describe an image, and hence one perfect translation may not exist.
- **Alignment**: The third challenge is to identify the direct relations between the elements of one more modalities. As an example, we may want to align the sequences of steps in a particular movie scene with the subtitles displayed. To tackle this challenge, we

should have a mechanism to measure the similarity between different modalities, and deal with the possible long-range dependencies and ambiguities.

- **Fusion:** The fourth challenges in multimodal learning is to join information from two or more modalities to perform a prediction. As an example, for audio-visual speech recognition, the visual description of the lip motion is fused with the speech signal to predict spoken words. This information is coming from different modalities, and may have varying predictive power and noise topology, with possibly missing data in at least one of the modalities.
- **Co-learning**: Another challenge in multimodal learning is to transfer knowledge between the modalities, their representations, and their predictive models. This can be achieved by the algorithms of co-training, conceptual grounding, and zero shot learning. The concept of co-learning investigates how knowledge learning from one modality can help a computational model trained on a different modality.

Some of the important technologies may be developed in the broad areas of Natural Language Processing, Multimodal Artificial Intelligence; and their applications to Education, Healthcare, Agriculture, Tourism, Law, and National Security etc. We list these point-wise here, and explain the details in the subsequent paragraphs and sub-sections.

- Machine Translation techniques (unimodal and multimodal) for low-resource languages
- Machine Translation for specialized domains, such as education, law/judiciary, social media contents, tourism etc.
- Summarization (unimodal and multimodal) of health, scientific documents etc.
- Multilingual Chatbot (unimodal and multimodal) for agriculture, health, tourism, education
- Multimodal Sentiment and Emotion Analysis
- Smart healthcare
- Multimodal Analytics for national security applications

[1]. Zhang, Z., Patras, P., & Haddadi, H. (2019). Deep learning in mobile and wireless networking: A survey. *IEEE Communications Surveys and Tutorials*, *21*(3), 2224–2287.

[2]. Meng,W., Li,W., Zhang, &Zhu, L. (2019). Enhancingmedical smartphone networks via blockchain-based trust management against insider attacks. *IEEE Transactions on Engineering Management*. doi:10.1109/TEM.2019.2921736

[3]. Qiu, T., Chen, N., Li, K., Atiquzzaman, M., & Zhao, W. (2018). How can heterogeneous Internet of things build our future: ASurvey. *IEEE Communications Surveys and Tutorials*, *20*(3), 2011–2027.

[4]. Gao, J., Li, P., & Chen, Z. (2019). A canonical polyadic deep convolutional computation model for big data feature learning in Internet of Things. *Future Generation Computer Systems*, *99*, 508–516.
[5]. Lv, Z., Song, H., Val, P. B., Steed, A., & Jo, M. (2017). Next-generation big data analytics: State of the art, challenges, and future research topics. *IEEE Transactions on Industrial Informatics*, *13*(4), 1891–1899.
[6]. Li, Y., Yang, M., & Zhang, Z. (2019). A survey of multi-view representation learning. *IEEE Transactions on Knowledge and Data Engineering*, *31*(10), 1863–1883.
[7]. Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., & Lew, M. S. (2016). Deep learning for visual understanding: A review. *Neurocomputing*, *187*, 27–48.
[8]. Tadas Baltruˇsaitis, Chaitanya Ahuja, and Louis-Philippe Morency (2017). Multimodal Machine Learning: A Survey and Taxonomy, arXiv:1705.09406v2
[9]. Chen, Z., Zhang, N. L., Yeung, D. Y., & Chen, P. (2017). Sparse Boltzmann machines with structure learning as applied to text analysis. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence* (pp. 1805–1811). Palo Alto, CA: AAAI.

3.1.3. Challenging problems: Core and of National needs

In this section, we describe the challenging problems. While on the one hand, the focus will be to take up some interesting core problems related to text, video and speech analytics; we will develop a set of prominent technologies that will serve the purpose of national needs. Proposals have been conceptualized in consultation with the collaborators and/or industry partners, such as Accenture, TCS Innovation Lab, Wipro Ltd., IBM Research etc.

3.1.3.1. Core Algorithms

We will take up a few interesting problems that have immense potential to advance the fields of text, video and speech analytics. Language models, Multimodal representations, Knowledge infusion into deep neural networks, Multimodal machine learning, Multitasking, Efficient techniques to handle noisy data, Meta and Few shot learning for quick domain adaptation, Multilingual and Cross-lingual learning, Unsupervised and Semi-supervised machine learning, will be the focus to develop prototypes, technologies and products.

3.1.3.1. Prominent Technologies to focus: *NLP and Multimodal Artificial Intelligence*

A. Machine Translation in Education, Law, Tourism and Noisy data

India is a multilingual country with 22 officially spoken languages. Majority of the population (almost 80%) do not speak in English, and therefore, developing machine translation system to make these various contents available in different Indian languages will play an important role towards building a digitally literate society. Education, judiciary, tourism are two important domains, where a large volume of texts is generated in English.

Making this information available in several Indian languages will be beneficial to the society at large to meet the goals of "*Education for All*" and "*Justice for All*" . Speech to Speech Translation of lectures and videos from English to regional languages like Hindi, Bengali, Marathi and Telugu will be tried. Apart from these, educational contents (books, web documents etc) available in English will be translated into different languages, such as Hindi, Bengali, Maratahi, Telugu.

Tourism is another area where translation could play an important role. Many tourists from Japan travel to India, especially the region of Bihar for visiting Buddhists temples. Machine Translation system from English-Japanese, Hindi-Japanese, Japanese-English will provide important support to these tourists.

Social media, on the other hand, is the source that produces enormous amount of information daily, but majorly in English. Translating this information into vernacular languages will facilitate various e-commerce services.

We will take up a few interesting problems on Machine Translation to address the problems of low-resource scenario (as Indian languages are *resource-constrained in nature*): unsupervised neural machine translation under low-resource scenario; domain adaptation and transfer learning involving low-resource languages; domain dictionary creation; parallel corpus filtering, handling noise etc.

Challenges: Developing MT system for these domains is challenging due to following reasons:

- **Unavailability of data:** One of the major challenges is unavailability of good quality parallel corpus. The data scarcity problem can be addressed through use of synthetic corpus but it might lead to inadequate translations. This is true for Education, Judiciary, Tourism or Nosiy data obtained from social media.
- **Nature of subtitle data and speech transcription** (for Speech-to-Speech translation)**:** The length of subtitles and text data generated from speech vary drastically from 1-2 words to 40-50 words per sentence. Most of the time the speech data, when transcribed, contains a lot of spontaneous words and phrases (such as 'ok! let's look at this' or 'good morning' etc). This might lead to erroneous translations when translating longer sentences from in-domain data.
- **Translation of domain specific terms:** Translation of domain specific terms is challenging due to two reasons. One is, domain terms may not appear frequently in the

corpus which might lead to wrong translation. Second is, domain terms may not always have one translation (for eg., translation of the word 'tree' will differ from Computer Science domain to general domain).

- **Code-Mixing:** Code-Mixing (CM) is mixing of two or more languages in a single sentence. This phenomenon is predominant in e-commerce, and also quite common in the other domains. A code-mixed sentence can also contain foreign words written in native script (transliteration of other language words in native script). The challenge is how to handle the CM text. In most of the cases the CM words should be kept as it is (as they might be domain specific terms) and in some cases, they need to be transliterated in target script.

Possible approaches to address the challenges: The challenges can be addressed by following approaches:

- **Unavailability of data:** Small amounts of domain specific data can be created via crowdsourcing. Using this data, synthetic data that is generated from available MT models, can be post-edited/cleaned and can be used to train models. Another approach is dynamic data selection. In this approach, the model is learned to choose relevant data from unlabelled data and used in the training.
- **Multilingual and Pivot based MT approaches:** Multilingual MT models (single model capable of translating multiple language pairs) and Pivot based MT models (training source to target model via source to pivot and pivot to target) have shown improvements when adapting model to a specific domain. The translation of domain specific terms are also shown improvement with these approaches.
- **MT training by noise augmentation:** MT models can be made robust to noise by adding the noise to training data. The artificial noise can be generated in many ways such as randomly dropping the words and replacing words with their similar counterparts etc. Adding noise to training data has shown improvements when the model is tested with noisy data.

Impact of this research:

- Creating MT models in education, law, tourism and e-commerce domain is beneficial in a multilingual country like India as everyone can access the information in their native language.
- Large amount of linguistic resources can be created such as parallel corpora, domain specific dictionaries etc, which would be useful for other allied disciplines.
- A lot of time and human efforts will be saved through this process, which, otherwise will take lot of time and effort from translating from scratch.

References:

- Kordoni, V., van den Bosch, A. P. J., Kermanidis, K. L., Sosoni, V., Cholakov, K., Hendrickx, I. H. E., & Huck, M. (2016). Enhancing access to online education: Quality machine translation of MOOC content.
- Castilho, S., Moorkens, J., Gaspari, F., Sennrich, R., Way, A., & Georgakopoulou, P. (2018). Evaluating MT for massive open online courses. Machine translation, 32(3), 255-278.
- Sosoni, V., Kermanidis, K. L., Stasimioti, M., Naskos, T., Takoulidou, E., Van Zaanen, M., ... & Egg, M. (2018, May). Translation crowdsourcing: Creating a multilingual corpus of online educational content. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*.
- Niu, X., Denkowski, M., & Carpuat, M. (2018). Bi-directional neural machine translation with synthetic parallel data. *arXiv preprint arXiv:1805.11213*.
- Abdelali, A., Guzman, F., Sajjad, H., & Vogel, S. (2014, May). The AMARA Corpus: Building Parallel Language Resources for the Educational Domain. In *LREC* (Vol. 14, pp. 1044-1054).
- Behnke, M., Miceli Barone, A. V., Sennrich, R., Sosoni, V., Naskos, T., Takoulidou, E., ... & Kermanidis, K. L. (2018). Improving machine translation of educational content via crowdsourcing.

B. Code-mixed Machine Translation: One of the core problems of noisy data translation

Challenges: Developing MT system for code-mixed data is challenging due to following reasons:

- **Unavailability of data:** Adapting MT models which are trained on non-CM data to CM inputs can be very difficult when there is no prior CM data for training. Generation of CM data for language pairs which do not have linguistic tools is very difficult. Randomly generated CM data will not work effectively as it might lead to poor translations even for non-CM inputs.
- **Types of CM data:** CM data can be generated in many ways but MT models trained on one type of CM data might not work on other CM data. The main reason is, synthetically generated CM data may not capture all aspects of code-mixing. CM data that is available not only contain code-mixing but also noise in the form of spelling and slang etc. This makes the model perform poorly on such inputs.
- **Translation performance on non-CM data:** Adding CM data may sometimes degrade the performance of the model on non-CM inputs. Even though it makes the model robust to CM inputs, adding synthetic data has shown performance degradation for non- CM inputs after training the model on CM data.
- **Generation of CM data:** Most of the cases, synthetic CM data generation depends on the linguistic resources of the languages involved. But these resources may not always be available. Unsupervised CM data generation is more useful for all languages since it does not require any language dependent resource (except monolingual corpus).

However, prediction of code-switching points is difficult and may not be common since these points depend upon all the languages involved in the code-switching.

Possible approaches to address the challenges: The challenges can be addressed by following approaches:

- **Bidirectional MT:** MT model can handle CM inputs even though it is trained only on non CM corpus. This can be achieved by training the model in both directions (such as training a single on source-target and target-source pairs). This makes the model to learn both language properties independently and when a CM sentence comes, it can handle effectively as it has already seen both the languages.
- **Multi-Task MT:** Multi-tasking is making a single model to learn two different tasks. In the case of CM translation, making a model to learn properties of both languages will make the model robust to CM inputs. Another objective is, cleaning the CM sentence by converting it into native language by automatically identifying the code-switched parts and translating them.
- **Unsupervised CM data generation:** CM data can be generated in unsupervised settings with the help of a Multi-tasking model.
- **CM to CM translation:** Sometimes the translation should contain specific terms as they are appearing in the input text (for eg., scientific formulae). This type of translation is required in domains such as educational domain. Making the model to copy specific text from source to target can be very challenging as there might be so few of such instances.

Impact of this research:

- Creating MT models for code-mixed scenarios is beneficial in translation of content from user reviews domain, educational domain, tourism etc.
- Since CM can be considered as noise in data, making MT models for CM data overall improves the robustness.
- CM to CM translation models will help in preserving the information which should not be translated. This might reduce the post-editing time of the translation.

References:

- Gupta, D., Ekbal, A., & Bhattacharyya, P. (2020, November). A semi-supervised approach to generate the code-mixed text using pre-trained encoder and transfer learning. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings* (pp. 2267-2280).
- Yang, Z., Hu, B., Han, A., Huang, S., & Ju, Q. (2020, November). CSP: Code-switching pre-training for neural machine translation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 2624-2636).

- Dhar, M., Kumar, V., & Shrivastava, M. (2018, August). Enabling code-mixed translation: Parallel corpus creation and MT augmentation approach. In *Proceedings of the First Workshop on Linguistic Resources for Natural Language Processing* (pp. 131-140).
- Pratapa, A., Bhat, G., Choudhury, M., Sitaram, S., Dandapat, S., & Bali, K. (2018, July). Language modeling for code-mixing: The role of linguistic theory based synthetic data. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 1543-1553).

C. Low-resource Neural Machine Translation

1. State-of-the-art methods in Low-resource NMT

a. **Attention is all you need** [1]: The Transformer – a model that uses attention to boost the speed with which these models can be trained. The Transformers outperforms the Google Neural Machine Translation model in specific tasks. The biggest benefit, however, comes from how The Transformer lends itself to parallelization.

b. **Meta-Learning for Low-Resource Neural Machine Translation** [2]: This method uses a model-agnostic meta-learning algorithm (MAML) to solve the problem of low-resource machine translation. In particular, many high-resource language pairs are used to find the initial parameters of the model. This initialization allows them to train a new language model on a low-resource language pair using only a few steps of learning.

c. **Adapting High-resource NMT Models to Translate Low-resource Related Languages without Parallel Data** [3]: Related or similar low resource languages share linguistic and semantic structures. Authors exploit this linguistic overlap to facilitate translating to and from a low-resource language with only monolingual data, in addition to any parallel data in the related high-resource language. This method combines denoising autoencoding, back-translation and adversarial objectives to utilize monolingual data for low-resource adaptation.

d. **Uncertainty-Aware Semantic Augmentation for Neural Machine Translation** [4]: As a seq-to-seq task, NMT naturally contains intrinsic uncertainty, where a single sentence in one language has multiple valid counterparts in the other. However, the dominant methods for NMT only observe one of them from the parallel corpora for the model training but have to deal with adequate variations under the same meaning at inference. This leads to a discrepancy of the data distribution between the training and the inference phases. First, a proper number of source sentences are synthesized to play the role of intrinsic uncertainties via the controllable sampling for each target sentence. Then, a semantic constrained network is developed to summarize multiple source inputs into a closed semantic region which is then utilized to augment latent representations.

e. **Meta Back-translation** [5]: Back-translation is an effective strategy to improve the performance of Neural Machine Translation (NMT) by generating pseudo-parallel data. However, it is found that better translation quality of the pseudo-parallel data does not necessarily lead to a better final translation model, while lower-quality but diverse data often yields stronger results instead. Meta back-translation model learns to match the forward-translation model's gradients on the development data with those on the pseudo-parallel (back-translated) data.

2. Gaps and Challenges to Address

a. **Rare word translation/Open vocabulary:** NMT systems have low quality when translating out-of-vocabulary words (OOVs). These are the low frequency word or unknown (UNK) words which are not seen by the NMT model during training. Subword unit [6], [7] is one of the popular methods to deal with UNK words but still it is not fully accurate. especially because they have a fixed modest sized vocabulary due to memory limitations.

b. **Robustness:** NMT models are sensitive towards the noise present in the source text [8] which decreases their performance. User generated content (social media, user reviews etc.) are the most common domains which contain the noisy text. The noises can be present in various forms like spelling, grammar, punctuation, abbreviations, code-mixed text, slang etc.

c. **Multiple domain NMT:** Multi-domain NMT intended to translate text from multiple domains having different syntax, domain specific vocabulary etc. Achieving unbiasedness and generating domain specific vocabulary at the target side are the significant challenges for multi-domain NMT.

d. **Lexical constrained decoding:** Decoder in the NMT system generates the output tokens from left to right depending on the previous context (previously generated output tokens). Lexical constrained decoding forces the decoder to generate domain specific output tokens [9].

e. **Multilingual NMT:** Multilingual neural machine translation (NMT) has led to impressive accuracy improvements in low-resource scenarios by sharing common linguistic information across languages. However, the traditional multilingual model fails to capture the diversity and specificity of different languages, resulting in inferior performance compared with individual models that are sufficiently trained [10].

f. **Self supervision:** Self-supervised NMT consumes the comparable corpus by selecting the useful samples from the corpus. The samples are selected and used to update the NMT model parameters through incremental learning [11].

g. **Code mixed translation:** Code-mixed text consists of the words from different languages. The words can either be from different script or common script. Lack of code-mixed parallel corpus and mixing of language specific syntax are challenges in code-mixed NMT.

h. **Gender neutrality:** When translating from one language into another, original author traits are sometimes partially lost. This results in morphologically incorrect variants due to a lack of agreement in number and gender with the subject. Such errors harm the overall fluency and adequacy of the translated sentence [12].

i. **Utilizing monolingual data:** Language pairs like English-Gujarati, Tamil-English etc. are considered as low resource pairs because of the absence of a huge amount of parallel training data. Utilizing correct language specific and domain specific monolingual data is very much helpful in improving the performance of the NMT model.

j. **Document level NMT:** Document level NMT model uses document level context. Document-level contexts denote the surrounding sentences of the current source sentence [13]. Efficiently utilizing the sentence level context is a challenge in document level NMT.

3. Methods

a) **Self-Supervised Learning:** In the absence of large in-domain parallel corpus for low resource language pairs, the monolingual corpus in both source and target is used to update the initial NMT model. The preliminary NMT model may be trained on either low resource in-domain parallel corpus or out-domain parallel corpus. Self supervision [11], [14] is a technique that focuses on utilising comparable/monolingual corpora.

b) **Joint Training for Multilingual NMT:** Johnson's zero shot approach [15] introduced a joint training of a single NMT model by merging the parallel data for each language pair by appending some special tokens to categorize them uniquely. This approach was very efficient in improving translation quality of low-resource language pairs because jointly training high resource and low resource language pairs help each other to increase the accuracy.

c) **Domain Adaptation and Transfer Learning:** Domain adaptation [16] and transfer learning [17] uses the weights of NMT models trained on out-of-domain or high resource language pairs to fine tune the model for in-domain or low resource language pairs.

d) Pivot based NMT: Pivot based machine translation [18], [19] uses a pivot language to train models on source–to–pivot and then pivot–to–target languages. It is used in the absence of direct source-target parallel corpus.

e) **Data Augmentation and Synthetic data creation (Back-Translation):** Data augmentation is needed to enrich the training data by augmenting original parallel corpus using synthetic parallel samples. Generation of synthetic parallel samples can be done by replacing words/phrases of similar context in the original parallel data.

f) **Teacher Student model for zero-shot translation:** With the help of a pivot language, without decoding twice like pivot translation, teacher student model for zero-shot translation [20] utilizes the knowledge distillation.

g) **Translating Code-mixed sentences:** Bilingual and multilingual users often use mixed languages while writing in social media, blogs, or in review sites. Language identification for converting romanized text into language specific script and finally translating it into the target language can be helpful. Text transliteration can be used to create the Romanized text.

h) **Phonetic-based token mapping and handling for noisy input:** In case of related languages , vocabulary overlap is possible which tends to train the encoder/decoder in a vocabulary-shared manner. [21] used a Romanized form of vocabulary at the target side, created from the different languages. This method is used for transfer learning from the parent to child model. In a similar way, each language can be splitted into phoneme based subwords, and shared vocabulary can be generated to adapt to the NMT model built for the language pairs for which sufficient data is available.

i) **Subword (BPE and Regularization):** To deal with a constrained number of vocabulary and Out-Of-Vocabulary (OOV) tokens, subword tokens will be used. In morphologically rich languages like Indian languages, use of subword units [6], will reduce the total vocabulary size at training and increase the coverage.

j) **Cross-lingual Embeddings:** In case of unsupervised NMT [22], [23], cross-lingual embeddings will help to create a shared space where the same sentence from each language will be represented by similar kinds of vector representations. Monolingual data of Indian languages will be used to train and map the embedding vectors. In mapping [24], vectors of words having similar sense in different languages will be kept closer in multidimensional space.

4. Applications

a) **Health domain:** In the medical field, machine translation systems are useful to translate medial phrases written in reports and prescriptions. ‘Canopy speak’ is one such mobile application which translates medical text from the health domain between English-Spanish language pairs.

b) **Legal domain:** Machine translation in the legal domain translates judgement, orders etc. HEMAT[1] is a machine translation system for translating legal documents between English and Indian languages.

c) **Finance:** In the financial domain, machine translation systems like VERTO are used to translate financial documents, fund sheets, company annual reports etc.

d) **Social media content:** Social media platforms are used by users across the world who speak different languages. Machine translation systems are helpful in translating these social feeds in multiple languages for user convenience.
e) **Product Review translation:** Product reviews are user generated content which also consists of various lexical and syntactical noises. Translating user reviews in vernacular languages provides valuable information about the products to the users.
f) **Subtitle translation:** Subtitle translator is useful in translating subtitles from one language to another language for users from different language backgrounds.
g) **Tourism:** Translating from English-low resource languages will play an important role for promoting tourism.

References

[1] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in Advances in neural information processing systems, 2017, pp. 5998–6008.

[2] J. Gu, Y. Wang, Y. Chen, V. O. K. Li, and K. Cho, "Meta-learning for low-resource neural machine translation," in Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Brussels, Belgium: Association for Computational Linguistics, Oct.-Nov. 2018, pp. 3622–3631. [Online]. Available: https://aclanthology.org/D18-1398

[3] W.-J. Ko, A. El-Kishky, A. Renduchintala, V. Chaudhary, N. Goyal, F. Guzm ́an, P. Fung, P. Koehn, and M. Diab, "Adapting high-resource NMT models to translate low-resource related languages without parallel data."

[4] X. Wei, H. Yu, Y. Hu, R. Weng, L. Xing, and W. Luo, "Uncertainty-aware semantic augmentation for neural machine translation," in Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). Online: Association for Computational Linguistics, Nov. 2020, pp. 2724–2735.

[5] H. Pham, X. Wang, Y. Yang, and G. Neubig, "Meta back-translation," in International Conference on Learning Representations, 2021

[6] R. Sennrich, B. Haddow, and A. Birch, "Neural machine translation of rare words with subword units," in Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Berlin, Germany, August 2016, pp. 1715–1725.

[7] T. Kudo, "Subword regularization: Improving neural network translation models with multiple subword candidates," in Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Melbourne, Australia: Association for Computational Linguistics, Jul. 2018, pp. 66–75.

[8] V. Vaibhav, S. Singh, C. Stewart, and G. Neubig, "Improving robustness of machine translation with synthetic noise," in Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies,

Volume 1 (Long and Short Papers). Minneapolis, Minnesota: Association for Computational Linguistics, Jun. 2019, pp. 1916–1920.

[9] G. Dinu, P. Mathur, M. Federico, and Y. Al-Onaizan, "Training neural machine translation to apply terminology constraints," in Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy: Association for Computational Linguistics, Jul. 2019, pp. 3063–3068. [Online].

[10] C. Zhu, H. Yu, S. Cheng, and W. Luo, "Language-aware interlingua for multilingual neural machine translation," in Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Online: Association for Computational Linguistics, Jul. 2020, pp. 1650–1655. [Online].

[11] D. Ruiter, J. van Genabith, and C. Espa ̃na-Bonet, "Self-induced curriculum learning in self-supervised neural machine translation," in Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). Online: Association for Computational Linguistics, Nov. 2020, pp. 2560–2571.

[12] E. Vanmassenhove, C. Hardmeier, and A. Way, "Getting gender right in neural machine translation," in Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Brussels, Belgium: Association for Computational Linguistics, Oct.-Nov. 2018, pp. 3003–3008.

[13] S. Ma, D. Zhang, and M. Zhou, "A simple and effective unified encoder for document-level machine translation," in Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Online: Association for Computational Linguistics, Jul. 2020, pp. 3505–3511

[14] A. Siddhant, A. Bapna, Y. Cao, O. Firat, M. Chen, S. Kudugunta, N. Arivazhagan, and Y. Wu, "Leveraging monolingual data with self-supervision for multilingual neural machine translation," in Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics.

[15] M. Johnson, M. Schuster, Q. V. Le, M. Krikun, Y. Wu, Z. Chen, N. Thorat, F. Vi ́egas, M. Wattenberg, G. Corrado, M. Hughes, and J. Dean, "Google's multilingual neural machine translation system: Enabling zero-shot translation," Transactions of the Association for Computational Linguistics, vol. 5, pp. 339–351, 2017.

[16] M.-T. Luong, C. D. Manning et al., "Stanford neural machine translation systems for spoken language domains," in Proceedings of the international workshop on spoken language translation, no. IWSLT. Da Nang, Vietnam, 2015.

[17] B. Zoph, D. Yuret, J. May, and K. Knight, "Transfer learning for low-resource neural machine translation," arXiv preprint arXiv:1604.02201, 2016.

[18] Y. Kim, P. Petrov, P. Petrushkov, S. Khadivi, and H. Ney, "Pivot-based transfer learning for neural machine translation between non-English languages,"arXiv preprint arXiv:1909.09524, 2019.

[19] Y. Leng, X. Tan, T. Qin, X.-Y. Li, and T.-Y. Liu, "Unsupervised pivot translation for distant languages," arXiv preprint arXiv:1906.02461, 2019.

[20] Y. Chen, Y. Liu, Y. Cheng, and V. O. Li, “A teacher-student framework for zero-resource neural machine translation,” in Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Vancouver, Canada: Association for Computational Linguistics, Jul. 2017, pp. 1925–1935.
[21] C. Amrhein and R. Sennrich, “On romanization for model transfer between scripts in neural machine translation,” arXiv preprint arXiv:2009.14824, 2020.
[22] M. Artetxe, G. Labaka, E. Agirre, and K. Cho, “Unsupervised neural machine translation,” arXiv preprint arXiv:1710.11041, 2017.
[23] G. Lample, A. Conneau, L. Denoyer, and M. Ranzato, “Unsupervised machine translation using monolingual corpora only,” arXiv preprint arXiv:1711.00043, 2017.
[24] A. Conneau, G. Lample, M. Ranzato, L. Denoyer, and H. J´egou, “Word translation without parallel data,” arXiv preprint arXiv:1710.04087, 2017.

D. Multimodal Summarization Systems

Recent years have witnessed the dramatic increase of multimedia data (including text, image, audio and video), which makes it difficult for users to obtain important information efficiently. Multi-modal summarization (MMS) has gained immense popularity in the last couple of years, due to the increasing availability of multi-modal data on the Internet. It considers inputs in different modalities and produces outputs in multiple modalities. Multimodal outputs are necessary because of the following reasons: 1) It is much easier and faster for users to get critical information from the images 2) According to recent experiments [1], the multimodal output (text + image) increases users’ satisfaction by 12.4% compared to the single-modality output (text) 3) Images help users to grasp events better while texts provide more details related to the events. Thus, the images and text can complement each other, assisting users to gain a more visualized understanding of the events.

The MMS systems have wide range of applications in meeting record summarization, sport video summarization, movie summarization, pictorial storyline summarization, timeline summarization and social multimedia summarization. Videos of meeting recordings, sports events and movies, consist of synchronized voice, visual and captions. The inputs for summarization of pictorial storylines consist of a set of images with text descriptions. But to the best of our knowledge, in none of these application domains, summarization of multimedia data containing asynchronous information about general topics producing multi-modal outputs (text+ image + video clip) was considered.

Through this project, we will propose different novel multi-modal summarization approaches that will generate a multi-modal summary (including abstractive text summary, an image, and a video clip). Initially focus will be given in generating summary in English language but later we will focus on other popular languages like Hindi and Bengali.

E. **Multilingual and Multimodal Conversational Systems**: Agriculture, Legal Assistance, Health, Tourism and/or Education

This project aims at developing a Multi-lingual Chatbot in English and Indian languages for four important domains, namely Judiciary, Health, Tourism and Education. The Chatbot will be a pluggable and open-source engine, with the capability to accept both text and voice as input. The inputs will be in the following languages: English, Hindi and Bengali. This will also have the facility to accept code-mixed inputs, i.e. Hinglish (mixing of Hindi and English) and Bengalish (mixing of Bengali and English). One distinguishing characteristics of the bot will be to make it affect-aware, i.e. capable of dealing with emotion, sentiment, politeness and personalization.

Conversational Artificial Intelligence, nowadays, is one the most discussed technologies all over in the world. The Chatbot Report 2019 [5]reveals the following: Business Insider experts predict that by 2020, 80% of enterprises will use chatbots; By 2022, banks can automate up to 90% of their customer interaction using Chatbots; According to Opus Research, by 2021, 4.5 billion dollars will be invested in Chatbots. Several applications have emerged, such as SIRI, Cortana, Google Now etc. There is also a tremendous growth in the industry of Conversational Artificial Intelligence as this technology is being explored with top priority by the big five leading Artificial Intelligence (AI) driven companies - Facebook, Google, Microsoft, Apple, and Amazon.

The first Chatbot ELIZA was developed by Joseph Weizenbaum at MIT in the 1960s, and since then there has been a tremendous growth in this technology, with the aim of making it more human-like by incorporating empathy, sentiment, emotion and politeness. We are already observing many Chatbots in many different websites used for various purposes. But, unfortunately none of the available chatbots support Indian languages and are capable of dealing with human empathy.

The practice of law involves developing arguments based on the reasoning of courts in previous instances with similar circumstances. This reasoning, known as precedent, is invoked to justify a legal standpoint or discredit the opposition's arguments. Precedent can be reinforced or dismissed by subsequent court decisions, and is therefore the fundamental point of debate in the application of the law. Consequently, it is imperative for legal practitioners to be well versed in legal precedents relevant to their area of practice and remain up to date in their knowledge of such. The weight attributed to this information, found mainly through the study of previous cases, makes the ability to access the decisions essential. These processes are

[5] (https://chatbotsmagazine.com/chatbot-report-2019-global-trends-and-analysis-a487afec05b)

largely manual, imposing lengthy production delays on the industry. Furthermore, lawyers dedicate a significant portion of their time and energy by reading and analyzing the decisions for the purposes of client representation and developing legal documents. Common citizens suffer much due to their lack of knowledge about the implications of various legal matters, terminologies, and their implications. For Healthcare, the Multilingual Chatbot could be of great assistance, particularly in situations such as COVID-19 pandemic, and to provide other basic health related information. In Tourism, Chatbot can be an effective tool to provide support in many ways, such as choosing the appropriate travel destinations, gathering basic information about the places, availability of hotels, restaurants, easy access to Flights, Railways, and other transportation means etc. Education is the manifestation of mind, and it is the fundamental rights of the citizens. Chatbot in Indian languages would immensely help the educational sectors by providing students to choose their programmes, courses, topic of the lectures as well as for counseling services. Agriculture is a sector where conversational system or Chatbot will surely play an important role in the following ways: farmers will get relevant information on weather condition, crops information, soil condition, market condition in their own languages.

F. Multilingual and Multimodal Sentiment and Emotion Analysis

Sentiment and Emotion analysis are two prominent research in AI, NLP, Computer Vision nowadays. While sentiment focuses on coarse-grained analysis of affects (e.g. positive, negative, neutral), the emotion recognition concerns with fine-grained affect analysis in the form of sad, fear, anger, disgust, surprise, happy etc. Sentiment and Emotion analysis can be performed either at the document level, or at the sentence level or at the aspect level. Multimodal sentiment analysis and emotion recognition concern with the combining information from a variety of sources, such as text, video, image, audio etc. Deep learning based techniques are widely used for developing such systems. It has immense potentials in variety of sectors like e-commerce, security, mental health analysis, building human-like conversational systems etc.

G. Indian Language Mixed-code Voice Assistants for Functional Domains

Many Indian corporate and social enterprises (like Banks, Hospitals and other Health care services, Public Services, Utilities) are looking forward to changing their traditional IVR (Interactive Voice Response) systems to AI-powered Chatbots and voice bots. This shift will help them to have better customer interaction, knowing the customer better, better engagement and service. One of the key technical issues in wide-spread adoption of voice bots in Indian context is lack of mature Automated Speech Recognition (ASR) and comprehension and Text to Speech (TTS) models – especially for Indian regional languages and mixed-code (e.g., Hindi + English, Tamil + English) conversations. We propose that R&D effort be spent on creating appropriate thesauri, language models, machine and deep learning models to aid such AI-powered virtual assistants in Indian industry context.

H. **Code-mixed Language Models and Applications**

Language Models are models that impart the understanding of a language and its intricacies to a machine. Language models are typically built using statistical significance of words called *Statistical Language Models* (**SLMs**). In the recent past, language models trained using neural networks, called *Neural Language Models* (**NLMs**) have emerged as a major player in the AI domain. With the advent of the Transformer architecture (Google, 2017), NLMs have aided in building exceptionally high accuracy AI systems with language models like BERT (*Google, 2019*), GPT-3 (*Open AI, 2020*) powering them. These models have completely changed the manner in which NLP applications are built, bringing forth a revolution in the NLP-AI space.

The ***accuracy*** of a language model is measured in terms of how *confused* the model is in predicting the *next word*, given a *context*. This metric is called ***Perplexity*** (PPL). A *better language model* produces a *lower perplexity score.* The state-of-the-art language models handle only *single languages*. Code-mixed language models need to be constructed as a building block, in order to develop applications in code-mixed languages. Although language models have evolved over the past decade and have gained significant attention, code-mixed language modeling still remains a sparsely explored domain.

Given the spectrum of multilingual societies across the world, we address the relevant work in code-mix for the top *most spoken languages*, viz. *Mandarin Chinese, Spanish* and *Hindi*. For *Mandarin-English* code-mixed language models, Genta *et. al.* report a perplexity of **127** for *Mandarin* to *English*. For *Spanglish* (*Spanish-English* language pair) code-mixed language models the state-of-the-art is the work by Gonen and Goldberg, who report a perplexity of **40**. For *Hinglish* (*Hindi-English* language pair) there are only a handful of significant contributions, among which the best perplexity is reported by Pratapa *et. al.* as **772.**

Once code-mixed language models are built, a plethora of AI applications can be built from these models. A few of such applications are outlined below:

1. **Chatbot** - Chat interfaces across domains in code-mixed languages
2. **Speech Recognition** - Converting audio signals to code-mixed text
3. **Machine Translation** - Translating code-mixed languages to matrix language

4. **Natural Language Generation** - Generating code-mixed text
5. **Text Summarization** - Creating summary of large body of code-mixed text
6. **Intelligent Personal Assistant** - Virtual assistants like *SIRI*, *Alexa* which can understand and converse in code-mixed language and context
7. **Smart Keyboard** - Code-mixed keyboards typically used for typing on smart devices, equipped with word completion and next word suggestions
8. **Spell/Grammar Checker** - Automatic detection and correction of incorrect spelling and grammar for code-mixed languages

Challenges in code-mixing

In addition to mixing languages at the sentence level, it is also fairly common to find code-mixing behavior at the word level. This linguistic phenomenon poses a great challenge to conventional NLP systems. The major challenges that lie when trying to build models and applications in code-mixed languages are as follows:

Mixed words across languages: As code-mixing has no predefined set of rules, word mixtures are a highly observed occurrence. For example, words in *Hindi* are used with *English* inflections like *darofy* - '*dar*' (fear) + '*fy*' (inflection in *English*)

Mixture of grammar of constituent languages: As mixing of languages is informal in nature, users tend to mix sentence structures of the member languages. For example:

- **Code-mix***:* *Main khatam karunga job*
- **English translation:** I will finish the job
- **Correct form:** In the above example, the sentence "*I will finish the job*" is written in code-mixed *Hinglish* with the structure of the *English* language. The correct form with *Hindi* grammar would have been: "*Main job khatam karunga*"

Multiple word forms: When languages with native scripts are code-mixed with *English* particularly, the transliteration of words in the native language to *English* varies in form. This happens because of the *unavailability* of a *standard romanized form* of words of such languages. This is observed especially when Indian languages like *Hindi* and *Bengali* are mixed with *English*. For example, the *Hindi* word 'है' (English translation: '*is*') in romanized form may have the variations: *hain, hai, hei, hein, he*

Switching points: Switching Points are the tokens in the text, where the language switches. Switching points have rare occurrences in the corpus. Such sparse occurrences of switching points makes it difficult for any Language Model to learn their probabilities and context.

Obviously, code-mixed language models fail at switching points. This is the **primary** bottleneck for code-mixed models.

Dataset: As research in the code-mixed domain is very limited, datasets are not available, especially large ones, that can be used to train language models for code-mixed languages.

3.1.3.2. Prominent Technologies to focus: *Speech, Video and Text Analytics in Healthcare*

A. Development of Multi-modal Techniques for Pancancer Prognosis Prediction

The high-dimensional nature of cancer-related data makes it hard for physicians to manually interpret these multimodal biomedical data to determine treatment and estimate prognosis. The pancancer analysis of large-scale data consisting of twenty different types of cancers has the potential to improve disease modeling by exploiting these pancancer similarities. The shared representation based on various cancer-related information generated from multiple modalities might help in finding the underlying similarity between various cancer patients. This might be beneficial for the physicians in the decision making of treatment and prognosis of cancer patients.

As we all know that adding multiple modalities to any artificial intelligence-based system generates a more comprehensive description of data and improves the prediction power of the model. But, in real-time scenario some modalities could be missing for some data samples. For instance, survival prediction of cancer patients using multi-modal data (clinical data, mRNA expression data, microRNA expression data and histopathology whole slide images (WSIs)) could benefit the oncologist in their prognosis and diagnosis but only a part of patients contain information from all possible modalities causing the prediction model to fail for these patients.

B. Speech, Video and Text Analytics for Smart Healthcare Systems

From physical devices to smart systems powering medical devices, new technological advances are helping doctors and patients connect in new ways, transmit vital data in real time, and identify and treat life-threatening events faster than ever before. The vision of "anywhere, anytime healthcare" is changing consumer expectations and fueling the next wave of innovation growth. In today's smartphone age, more and more consumers are getting comfortable with the idea of video consultations with their physician, remote monitoring via health apps, and using personalized diagnostic tools in smartphones as a ready reckoner.

We have heard about Internet of Things (IoT) implementation in medical devices, but mostly in the diagnostics area. However, IoT devices are also managing the sudden rush to user-centric environments for growing applications in self-monitoring, rather than being available in hospitals and offices alone. This goes hand-in-hand with the concept of tele-healthcare with wireless monitoring services. The main benefit of IoT for patients is convenience and quick access to vital information to avoid emergency situations (The time to conduct vitals at home vs. finding the time to go to a doctor). People are more willing and likely to take control of and monitor their health if they feel it is easy, convenient and fits into their busy schedule.

The recent development of big data-oriented wireless technologies in terms of emerging 5G, edge computing, interconnected devices of the IoT and data analytics have enabled healthcare services for a happier and healthier life. Although, the quality of the healthcare services can be enhanced through big data-oriented wireless technologies, however, the challenges remain for not considering emotional care, especially for children, elderly, and mentally ill people. Based on the above problem, we see that there is a need of emotion-aware healthcare framework, where the emotion will be an indicator of the health situation and the satisfaction of the patient or the doctor regarding the healthcare service.

The popular healthcare services in 5G include remote diagnosis and intervention and long-term monitoring for chronic diseases. The emerging applications include robot-assisted remote patient care, care services empowered by AR, automation and optimization of hospital logistics, remote surgery, etc.

5G enables a large number of devices to be connected via various protocols. The obvious benefit of 5G is its low latency and ability to transmit large data sets, such as image data in remote surgery or assisting patients with disability via robot where high-quality pictures may need to be transmitted.

Telemedicine continues to shift from the edge of healthcare to the mainstream. With 5G wireless networking, it is becoming more possible now. Telemedicine in such conditions when patients need immediate care but no doctor is available can be life saver. In all these applications, speech, video and text analytics play a vital role to make these healthcare systems more efficient, automated and reliable.

C. Speech, Video and Data Analytics in Healthcare

In this project, we will develop a smart configurable healthcare system for neonatal monitoring. The major problems in these modules are as follows:

(a) Development of contact based/contactless (camera based) vital sign monitoring module: The system can be configured to the contact-based monitoring mode in case of poor light condition, camera position, different sleep position, infant with parents, and the non-contact–based monitoring mode in case of loose electrode condition and very delicate skin of the infants.

(b) Development of cry detection and pain analysis module for neonates: Cry detection and pattern analysis module will be developed for detecting the neonatal cry and pain analysis non-contact microphone sensors

(c) Development of event-triggered, signal quality-aware Internet of Things (IoT)-enabled physiological telemetry module: This module will transmit clinical along with the recorded physiological signals and visual images via different wireless mediums subject to any event detection.

D. Edge-AI based Social-Distance Tracker IoT-Camera

The drug discovery for SARS-CoV-2 is still on research labs, and the growth of the infection rate of the COVID19 is showing exponential. Therefore, to break the transmission rate of the COVID19, the most favorable approach suggested by WHO is maintaining social distance. Most of the countries are applying this approach to slow down the spread of this disease. It is challenging to maintain social distancing in an overpopulated country like India. The protocol of social distancing can be violated in the open market, supermarket, and other congested places where huge people come to buy essential goods. It is quietly unmanageable to track every person whether they are violating the social distancing. Disobeying this protocol can lead to community spread of the disease. So in this project, we have figure out the following problems that we are going to solve using EdgeAI and Computer Vision.

A. How to track a mob violating social distancing.
B. Using video analytics and real-time image data, suggest people a suitable time to go to the market.

E. **Multimedia Lifelog: Foodlog**

This research work aiming is to do develop a foodlog website and also application software for calorie identification in order to dietary control which has its social impact for development, and make a system called FoodLog. This value for users lies in personal enjoyment, in supervision their health, in making a social contribution, depending on how they choose to use it. Being able to generate such additional applications may be a key factor in encouraging users to change their lifestyles. We are focusing on analysis of trend estimation between users and individuals and image recognition using large scale data. Our aim is to keep Food Record for Health Management using multimedia technology. FoodLog: An Easy Way to Record and Archive What We Eat. An image processing engine analyzes the content ofthe meals, divide these into different meal category based on calorie value contained. Next is to determine what food types appear in the picture and how they fit into the dietary balance. It then estimates the dietary balance values which helps us to monitor our health.

Outcomes:

- Datasets (Foodlog for training and testing) for dietary control using calorie identification and verification
- Selection of appropriate computer vision technique and deep learning based techniques for above said applications.
- The algorithmic development in running coded form for these (above said) applications
- Foodlog website and application software for calorie identification using foodlog image.

F. Breast cancer detection and classification from tomosynthesis dataset.

Mammography is the most popular technology used for early detection of breast cancer. Manual classification of mammogram data is a difficult task. We aim to develop a CNN structure for tomosythesisdata. We plan to collect the 3D tomosynthesis data from hospitals and use different artificial intelligence and deep learning techniques for classifying them into benign and malignant. We also aim to work on different machine learning and deep learning techniques to locate the tumour, if present, in the tomosynthesis images. For that we need to have an organized dataset. We also aim to contribute in a state of art dataset for researchers working in this field.

G. Multi-modal AI for Telehealth

AI based telehealth systems (moving beyond Telemedicine) is destined to be a part of better healthcare delivery mechanism – especially in post COVID context. This is especially visible in areas of telehealth innovations where AI applications are used to support, supplement or develop new remote healthcare models and increase access to millions. According to WHO’s eHealth observatory survey, AI in the telemedicine field is directly supplementing innovations in these areas: Tele-radiology, Tele-pathology, Tele-dermatology, and Tele-psychiatry.

We propose to create a framework for a multi-modal AI system (Text, image and video, voice and sensor based) to augment the telehealth capability for leading Indian hospitals. This has go beyond remote patient monitoring to provide truly intelligent and interactive healthcare intervention, assistive guidance and alerts.

3.2. Other Problems

In this section we include the problems which have been conceived in consultation with industry experts, academic collaborators; and few have been proposed by our faculty members.

We represent the problems in terms of a matrix shown in Table 1, where each row of the matrix denotes the three broad themes or technology verticals of our proposed i-Hub, *i.e.* Speech, Video and Text; and each column corresponds to the application verticals, such as Tourism, Judiciary, Health, Education, Railways; Housing and Urban Affairs, Road Transport and Highways; Border Management, Security; Environment, Forest and Climate Change" and Electronics and IT. Each cell of the matrix corresponds to a set of problems that spans across the multiple disciplines. These problems correspond to the: (i). projects submitted by the Faculty members of IIT Patna (*24 projects*); and (ii). problems which we would like to put up as open calls for participation (*43 problems*).

Application Vertical→ **↓** **Technology Verticals**	**Tourism, Judiciary, Railways**	**Border Management/ Security**	**Environment, Forest and Climate Change**	**Education and Health**	**Electronics and IT**	**Road Transport, Housing and Urban Affairs**
Speech	O2, O3,O9, O13	040	037	O1,O3, O10,011, 033,034,035,036, 037,038	P2,P3,P5 , O31	P8, O13
Video	015	P6,P14,P16,O1	P9,	P15,P17,	P7,P8,P9	P7,P8,P10,P1

		5,O16,017,O19,O20,O25,O26, O30,040	O13,O22,028, 037	P18,O9,O21,023, O29, 033,034,035,037, O38, O41		1,014
Text	O8,09	P19, P24,040	P13,037	P24,O9, O10, 011, O32,035,036,037,038, 039, O41	P8,P9,P18, O31, O43, O44	P1,P12,P16

Table 1: Problems to be addressed. Pi denotes the project to be taken by IIT Patna faculty members; Oi: denotes a set of the problems reserved for open "Call for Proposal". Apart from these, the following projects, kept for open calls: O4 (Speech), O5 (Speech), O6 (Speech and Text),O7(Video),O24(Video), O27 (Video) aim at contributing to the basic research of "Speech, Video and Text Analytics", and can be applied to more than one application verticals as mentioned in the table.

A. **Project Proposals:** Conceived by IIT Patna Faculty members

1. IITP1: Deep learning-based models for leveraging data from heterogeneous sources for improved traffic prediction

Handling the heterogeneity and dimensionality of data from the diverse sources becomes a challenging task. Apart from these, the huge structure of the road networks and the variance in spatio-temporal properties across different cities makes traffic modeling a difficult task. Dealing with these hurdles in the traffic demands prediction, and our proposal involves (a) collecting multi-featured data from different cyber-physical sources (b) developing high level Hybrid Deep Neural Network models that can exploit these data to learn the spatio temporal behavior of traffic time series of city with different mobility patterns for improved traffic prediction and (c) integrate the cyber-physical data with social sensing and textual data obtained from location enabled social media feeds for anomaly detection.

In the context of the public transportation system, prediction of traffic demand plays a crucial role in a smart city traffic network. Traffic prediction not only provide the administrative authorities vital cues necessary for better management of transport resources but can also help in better preparation to meet a sudden increase in traffic demands. Although recent techniques for traffic demand prediction rely on machine learning models like tensors and Artificial

Neural Networks (ANNs) trained using data obtained from different cyber-physical sources, however there are several issues that make prediction tasks quite challenging. The increase in cyber-physical sources like GPS, traffic sensors, road and weather sensors, video from CCTV cameras installed at vehicles and other strategic positions along with user-generated textual contents provide opportunities in improving prediction accuracy by exploiting the plethora of information that may be made available.

Many research groups are working on different aspects of the Intelligent Transportation System. Despite working with modern technologies for enhancement of their results, a cross domain platform has always shown improved results. We highlight a few of these groups with their major objectives in this area:

MIT Center of Transportation & Logistics: The group works on diversified areas of ITS. The group contributes also in the field of impact of urban road transportation and trips efficiency.

The Future Urban Mobility IRG: One of the objectives of this MIT alliance group is to provide traffic demand prediction in transportation. It also works in the field of safe and environmental trends associated with transportation.

Other groups are **MIT-Portugal: Sustainable Energy and Transportation System**, **Intelligent Transportation Research Center**, **CSAIL**, **Intelligent Transportation System** and many more. The Initial phase of their program starts with the data collections which most often carried through sensors and other devices of IOTs. Also artificial neural network ANN forms a staple technology of the traffic demand prediction and analysing the trend of mobility pattern. Besides, many other mathematical tools are being coupled with ANN to obtain enhanced hybrid models of traffic analysis.

6. **IITP2:** Deep Learning Audio Signal Processing

We plan to work on deep learning techniques for real-time audio signal processing. We plan to make the datasets and also use the publicly available datasets to study the impact of features and raw audio waveforms for different audio applications such as blind source separation, audio enhancement and sound morphing. Raw waveform does not require hand-crafted features; however, this can increase the complexity of the problem. We need to have a detailed study for various applications as these are challenges in the literature. We also need to investigate the selection of appropriate deep learning models from CNN, RNN, CRNN amongst others for sequence classification. Some of the works can be found in [2-4].

3. **IITP3: Embedded System Design for Text, Speech and Video**

With the advancement of artificial intelligence (AI) and machine learning (ML) there is huge demand for development of application specific processors or embedded systems that can work in real time, specifically for, speech, video and text. Facebook, Amazon and Google are racing each other for the development of AI inspired integrated circuits (processors) that can infer the data flow, web services, and traffic mapping in real time. In this proposal, the focus is to develop (a) an embedded system using off-the-shelf components, and (b) an application specific integrated circuit (ASIC) with latest technology node (10/12 nm).

As modern computers are based on von-Neumann architecture in which memory and processing units are physically separated, hence, it introduces significant latency, as well as shuttling the data between them burnt most of the energy. This calls for a radical departure from the von-Neumann architecture to one such non-von Neumann computational approach, where massive parallel processing can be performed efficiently both in terms of latency and energy per operation. In this line, the project aims to initially explore the innovative device structures that can emulate the behaviour of biological neuron to infer the data in real time. These device structures should be CMOS process compatible, energy efficient, and having small intrinsic delay. Based on that different AI algorithms will be implanted for real time video analytics, text recognition and speech translation.

4. IITP4: **Physical Layer and Cryptographic Security for Video and Speech**

In recent years, video and speech are often communicated through a wireless medium. Since in the 5G communication spectrums are overlay by the use of non-orthogonal multiple access (NOMA) and cognitive radio networks, the high-speed video and speech data may be vulnerable [5][6]. Due to the recent development of the internet of things (IoT), the number of users increases day by day [10]. Eavesdroppers can wiretap the high-speed video and speech while it passes through wireless medium and can be decrypted the video by using superpower computer facility. In conventional video and speech communication, the data is mainly secured by encrypted key[11]. In the quantum computing era, the security key can be obtained in a fraction of time, hence video and speech can be decrypted [12].

To secure the video and speech from the eavesdroppers, the 5G technology is looking for an alternative solution through a double layer authentication in the physical layer which is known as physical layer security (PLS). The PLS is widely investigated for wireless communication in our previous work [7][8][9]. Several methods such as anti-jamming approaches, coding techniques, artificial noise generation, physical layer parameter adaptation, beamforming techniques available in the literature [5][6]. But how a PLS can be adopted for high data rate video and speech communication not yet investigated thoroughly. Most of the research on PLS is limited to how the legitimate link can be improved over the eavesdropper.

In this work, we will first investigate how an advanced PLS can be applied for high data rate 5G communication such as for massive multiple-input and multiple-output (MIMO) or millimeter-wave (mmWave) communication. Secondly, we will study advanced cryptographic security and combined it with PLS and see the overall security performance for video and speech communication which is totally a novel idea. Still dated, there is no research available in this direction. Finally, we will investigate the source and channel coding mechanism to improve the video and speech quality for the legitimate receiver in the presence of PLS constraint and eavesdroppers. This is also a novel idea and no study has been done in this direction. We will design a testbed prototype and measure performance over real-time scenarios. The research group at Electrical Engineering Department is working on PLS for the last ten years and has several high-quality publications [5-10].

5. IIT**P9: Remote Monitoring and maintenance of micro cyber physical machine tools**

In this project, intelligent predictive analytics integrated with communication technologies through video, speech and text in conjunction with the physical CNC machines are going to remotely monitor and control the micro machining process for achieving sustainable advancement of the current technology, i.e, achieving smart micromachining system with wide accessibility and more adaptability. Nowadays, the miniaturization of many consumer products is extending the use of micro-machining operations with high-quality requirements. However, preventing tool failure and reducing the impacts of cutting tool wear on part dimensions and surface integrity for minimum production cost and improvement of product quality are challenging. In fact, industrial practices usually set conservative cutting parameters and early tool replacement policies in order to minimize the impact of tool wear on part quality. Most of the monitoring and controlling strategies in micromachining reported in the literature are applied locally and are individual machine centric. Aiming to achieving higher level of intelligent remote monitoring and adaptive controlling in micromachining, i.e., smart micromachining, CNC machine tools must be integrated with cyber space through applications of ICT tools.

6. IIT**P10: Decentralized Real-time Video Analytics for Robotic Swarm Border Surveillance System**

The border security scenario in India is marked by many threats, with different sectors of the border posing different challenges and complexities. The threats to India are arguably increasing, with principal threats coming from various terrorist organizations using lands of its neighboring nations. For example, the Jammu and Kashmir state in India is the most affected. Moreover, illegal immigration, smuggling, armed intrusions, etc., always remain a concern in this context.

Although various hi-tech systems such as Hand Held Thermal Imagery (HHTI) systems, Long Range Reconnaissance Observation Systems (LORROS), and Battle Field Surveillance Radars (BFSR) greatly enhanced the detection ability of security personnel, yet they fail to provide a complete all-round surveillance due to various factors namely intensive human involvement, unfavourable climatic conditions such as heavy fog and snowfall, difficult topography, high attitude area like Siachen Glacier, riverine areas, dense forests, etc.

To overcome the above challenges, we propose to develop a surveillance system based on a heterogeneous robotic swarm consisting of both unmanned ground robots and aerial robots. Use of heterogeneous swarm of robots for surveillance combines the capability of acquiring the bird's-eye-view of the aerial robots to the close range inspection and direct manipulation of the ground robots. The key challenges involved in the use of such heterogeneous swarms are (1) designing of motion planning algorithm that can handle the heterogeneity of dynamical constraints and cover larger geographical areas in an optimal way, (2) Real-time Video Data processing for intelligence gathering, (3) Achieving critical system attributes (reliability, availability, safety and security). Additional challenges include scalability, efficiency, and real-time response. Notably, we consider video analytics to play an important role in our proposed system.

7. IIT**P11**: **Hyperspectral Video Processing Assisted Automated Segregation of Recyclables from Solid Waste Streams in a Smart City**

Solid waste is broadly consisting of two prominent streams namely, municipal solid waste (MSW) and electronic waste (e-waste). There has been a significant rise in MSW generation in the last few decades due to rapid urbanization and industrialization. Similarly, the declining obsolescence period of electrical and electronic equipment (EEE) as a consequence of rapid technological progress has led to the massive generation of waste electrical and electronic equipment (WEEE) or electronic waste (e-waste) worldwide. MSW comprises many useful recyclable materials such as metal, plastic, and paper. The e-waste is heterogeneous due to its composition, which comprises of many useful recyclable materials such as metallic fractions (MFs), like aluminum and copper and non-metallic fractions (NMFs), like plastic, printed circuit boards (PCB) and glass. There is a need for automated segregation of recyclables from source-separated MSW in a material recycling facility (MRF) as a part of MSW management as well as e-waste in a recycling facility in the developing countries for reducing human drudgery and improving the sorting efficiency for material recycling.

This project aims at development of an automated approach for classification and robot-based sorting of recyclable materials from solid waste comprising of source-separated MSW and e-waste. The main challenges involved in the development of such an automated system are: (1) the high fidelity recognition and classification of recyclable objects from the waste stream using machine learning techniques to effectively handle material inhomogeneity, moisture, dusty environment, and variable illumination, (2) efficient motion planning for the robotic manipulators in the presence of sensing noise to minimize the sorting time, and (3) variations in shape, size, and inertia of the recyclable objects. We plan to employ a combination of hyperspectral and thermal imaging as a sensing mechanism for acquiring the data of the waste stream that will be fed into the proposed system via a conveyor belt. The hyperspectral and thermal images will then be processed intelligently to extract the key features. After this, machine learning based algorithms will be developed to use the obtained feature representation for the classification of the recyclable objects. Intelligent robot motion planning algorithms to bin the classified recyclable objects into the identified categories will then be developed that can handle the variations in shape, size, and inertia of the recyclable objects. It is envisaged that the proposed system can contribute towards intelligent solid waste management for Smart City employing machine learning techniques.

8. **IITP12: Modern Application of Audio and Visual Sensing in Structural Health Monitoring**

There are various challenges faced during structural health monitoring as the system is mainly dependent on human skills starting from the installation of systems till the decision making at the end of the process. Sometimes, the structures are placed in inaccessible regions where installing, monitoring, data collection and maintenance of sensors are difficult. This can be overcome if the system be wireless, non-contact and non-invasive. All these systems are based on human judgement whether it is identifying the appropriate locations of sensors to the decision making for repairing or retrofitting of any structure.

9. **IITP13: AI based Water Level Prediction**

Heavy rains in the mountainous regions of Nepal cause huge amount of water that flows in to the major drainages of Narayani, Bagmati, and Kosi rivers. As these rivers cross into India, they flow into the plains and lowlands of Bihar and break their banks. To protect the Kosi River dam as well as the Kosi Barrage Pool's embankments, Indian engineers who are in charge of the dam in Nepal, further open the dam's gates, which can cause flooding down river in the state of Bihar. A breach in the East Kosi afflux embankment above the dam occurred in 2008 and the Kosi River picked up an old channel it had abandoned for over 100 years near the border with Nepal and India. Approximately 2.7 million people were affected as the river broke its embankment at Kusaha in Nepal, submerging several districts of Nepal and India. About 95% of the Kosi's total water flowed through the new course.

10. **IITP14: Semi-Automated Preliminary Health Assessment of Structures through video processing and deep learning algorithm**

Routine inspection and monitoring of infrastructures must be performed to prevent structural failures which may cause loss of life and properties. A delay in the detection of damage and distress in structure can lead to large maintenance and repair costs, e.g., in Europe, €4–6 billion are spent annually on maintenance of concrete infrastructure. Typically, visual inspection is conducted for structural health monitoring (SHM) which refers to the process of implementing a damage detection and characterization strategy for engineering structures such as bridges and buildings. Visual inspections are very effective but are prone to human error and inspection of large structures such as dams, bridges and tall buildings can be risky, expensive and time consuming.

The SHM process also involves the observation of a system over the time using periodically sampled response measurements from an array of sensors (often inertial accelerometers and optical fibres), the extraction of damage-sensitive features from these measurements, and the statistical analysis of these features to determine the current state of system health. For long term SHM, the output of this process is periodically updated information regarding the ability of the structure to perform its intended function in light of the inevitable aging and degradation resulting from operational environments. However, for real-time monitoring, various sensors need to be installed throughout the structure which has a high installation cost. For example, the wind and structural health monitoring system, costing US$1.3 million, was used by the Hong Kong Highways Department to ensure road user comfort and safety. Further, such monitoring systems are common to modern infrastructure, i.e., much of the existing ageing concrete structures ought to be inspected in the traditional way, i.e., visual inspection followed by non-destructive testing.

The detection of cracks is an important aspect in health monitoring and its early detection provides a better chance to avoid major distress in any structure. In this project, we propose the use of unmanned aerial vehicles (UAV) or drones for visual inspection which can significantly improve the efficiency of the inspection team. Although the use of drones for visual inspection is easier and not entirely new, damage assessment from collected data is not easy. Extracting meaningful information from such collected extensive data can be time consuming and cumbersome, so the time saved during data collection from a structure can easily be spent on the analysis of data. Several automated crack detection techniques are available in the literature. Most of these techniques are based on conventional image processing and have shown unreliable results due to complex features of concrete surfaces such as different light condition, surface finish and roughness. The proposed methodology in this study will use deep learning algorithms, such as Convolutional Neural Networks (CNN), to overcome the existing limitations in crack detection. CNN has successfully been applied to image classification while featuring a great level of abstraction and learning capabilities.

The successful development of the proposed methodology and final product will assist experts to provide a safer, faster and productive inspection. The developed technique will make periodic structure monitoring and/or damage assessment feasible and can help various stakeholders in creating effective infrastructure asset management.

11. **IITP11: Real-time Anomaly Detection in Traffic Video Streams**

Anomaly detection from traffic video streams is an important task for many real world applications such as mobility and surveillance. For example, in mobility applications there is a huge amount of data regarding typical mobility patterns like vehicle turning, pedestrian crossing and so on. Abnormalities or anomalies in such an application correspond to rare patterns deviating from the normal/typical ones, such as the illegal U-turn, crossing vehicles while there is pedestrian crossing, going through a red light and so on. Such abnormalities do not occur often as they are illegal, however their detection is of paramount importance. Moreover, with new driving systems, like self-driving cars, new patterns might emerge that were before unknown.

12. IITP12:_IoT based Condition Monitoring and Fault Diagnosis of Gearbox

In today's industry 4.0 era, the industrial gearboxes are getting used in almost every industry or manufacturing unit for handling different industrial and mending (precision) functionalities. Any sudden failure within these gearboxes can create fatal harm, industry breakdown and massive economic loss. Therefore, Condition monitoring and fault diagnosis of industrial gearbox have been researched and developed in the last few decades at a very rapid rate. Due to the high complexity of industrial gearbox, research on improving the accuracy and reliability of intelligent fault diagnosis expert system via data mining remains a prominent issue in this field. Therefore, the industrial sector is optimizing the usage of IoT for better performance. This research work investigates an intelligent fault diagnosis expert system of industrial gearbox based on data mining approaches. This expert system is formulated in a systematic manner by using the concept of Industry 4.0, which includes five modules: (1) Sensor Selection and Data Acquisition (DAQ), (2) Data Preprocessing, (3) Data Mining, (4) Decision making, and (5) Maintenance Implementation.

13. IITP13: Development of Machine Learning based IDS (ML-IDS)

Malware have become increasingly sophisticated in the recent times with the advancement in software, hardware and network technologies. Traditional malware upon inflicting the target system often lead to sudden rise in CPU, memory or network usage, hampering vital system functions or hogging system resources. Although this often lead to the target system being unusable, their detection was often easy due to the abnormal system behavior inflicted by them. Consequently, malware defense system (Intrusion Detection System (IDS)) that were developed wrote various rules and patterns for such abnormal system behavior in order to detect such malware.

This lead to the change in strategy adopted by the malware developers. Overtime malware developers became more and more aware of the defense techniques applied by the IDS. The malware developers now focus on the development of clandestine malware – a malware that is characterized by its stealthy behavior. Such clandestine malware stay in the target system occupying minimal system resources. These malware either exfiltrate critical system data to command and control server (C&C servers), exploit system vulnerabilities to conceal their presence, alter operating system functions like patching task manager process or system libraries at runtime to prevent its enumeration from listing the malware process etc. Such clandestine malware often incorporate various anti detection and anti-analysis methods like anti-VM (prevent from analysis in virtual machine), process hiding (patch task manager or ps utilities), file/directory hiding (prevent from displaying), obfuscation, anti-debug (prevent debugging) etc. In fact a kernel level rootkit can easily conceal itself from user level programs. These characteristics make the detection of clandestine malware extremely difficult and complex. Examples of such clandestine malware are vlany, enyelkm, azazel etc...

In this research we focus on development of machine learning based IDS (ML-IDS) that can detect the presence of such clandestine malware.

14. IITP14: Secure Monitoring and Data Analysis Management tool

The advent of Internet of Things (IoT) and information society has highlighted the need for reliable and secure data monitoring and its analysis. This part of the work aims at developing the first general-purpose secure computational framework and supporting software tools that will develop a sensor platform for environment sensing and activity monitoring in a domestic setting. The system will collect structured data from multiple sensors. These will be used to build linked data sets that will enable clinical and non-clinical studies requiring home environment data, such as indirect activity recognition or investigating links between environment variables (e.g. air quality) to health monitoring and/or development, including onset, of medical conditions. A related use will enable automatic detection of human behaviour related insecurity at the HCI (Human Computer Interaction) level without the need to involve real human users.

15. IITP15: Edge-AI based Social-Distance Tracker IoT-Camera

The drug discovery for SARS-CoV-2 is still on research labs, and the growth of the infection rate of the COVID19 is showing exponential. Therefore, to break the transmission rate of the COVID19, the most favorable approach suggested by WHO is maintaining social distance. Most of the countries are applying this approach to slow down the spread of this disease. It is challenging to maintain social distancing in an overpopulated country like India. The protocol of social distancing can be violated in the open market, supermarket, and other congested places where huge people come to buy essential goods. It is quietly unmanageable to track every person whether they are violating the social distancing. So disobeying this protocol can lead to community spread of the disease. So in this project, we have figure out the following problems that we are going to solve using EdgeAI and Computer Vision.

A. How to track a mob violating social distancing.
B. Using video analytics and real-time image data, suggest people a suitable time to go to the market.

16. IITP16: Real Time Video Stabilization

Video stabilisation technique is essential for most hand held captured videos due to high frequency shakes. We aim to build up a Video stabilisation model using deep learning techniques to handle these high frequency shakes. The dataset for video stabilisation requires two sets of synchronised video sequences of same scene, one shaky video and another stable video. Due to unavailability of such dataset we aim to locally develop such video dataset and work on to develop a real time video stabilisation system using deep learning techniques. A study on various available networks like StabNet [13,14] is to be studied and necessary modifications for making the system real time is to be done.

17. IITP17: Combating Misinformation using NLP and Deep Learning

Fake news, defined by the New York Times as “a made-up story with an intention to deceive”, often for a secondary gain, is arguably one of the most serious challenges facing the news industry today. Misinformation on social media is a real problem now, and having a deep social, economic, and political impact resulting in visible activities like election interference, polarisation, and violence. This problem is more profound in developing countries where literacy levels are low, understanding and exposure to technology are limited, but increasing access to cheap internet makes the mass more susceptible to believing and acting upon misinformation. Research on fake news and misinformation [15,18] show that “Novelty” is a key attribute for misinformation or fake news and contributes significantly to its virality and penetration in the society. Novelty attracts human attention and acts as a stimuli for information sharing and decision-making. Findings from a massive study on Twitter conducted by researchers of MIT [15] suggest that news categorized as false or fake was 70 percent more likely than true news to receive a retweet. False news was more novel than true news, and users were far more likely to retweet a tweet that was “measurably more novel.” The emotional response a tweet generated also played a role in user engagement. Fake news generated replies showing different emotional affective information such as fear, disgust, and surprise. True news, on the other hand, inspired anticipation, sadness, joy, and trust emotion. With the rise of internet and social media usage across India, the menace of misinformation especially in regional languages are rampant [16]. There is an imminent need for reliable regional language “fact-checking services' ', especially in this age of pandemics where misinformation among the less privileged masses could result in severe health consequences. Misinformation via fake news gets easily percolated into the society and is taken on the surface value, adding fuel to the viral spread (especially in Health, Politics, Religion domain in Indian context). Our research aims to build NLP models to explore the role of “textual novelty” and “emotions” to identify potential news items/posts for misinformation. The initial investigation would be on English language data. The models we would develop for English would be extended for popular Indic languages, namely Hindi and Bengali.

B. **Project Proposals for Open Calls:** These problems have been conceived on consultations with industry experts and academic collaborators.

1. **OpenP1:** Conversational Speech Recognition and Synthesis

While most of the advancements in speech science have focused on read speech, there is a lot potential for progress in the area of conversational and spontaneous speech. This applies to both speech recognition and speech synthesis. Such capabilities hold a lot of promise to understand and interpret everyday conversations between humans for various practical applications in areas such as education, agriculture, finance, telecommunications, and several other domains. One of the most important challenges is around standardized data collection and methodological creation of large scale benchmarks (such as SQUAD for Question-Answering in NLP, or ImageNet for object detection in images). Further, there are important problems that address characteristics particularly challenging in conversational speech such as handling disfluencies, multi-party conversations, speaker overlap, exploiting non-verbal cues such as sentiment and emotion, and addressing topical coherence. Novel techniques and methodologies for these and other such challenges continue to remain important. Finally, novel evaluation metrics for conversational speech that go beyond traditional measures such as Word Error Rate, etc will be critical.

2. **OpenP2:** Spoken Language Understanding in Indian Languages

Speech science in Indian languages holds enormous potential thanks to both the diversity of Indian languages, as well as large number of speakers for each language. Almost half of the top-20 most spoken languages in the world are from India. Further, with the deep penetration of mobile phones in the country today, almost 700 million people now have access to technology and the web. This is a perfect combination to unlock the potential for speech science and technology in the country. Speech understanding in Indian languages presents several unique opportunities for speech research not seen as much elsewhere in the world. These include studying and exploiting the genealogical information across language families in India and using that to overcome data challenges that might be hindering progress in low-resource languages. Variations in dialects and accents with substantial speakers in each across all major languages present interesting problems in stress, tone and pronunciation modelling. Further, being a country with such a large number of languages coupled with migration patterns naturally results in a significant percentage of population being multilingual and therefore leading to multilingual conversations exhibiting characteristics of code-mixing.

3. **OpenP3:** Combined Speech and Natural Language Processing

Traditionally, speech science and NLP have mostly been tackled by separate communities in silos. However, both areas are critical and highly inter-related to the larger goal of human language understanding. Thanks to recent advancements in hardware and GPU technologies, there has been recent interest in combining problems across these areas. With that in mind, there are several open challenges when viewed from the lens of combining speech and NLP. These include identifying important parts-of-speech and tackling shallow parsing of say noun phrases and verb phrases directly from the audio signal without having to perform full ASR. Going further up the NLP stack, identifying named entities, performing sentiment analysis, as well as inferring meaning and intents directly from the speech signal holds a lot of potential. Further, audio signals carry a lot of non-verbal information that can be crucial to understanding tone, emotion, meaning, that go well beyond the spoken words. Additionally, there are several open problems in both representation learning as well as exploitation of these for solving specific tasks. Finally, there are a large set of similar problems on the synthesis side looking at combined textual language generation and speech synthesis.

4. OpenP4: Speech Databases Development

India is a multilingual country with many languages, each having a variety of dialects. Further Hindi is our national language and English is being used pan India. Except for a few languages like Indian English, Hindi, Bengali, Tamil, etc, there are no speech databases available for speech technology development. Even though some funded projects of TDIL several attempts have been made, it is only for a few languages.

a. Need serious attempts to collect speech databases in a war footing and make it available to academia, startups and companies for technology development. Every smaller academic institute or start up may not have the knowhow, resources and bandwidth to collect speech databases on their own.

b. TIH can take lead in this direction and come up with a framework for speech databases collection, organization and distribution.

c. TIH may adapt the model for Indian languages that is being run by LDC, University of Pennsylvania, USA. This may become a revenue earning model for TIH.

d. Inviting ideas from academia and industry persons through all possible means, including social media, and brainstorming and generation of white paper on

i. Types of speech databases to be collected

ii. Framework for database collection, organization and distribution

iii. How to collect speech data and prepare a database

iv. Tools for speech database collection

v. Platform for speech data collection and organization.

5. **OpenP5:** Development of Speech Technologies for Indian Languages

There are several core speech technologies that need to be developed for Indian languages. Among these, at the first level speech recognition and speech synthesis systems need to be developed. After this focus may be on other speech technologies enlisted below. TIH may come up with general framework and toolkits which can be used by smaller places and start ups to develop speech recognition and speech synthesis systems.

a. Speech Recognition: Speech to text conversion, monolingual, multilingual, code switching, low resource and zero resource cases.
b. Text to Speech Synthesis: Text to speech conversion, monolingual, multilingual, code switching, low resource and zero resource cases.
c. Speaker Recognition: Speech based person authentication.
d. Language Identification: Identifying spoken language.
e. Dialect Identification: Identifying dialect of a language.
f. Pathological Speech: Processing of pathological speech.
g. Forensic Speech: Processing speech for forensic application.
h. Applications: Using one or more of the above speech technologies.
i. Toolkits: Toolkits for development of different speech technologies.
j. Call for proposals to academia, start ups and industry to do one or more of the above tasks.

6. **OpenP6:** Speech Technologies at Crossroads with Natural Language Processing

a. Linguistic Resources using NLP for Speech Technologies: There are several linguistic resources needed for speech technology development like pronunciation dictionaries, language models and grammar. The sane can be developed using NLP.
b. Speech to Speech Translation: This is one important technology that involves translation of speech from source language to target language. The same needs speech recognition, natural language processing and then text to speech synthesis. Both Speech and NLP teams can work together to realize this dream of developing speech translators for several Indian languages.
c. Language Assessment: The assessment of language proficiency and fluency can be done better by combining both speech recognition and natural language processing.
d. Emotion and Sentiment Analysis: The assessment of emotions and sentiments can be done better by combining Speech and Natural Language Processing
e. Call for proposals to academia, start ups and industry to do one or more of the above tasks.

4. **OpenP7:** Speech Technologies for Video Processing

a. Audio Visual Speech Recognition: Speech recognition, especially under degraded condition, seems to benefit from video processing.
b. Audio Visual Person Authentication: Person authentication can be mode robust by employing both audio and video modalities.
c. Audio Visual Processing for Analytics: The video is processed to derive some analytics. Since speech and audio is at a very low data rate, and then can be used as a preprocessor to identify regions of interest for video analytics.
d. Audio Visual Diarisation for Analytics: Nowadays, there are multiparty conversations like news debates arranged in many TV channels. The diarisation of such videos for analytics. Such things can be done better by using information from both modalities.

8. OpenP8: Machine Translation Indian Languages to Indian Languages

Mostly due to unavailability of large datasets MT systems on Indian languages to Indian languages are still an unsolved problem. Largely BLEU scores are in the range of 0.23-0.28 ranges. However, there are significant developments in recent times on making English as a pivot language to translate one language to the another using NMT techniques. In coming 5 years, we have to work together to bring MT technology to at least 0.5-0.6 BLEU ranges for major languages like Hindi, Bengali, Tamil, Telugu, Punjabi, Marathi, Kannada etc.

Considering social media cases code-mixing is yet another unseen challenge for Indian subcontinent.

9. OpenP9: Sentiment Analysis Techniques for Indian languages

Sentiment Analysis has become the holy grail for almost any e-commerce organization, or for political analyst, or for a market surveyor. Although there are dozens of readymade solutions available for English, but there is almost none for Indian languages, except some datasets available for Hindi. The project aims at developing solutions with robust accuracy levels, in the range of 0.8-0.9 F1-score level, for major languages like Hindi, Bengali, Tamil, Telugu, Punjabi, Marathi, Kannada etc. Considering social media cases code-mixing is yet another unseen challenge for Indian subcontinent.

10. **OpenP10: Automatic Speech Recognition**

ASR – although Indian languages ASR has been a research topic since almost last 3 decades, but still the availability of usable ASR modules for Indian languages are not quite readily available. While Hindi and Hinglish ASR is quite wonderful from the solution available from Google as APIs, but for other languages ASR is still not usable. Although, there are plethora of research available, but now the systems have started to emerge. Finally, ASR working on mobile devices is the need of the hour.

11. **OpenP11: Speech-to-Speech translation**

With the advancement of several Indian language content on YouTube or on any other video streaming platforms like Udacity, NPTEL or news media it has almost become imperative to use speech-to-speech translation system for Indian language contents, but there is no existing solution that area readily available. Translation of video lectures from English to Indian languages will have important usages in a country like India.

12. **Open12: Emotion Recognition from Speech**

Although it's a well-studied problem, but no solution is readily available. We urge to researchers to seriously work on this problem and make some useful solution available for real life use.

13. Open13: Video-Captioning for Indian Languages

This area needs serious attention. With the increasing amount of video content available online, it is yet another need of the hour. We put thrust on the major languages like Hindi, Bengali, Tamil, Telugu, Punjabi, Marathi, Kannada etc.

14. Open14: Indian Road Traffic Analysis

While the world is actually super-excited about autonomous cars, Indians are ready? The answer might not be easy, but serious developments over Indian road safety through CCTV footages using computer vision techniques is an area need major developments.

15. Open15: Hate/Fake Videos Detection

Indians are highly political and given the nature of the country hate/fake videos are being circulated and become viral soon. Identification of such videos quickly over social media is an essential problem to solve.

16. Open16: Online Human Behavior Detection

Online Human behaviour detection and recognition in untrimmed videosare very challenging computer vision task. In traditional offline action detection and recognition approaches where the evaluation metrics are clear and well established. But, in online behaviour detection it is observed that no consensus of the evaluation protocols is used. Therefore, this problem is aimed get a novel online metric that exhibits an online behaviour, solving most of the limitations of the previous (offline) metrics. Also, we need to develop analgorithm for activity classification and detection. A strong classifier is required to achieve better performance than the state-of-the-art techniques.

Outcomes: Success criterion to assess whether the development objectives have been achieved should be spelt out in measurable terms. Baseline data should be made available to assess the impact of the project at the end. It is essential that baseline surveys be undertaken in case of large, beneficiary-oriented schemes. Success/Evaluation criterion for project deliverables/outcomes should be specified in measurable terms to assess achievements against the projected goals(s).

- Datasets are required for video processing.
- Use optimal features for online behaviour/activity detection and classification using fewer numbers of frames.

- Selection of appropriate deep learning/video processing models for these different applications
- The algorithmic development in running coded for these (above said) applications

17. Open17: **Active Authentication on mobile devices**

In recent years, we have witnessed significant growth in the use of mobile devices such as smartphones etc. In this context, security and privacy in mobile devices becomes very important as the loss of a mobile device could compromise personal data of the user. To deal with this problem, Active Authentication (AA) systems needs to be developed which users can continuously monitor the initial access to the mobile device. Therefore, this project aims to develop more robust algorithms for active authentication on smartphones using various modalities such as

- Face-based Active Authentication Methods
- Attribute-based Active Authentication Method
- Behavioural biometric traits such as gait, touch gestures, and hand movement transparently.

Also, compare the response of the above methods with the performance of the state-of-the-art techniques. A strong classifier is required to achieve comparatively better response.

Outcomes:

- Datasets for pose, gaits etc. are required for testing and verification of mobile authentication.
- Use optimal features for online face, attributes and gait, touch gestures and hand movement detection
- Selection of appropriate deep learning/mobile data based video processing models for above said applications
- The developed algorithm should have running code for the above said problem.

18. Open18: Independent moving object detection from dense scene using moving stereo camera system.

Classically, moving objects are separated from the stationary background by change detection. But if the camera is also moving in a dynamic scene, motion fields become rather complex.

Thus, the classic change detection approach is not suitable. So the goal of the research work is to derive segmentation of moving objects for this general dynamic setting. Also, develop some approaches for identifying and segmenting independently moving objects from dense scene flow information, using a moving stereo camera system. Study the disparity and the optical flow in the image domain and the three-dimensional motion is inferred from the binocular triangulation of the translation vector. Also, demonstrate the improvement using reliability measures for the scene flow variables. Furthermore, compare the binocular segmentation of independently moving objects with a monocular version, using solely the optical flow component of the scene flow. Also, compare the response of the proposed methods with the performance of the state-of-the-art techniques.

Outcomes (Expected):

- Datasets (using moving stereo camera system) for required stereo imaging
- Use optical flow/other approach for independent moving object detection.
- Selection of appropriate deep learning based stereo imaging data processing models for above said applications.
- The algorithmic development in running coded form for these (above said) applications

19. Open19: Automatic target verification and identification for air borne surveillance video

Nowadays, with the rapid development of consumer Unmanned Aerial Vehicles (UAVs), visual surveillance by utilizing the UAV platform has been very attractive. Most of the research works related to UAV captures visual data, mainly focused on the tasks of object detection and tracking. However, limited attention has been paid to the task of person identification which has been widely studied in ordinary surveillance cameras.

- The objective of this research project is to develop automatic target identification and verification techniques for airborne surveillance video along with person identification and verification.
- The other issues required to be addressed in this project are target detection from airborne surveillance video (moving platform). A proper reliable target tracking system is needed.
- Shadow detection, verification by synthesis algorithm using homography and template matching is also a part of the project.
- Alternative verification systems accomplishing tracking and *recognition* simultaneously are the final expectation of the project.

Outcomes (expected):

1. Datasets (Unmanned Aerial Vehicles captured visual data focusing on human) are required for target & person identification and verification
2. Selection of appropriate deep learning and other video processing based techniques for above said research works.
3. The algorithmic development in running coded form for these (above said) applications

20. Open20: Novel Technique for Capture, Analysis and Visualisation of Human body movement using distributed camera.

Development of next generation distributed video sensing systems for understanding human body movements is the aim of this research work. New models of human body movement and structure movement will be used for modelling the movements of singe-joint and whole bodies with applications to animation, biomotion, and gait analysis for diagnosing and treating movement-related disorders. The given research efforts enable novel approaches for realistic animation and the detection of indirect variations in movement, leading to better diagnostic tools and personalized programs for rehabilitation of movement disorders. Strong educational and industrial outreach programs also enhance the research program. The objective is to perform markerless motion capture using multiple calibrated cameras. Shape models such as super-quadrics can be used to represent the humans. Such models are essential for tracking the articulated motion accurately.

Outcomes (Expected):

1. Datasets (capture human movement using distributed camera) for target & person identification and verification
2. Selection of appropriate computer vision technique as well as deep learning and video processing based techniques for above said applications.
3. The algorithmic development in running coded form for these (above said) applications

21. Open21: Multimedia Lifelog: Foodlog

This research work aims to do develop a foodlog website and also application software for calorie identification in order to dietary control which has its social impact for development, and make a system called FoodLog. This value for users lies in personal enjoyment, in supervision their health, in making a social contribution, depending on how they choose to use it. Being able to generate such additional applications may be a key factor in encouraging users to change their lifestyles. We are focusing on analysis of trend estimation between users and individuals and image recognition using large scale data. Our aim is to keep Food Record for Health Management using multimedia technology. FoodLog: An Easy Way to Record and Archive What We Eat. An image processing engine analyzes the content ofthe meals, divide these into different meal category based on calorie value contained. Next is to determine what food types appear in the picture and how they fit into the dietary balance. It then estimates the dietary balance values which helps us to monitor our health.

Outcomes:

- Datasets (Foodlog for training and testing) for dietary control using calorie identification and verification
- Selection of appropriate computer vision technique and deep learning based techniques for above said applications.
- The algorithmic development in running coded form for these (above said) applications
- Foodlog website and application software for calorie identification using foodlog image.

22. 3D & Omnidirectional Image Processing: Gathering information about different types of trees.

It is really important to our country and for nature to know the location of trees and measure tree structures attributes such as trunk diameter and height. The accurate measurement of these parameters will lead to efficient forest resource utilization, maintenance of trees in urban cities, and feasible afforestation planning in the future. Therefore, here aim is to design algorithms for automatic detection of tree from 3-D images reconstructed from 360^0 spherical camera which takes omnidirectional images. Proper study and analysis of 3D images are required especially for conducting research on various tree and city images.

Outcomes:

- Datasets (3-D image using 360^0 spherical camera) for calculating the attributes of tree such as trunk diameter and height is required.
- Selection of appropriate computer vision technique and 3-D imaging or Stereo imaging based techniques for above said applications.

- The algorithmic development in running coded form for these (above said) applications

23. Open23: Personalization of Saliency Estimation

Most existing saliency models use low-level features or task descriptions when generating attention predictions. However, the link between observer characteristics and gaze patterns is rarely investigated. Here aim is to develop a novel saliency prediction technique which takes viewers' identities and their individual traits into consideration while modeling human attention. Instead of only computing image salience for average observers, the interpersonal variation in the viewing behaviors of observers with different individual traits and backgrounds can also be considered. Here, objective is to personalizing the saliency model, which may consider the characteristics of the observer. The influence of observer's behavior and building a personalized model can also be investigated.

Outcomes

1. Datasets (different aged group person) for predicting the personalized saliency.
2. Deep Convolutional Generative Adversarial Network (DCGAN) for generating personalized saliency predictions.
3. The algorithmic development in running coded form for these (above said) applications

24. Open24: Multimedia artworks and attractiveness computing

Aim of this research work is to synthesize a targeted painting image from a painting/comic images having similar texture/others characteristics to the targeted painting while visual content of the given painting/comic is maintained. The same research work can be extended to mapping multiple styles of painting and comic at the same time to a single synthesized image with lower computation cost. Analyzing how much consumers would perceive attractiveness of multimedia artworks has significant potential from the viewpoints of both research and business. The important approaches for predicting attractiveness of multimedia content has been discussed in the computer vision and multimedia communities. Here we are interested in analysing why and how we are attracted to specific persons, content, and services. We need to analyze and enhance such “attractiveness” using multimedia big data.

Outcomes:

- Datasets for Multimedia artworks and attractiveness computing is required for training and testing.
- Machine learning/Deep neural Network is required for both Multimedia artworks and attractiveness computing measures
- The algorithmic development in running coded form for these (above said) applications

25. Open25: Detecting people looking at each other in videos

Capturing the 'mutual gaze' of people is essential for understanding and interpreting the social interactions between them. In this research work, we address the problem of detecting people looking at each other in video sequences. We need to focus on some of the important problems related to mutual gaze. (a) Two people talking to each other but not looking each other. (b) Looking at each other with eye occluded and (c) Looking at each other with very close eye.

Outcomes:

1. Video datasets containing mutual gaze is required for training and testing.
2. Computer vision technique and Machine learning/Deep neural Network are required for identification of mutual gaze persons.
3. The algorithmic development in running code form for these (above said) applications

26. Open26: Pose, gait and activity based exact human and their activity detection from Video

The objective of this work is to detect human and their activity from video using different pose, gait and activity. These three descriptions aim to recognize exact human and their activity from a normal and crowded video. The vision-based human detection research is the basis of many applications including video surveillance, health care, and human-computer interaction.Based on the applicability, some of the important problems are mentioned below.

- Human and their activity detection from Complex and Various Backgrounds (realistic videos prosper with occlusions, illumination variance, and viewpoint changes, which make it harder to recognize activities in such complex and various conditions).
- Multi-subject Interactions and Group Activities:Previous problem is concentrated on low-level human activities such as jumping, running, and waving hands. One typical characteristic of these activities is having a single subject without any human-human or

human-object interactions. However, in the real world, people tend to perform interactive activities with one or more persons and objects.

Outcomes

- Video datasets containing different pose, gait and activity are required for training and testing.
- Computer vision technique and Machine learning/Deep neural Network based techniques are required for solution of the above problem.
- The algorithmic development in running coded form for these (above said) applications

27. Open27: Deep Audio-Visual Speech Enhancement and Recognition

This project aim is to develop new techniques to isolate individual speakers from multi-talker simultaneous speech in videos. The present research problem has focussed on trying to segregate sounds (noise) from different unknown speakers. Here our aim is to develop a deep audio-visual speech enhancement network that can be able to segregate individual speaker's voice from many voices and allow speaking one, keeping other speaker’s voice in mute mode. The proposed algorithm should be applicable for extremely challenging real-world videos. We also extend this problem to identify the speaker’s based on speech/image recognition.

28. Open28: Video Processing/Deep Learning based Tools for the Measurement of Animal Behaviour

In computer vision, most of the researches are related to human behavior/activity recognition. But, measurement of animal behavior i.e. activity recognition based on pose or gait is also a challenging but very important and useful research for the development of our society. This problem obviously can be cracked by the help of computer vision techniques which can make accurate, fast and robust measurement of animal behavior a reality. Here we focus on the problem of how capturing the postures of animals - pose estimation.Pose estimation refer to methods for measuring posture, while posture denotes to the geometrical configuration of body parts. While there are many ways to record behavior, videography is a non-invasive way to observe the posture of animals. So, novel modern approaches are required for robust, fast, and efficient measurement of animal behavior. We also need to develop customized tracking approaches, which opens new avenues for more flexible and ethologically (behavior of animals in their natural habitats) relevant real-world neuroscience.

29. Open29: Video watermarking for authentication

The large availability and access of video data for education such as video lectures, tutorials etc., entertainment such as games and movies poses several challenges of copyright violation and illegal distribution of data. More specifically, pirated video distribution has been a key threat for such copyright violation and authentication. To handle these challenges, watermarking techniques can be used to provide copyright protection and authentication of video data. The wide applications of transform domain can be used for the implementation of video watermarking techniques. In the present research significant frame selection procedure based on mathematical relationship between number of frames, block size and watermark size can be incorporated for better authenticity. The watermark embedding steps should be enough robust so that there should not be any difference between original and watermarked data. At the same time detection steps should also be robust enough so that the attacker can't extract the meaningful information. Also the proposed algorithm should be robust for all types of attacks.

30. Open 30: Neural Machine Translation for Extremely Low-resource Languages

Neural machine translation (NMT) has recently shown highly promising results on publicly available benchmark datasets and is being rapidly adopted in various production systems. However, standard NMT systems need huge amount of parallel corpora: hundreds of millions of sentences. Such amount of data are not available for many languages (e.g. Indic languages) and domains (e.g. medical, tourism, judicial, social media and e-commerce contents etc.).

This project aims at developing effective Unsupervised (and Semi-supervised) Neural Machine Translation (bilingual and multilingual) models keeping Indic Languages in focus. This is targeted for Health, Judiciary and Education domains.

31. Open 31: Mimicking the nuances of a human voice is a major challenge for modern text to speech (TTS) and vice versa (STT) systems. In order to produce human like audio of one second speech, a TTS system require to produce as many as 24K samples. As it involves massive parallelism and complex computations, real time applications which often require GPUs (Graphic Processing Units) or other application specific hardware. Similarly, for video analytics which also involves huge computations and may require some specialized processors with superior computing power to transform video content into actionable intelligence.

32. **Open32:** Procedure extraction from Technical Text Documents

Extracting experimental procedures, manufacturing procedures, treatment procedures as mentioned in text. The need can be to train for a task or compare multiple similar procedures. These can also be linked to video documentation of similar tasks. The above systems can be further enhanced with conversation systems.

33. **Open 33**: Building a General purpose Speech enabled conversation systems for Teaching Learning Systems

This project aims at developing conversational systems that would accept speech input and produces speech output. This bot can assist the students and teachers for various purposes.

34. **Open 34**: Video and Speech Analytics for Elderly Health Care and Teaching specially abled students

This project aims at developing speech and video based analytics models for assisting the elderly people.

35. **Open 35**: Predictive Analytics with News and Structured Business data

Connecting the dots spread across multiple documents - building knowledge networks - applying predictive techniques on these. Thus would be applicable for Supply Chain Reason, better demand forecasting etc.

36. Open36: Indian Language Mixed-code Voice Assistants for Functional Domains

Many Indian corporate and social enterprises (like Banks, Hospitals and other Health care services, Public Services, Utilities) are looking forward to changing their traditional IVR (Interactive Voice Response) systems to AI-powered Chatbots and voice bots. This shift will help them to have better customer interaction, knowing the customer better, better engagement and service. One of the key technical issues in wide-spread adoption of voice bots in Indian context is lack of mature Automated Speech Recognition (ASR) and comprehension and Text to Speech (TTS) models – especially for Indian regional languages and mixed-code (e.g., Hindi + English, Tamil + English) conversations. We propose that R&D effort be spent on creating appropriate thesauri, language models, machine and deep learning models to aid such AI-powered virtual assistants in Indian industry context.

37. Open 37: AI + Robotics

Industrial and Social robotics are coming of age and increasingly being adopted, both in large industry, MSME and in household sectors. The applications can be of operating in hazardous environment (such as infectious disease treatment, disaster recovery), remote operations (mining, agriculture), social robotics (educational assistants, robotic companions for elders). The proposal is to create appropriate AI techniques (such as visual recognitions, spatial reasoning, reinforcement learning) for such robotic firmware to impart improved learning, cognitive and functional capabilities in the fields.

38. Open 38: Multi-modal AI for Telehealth

AI based telehealth systems (moving beyond Telemedicine) is destined to be a part of better healthcare delivery mechanism – especially in post COVID context. This is especially visible in areas of telehealth innovations where AI applications are used to support, supplement or develop new remote healthcare models and increase access to millions. According to WHO's eHealth observatory survey, AI in the telemedicine field is directly supplementing innovations in these areas: Tele-radiology, Tele-pathology, Tele-dermatology, and Tele-psychiatry.

We propose to create a framework for a multi-modal AI system (Text, image and video, voice and sensor based) to augment the telehealth capability for leading Indian hospitals. This has go beyond remote patient monitoring to provide truly intelligent and interactive healthcare intervention, assistive guidance and alerts.

39. **Open 39:** AI based mis-information detection and prevention system

Rather than pandemics, misinformation spreading kills more people and create social discord world over. There is an urgent need to detect and prevent misinformation spreading through social media through effective AI intervention in Indian context (keeping in view our social, religious and cultural sensitivity). In essence, this will have three primary parts –

- Detect if a particular social network post is fake or un-trustworthy
- Detect virality and spreading potential of the content and the "super-spreaders" in the network
- Analyse veracity / claims in non-reviewed or general publications or blogs.

40. Open40: AI based personalized learning augmentation

AI can be a great enabler for personalized learning through higher adaptiveness, audio-visual interactivity, AI based assessment for individual learning proclivity and path. It is not just for curriculum learning, but for social and behavioral skills as well. We propose that a set of projects be initiated on various aspects of AI assisted learning through – AI based course curation, AI based assessment (questionnaire generation, rating and ranking), AI based personalized educational coach, AI based life skill coaching and guidance.

41. **Open 41**: Multi-modal approaches for solving biomedical problems

Bioinformatics is an interdisciplinary field which deals with the methods and software tools for understanding the biological data. In general, the data present in this field are large and complex. This is an amalgamation of three sub-fields, namely biology, computer science, mathematics/statistics. The techniques developed as a part of this field is used to analyze and interpret biological data. There are several challenging problems of bioinformatics like protein structure prediction, homology searches, multiple alignments and phylogeny construction, genomic sequence analysis, gene-finding, and many more.

With the exploration of genomics technologies, researchers are able to collect high-throughput biomedical data. The explosion of these new frontier genomics technologies produces diverse genomic datasets such as microarray gene expression, miRNA expression, DNA sequence, 3D structures, etc. These different representations (modality) of the biomedical data contain distinct, useful, and complementary information of different samples. As a consequence, there is a growing interest in collecting "multi-modal" data for the same set of subjects and integrating this heterogeneous information to obtain more profound insights into the underlying biological system. In recent years different machine learning and deep learning-based approaches become popular in dealing with multimodal data.

The current call aims to grant some recent works in the field of biomedical and health informatics which utilize different DL techniques as solution frameworks. The proposals addressing the following topics will be considered in this special issue:

- Deep learning for protein function prediction, protein-protein interaction detection
- Deep learning for imaging informatics and large-scale mining/classification
- Deep learning for translational bioinformatics and drug discovery
- Deep learning for medical informatics and public health
- Deep learning for solving different bio-NLP problems
- Deep learning techniques for computational genomics

- Deep learning techniques for biomedical signal processing

42. **Open 42**: AI-based Smart Grid Data Analytics

Smart grid is a vital part of energy because it allows energy providers to draw full value from the Smart grid. It allows for a layer of communication between local actuators, central controllers and logistic units, which enables better response during emergencies and more efficient use of resources. It is proposed to use an advanced data analytics platform that can capture and analyse data from different endpoints which enables utility companies to distribute resources more efficiently, cut costs and discover better ways to server their customers. The Data Analytics platform would have the capability to collect large volumes of Data either in structured, semi-structured or unstructured format. Different data ingestion mechanisms will be supported like streaming, micro-batches, logs, bulk etc. These data will be analysed in real-time. AI would be used at all layers - data ingestion, data cleansing and data visualization. The Platform will be optimized on High Performance Compute and High-Performance Data Storage drastically reducing data access, model training and inference time.

This project's objectives are to develop Predictive analytics giving insights to: Demand & Supply; Consumption Patterns; Forecasting renewal energy production; Decide on a real-time basis, which energy source to use and in which proportion; Early warning system to help in better planning and response of service provider; Potential cost savings along with uses of different energy sources; Improved revenue cash flow due to intelligent switchover between available energy sources; Flexibility to add more renewable solution.

43. **Open-43:** Advanced Data Analytics Solutions to the Smart & Integrated Local Energy Systems

Telecom companies that want to be innovative and maximise their revenue potential must have the right solution in place so that they can harness the volume, variety and velocity of data coming to their organization and leverage on actionable insights from that data. Telecom companies are sitting on terabytes of data that are stored in silos and scattered across the organization. For simpler and faster processing of only relevant data, telcos need an advanced analytics driven data solution that will help them to achieve timely and accurate insights using data mining and predictive analytics. The massive amount of data when captured wisely and analysed professionally can reveal powerful insights. Big data and advanced analytics provide telcos with the tools and techniques to harness and integrate new sources and new types of data in larger volumes; and can help operators enhance the overall value of their business in regards to service optimization, customer satisfaction and revenues.

4. Section-4: Aims and Objectives

The proposed "***IIT Patna Vishlesan i-Hub Foundation***" on "Speech, Video and Text Analytics" at IIT Patna will work towards the following aims and objectives with utmost commitment to integrity, fairness and respect to individuals.

1. **Aims**
   a. To make our nation a player in CPS technologies, especially in the areas of "Speech, Video and Text Analytics".

   b. To publish in the reputed journals and top-tier conferences in the areas of "Speech, Video and Text Analytics".

   c. To impact the nation by facilitating the creation of national competence in essential technologies related to "Speech, Video and Text Analytics" of the future and by catalyzing the translation of that technology into usable applications for greater welfare of the society.

   d. Achieve translation of CPS technologies in the areas of "Speech, Video and Text Analytics" for societal and commercial use, nurture startups and increase the job market.

   e. Produce next generation technocrats in CPS technologies, especially in the areas of "Speech, Video and Text" analytics.

2. **Objectives**
   a. To promote translational research in CPS technologies, especially in "Speech, Video and Text Analytics".
   b. To take up foundational, theoretical and applied research in "Speech, Video and Text Analytics".

   c. Development of technologies, prototypes and products by using the latest techniques using Artificial Intelligence (AI), Machine Learning (ML), Deep Learning (DL), Natural Language Processing (NLP), Computer Vision, and Speech Processing.

d. To Take up foundational research in the broad areas of natural language processing, computer vision, speech processing and multimodal information analysis.

e. To work on the fundamental research problems, *viz.* developing robust techniques for multilingual representation learning, multimodal representation learning, scalable multimodal learning, multitasking models, meta learning and few-shot learning for domain adaption, unsupervised and semi-supervised learning, knowledge infused machine learning models, investigating methods for low-resource scenario, efficient techniques for handling noisy, language models for code-mixing etc.

f. To use the developed technologies for solving some interesting problems of national importance, relevant to education, health, tourism, judiciary, railways, border management, security, environment, forest and climate change, electronics and IT, road transport, housing and urban affairs etc.

g. Developing methodologies, technologies and products involving "Speech, Video and Text Analytics" to solve various problems in Indian context.

h. To develop technologies, prototypes and demonstrate associated applications pertaining to national priorities and competence in "Speech, Video and Text Analytics" by

   i. carefully selecting the impactful and innovative technologies for the future to work on;

   ii. crafting harmonious collaboration within and outside IIT Patna for knowledge creation and dissemination;

   iii. catalyzing the conversion of such knowledge into tools, platforms, products for wider use;

   iv. creating and sustaining commercial viability for the long run

i. To nurture and scale up high-end researchers' base, Human Resource Development (HRD) and skill-sets in the emerging areas of "Speech, Video and Text Analytics".

j. To enhance core competencies, capacity building and training to nurture innovation and start-up ecosystems.

k. To create the world-class multi-disciplinary Technology Innovation Hub in "Speech, Video and Text Analytics", which will serve as the focal point for technology inputs for the industry and policy advice for the government in the allied disciplines.

l. To actively pursue engaging Government and Industry R&D labs as partners in the proposed TIH.

m. To incentivise private participation to encourage professional execution and management of pilot scale research projects.

n. To set up a CPS-TBI with a focus on "Speech, Video and Text Analytics".

o. To tie up with the existing facilities at IIT Patna (Incubation Centre, IIT Patna and/or TBI, IIT Patna) to foster close collaboration with the entrepreneurship ecosystem.

The further sub-objectives to enhance core competencies, capacity building and training to nurture innovation and Start-up ecosystem:

1. Equipping the incubate entity with all the world class facility, equipment and services that are essential to convert the idea/ concept to successful business.
2. Providing techno business mentorship to prune and refine the idea from the concept board level to an organizational setup.
3. Encouraging fail-fast to ensure efficient utilization of high-tech resources made available
4. Creation of a holistic ecosystem for encouraging R&D, innovation, and Entrepreneurship in the CPS domain.
5. Enabling creation of IPR within the country for maximizing the domestic value add and diminishing the external dependency in CPS domain providing assistance during prototyping, development and commercialization for the products produced through the scheme for India and other growth markets.
6. Creation of the employments at various levels.
7. Creation of long-term partnership with strategic sectors.
8. Emphasis will be on IP creation and product development to result in increased domestic value addition.
9. i-Hub should demonstrate unique integration of academia, industry, government and Incubation eco systems.

Other objectives/goals in terms of quantifiable measures are shown in the following table.

Sl no.	Components	Activity	Targets
1	**Technology Development**	No of Technologies (IP, Licensing, Patents etc etc)	25
		Technology Products	20
		Publications (Journals and Conferences)	55
2	**HRD and Skill Development**		
		Skill Development	420
		Graduate Fellowships	250
		Post-Graduation Fellowships	50
		Doctoral Fellowships	25
		Post-Doctoral Fellowships	25
		Faculty Fellowships	3
		Chair Professors	3
3	**Increase Research base in CPS Technology**	Researchers, Doctoral, Post-doctoral	75
4	**Centre of Excellences**		
5	**Innovation and Startup Ecosystem**	CPS-GCC-Grand Challenge and Competition	1 (***10 events***)
		CPS-Promotion and Acceleration of Young and Aspiring technology entrepreneurs (CPS-PRAYAS)	1
		CPS-Entrepreneur In Residence (CPS-EIR)	25
		CPS-Start-ups & Spin-off companies	40
		CPS-Technology Business Incubator (TBI)	1
		CPS-Dedicated Innovation Accelerator (CPS-DIAL)	1

		CPS-Seed Support System (CPS- SSS)	1
6	**Job Creation**	Through startups, academic programs, skill training etc.	8750
6	**International Collaboration**		1
7	**Mission Management Unit**		1

Mission Components	**TIH Goals/Startups**
CPS-GCC - Grand Challenges and Competitions	**Organize** 10 challenges in 2 years for scouting ideas for support and incubation under various schemes. 5 winner startups will get 5 Lakhs and 20 Lakhs as Seed Grant
CPS-PRomotion and Acceleration of Young and Aspiring technology entrepreneurs (CPS-PRAYAS)	**Set up PRAYAS Centre** focused on CPS (by augmenting existing prototyping capacity). **Support 10 teams** under PRAYAS scheme in Speech, Video and Text Analytics based products/solutions development, with a **seed fund support of Rs 10 Lakhs**. Duration 1 Year **Migrate 40% of PRAYAS Teams to Startups**
CPS-Entrepreneur In Residence (CPS-EIR)	**Support 50-60 Individuals** in Speech, Video and Text based product/solution development **Migrate 25% EIRs to Startups**
CPS- Start-up	Support 30 Startup companies in the broad area of “Speech, Video and Text Analytics” with a seed fund support of up to Rs 10 Lakhs (from CPS -SSS).

	Migrate 30% Startups to Accelerator
CPS-Technology Business Incubator (TBI)	CPS-TBI of I-Hub will tie up with IC IITP, and work together.
CPS-Dedicated Innovation Accelerator (DIAL)	**Set up the accelerator** (First year) **Support 10-12 Companies** in the accelerator during the remaining mission period with **early stage investment of up to Rs 25 lakh (from CPS -SSS)**
CPS-Seed Support System (CPS-SSS)	Refer above : - 30 Start-up supported @Rs 10 lakhs - 10-12 DIAL Companies invested in @ Rs 25 lakhs

The supported individuals, teams and startup companies will be working in the broad areas of "Speech, Video and Text" Analytics. Incubation period will be **12 months**, extensible up to **24 months** based on need and progress.

5. Section-5: Strategy

The following strategy will be adopted to fulfill the mission's objectives.

1. ***"IIT Patna Vishlesan i-Hub Foundation"*** will focus on areas in technology stack of Cyber Physical Systems related to "Speech, Video and Text Analytics" or the allied areas where IIT Patna has core strength in terms of manpower & facilities. This is considered essential to achieve the mission.

"This will subsequently guide proposal initiation and selection for research and funding."

2. i-Hub will ensure that its collaborations are carefully selected and developed. Primary criteria for selection of partner institutions, companies or people will be the strategic alignment of the objectives of the partnership to TIH mission & technology focus areas and also the ability of the partnership to deliver synergies (talent, skills, infrastructure, industry inputs) leading to enhanced results.

   "This will guide the criteria for evaluating and finalizing collaborative partners in industry and academia."

3. i-HUB will ensure collaboration with the universities abroad., especially from USA, Europe, Singapore and Japan etc. Primary criteria for selection of partner institutions will be the strategic alignment of the objectives of the partnership to TIH mission & technology focus areas and also the ability of the partnership to deliver synergies (talent, skills, infrastructure) leading to enhanced results.

4. i-Hub will mandate industry or startup partnership for all the proposals to be funded. This will ensure that problem statements are application oriented, inputs are need driven and outcomes are implementable and test beds are readily available. This will also ensure that the barrier between academic thought process and industrial expectations are removed and research is more implementation focused.

   Support from the industries and/or the startups in the forms of *In-kind* or *In-cash* will be an important selection criterion. This will ensure the self-sustainability of the projects undertaken. The support will range at least in the range of 15%-25% of the overall outlay.

5. Proposal generation will be a *two-fold* activity. Proposals obtained through general calls will require industry/startup partnership in place, whereas I-Hub will also facilitate calls, where research partners from academia and industry can submit proposals in the thematic areas. The academia, in collaboration with industry, or startup and/or industry in collaboration with academia, can submit the R&D proposals.

   In order not to turn away proposals which presently do not have any industry/academic partnership, such proposals will be required to declare intent for funding or commercialization. Funding will be subject to compliance of the same within stipulated timelines.

“This will guide the selection and funding of proposals"

6. TBI of i-Hub named as ***“IIT Patna Vishlesan TBI”*** on “Speech, Video and Text Analytics” will open several calls which are viable for commercialization, ask for participations from industries and/startups, and a careful selection mechanism will be adopted to select the proposals of high quality.

7. While selecting the proposals, priority will be given to the proposals, which have very clearly defined deliverables, with industry or startup engagement. Support from the industries and/or the startups in the forms of In-kind or In-cash will be an important selection criterion. The support should be at least 15-25% of the overall budget outlay.

   This will ensure the self-sustainability of the projects undertaken.

8. I-Hub will focus on Human Resources Development at all the levels, starting from Undergraduate, Post-graduate, Doctoral, Post-doctoral as well as Faculty levels.

   The proposed course on “B.Tech in Artificial Intelligence and Data Science” at IIT Patna is aligned to the proposed theme of the I-Hub. Apart from in-house, we shall also offer internships opportunities; conduct training programmes, summer schools and winter schools for the undergraduate students. The earnings from these training programs will contribute towards the sustainability of the program.

   The i-Hub will conduct various training programmes in the specialized areas of “Speech, Video and Text Analytics” for the postgraduates, doctoral researchers, faculty members, and professionals. The earnings (Industry: Rs. 30-40k/one-week course; Academia: Rs. 15-20k/one-week course; Students: Rs. 3-5k/one-week course) from these training programs will contribute towards the self-sustainability of TIH.

   The i-Hub will offer MTech in Artificial Intelligence to train the students at the post-graduate levels. The proposed in-take is 50-60, and the fees @Rs. 75k/semester will contribute towards the self-sustainability of the program.

The i-Hub will recruit dedicated and motivated PhD students to conduct research and development in the cutting-edge research areas of "Speech, Video and Text Analytics". There will be following categories: Self-Sponsored, Sponsored, Project-funded (*Research Assistantship*), Regular and Full-time (Teaching Assistantship with *fellowship from the Centre*), Part-time. Proposed tuition fees/semester: Self-sponsored- 40k; Sponsored-75k; Project funded- 30k; Regular and Full-time- 30k.

These earnings will be used for the self-sustainability of the proposed program.

The i-Hub will recruit postdoctoral research fellows who have finished their PhDs in the relevant areas with publications in the reputed foras.

The i-Hub will engage Faculty members with good research experience in "Speech, Video and Text Analytics".

The i-Hub will engage renowned experts with demonstrable research and technology development experiences as the Chair Professors to mentor the various activities in the Centre.

9. TIH considers IP management and technology transfer as a core competence essential for its mission. Hence TIH will prioritize setting up a technology transfer cell that will primarily serve TIH and can also offer consultancy and support to other institutions to facilitate technology transfer.

   "This will smoothen the commercialization drive and self sustainability initiatives."

10. TIH will focus on eliminating redundancy in the system as much as possible by extending existing infrastructure and systems.

    - IIT Patna Labs related to Cyber Physical Systems will be brought under to TIH
    -Commercialization and Acceleration will be done through the CPS-TBI and Incubation Centre
    - Facilities of several Laboratories under the "Centre of Excellence for Artificial Intelligence", such as Artificial Intelligence-Natural Language Processing-Machine Learning (AI-NLP-ML), Data Analytics and Network Science Lab, Centre for Endangered Language Studies and groups working in the areas of Computer Vision,

Image Processing, IoT, Big Data Analytics, Robotics etc. in the Department of Computer Science and Engineering, Electrical Engineering, Mechanical Engineering, Mathematics, Civil and Environmental Engineering, Humanities and Social Sciences will be used for the R&D activities of the TIH. The facilities of the existing Incubation Centre at IIT Patna will be augmented to support the TIH activities.

"This will optimize the utilization of resources, quickly get parts of the system into operation and take off skill and manpower deficiencies in those areas."

11. People are key in achieving the objectives of the mission and continuity of key personnel is critical to its long-term success. Hence, TIH will invest and develop manpower with a long-term focus. Systems will be put in place with the objective of retention of good talent and terms and conditions will be offered at par with similar jobs in the ecosystem. While project rules are complied, options will be created for longer term career growth similar to that of performance driven organizations.

    "This will attract right talent and also is expected to reduce the churn of good talent. "

12. TIH will create a smooth experience for all its stakeholders. It will strive to create simplified operating procedures, clear communication channels, minimal paperwork, defined service turn around targets etc to ensure that the stakeholder expectations are met, while ensuring that the required diligence is maintained.

    "This will ensure that operations are lean and productivity levels are high"

13. TIH will emphasize on accountability from all the beneficiaries and partners. Necessary checks and balances will be established to ensure that accountability is maintained and results are achieved. Partnerships will be evaluated periodically to ensure that it continues to be effective. Ineffective partnerships will be dropped so as not to burden the system with unproductive stakeholders.

    "This will ensure that operations are lean and productivity levels are high"

14. TIH scope is wide and it is practically impossible to have in-house expertise in all areas of technology dealt with. Hence, TIH will primarily operate through expert panels for

decision making with respect to technology operations. For other areas, committees will assist TIH management to run its operations.

“This will ensure that operations are lean and productivity levels are high"

15. TIH will have minimum in-house capability for PR only and it will depend on specialized third parties or existing units of IIT Patna that has a marketing and publicity competence for its major marketing campaigns and events. Marketing and Publicity of TIH activities will be on an outsourced basis. The effectiveness of the same may be revisited mid-term and necessary corrections may be made

“This will ensure that operations are lean, focused on core activities and productivity levels are high".

16. While profit is not its supreme motive, TIH will be carefully selecting proposals with a potential for commercial application. It ensures its own viability in the long term. TIH will look at technology transfers, industry funding and returns on investments made in commercialization projects as its primary sources of revenue.

“ This will guide project selection and funding, industry and startup partnerships.”

17. TIH will incentivise the key contributors to its mission. IIT Patna will also create incentives for its personnel contributing significant time and effort to TIH mission.

“This ensures commitment towards TIH by IIT Patna functionaries who are deeply engaged in its activities.”

18. Government stakeholders (various departments in state and central government) are key to achieving the mission in more than one way. TIH will leverage Govt. stakeholders to create wider adoption of technology, sustainability and visibility.

19. TIH will be self-sustainable through the earnings from the Educational Programmes, Technology Transfers.

20. TIH will have the flexibility to decide the fees structures of the various academic programmes (*Full-time and/or Part-time*).

21. TIH will be an independent body to manage all the innovations coming from the institute and their monetizing.

6. Section- 6: Target Beneficiaries

Below we present the details of beneficiaries for our proposed *"IIT Patna Vishlesan i-Hub Foundation"* on "Speech, Video and Text Analytics"**.**

(6.1). Stakeholders Consultations

To prepare this Detailed Project Report (DPR) comprehensively so that the Mission objectives are realized in full measures, we consulted with the stakeholders, leaders and enthusiasts of "Speech, Video and Text Analytics" in the country and abroad. The interactions happened through emails, skype and other virtual online mediums. We conducted a series of interactions with experts from academia, R&D organizations, industries and other government institutions through the internet and other online platforms. Details of these institutions are mentioned in the following table.

Sl No	**Academic/Industry/Govt. bodies Consulted /Have already collaboration in the related areas**	**Country**	**Remarks**
1	IBM Research	India	Participated in designing a set of problems
2	Microsoft Research	India	Participated in designing a set of problems
3	Accenture Ltd, India	India	Participated in designing a set of problems

4	TCS Innovation Lab	India	Participated in designing a set of problems
5	Wipro Ltd	India	Participated in designing a set of problems
7	Samsung R&D	India	Discussion going on for a joint project
8	Flipkart Internet Pvt Ltd	India	Conceived a project
9	Prithvi.ai	India	Conceived a project
10	Vidhya Sangha Technologies	India	Consulted for project conceptualization
11	Ara Municipal Corporation, Bihar	India	Consulted for project conceptualization
12	Prof. Neeraj Kumar Singh, IRIT	France	Consulted for project conceptualization
13	Dr. Szalay Tibor, Budapest University of Technology and Finance	Hungary	Consulted for project conceptualization
14	Prof. S. R. Mahadeva Prasanna, IIT Dharwad	India	Participated in designing a set of problems
15	Prof. Sadao Kurohashi	Kyoto University, Japan	Already collaborating in the related areas
16	Prof. Massimo Poesio	Queen Mary University of London, UK	Already collaborating in the related areas
17	Prof. Erik Cambria	NTNU, Singapore	Already collaborating in the related areas
18	Prof. Gael Dias	University of Caen, France	Already collaborating in the related areas

19	Prof. Chris Biemann	Hamburg University, Germany	Already collaborating in the related areas
20	Prof. Dieogo Molla-Aliod	Macquire University, Australia	Already collaborating in the related areas
21.	Prof. Stefan Dietze	Heinrich-Heine-University Düsseldorf, Germany	Already collaborating in the related areas
22	Dr. Soujanya Poria	Singapore University of Technology and Design, Singapore	Already collaborating in the related areas
23	Dr. Adam Jatowt	Kyoto University, Japan	Already collaborating in the related areas
24	Dr. Jose Moreno	University of Toulouse - IRIT, France	Already collaborating in the related areas
25	Dr. Roman Klinger	University of Stuttgart, Germany	Already collaborating in the related areas
26	Prof. Stefan Kramer	University of Mainz, Germany	Already collaborating in the related areas
27	Prof. Stefan Bafna	University of California San Diego, USA	Already collaborating in the related areas
28	Prof. Sara Tonelli	FBK, Italy	Already collaborating in the related areas
29	Prof. Andy Way	Dublin City University	Already collaborating in the related areas
30	Prof. Daisuke Kawahara	Waseda University, Japan	Already collaborating in the related areas
31	Prof. Matt Lease	University of Texas at Austin	Already collaborating for a joint project in the areas of multimodality
32	Prof. Amit Seth	University of South	Already collaborating

		Carolina	for a joint project in the areas of AI
33	NatureSense Technologies Pvt Ltd, Kanpur	India	Consulted for project conceptualization
34	Calligo Technologies, Bangalore	India	Consulted for project conceptualization
35	Kiprango Technologies	India	Consulting for a project
35	NTPC	India	Consulting to solve a problem
37	Department of Agriculture	India	Consulted for project conceptualization
38	IIT Bombay, IIT Delhi, IIT Kharagpur, IIIT Hyderabad, CDAC Kolkata, CDAC Pune, Jadavpur University, Au-KBC, IIIT Allahabad, CDOT	India	Already have been collaborating in various projects related to "Text and Speech"
39	Kyoto University, Japan; NTNU, Singapore; Dublin City University, Ireland; University of Caen, France; Humburg University, Germany; Darmstadt University, Germany; Wright State University, USA	Foreign	Already have formal and/or research collaboration in the areas of "Speech, Video and Text Analytics". Their expertise will be utilized.
40	LG Soft, Honeywell, Skymap	India	Already have collaboration with IIT Patna. Their expertise will be used at the various levels including joint projects
41	TIH of other Institutions in the related areas, such as IITKGP, IIT Jodhpur, IIIT Delhi, ISI Kolkata etc.	India	Will be explored to tie-up

(6.2). **Beneficiaries**

The proposed TIH Centre on "Speech, Video and Text Analytics" rightly fits into the major National initiatives like Digital India, Swachh Bharat, Make-in-India, Industry 4.0, SMART Society 5.0, Skill India and Start-Up India. The TIH centre will act as the facilitator and cater to these national initiatives by (i). Taking up research in the core foundational areas as well as applications areas; (ii). developing India specific core technologies in "Speech, Video and Text Analytics"; (iii). creating human resources in the broad areas of "Speech, Video and Text" at the undergraduate, post-graduate, doctoral, postdoctoral and faculty levels; (iv). creating an innovation and start-up ecosystem in line with priorities of our nation.

TIH will create a smooth experience for all its stakeholders. It will strive to create simplified operating procedures, clear communication channels, minimal paperwork, defined service turn around targets etc to ensure that the stakeholder expectations are met, while ensuring that the required diligence is maintained. The following are some of the application verticals (of national priorities) wherein our proposed TIH on "Speech, Video and Text" Analytics will focus on.

1. **Healthcare for all:** In order to provide accessible and economical healthcare solutions to all, it is very crucial to make use of the recent advances of "Speech, Video and Text" technologies to build the robust systems. We will focus on developing technologies, for examples, Chatbot for assisting people in Pandemic situations, to provide useful information for health, and assisting the patients in an empathetic way; developing Edge AI based Social distance Tracker to combat the COVID-19 pandemic situations; developing a smart configurable healthcare system for neonatal monitoring; developing smart healthcare systems based on 5G network platforms; and the development of multimodal pancancer prognosis prediction etc.
2. **Judiciary, Railways, Tourism:** Education, Judiciary, Railways and Tourism are the four important components of our nation. The use of Artificial Intelligence, Natural Language Processing and Machine Learning technologies can facilitate at the various phases of the Judiciary system (finding relevant precedence, accessing the important and relevant information very quickly, making the information available to the

appellants and other stakeholders); Railways (searching for appropriate trains, availability of reservations, interacting with people through Chatbot which can understand user's needs and able to provide very specific information); and tourism (finding relevant destinations, recommending hotels, restaurants, conveyances etc.). Some of the projects aim at developing "Speech, Video and Text" based technologies for Information Extraction, Chatbots, Machine Translation, Information Retrieval etc.

3. **Education:** Education is one of the most important sectors for nation building. We will take up projects to develop solutions such as developing multilingual Chatbots to assist the students for various activities, namely choosing appropriate courses, lectures, topics and also mentoring at the time of distress; translating video lectures from English to Indian languages; translating educational contents in the regional languages, authenticating online video lectures, helping abled students for learning etc.

Some of the other areas, that may be explored include the following:

1. **Environment, Forest and Climate Change:** We will take up a few problems that will address core issues related to the environment, forest and climate change. Some of the specific tasks will be predicting the water level using AI techniques; Multi-lingual video/image captioning of forest, land weather conditions; Predicting the location of trees and their different attributes; Summarizing multi-modal information in a disaster situations.
2. **Electronics and Information Technology:** We will take up a few problems that relate to the broad areas of Electronics and Information Technology. Some of these problems include designing novel embedded system for text, speech and video processing using AI and Deep Learning; designing advanced cryptographic security for video and speech communication with 5G network platforms; designing intelligent predictive analytics tools for remote monitoring and maintenance of micro cyber physical machine tools; designing technologies for real-time audio signal processing; developing machine learning (ML) based Intrusion Detection System (IDS) etc.
3. **Road Transport:** The TIH will take up a few projects to solve some of the critical issues related to the road transportation, such as prediction of traffic demands using machine learning and data analysis; road safety analysis using computer vision and machine

learning; real-time anamoly detection of traffic video streams, condition monitoring and fault diagnosis of Gearbox etc.

4. **Border Management and Security:** Developing technologies, prototypes and tools and their robustness are extremely crucial as these are related to our nation’s security and sovereignty. In our centre, we will take up a few problems with a focus to develop technologies to combat with fake information; video analysis using robotic swarm for border survelliance; secure monitoring and data management; video stabilization model using deep learning; human activity detection from the videos; mobile device authentication etc.

(6.3). HRD and Skill Development

- Skilled PhD students will be produced through this TIH
- MTech and BTech students will be trained with CPS and its allied areas through thesis and projects, especially in the areas of “Speech, Video and Text Analytics”.
- Courses for the upcoming BTech in AI and Data Science are aligned to the overall theme of our proposed TIH. Hence, the students will be trained in the CPS and its related disciplines, such as Artificial Intelligence, Machine Learning, Computer Vision, Data Science etc.
- One MTech program in “Speech, Video and Text Analytics” will be started under this TIH to create a pool of skilled manpower.
- The Engineering students from the other institutions will be encouraged to participate in the process of applying for project funding and setting up common facilities that can enhance the learning experience of the students.
- Faculty Development Programmes in the thematic areas will be organized at the regular interval to train the faculty members from the other institutions. The trained faculty members will, in turn, teach the undergraduate and post-graduate students in their respective institutions.

- Certification Courses on the thematic areas will be jointly organized with industries for the students and faculty members.

- Summer and Winter schools on CPS, for faculty members, researchers, and students will be organized.

- Through prizes and/or certifications, undergraduate students will be encouraged to take up a project in the CPS and its allied areas from the 2nd year itself, and appropriate facilities will be provided through TBI and/or IC to incubate the product and patenting.

(6.4). Innovation, Entrepreneurship Start-up ecosystem

- The Hub will focus on fundamental research in the broad areas of text, video and speech analytics and publish its research findings in the journals and conferences of international repute.

- All the projects in the i-Hub will asked to tie-up with the industries and/or startups. Each project's deliverables will be connected to an existing startup or industry.
- Open challenges will be launched to boost start-up culture in the areas of "Speech, Video and Text Analytics". This will create an ecosystem to build demonstrable solutions in some of the verticals (of national needs) in a very short span of time.
- Facilities of Incubation Centre at IIT Patna will be exploited to incubate and commercialize the products which will emerge from the projects given to the faculty members of IIT Patna. Participations will also be invited from the interested researchers and/or developers to start the start-ups.

(6.5). International Collaborations

Appropriate provision will be kept to ask the investigators to interact with the foreign institutions working in the broad areas of CPS, since from the very beginning of project inceptions. IIT Patna has already signed general and specific MoUs with several institutions in USA, Europe and Asia, such as University of Houston, USA; Kyoto University, Japan; Dublin City University, Ireland etc. Some of the other potential universities or organizations to explore for collaboration: NTNU, Singapore; Darmstadt University, Germany; University of Caen, France; Hamburg University, Germany; Edinburgh University, Scotland; Macquarie University, Australia; Allen Institute for AI, USA; University of Toronto, Canada; University of Pennsylvania, USA etc. TIH will focus on collaborating through joint programs, joint projects with the international institutions.

A few of the projects have been conceptualized after the discussion with the international collaborators. The existing foreign collaborators all over in the world were also contacted for their explicit consents.

7. Section-7: Legal Framework

The ***"IIT Patna Vishlesan i-Hub Foundation"*** on "Speech, Video and Text Analytics" will be registered as a Section 8 Company. Section 8 is opted over a non-profit society considering the transparency in operational and financial matters, accountability, operational control and scope & range of possible operations, as the wide project mandate and range of stakeholders involved. Section-8 Company will be governed by the Hub Governing Body (HGB), TIH, IIT Patna.

To facilitate the creation and execution of the documentation and also to oversee legal matters of TIH, a person with legal background (Law/CS/CA) will be recruited as part of the team. A panel of law / CA firms will be empanelled to render legal services as required.

Given below is the overall legal framework proposed to operationalize TIH. This will be further verified by a legal expert and required changes in the nature of the document will be brought in.

TIH - Proposed Legal Framework (Prepared based on mission strategy components)

Sl No	**"Stakeholder Pairing Benefactor - Beneficiary"**		**Scope of the Engagement**	**Legal Documentation**	**Coverage of Legal Document**	**Remark**
1.	IIT Patna	TIH	Hosting and Supporting TIH	MoU-Alpha	1. Governance of and support to TIH 2. Facility 3. TIH Operational framework	
2.	TIH	Educational Partners	1. UG Student Engagement; 2. PG, PhDs Students & Postdoc Engagement; 3. CPS Infrastructure Development Fund; 4. Upgrading PG Program	i. DPR Received ii. Admin Approval Issued iii. UC Submitted	1. Rules and regulations for the Scheme 2. Financial utilization guidelines, schedules 3. Monitoring and Evaluation framework 4. Remedial measures in	

					terms of non-compliance	
3.	TIH	Educational Partners	1. UG Student Engagement: Internship 2. PG, PhDs and Postdocs Engagement: Fellowship/Salary	MoU - Educational Partner	**1.** Industry/Start up connects **2.** Selection Criteria **3.** Financial Engagement Rules **4.** Supervision and monitoring **5.** IP Ownership Rules **6.** Degree award and compliance to university rules	Either the partner may be entrusted with money with provision for UC OR TIH directly handle fellowships
4	TIH	IIT Patna	1. UG Student Engagement and Fellowship; 2. PG, PhDs Students & Postdoc Engagement and Fellowship; 3. CPS Infrastructure Development Fund; 4. New PG Program 5. Faculty Engagement 6. Faculty Chair	MoU	1. Terms of internship/fellowship support, including research area, selection criteria etc 2. Overheads for IITP 3. Financial Utilization and UC 4. Outcome expectations 5. Feedback and review mechanism 6. Supervision, if work place is Industry or Startup	IIT Patna may be entrusted with money with provision for UC; Regular R&D channel will be followed; Separate legal documentation with the participant may not be needed as the rules of institute will be carried forward and agreed in

					7. IP rights 8. TIH Returns 9. Remedies	MoU
5	TIH	Industry	1. UG Student Engagement: Internship 2.PG, PhDs/Postodc Engagement: (i). PG - Sponsored and/OR located at Industry (ii) PhD - Sponsored and/or located at Industry (iii). PDF – Sponsored or located at Industry 3.Faculty Engagement 4.Faculty - Sponsored projects	MoU	1. Terms of internship/financial support 2. Supervision, monitoring and reporting if work place is Industry 3. Academic requirements of the engagement 4. IP rights 5. Remedies	
6	TIH	Student /Researcher	1. UG Student Engagement: Internship 2. PG and Above Student Engagement 3. PG Fellowship	Prospectus - Internship Prospectus - PG Rules - PhD, PDF	1. Terms of internship/study/fellowship support 2. Outcome expectations if industry project 3. Feedback and review mechanism 4. Supervision, if work place is Industry or Startup 5. Remedies	
7	TIH	Faculty/Chair Professor	1. Industry sponsored research / Tech development 2. TIH Sponsored research / Tech development	1. DPR Received 2. Admin Approval Issued 3. UC	1. Rules and regulations for the Scheme 2. Financial utilization guidelines,	

				Submitted	schedules 3. Monitoring and evaluation framework 4. Remedial measures in terms of non-compliance	
8	TIH	IC, IIT Patna	DIAL Strategic Information Services For Entrepreneurship	1. DPR Received 2. Admin Approval Issued 3. UC Submitted	1. Rules and regulations for the Scheme 2. Financial utilization guidelines, schedules 3. Monitoring and evaluation framework 4. Remedial measures in terms of non-compliance	
9	TIH	IC, IIT Patna	Grand Challenges	1. DPR Received 2. Admin Approval Issued 3. UC Submitted	1. Rules and regulations for the Scheme 2. Financial utilization guidelines, schedules 3. Monitoring and evaluation framework 4. Remedial measures in terms of non-compliance	
10	TIH	IC	EIR Prayasee Startup Seed support system	MoU/DPR	1. Terms of support, selection criteria etc 2. Overheads for IC 3. Financial Utilization and UC	

					4. Outcome expectations 5. Feedback and review mecahnism 6. IP rights 7. TIH Returns 8. Remedies	
11	IC	Aspriring Entrepreneur/ Innovator	EIR PRAYAS	Agreement	1. Terms of support 2. Overheads for IC 3. Financial Utilization and UC 4. Outcome expectations 5. Feedback and review mechanism 6. IP rights 7. Remedies	
12	IC	Startup Company	DIAL Startup	Agreement	1. Terms of support 2. Overheads for IC 3. Financial Utilization and UC 4. Outcome expectations 5. Feedback and review mechanism 6. IP rights 7. IC / TIH Returns 8. Remedies	
13	TIH	State and Central Govt	1.Financial/Non financial collaboration	MoU	1.Financial/Non financial collaboration	

		Agencies	2. Field tests of the technologies developed 3. Adoption of core technologies/tools		2. Field tests of the technologies developed 3. Adoption of core technologies/tools	

8. Section-8: Environmental Impact

Proposed proposal has NO impact on the environment as far as land acquisition, diversion of forest land, wildlife clearances, rehabilitation and resettlement issues are concerned. Further, development of technologies relating to speech, video, and text are concerned, the host institute has adequate infrastructure for accomplishment of the set objectives without impacting the ecosystem.

Research and Development (R&D) unit of IIT Patna has setup a committee for looking after the "Ethical Standards". The Institute Ethics Committee (IEC) has been constituted as prescribed by the Indian Council of Medical Research (ICMR) to review all types of biomedical health related research proposals, proposals involving biological samples, vulnerable population, and/or sharing of confidential information involving human being with a view to safeguard the dignity, rights, safety and well-being of all actual and potential research participants.

9. Section-9: Technology

The proposed ***"IIT Patna Vishlesan i-Hub Foundation"*** on "Speech, Video and Text Analytics" has wide-ranging activities ranging from the fundamental basic research, to translational research and development of cutting-edge technologies for creating solutions to serve the technological requirements of the common man through the development of appropriate skills and technologies. The proposed mission aims at supporting research and development activities through a large number of schemes, programmes and missions.

The proposed *i-Hub* would take up research and developments into the following three sub-categories to remain consistent with the sub-mission of NM-ICPS, and based on the Technology Readiness Level (TRL):

a. **Expert-driven new knowledge generation /Discovery**

We will take up projects that focus on creating new knowledge in the broad areas of "Speech, Video and Text Analytics". The projects mentioned below: IITP1-IITP24 have promise of creating new knowledge (*please refer to the Novelty part of each project*).

The projects reserved under the "Open Call" are conceptualized in consultation with the industry experts, and have novelty components as these all focus towards building new solutions.

b. **Development of products /prototypes from existing knowledge (by experts or teams)**

Most of the projects mentioned below IITP1-IITP24 have the mandates to generate a prototypes, either through new knowledge or by using the existing knowledge.

We have also designed a set of problems which we intend to open up as "*Call for Proposals*". These have been designed in consultations with the industry collaborators, such as Accenture, Microsoft, Wipro, TCS Innovation Lab, IBM Research etc; and academic collaborators form abroad and India.

c. **Technology /product delivery in specific sectors, i.e., projects that involve knowledge generation and also conversion to technology, demonstration of full working technology (by experts or teams)**

All the projects, IITP1-IIITP24 will have a demonstrable product at the end. Most of these projects have been prepared in consultation with the industry experts, PSUs or collaborators from academics, both national and international. The i-Hub also emphasizes for the deployment of the products at the end of the project tenure. TIH will ensure that each project yields with a prototype, tool or product, deployed to the stakeholders, and/or commercialized through startups and industries.

9.1. Prominent Technologies to focus: *NLP and Multimodal Artificial Intelligence*

9.1.1 Machine Translation in Education, Law, Tourism and Noisy data

India is a multilingual country with 22 officially spoken languages. Majority of the population (almost 80%) do not speak in English, and therefore, developing machine translation system to make these various contents available in different Indian languages will play an important role towards building a digitally literate society. Education, judiciary, tourism are two important domains, where a large volume of texts is generated in English.

Making this information available in several Indian languages will be beneficial to the society at large to meet the goals of "*Education for All*" and "*Justice for All*" . Speech to Speech Translation of lectures and videos from English to regional languages like Hindi, Bengali, Marathi and Telugu will be tried. Apart from these, educational contents (books, web documents etc) available in English will be translated into different languages, such as Hindi, Bengali, Maratahi, Telugu.

Tourism is another area where translation could play an important role. Many tourists from Japan travel to India, especially the region of Bihar for visiting Buddhists temples. Machine Translation system from English-Japanese, Hindi-Japanese, Japanese-English will provide important support to these tourists.

Social media, on the other hand, is the source that produces enormous amount of information daily, but majorly in English. Translating this information into vernacular languages will facilitate various e-commerce services.

We will take up a few interesting problems on Machine Translation to address the problems of low-resource scenario (as Indian languages are *resource-constrained in nature*): unsupervised neural machine translation under low-resource scenario; domain adaptation and transfer learning involving low-resource languages; domain dictionary creation; parallel corpus filtering, handling noise etc.

Challenges: Developing MT system for these domains is challenging due to following reasons:

- **Unavailability of data:** One of the major challenges is unavailability of good quality parallel corpus. The data scarcity problem can be addressed through use of synthetic corpus but it might lead to inadequate translations. This is true for Education, Judiciary, Tourism or Nosiy data obtained from social media.
- **Nature of subtitle data and speech transcription** (for Speech-to-Speech translation)**:** The length of subtitles and text data generated from speech vary drastically from 1-2 words to 40-50 words per sentence. Most of the time the speech data, when transcribed, contains a lot of spontaneous words and phrases (such as 'ok! let's look at this' or 'good morning' etc). This might lead to erroneous translations when translating longer sentences from in-domain data.
- **Translation of domain specific terms:** Translation of domain specific terms is challenging due to two reasons. One is, domain terms may not appear frequently in the corpus which might lead to wrong translation. Second is, domain terms may not always have one translation (for eg., translation of the word 'tree' will differ from Computer Science domain to general domain).

- **Code-Mixing:** Code-Mixing (CM) is mixing of two or more languages in a single sentence. This phenomenon is predominant in e-commerce, and also quite common in the other domains. A code-mixed sentence can also contain foreign words written in native script (transliteration of other language words in native script). The challenge is how to handle the CM text. In most of the cases the CM words should be kept as it is (as they might be domain specific terms) and in some cases, they need to be transliterated in target script.

Possible approaches to address the challenges: The challenges can be addressed by following approaches:

- **Unavailability of data:** Small amounts of domain specific data can be created via crowdsourcing. Using this data, synthetic data that is generated from available MT models, can be post-edited/cleaned and can be used to train models. Another approach is dynamic data selection. In this approach, the model is learned to choose relevant data from unlabelled data and used in the training.
- **Multilingual and Pivot based MT approaches:** Multilingual MT models (single model capable of translating multiple language pairs) and Pivot based MT models (training source to target model via source to pivot and pivot to target) have shown improvements when adapting model to a specific domain. The translation of domain specific terms are also shown improvement with these approaches.
- **MT training by noise augmentation:** MT models can be made robust to noise by adding the noise to training data. The artificial noise can be generated in many ways such as randomly dropping the words and replacing words with their similar counterparts etc. Adding noise to training data has shown improvements when the model is tested with noisy data.

Impact of this research:

- Creating MT models in education, law, tourism and e-commerce domain is beneficial in a multilingual country like India as everyone can access the information in their native language.
- Large amount of linguistic resources can be created such as parallel corpora, domain specific dictionaries etc, which would be useful for other allied disciplines.
- A lot of time and human efforts will be saved through this process, which, otherwise will take lot of time and effort from translating from scratch.

References:

- Kordoni, V., van den Bosch, A. P. J., Kermanidis, K. L., Sosoni, V., Cholakov, K., Hendrickx, I. H. E., & Huck, M. (2016). Enhancing access to online education: Quality machine translation of MOOC content.

- Castilho, S., Moorkens, J., Gaspari, F., Sennrich, R., Way, A., & Georgakopoulou, P. (2018). Evaluating MT for massive open online courses. Machine translation, 32(3), 255-278.
- Sosoni, V., Kermanidis, K. L., Stasimioti, M., Naskos, T., Takoulidou, E., Van Zaanen, M., ... & Egg, M. (2018, May). Translation crowdsourcing: Creating a multilingual corpus of online educational content. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*.
- Niu, X., Denkowski, M., & Carpuat, M. (2018). Bi-directional neural machine translation with synthetic parallel data. *arXiv preprint arXiv:1805.11213*.
- Abdelali, A., Guzman, F., Sajjad, H., & Vogel, S. (2014, May). The AMARA Corpus: Building Parallel Language Resources for the Educational Domain. In *LREC* (Vol. 14, pp. 1044-1054).
- Behnke, M., Miceli Barone, A. V., Sennrich, R., Sosoni, V., Naskos, T., Takoulidou, E., ... & Kermanidis, K. L. (2018). Improving machine translation of educational content via crowdsourcing.

9.1.2. Code-mixed Machine Translation: One of the core problems of noisy data translation

Challenges: Developing MT system for code-mixed data is challenging due to following reasons:

- **Unavailability of data:** Adapting MT models which are trained on non-CM data to CM inputs can be very difficult when there is no prior CM data for training. Generation of CM data for language pairs which do not have linguistic tools is very difficult. Randomly generated CM data will not work effectively as it might lead to poor translations even for non-CM inputs.
- **Types of CM data:** CM data can be generated in many ways but MT models trained on one type of CM data might not work on other CM data. The main reason is, synthetically generated CM data may not capture all aspects of code-mixing. CM data that is available not only contain code-mixing but also noise in the form of spelling and slang etc. This makes the model perform poorly on such inputs.
- **Translation performance on non-CM data:** Adding CM data may sometimes degrade the performance of the model on non-CM inputs. Even though it makes the model robust to CM inputs, adding synthetic data has shown performance degradation for non- CM inputs after training the model on CM data.
- **Generation of CM data:** Most of the cases, synthetic CM data generation depends on the linguistic resources of the languages involved. But these resources may not always be available. Unsupervised CM data generation is more useful for all languages since it does not require any language dependent resource (except monolingual corpus). However, prediction of code-switching points is difficult and may not be common since these points depend upon all the languages involved in the code-switching.

Possible approaches to address the challenges: The challenges can be addressed by following approaches:

- **Bidirectional MT:** MT model can handle CM inputs even though it is trained only on non CM corpus. This can be achieved by training the model in both directions (such as training a single on source-target and target-source pairs). This makes the model to learn both language properties independently and when a CM sentence comes, it can handle effectively as it has already seen both the languages.
- **Multi-Task MT:** Multi-tasking is making a single model to learn two different tasks. In the case of CM translation, making a model to learn properties of both languages will make the model robust to CM inputs. Another objective is, cleaning the CM sentence by converting it into native language by automatically identifying the code-switched parts and translating them.
- **Unsupervised CM data generation:** CM data can be generated in unsupervised settings with the help of a Multi-tasking model.
- **CM to CM translation:** Sometimes the translation should contain specific terms as they are appearing in the input text (for eg., scientific formulae). This type of translation is required in domains such as educational domain. Making the model to copy specific text from source to target can be very challenging as there might be so few of such instances.

Impact of this research:

- Creating MT models for code-mixed scenarios is beneficial in translation of content from user reviews domain, educational domain, tourism etc.
- Since CM can be considered as noise in data, making MT models for CM data overall improves the robustness.
- CM to CM translation models will help in preserving the information which should not be translated. This might reduce the post-editing time of the translation.

References:

3. Gupta, D., Ekbal, A., & Bhattacharyya, P. (2020, November). A semi-supervised approach to generate the code-mixed text using pre-trained encoder and transfer learning. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings* (pp. 2267-2280).
4. Yang, Z., Hu, B., Han, A., Huang, S., & Ju, Q. (2020, November). CSP: Code-switching pre-training for neural machine translation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 2624-2636).
5. Dhar, M., Kumar, V., & Shrivastava, M. (2018, August). Enabling code-mixed translation: Parallel corpus creation and MT augmentation approach. In *Proceedings of the First Workshop on Linguistic Resources for Natural Language Processing* (pp. 131-140).

6. Pratapa, A., Bhat, G., Choudhury, M., Sitaram, S., Dandapat, S., & Bali, K. (2018, July). Language modeling for code-mixing: The role of linguistic theory based synthetic data. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 1543-1553).

9.1.3. Low-resource Neural Machine Translation

1. State-of-the-art methods in Low-resource NMT

1. **Attention is all you need** [1]: The Transformer – a model that uses attention to boost the speed with which these models can be trained. The Transformers outperforms the Google Neural Machine Translation model in specific tasks. The biggest benefit, however, comes from how The Transformer lends itself to parallelization.
2. **Meta-Learning for Low-Resource Neural Machine Translation** [2]: This method uses a model-agnostic meta-learning algorithm (MAML) to solve the problem of low-resource machine translation. In particular, many high-resource language pairs are used to find the initial parameters of the model. This initialization allows them to train a new language model on a low-resource language pair using only a few steps of learning.
3. **Adapting High-resource NMT Models to Translate Low-resource Related Languages without Parallel Data** [3]: Related or similar low resource languages share linguistic and semantic structures. Authors exploit this linguistic overlap to facilitate translating to and from a low-resource language with only monolingual data, in addition to any parallel data in the related high-resource language. This method combines denoising autoencoding, back-translation and adversarial objectives to utilize monolingual data for low-resource adaptation.
4. **Uncertainty-Aware Semantic Augmentation for Neural Machine Translation** [4]: As a seq-to-seq task, NMT naturally contains intrinsic uncertainty, where a single sentence in one language has multiple valid counterparts in the other. However, the dominant methods for NMT only observe one of them from the parallel corpora for the model training but have to deal with adequate variations under the same meaning at inference. This leads to a discrepancy of the data distribution between the training and the inference phases. First, a proper number of source sentences are synthesized to play the role of intrinsic uncertainties via the controllable sampling for each target sentence. Then, a semantic constrained network is developed to summarize multiple source inputs into a closed semantic region which is then utilized to augment latent representations.
5. **Meta Back-translation** [5]: Back-translation is an effective strategy to improve the performance of Neural Machine Translation (NMT) by generating pseudo-parallel data. However, it is found that better translation quality of the pseudo-

parallel data does not necessarily lead to a better final translation model, while lower-quality but diverse data often yields stronger results instead. Meta back-translation model learns to match the forward-translation model's gradients on the development data with those on the pseudo-parallel (back-translated) data.

7. Gaps and Challenges to Address

a. **Rare word translation/Open vocabulary:** NMT systems have low quality when translating out-of-vocabulary words (OOVs). These are the low frequency word or unknown (UNK) words which are not seen by the NMT model during training. Subword unit [6], [7] is one of the popular methods to deal with UNK words but still it is not fully accurate. especially because they have a fixed modest sized vocabulary due to memory limitations.

b. **Robustness:** NMT models are sensitive towards the noise present in the source text [8] which decreases their performance. User generated content (social media, user reviews etc.) are the most common domains which contain the noisy text. The noises can be present in various forms like spelling, grammar, punctuation, abbreviations, code-mixed text, slang etc.

c. **Multiple domain NMT:** Multi-domain NMT intended to translate text from multiple domains having different syntax, domain specific vocabulary etc. Achieving unbiasedness and generating domain specific vocabulary at the target side are the significant challenges for multi-domain NMT.

d. **Lexical constrained decoding:** Decoder in the NMT system generates the output tokens from left to right depending on the previous context (previously generated output tokens). Lexical constrained decoding forces the decoder to generate domain specific output tokens [9].

e. **Multilingual NMT:** Multilingual neural machine translation (NMT) has led to impressive accuracy improvements in low-resource scenarios by sharing common linguistic information across languages. However, the traditional multilingual model fails to capture the diversity and specificity of different languages, resulting in inferior performance compared with individual models that are sufficiently trained [10].

f. **Self supervision:** Self-supervised NMT consumes the comparable corpus by selecting the useful samples from the corpus. The samples are selected and used to update the NMT model parameters through incremental learning [11].

g. **Code mixed translation:** Code-mixed text consists of the words from different languages. The words can either be from different script or common script. Lack of code-mixed parallel corpus and mixing of language specific syntax are challenges in code-mixed NMT.

h. **Gender neutrality:** When translating from one language into another, original author traits are sometimes partially lost. This results in morphologically incorrect variants due to a lack of agreement in number and gender with the subject. Such errors harm the overall fluency and adequacy of the translated sentence [12].

i. **Utilizing monolingual data:** Language pairs like English-Gujarati, Tamil-English etc. are considered as low resource pairs because of the absence of a huge amount of parallel training data. Utilizing correct language specific and domain specific monolingual data is very much helpful in improving the performance of the NMT model.

j. **Document level NMT:** Document level NMT model uses document level context. Document-level contexts denote the surrounding sentences of the current source sentence [13]. Efficiently utilizing the sentence level context is a challenge in document level NMT.

3. Methods

a) **Self-Supervised Learning:** In the absence of large in-domain parallel corpus for low resource language pairs, the monolingual corpus in both source and target is used to update the initial NMT model. The preliminary NMT model may be trained on either low resource in-domain parallel corpus or out-domain parallel corpus. Self supervision [11], [14] is a technique that focuses on utilising comparable/monolingual corpora.

b) **Joint Training for Multilingual NMT:** Johnson's zero shot approach [15] introduced a joint training of a single NMT model by merging the parallel data for each language pair by appending some special tokens to categorize them uniquely. This approach was very efficient in improving translation quality of low-resource language pairs because jointly training high resource and low resource language pairs help each other to increase the accuracy.

c) **Domain Adaptation and Transfer Learning:** Domain adaptation [16] and transfer learning [17] uses the weights of NMT models trained on out-of-domain or high resource language pairs to fine tune the model for in-domain or low resource language pairs.

d) Pivot based NMT: Pivot based machine translation [18], [19] uses a pivot language to train models on source–to–pivot and then pivot–to–target languages. It is used in the absence of direct source-target parallel corpus.

e) **Data Augmentation and Synthetic data creation (Back-Translation):** Data augmentation is needed to enrich the training data by augmenting original parallel corpus using synthetic parallel samples. Generation of synthetic parallel samples can be done by replacing words/phrases of similar context in the original parallel data.

f) **Teacher Student model for zero-shot translation:** With the help of a pivot language, without decoding twice like pivot translation, teacher student model for zero-shot translation [20] utilizes the knowledge distillation.

g) **Translating Code-mixed sentences:** Bilingual and multilingual users often use mixed languages while writing in social media, blogs, or in review sites. Language identification for converting romanized text into language specific script and finally translating it into the target language can be helpful. Text transliteration can be used to create the Romanized text.

h) **Phonetic-based token mapping and handling for noisy input:** In case of related languages , vocabulary overlap is possible which tends to train the encoder/decoder in a vocabulary-shared manner. [21] used a Romanized form of vocabulary at the target side, created from the different languages. This method is used for transfer learning from the parent to child model. In a similar way, each language can be splitted into phoneme based subwords, and shared vocabulary can be generated to adapt to the NMT model built for the language pairs for which sufficient data is available.

i) **Subword (BPE and Regularization):** To deal with a constrained number of vocabulary and Out-Of-Vocabulary (OOV) tokens, subword tokens will be used. In morphologically rich languages like Indian languages, use of subword units [6], will reduce the total vocabulary size at training and increase the coverage.

j) **Cross-lingual Embeddings:** In case of unsupervised NMT [22], [23], cross-lingual embeddings will help to create a shared space where the same sentence from each language will be represented by similar kinds of vector representations. Monolingual data of Indian languages will be used to train and map the embedding vectors. In mapping [24], vectors of words having similar sense in different languages will be kept closer in multidimensional space.

4. Applications

a) **Health domain:** In the medical field, machine translation systems are useful to translate medial phrases written in reports and prescriptions. ‘Canopy speak’ is one such mobile application which translates medical text from the health domain between English-Spanish language pairs.

b) **Legal domain:** Machine translation in the legal domain translates judgement, orders etc. HEMAT[1] is a machine translation system for translating legal documents between English and Indian languages.

c) **Finance:** In the financial domain, machine translation systems like VERTO are used to translate financial documents, fund sheets, company annual reports etc.

d) **Social media content:** Social media platforms are used by users across the world who speak different languages. Machine translation systems are helpful in translating these social feeds in multiple languages for user convenience.

e) Product Review translation: Product reviews are user generated content which also consists of various lexical and syntactical noises. Translating user reviews in vernacular languages provides valuable information about the products to the users.

f) Subtitle translation: Subtitle translator is useful in translating subtitles from one language to another language for users from different language backgrounds.

g) Tourism: Translating from English-low resource languages will play an important role for promoting tourism.

References

[1] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in Advances in neural information processing systems, 2017, pp. 5998–6008.

[2] J. Gu, Y. Wang, Y. Chen, V. O. K. Li, and K. Cho, "Meta-learning for low-resource neural machine translation," in Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Brussels, Belgium: Association for Computational Linguistics, Oct.-Nov. 2018, pp. 3622–3631. [Online]. Available: https://aclanthology.org/D18-1398

[3] W.-J. Ko, A. El-Kishky, A. Renduchintala, V. Chaudhary, N. Goyal, F. Guzm ́an, P. Fung, P. Koehn, and M. Diab, "Adapting high-resource NMT models to translate low-resource related languages without parallel data."

[4] X. Wei, H. Yu, Y. Hu, R. Weng, L. Xing, and W. Luo, "Uncertainty-aware semantic augmentation for neural machine translation," in Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). Online: Association for Computational Linguistics, Nov. 2020, pp. 2724–2735.

[5] H. Pham, X. Wang, Y. Yang, and G. Neubig, "Meta back-translation," in International Conference on Learning Representations, 2021

[6] R. Sennrich, B. Haddow, and A. Birch, "Neural machine translation of rare words with subword units," in Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Berlin, Germany, August 2016, pp. 1715–1725.

[7] T. Kudo, "Subword regularization: Improving neural network translation models with multiple subword candidates," in Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Melbourne, Australia: Association for Computational Linguistics, Jul. 2018, pp. 66–75.

[8] V. Vaibhav, S. Singh, C. Stewart, and G. Neubig, "Improving robustness of machine translation with synthetic noise," in Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). Minneapolis, Minnesota: Association for Computational Linguistics, Jun. 2019, pp. 1916–1920.

[9] G. Dinu, P. Mathur, M. Federico, and Y. Al-Onaizan, “Training neural machine translation to apply terminology constraints,” in Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy: Association for Computational Linguistics, Jul. 2019, pp. 3063–3068. [Online].

[10] C. Zhu, H. Yu, S. Cheng, and W. Luo, “Language-aware interlingua for multilingual neural machine translation,” in Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Online: Association for Computational Linguistics, Jul. 2020, pp. 1650–1655. [Online].

[11] D. Ruiter, J. van Genabith, and C. Espa ̃na-Bonet, “Self-induced curriculum learning in self-supervised neural machine translation,” in Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). Online: Association for Computational Linguistics, Nov. 2020, pp. 2560–2571.

[12] E. Vanmassenhove, C. Hardmeier, and A. Way, “Getting gender right in neural machine translation,” in Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Brussels, Belgium: Association for Computational Linguistics, Oct.-Nov. 2018, pp. 3003–3008.

[13] S. Ma, D. Zhang, and M. Zhou, “A simple and effective unified encoder for document-level machine translation,” in Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Online: Association for Computational Linguistics, Jul. 2020, pp. 3505–3511

[14] A. Siddhant, A. Bapna, Y. Cao, O. Firat, M. Chen, S. Kudugunta, N. Arivazhagan, and Y. Wu, “Leveraging monolingual data with self-supervision for multilingual neural machine translation,” in Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics.

[15] M. Johnson, M. Schuster, Q. V. Le, M. Krikun, Y. Wu, Z. Chen, N. Thorat, F. Vi ́egas, M. Wattenberg, G. Corrado, M. Hughes, and J. Dean, “Google’s multilingual neural machine translation system: Enabling zero-shot translation,” Transactions of the Association for Computational Linguistics, vol. 5, pp. 339–351, 2017.

[16] M.-T. Luong, C. D. Manning et al., “Stanford neural machine translation systems for spoken language domains,” in Proceedings of the international workshop on spoken language translation, no. IWSLT. Da Nang, Vietnam, 2015.

[17] B. Zoph, D. Yuret, J. May, and K. Knight, “Transfer learning for low-resource neural machine translation,” arXiv preprint arXiv:1604.02201, 2016.

[18] Y. Kim, P. Petrov, P. Petrushkov, S. Khadivi, and H. Ney, “Pivot-based transfer learning for neural machine translation between non-English languages,”arXiv preprint arXiv:1909.09524, 2019.

[19] Y. Leng, X. Tan, T. Qin, X.-Y. Li, and T.-Y. Liu, “Unsupervised pivot translation for distant languages,” arXiv preprint arXiv:1906.02461, 2019.

[20] Y. Chen, Y. Liu, Y. Cheng, and V. O. Li, “A teacher-student framework for zero-resource neural machine translation,” in Proceedings of the 55th Annual Meeting of the Association for

Computational Linguistics (Volume 1: Long Papers). Vancouver, Canada: Association for Computational Linguistics, Jul. 2017, pp. 1925–1935.
[21] C. Amrhein and R. Sennrich, “On romanization for model transfer between scripts in neural machine translation,” arXiv preprint arXiv:2009.14824, 2020.
[22] M. Artetxe, G. Labaka, E. Agirre, and K. Cho, “Unsupervised neural machine translation,” arXiv preprint arXiv:1710.11041, 2017.
[23] G. Lample, A. Conneau, L. Denoyer, and M. Ranzato, “Unsupervised machine translation using monolingual corpora only,” arXiv preprint arXiv:1711.00043, 2017.
[24] A. Conneau, G. Lample, M. Ranzato, L. Denoyer, and H. J ́egou, “Word translation without parallel data,” arXiv preprint arXiv:1710.04087, 2017.

10. Multimodal Summarization Systems

Recent years have witnessed the dramatic increase of multimedia data (including text, image, audio and video), which makes it difficult for users to obtain important information efficiently. Multi-modal summarization (MMS) has gained immense popularity in the last couple of years, due to the increasing availability of multi-modal data on the Internet. It considers inputs in different modalities and produces outputs in multiple modalities. Multimodal outputs are necessary because of the following reasons: 1) It is much easier and faster for users to get critical information from the images 2) According to recent experiments [1], the multimodal output (text + image) increases users’ satisfaction by 12.4% compared to the single-modality output (text) 3) Images help users to grasp events better while texts provide more details related to the events. Thus, the images and text can complement each other, assisting users to gain a more visualized understanding of the events.

The MMS systems have wide range of applications in meeting record summarization, sport video summarization, movie summarization, pictorial storyline summarization, timeline summarization and social multimedia summarization. Videos of meeting recordings, sports events and movies, consist of synchronized voice, visual and captions. The inputs for summarization of pictorial storylines consist of a set of images with text descriptions. But to the best of our knowledge, in none of these application domains, summarization of multimedia data containing asynchronous information about general topics producing multi-modal outputs (text+ image + video clip) was considered.

Through this project, we will propose different novel multi-modal summarization approaches that will generate a multi-modal summary (including abstractive text summary, an image, and a video clip). Initially focus will be given in generating summary in English language but later we will focus on other popular languages like Hindi and Bengali.

11. Multilingual and Multimodal Conversational Systems: Agriculture, Legal Assistance, Health, Tourism and/or Education

This project aims at developing a Multi-lingual Chatbot in English and Indian languages for four important domains, namely Judiciary, Health, Tourism and Education. The Chatbot will be a pluggable and open-source engine, with the capability to accept both text and voice as input. The inputs will be in the following languages: English, Hindi and Bengali. This will also have the facility to accept code-mixed inputs, i.e. Hinglish (mixing of Hindi and English) and Bengalish (mixing of Bengali and English). One distinguishing characteristics of the bot will be to make it affect-aware, i.e. capable of dealing with emotion, sentiment, politeness and personalization.

Conversational Artificial Intelligence, nowadays, is one the most discussed technologies all over in the world. The Chatbot Report 2019 [6]reveals the following: Business Insider experts predict that by 2020, 80% of enterprises will use chatbots; By 2022, banks can automate up to 90% of their customer interaction using Chatbots; According to Opus Research, by 2021, 4.5 billion dollars will be invested in Chatbots. Several applications have emerged, such as SIRI, Cortana, Google Now etc. There is also a tremendous growth in the industry of Conversational Artificial Intelligence as this technology is being explored with top priority by the big five leading Artificial Intelligence (AI) driven companies - Facebook, Google, Microsoft, Apple, and Amazon.

The first Chatbot ELIZA was developed by Joseph Weizenbaum at MIT in the 1960s, and since then there has been a tremendous growth in this technology, with the aim of making it more human-like by incorporating empathy, sentiment, emotion and politeness. We are already observing many Chatbots in many different websites used for various purposes. But, unfortunately none of the available chatbots support Indian languages and are capable of dealing with human empathy.

The practice of law involves developing arguments based on the reasoning of courts in previous instances with similar circumstances. This reasoning, known as precedent, is invoked to justify a legal standpoint or discredit the opposition's arguments. Precedent can be reinforced or dismissed by subsequent court decisions, and is therefore the fundamental point of debate in the application of the law. Consequently, it is imperative for legal practitioners to be well versed in legal precedents relevant to their area of practice and remain up to date in their knowledge of such. The weight attributed to this information, found mainly through the study of previous cases, makes the ability to access the decisions essential. These processes are largely manual, imposing lengthy production delays on the industry. Furthermore, lawyers

[6] (https://chatbotsmagazine.com/chatbot-report-2019-global-trends-and-analysis-a487afec05b)

dedicate a significant portion of their time and energy by reading and analyzing the decisions for the purposes of client representation and developing legal documents. Common citizens suffer much due to their lack of knowledge about the implications of various legal matters, terminologies, and their implications. For Healthcare, the Multilingual Chatbot could be of great assistance, particularly in situations such as COVID-19 pandemic, and to provide other basic health related information. In Tourism, Chatbot can be an effective tool to provide support in many ways, such as choosing the appropriate travel destinations, gathering basic information about the places, availability of hotels, restaurants, easy access to Flights, Railways, and other transportation means etc. Education is the manifestation of mind, and it is the fundamental rights of the citizens. Chatbot in Indian languages would immensely help the educational sectors by providing students to choose their programmes, courses, topic of the lectures as well as for counseling services. Agriculture is a sector where conversational system or Chatbot will surely play an important role in the following ways: farmers will get relevant information on weather condition, crops information, soil condition, market condition in their own languages.

12. Multilingual and Multimodal Sentiment and Emotion Analysis

Sentiment and Emotion analysis are two prominent research in AI, NLP, Computer Vision nowadays. While sentiment focuses on coarse-grained analysis of affects (e.g. positive, negative, neutral), the emotion recognition concerns with fine-grained affect analysis in the form of sad, fear, anger, disgust, surprise, happy etc. Sentiment and Emotion analysis can be performed either at the document level, or at the sentence level or at the aspect level. Multimodal sentiment analysis and emotion recognition concern with the combining information from a variety of sources, such as text, video, image, audio etc. Deep learning based techniques are widely used for developing such systems. It has immense potentials in variety of sectors like e-commerce, security, mental health analysis, building human-like conversational systems etc.

G. Indian Language Mixed-code Voice Assistants for Functional Domains

Many Indian corporate and social enterprises (like Banks, Hospitals and other Health care services, Public Services, Utilities) are looking forward to changing their traditional IVR (Interactive Voice Response) systems to AI-powered Chatbots and voice bots. This shift will help them to have better customer interaction, knowing the customer better, better engagement and service. One of the key technical issues in wide-spread adoption of voice bots in Indian context is lack of mature Automated Speech Recognition (ASR) and comprehension and Text to

Speech (TTS) models – especially for Indian regional languages and mixed-code (e.g., Hindi + English, Tamil + English) conversations. We propose that R&D effort be spent on creating appropriate thesauri, language models, machine and deep learning models to aid such AI-powered virtual assistants in Indian industry context.

H. **Code-mixed Language Models and Applications**

Language Models are models that impart the understanding of a language and its intricacies to a machine. Language models are typically built using statistical significance of words called *Statistical Language Models* (**SLMs**). In the recent past, language models trained using neural networks, called *Neural Language Models* (**NLMs**) have emerged as a major player in the AI domain. With the advent of the Transformer architecture (Google, 2017), NLMs have aided in building exceptionally high accuracy AI systems with language models like BERT (*Google, 2019*), GPT-3 (*Open AI, 2020*) powering them. These models have completely changed the manner in which NLP applications are built, bringing forth a revolution in the NLP-AI space.

The ***accuracy*** of a language model is measured in terms of how *confused* the model is in predicting the *next word*, given a *context*. This metric is called ***Perplexity*** (PPL). A *better language model* produces a *lower perplexity score.* The state-of-the-art language models handle only *single languages*. Code-mixed language models need to be constructed as a building block, in order to develop applications in code-mixed languages. Although language models have evolved over the past decade and have gained significant attention, code-mixed language modeling still remains a sparsely explored domain.

Given the spectrum of multilingual societies across the world, we address the relevant work in code-mix for the top *most spoken languages*, viz. *Mandarin Chinese, Spanish* and *Hindi*. For *Mandarin-English* code-mixed language models, Genta *et. al.* report a perplexity of **127** for *Mandarin* to *English*. For *Spanglish* (*Spanish-English* language pair) code-mixed language models the state-of-the-art is the work by Gonen and Goldberg, who report a perplexity of **40**. For *Hinglish* (*Hindi-English* language pair) there are only a handful of significant contributions, among which the best perplexity is reported by Pratapa *et. al.* as **772.**

Once code-mixed language models are built, a plethora of AI applications can be built from these models. A few of such applications are outlined below:

9. **Chatbot** - Chat interfaces across domains in code-mixed languages
10. **Speech Recognition** - Converting audio signals to code-mixed text
11. **Machine Translation** - Translating code-mixed languages to matrix language
12. **Natural Language Generation** - Generating code-mixed text

13. **Text Summarization** - Creating summary of large body of code-mixed text
14. **Intelligent Personal Assistant** - Virtual assistants like *SIRI*, *Alexa* which can understand and converse in code-mixed language and context
15. **Smart Keyboard** - Code-mixed keyboards typically used for typing on smart devices, equipped with word completion and next word suggestions
16. **Spell/Grammar Checker** - Automatic detection and correction of incorrect spelling and grammar for code-mixed languages

Challenges in code-mixing

In addition to mixing languages at the sentence level, it is also fairly common to find code-mixing behavior at the word level. This linguistic phenomenon poses a great challenge to conventional NLP systems. The major challenges that lie when trying to build models and applications in code-mixed languages are as follows:

Mixed words across languages: As code-mixing has no predefined set of rules, word mixtures are a highly observed occurrence. For example, words in *Hindi* are used with *English* inflections like *darofy* - '*dar*' (fear) + '*fy*' (inflection in *English*)

Mixture of grammar of constituent languages: As mixing of languages is informal in nature, users tend to mix sentence structures of the member languages. For example:

- **Code-mix***: Main khatam karunga job*
- **English translation:** I will finish the job
- **Correct form:** In the above example, the sentence "*I will finish the job*" is written in code-mixed *Hinglish* with the structure of the *English* language. The correct form with *Hindi* grammar would have been: "*Main job khatam karunga*"

Multiple word forms: When languages with native scripts are code-mixed with *English* particularly, the transliteration of words in the native language to *English* varies in form. This happens because of the *unavailability* of a *standard romanized form* of words of such languages. This is observed especially when Indian languages like *Hindi* and *Bengali* are mixed with *English*. For example, the *Hindi* word 'हैं' (English translation: '*is*') in romanized form may have the variations: *hain, hai, hei, hein, he*

Switching points: Switching Points are the tokens in the text, where the language switches. Switching points have rare occurrences in the corpus. Such sparse occurrences of switching points makes it difficult for any Language Model to learn their probabilities and context. Obviously, code-mixed language models fail at switching points. This is the **primary** bottleneck for code-mixed models.

Dataset: As research in the code-mixed domain is very limited, datasets are not available, especially large ones, that can be used to train language models for code-mixed languages.

5.2. Prominent Technologies to focus: *Speech, Video and Text Analytics in Healthcare*

A. Development of Multi-modal Techniques for Pancancer Prognosis Prediction

The high-dimensional nature of cancer-related data makes it hard for physicians to manually interpret these multimodal biomedical data to determine treatment and estimate prognosis. The pancancer analysis of large-scale data consisting of twenty different types of cancers has the potential to improve disease modeling by exploiting these pancancer similarities. The shared representation based on various cancer-related information generated from multiple modalities might help in finding the underlying similarity between various cancer patients. This might be beneficial for the physicians in the decision making of treatment and prognosis of cancer patients.

As we all know that adding multiple modalities to any artificial intelligence-based system generates a more comprehensive description of data and improves the prediction power of the model. But, in real-time scenario some modalities could be missing for some data samples. For instance, survival prediction of cancer patients using multi-modal data (clinical data, mRNA expression data, microRNA expression data and histopathology whole slide images (WSIs)) could benefit the oncologist in their prognosis and diagnosis but only a part of patients contain information from all possible modalities causing the prediction model to fail for these patients.

B. Speech, Video and Text Analytics for Smart Healthcare Systems

From physical devices to smart systems powering medical devices, new technological advances are helping doctors and patients connect in new ways, transmit vital data in real time, and identify and treat life-threatening events faster than ever before. The vision of "anywhere, anytime healthcare" is changing consumer expectations and fueling the next wave of innovation growth. In today's smartphone age, more and more consumers are getting comfortable with the idea of video consultations with their physician, remote monitoring via health apps, and using personalized diagnostic tools in smartphones as a ready reckoner.

We have heard about Internet of Things (IoT) implementation in medical devices, but mostly in the diagnostics area. However, IoT devices are also managing the sudden rush to user-centric environments for growing applications in self-monitoring, rather than being available in hospitals and offices alone. This goes hand-in-hand with the concept of tele-healthcare with wireless monitoring services. The main benefit of IoT for patients is convenience and quick access to vital information to avoid emergency situations (The time to conduct vitals at home vs. finding the time to go to a doctor). People are more willing and likely to take control of and monitor their health if they feel it is easy, convenient and fits into their busy schedule.

The recent development of big data-oriented wireless technologies in terms of emerging 5G, edge computing, interconnected devices of the IoT and data analytics have enabled healthcare services for a happier and healthier life. Although, the quality of the healthcare services can be enhanced through big data-oriented wireless technologies, however, the challenges remain for not considering emotional care, especially for children, elderly, and mentally ill people. Based on the above problem, we see that there is a need of emotion-aware healthcare framework, where the emotion will be an indicator of the health situation and the satisfaction of the patient or the doctor regarding the healthcare service.

The popular healthcare services in 5G include remote diagnosis and intervention and long-term monitoring for chronic diseases. The emerging applications include robot-assisted remote patient care, care services empowered by AR, automation and optimization of hospital logistics, remote surgery, etc.

5G enables a large number of devices to be connected via various protocols. The obvious benefit of 5G is its low latency and ability to transmit large data sets, such as image data in remote surgery or assisting patients with disability via robot where high-quality pictures may need to be transmitted.

Telemedicine continues to shift from the edge of healthcare to the mainstream. With 5G wireless networking, it is becoming more possible now. Telemedicine in such conditions when patients need immediate care but no doctor is available can be life saver. In all these applications, speech, video and text analytics play a vital role to make these healthcare systems more efficient, automated and reliable.

C. Speech, Video and Data Analytics in Healthcare

In this project, we will develop a smart configurable healthcare system for neonatal monitoring. The major problems in these modules are as follows:

(a) Development of contact based/contactless (camera based) vital sign monitoring module: The system can be configured to the contact-based monitoring mode in case of poor light condition, camera position, different sleep position, infant with parents, and the non-contact–based monitoring mode in case of loose electrode condition and very delicate skin of the infants.

(b) Development of cry detection and pain analysis module for neonates: Cry detection and pattern analysis module will be developed for detecting the neonatal cry and pain analysis non-contact microphone sensors

(c) Development of event-triggered, signal quality-aware Internet of Things (IoT)-enabled physiological telemetry module: This module will transmit clinical along with the recorded physiological signals and visual images via different wireless mediums subject to any event detection.

D. Edge-AI based Social-Distance Tracker IoT-Camera

The drug discovery for SARS-CoV-2 is still on research labs, and the growth of the infection rate of the COVID19 is showing exponential. Therefore, to break the transmission rate of the COVID19, the most favorable approach suggested by WHO is maintaining social distance. Most of the countries are applying this approach to slow down the spread of this disease. It is challenging to maintain social distancing in an overpopulated country like India. The protocol of social distancing can be violated in the open market, supermarket, and other congested places where huge people come to buy essential goods. It is quietly unmanageable to track every person whether they are violating the social distancing. Disobeying this protocol can lead to community spread of the disease. So in this project, we have figure out the following problems that we are going to solve using EdgeAI and Computer Vision.

A. How to track a mob violating social distancing.
B. Using video analytics and real-time image data, suggest people a suitable time to go to the market.

E. **Multimedia Lifelog: Foodlog**

This research work aiming is to do develop a foodlog website and also application software for calorie identification in order to dietary control which has its social impact for development, and make a system called FoodLog. This value for users lies in personal enjoyment, in supervision their health, in making a social contribution, depending on how they choose to use it. Being able to generate such additional applications may be a key factor in encouraging users to change their lifestyles. We are focusing on analysis of trend estimation between users and individuals and image recognition using large scale data. Our aim is to keep

Food Record for Health Management using multimedia technology. FoodLog: An Easy Way to Record and Archive What We Eat. An image processing engine analyzes the content ofthe meals, divide these into different meal category based on calorie value contained. Next is to determine what food types appear in the picture and how they fit into the dietary balance. It then estimates the dietary balance values which helps us to monitor our health.

Outcomes:

- Datasets (Foodlog for training and testing) for dietary control using calorie identification and verification
- Selection of appropriate computer vision technique and deep learning based techniques for above said applications.
- The algorithmic development in running coded form for these (above said) applications
- Foodlog website and application software for calorie identification using foodlog image.

F. Breast cancer detection and classification from tomosynthesis dataset.

Mammography is the most popular technology used for early detection of breast cancer. Manual classification of mammogram data is a difficult task. We aim to develop a CNN structure for tomosythesisdata. We plan to collect the 3D tomosynthesis data from hospitals and use different artificial intelligence and deep learning techniques for classifying them into benign and malignant. We also aim to work on different machine learning and deep learning techniques to locate the tumour, if present, in the tomosynthesis images. For that we need to have an organized dataset. We also aim to contribute in a state of art dataset for researchers working in this field.

G. Multi-modal AI for Telehealth

AI based telehealth systems (moving beyond Telemedicine) is destined to be a part of better healthcare delivery mechanism – especially in post COVID context. This is especially visible in areas of telehealth innovations where AI applications are used to support, supplement or develop new remote healthcare models and increase access to millions. According to WHO's eHealth observatory survey, AI in the telemedicine field is directly supplementing innovations in these areas: Tele-radiology, Tele-pathology, Tele-dermatology, and Tele-psychiatry.

We propose to create a framework for a multi-modal AI system (Text, image and video, voice and sensor based) to augment the telehealth capability for leading Indian hospitals. This has go beyond remote patient monitoring to provide truly intelligent and interactive healthcare intervention, assistive guidance and alerts.

5.3. Other Problems

Below we discuss some of the other projects that may be taken up, and highlight their novelty aspects.

1. IITP1: Deep Learning based Models for Leveraging Data from Heterogeneous Sources for Improved Traffic Prediction

Technology: In this project we use both software and hardware technologies. For data processing use mostly deep learning-based technologies. For data collection we use different kinds of sensors and GPS. Here are some of the main technologies that will be heavily using in the project:

a) Cameras, GPS devices installed in vehicles would provide individual vehicular data.
b) Traffic sensors (cameras, sound sensors etc) would provide traffic data.
c) Time series analysis for analysis of periodicity and pattern and detection of anomaly..
d) Hybrid neural network models (eg ConvLSTM with attention GCN, featured with Kalman Filter or DMD) for traffic demand prediction and anomalous behavior.

Novelty: In the context of the public transportation system, prediction of traffic demand plays a crucial role in a smart city traffic network. Traffic prediction not only provides the administrative authorities vital cues necessary for better management of transport resources but can also help in better preparation to meet a sudden increase in traffic demands. The knowledge about intelligent transport management gained from this project will have a great impact on the field of urban planning and management and specifically in smart city development. We shall have a better traffic prediction model which leverages the social network and social media inputs.

2. IITP2: A Multilingual Chabot for Legal Assistance, Health Care, Education and Tourism

Technology: The project aims at developing a multilingual Chatbot in Judiciary, Legal, Education, Healthcare and Tourism domains using Natural Language Processing and Deep Learning techniques.

The Chatbot will be a pluggable and open-source engine, with the capability to accept both text and voice as input. The inputs will be in the following languages: English, Hindi and

Bengali. This will also have the facility to accept code-mixed inputs, i.e. Hinglish (mixing of Hindi and English) and Bengalish (mixing of Bengali and English). One distinguishing characteristics of the Chatbot will be to make it affect-aware, i.e. capable of dealing with emotion, sentiment, politeness and personalization.

The key technologies involved in this project are:

(i). Automatic Speech Recognizer (ASR) in English and Indian languages

(ii). Designing crawlers for data collection, and pre-processing tools for data annotation

(iii). Adapting appropriate Chatbot framework for implementation

(iv). Deep Learning based Natural Language Understanding (NLU), Dialogue Management (DM) and Natural Language Generation (NLG) components

(v). Deep Learning based models for personalization and empathetic dialogue generation

Novelty of the Project: The project aims at developing an open-source, pluggable Chatbot engine supporting Indian languages and code-mixed English-Indian languages for five important technology verticals, namely Judiciary, Education, Health, Tourism and Railways.

Most of the existing Chatbots can converse in English language only. In a multilingual country like India, there is a necessity to provide this Chatbot service in Indian languages to reach to the common citizens. This Chatbot will provide the citizens the information related to healthcare, judiciary, tourism, railways and education.

The novelty of our proposed Chatbot lies in the following: (i). We provide a Chatbot service in multiple languages like English, Bengali, and Hindi; (ii). The Chatbot service will be enabled to deal with the code-mixed environment, i.e. to operate with mixed English-Hindi, English-Bengali languages; (iii). The Chatbot will be enabled with the empathetic attributes, i.e. it will understand human's sentiment, emotion etc. during conversation.; and (iv). Different modules of the Chatbot will be implemented using the recent deep learning and natural language processing techniques.

3. **IITP3:** Multimodal Abstractive Summarization Systems

Technology: As a baseline for the proposed model, the proposed approach of Zhu et al., 2018 [18] is used. A hierarchical recurrent neural network (RNN) will be used to encode the documents and speech transcriptions, an image encoder to encode the image set (comprising of images and key-frames extracted from the video footage). A global multi-modal attention mechanism will be used to guide the decoder to generate the summary. A visual coverage mechanism to assist the model in choosing the most relevant image, $I_{best,}$ and video V_{best} will also be proposed.

Text Encoder

The input text is of the form M_{text} : { D_1, D_2 , ..., $D_{|M\ text|}$}.
M_{text} is a multi-document data, where each document is defined as
D_i : { s_1, s_2, ..., $s_{|D_i|}$ } ; and each sentence is defined as s_i: {$x_{i,1}$, $x_{i,2}$, ..., $x_{i,|s_i|}$}.
Here $x_{i,j}$ is the word embedding of jth word in the sentence s_i.
Each sentence will be first tokenized with a start of sentence <sos> and an end of sentence <eos> token before feeding it to the encoder. A bi-directional LSTM/transformer based models/BERT models will be used to encode the sentence. The document will be encoded in a similar manner using these encoded sentences as the input to the hierarchical encoder.

Image Encoder: We proposed to use a Region Proposal Network (RPN) to extract low-level visual features from given images. In particular, Faster R-CNN which consists of two networks will be used for the task: RPN for generating region proposals and a network using these proposals to differentiate background with the foreground inside the image. We propose to use pre-trained CNN model namely Inception V3/VGG19 to extract high-level semantic features from image.

Audio and Video Encoder: An Automatic Speech Recognition (ASR) system will be used to perform the speech transcription of the audio signals to form pseudo-text documents (as it might contain some noise due to errors in the ASR systems). These pseudo documents will be encoded using the same hierarchical text encoder used to encode the text documents.In general videos are converted into a single vector. Several different encoding policies will be followed for this purpose. But here in order to capture local information, the video will be divided into small parts. The smaller video parts will be encoded separately to determine local objects. Finally local guided global embedding model will be developed for video encoding.

Decoder: Hierarchical LSTM based models will be developed for generating the summary. Later on some transformer based models coupled with reinforcement learning based reward models will also be utilized for generating the abstractive summary.

Attention: An extension of the global attention mechanism proposed by Bahdanau et al., 2015 [1] will be used in our multi-modal model. First uni-modal attention systems will be used to generate a modal specific context vector. Using these context vectors, a multi-modal context vector will be computed as a weighted sum over all of the datasets.

Coverage: The concept of coverage was first introduced by See et al., 2017 [18], where they used it on a Pointer-Generator Network to prevent the model from over attending a particular section of the document. Zhu et al., 2018 [19] expanded the coverage mechanism by introducing the concept of visual coverage that helps in preventing the model from over attending a particular image or a part of an image. Zhu et al., 2018 [19] also introduced the visual coverage score that guides the model to choose the most relevant image at the last time step. A similar concept will be used in this work to choose the most suitable video clip as an addition to our summary. For scoring a video for its salience, a normalized score for all the key-frames belonging to particular video footage is proposed. The clip with the best score will be chosen as the best video.

Microblog summarization system: The developed MMS systems will also be applied on Twitter data which may also consist of text and images. Techniques will be developed for handling code-mixed tweets.

Online micro-blog summarization system: As twitter data is online in nature, the developed summarization system has to be converted online. Whenever a new tweet arrives in the system, the existing summary will be updated by adding the new tweet in the summary.

Novelty: To the best of our knowledge, there are very few works in the field of multimodal summarization where inputs and outputs will be multimodal in nature. In recent years there are few works in English language, but we do not find any such works in Indian languages. The novelty of the current project lies in the following:

1) No data set exists for multi-modal summarization with multimodal inputs and multi-modal outputs. We would like to create a standard data set where inputs will be in the form: text, images, videos and outputs will be: summarized text (extractive and abstractive), a few images (summarizing the event) and a video clip (summarizing the event). Initially the data set will be created for English but later on we would also like to develop data sets in popular Indian languages like Hindi and Bengali.
2) Development of some frameworks using deep-learning to solve the multi-modal summarization problem producing multi-modal summary (including abstractive text summary, an image, and a video clip).
3) Development of some multi-modal micro-blog summarization system as multi-modal data is commonly available in social media.

4) Development of some online multi-modal summarization system: the social media data keeps on changing over time. Thus the summarization system developed on a static collection of data has to be updated with the arrival of new social media data.
5) Applications of the developed MMS systems in solving meeting record summarization, sport video summarization, movie summarization, pictorial storyline summarization, timeline summarization and social multimedia summarization problems.

4. IITP4: Development of Multi-modal Techniques for pancancer prognosis prediction

Technology: To generate shared representation, we propose to use unsupervised learning by optimizing various similarity measures like (cosine similarity, Euclidian distance, etc. between shared representation and different modalities of a particular sample). To generate missing samples of some modalities using shared representation, we propose to use advanced deep learning architectures based on encoder-decoder framework. The encoder might be consisting of deep neural networks or convolutional neural networks depending on the data modalities. For each modality, we propose to optimize separate encoding function which will generate a compact feature space for that modality. These compact features along with the shared representation are then passed to the decoder to generate the missing samples. We again regenerate the more representative and informative updated shared representation using newly generated missing samples. The complete dataset (newly generated missing samples + already available data) can be used by various multi-modal systems for prediction and classification tasks. We can also use the shared representation features for the classification. The models for classification can vary from simple deep neural networks to advanced deep neural networks. The stacked based ensemble frameworks can also be utilized for the classification task.

Novelty:

The shared representation of multi-modal data is generated using unsupervised learning in such a manner that the shared feature vector is similar to all the modalities of a particular sample and dissimilar to all the modalities of another sample. The main objective of unsupervised learning is to increase the similarity between shared representation and all the modalities for a particular sample and increase the dissimilarity between shared representation and all the modalities of another sample. Recent advancements of deep learning like deep highway networks, convolutional neural networks, attention-based deep neural networks etc. can be used to generate the shared representation in an encoder-decoder framework.

To generate the missing samples of some modalities using shared representation, we can proceed with the novel architecture "Adversarial Incomplete Multi-view Clustering". This architecture proved its significance in the clustering of multi-modal data consisting of videos, texts, and images from Reuters, BDGP and Youtube. It uses the encoder-decoder network to generate the shared representation and missing data. Once the missing data is generated, it uses the newly generated data and already available data to regenerate the updated shared representation.

The updated shared representation or the newly generated missing data along with already available data can be feed to any advanced deep learning-based model for cancer survival prediction or other classification tasks.

5. **IITP5:** Speech, Video and Text Analytics for Smart Healthcare Systems

Technology: The connected healthcare along with the connected devices and sensors generates a massive amount of datasets in terms of volume, variety, and velocity. Due to the massiveness, complexity, and multidimensionality of this wireless big data generated from the connected healthcare, there is a scope to apply machine learning algorithms for data computation before communication to the server. 5G introduces new concepts, such as network slicing to better support various applications with different performance requirements on data rate and latency, and edge and cloud computing that will be responsible for the leverage of computational requirements.

5G also introduces intelligence at the edges of the network, such as software-defined networking (SDN), network function virtualization (NFV) and multi-access edge computing (MEC). The next-generation big data aware wireless technology like the emerging 5G coupled with IOT has a tremendous potential to alleviate the challenge with a very fast response time, improved resource utilization along with the best accuracy. Health big data comes from various sources and sometimes these sources might be large repositories of data which have to be brought into a common platform for a unified analysis.

The aggregation challenge is related to high volume and variety of data that needs to be brought together from divergent data warehouses and real-time data. The solution can be use of high-speed file transfer technologies. We can use Burrows-wheeler transform (BWT) to compress DNA sequences. The BWT is a string compression algorithm that compresses the data by grouping similar characteristics in a series of strings. Map Reduce programming model and its Hadoop implementation provide robust analytical tools to do the analysis. Medical problems are usually complex and one learning algorithm might fail to yield an informative result. Hence, we can combine the learning techniques into a more effective and robust learning technique better than the constituent algorithms. An ensemble is a collection of various learning algorithm working together in parallel or in sequence to produce better results. Ensemble learning has shown to provide very efficient and favorable outcome compared to individual learning systems as it usually perceived as a process of consulting multiple experts before deciding. Bagging and Boosting is the most popular ensemble learning technique. Spark is a heavily used platform for healthcare big data analytics. It leverages its stream computing capabilities to perform faster analysis without the need to use other supportive frameworks.

Novelty:

We are going to apply Speech, Video and Text analytics for smart healthcare systems in order to make healthcare more reliable and advanced as healthcare is a data-intensive field. Health data has become ubiquitous because of improvements to recording systems in healthcare, the participation of patients in their treatment using social networks. Hence, the field of data analytics or Big data and computational intelligence promises a bright prospect in building a smart healthcare system.

5G Networks are able to provide diversified services of different performance needs, support co-existent accesses of multiple standards, such as 5G, LTE, and Wi-Fi, and coordinate different site sizes, including macro, micro and Pico base stations.

a. In short video clips, Convolutional networks are used to capture visual cue from face images, and a deep belief network is used to extract information from the audio stream.
b. Exploring and processing information in real time using AI techniques require a high level of computation power, therefore, a high level of computation ability is needed for the local nodes.
c. Implementation of onsite data analysis capability in the sensor node itself i.e. compression of raw data into features or decisions of much smaller volume to compensate the Bandwidth burden on the back-end servers.
d. Implementation of the Edge computing into the 5G architecture. Mobile Edge computing (MEC) offers cloud computing capabilities and an IT service environment at the edge of the network. It allows software applications to tap into the local contents and real-time information about the local-access network conditions.

e. Implementation of an offloading framework to solve the mobility of the connected devices.
f. Implementation of an emotion recognition system in the 5G emotion-aware healthcare framework using input modalities as Speech, Video and Text.

6. **IITP6:** Deep Learning Audio Signal Processing

Technology: The deep learning can be applied in any field for improved performance without the requirement of handcrafted features. The application chosen is audio signal processing for automatic audio recognition, environmental sound detection, localization and tracking, source separation, and audio enhancement

Novelty: Selection of appropriate features or raw audio waveforms, and appropriate deep learning models for audio applications such as blind source separation, audio enhancement, sound detection, and localization.

7. **IITP7:** Embedded System Design for Speech, Video and Text

Technology: For development of such solutions there are two technology choices are available (a). Field Programmable Gate Arrays (FPGAs), and (b). Application Specific Integrated Circuits (ASICs). The FPGA provides quick hardware implementation of algorithms for video, speech and text, however, there is less room for performance optimizations of these algorithms at the same time is cost effective. The ASIC can be realized with implementation state-of-the-art technology available in the market and optimized for the best performance in terms of latency and energy requirements. However, development of ASIC requires more time and expensive at the early stage.

Novelty: Till date there is no such specialized integrated circuits are available, most of them use the standard processors or ICs for processing of video, speech, and text. Even Facebook, Amazon and Google are desperately working in this direction with Intel and TSMC (Taiwan Semiconductor Manufacturing Company).

8. **IITP8:** Physical Layer and Cryptographic Security for Video and Speech

Technology:

The individual security mechanism such as PLS or cryptographic security has been studied, but how to double-layer authentication secure the video and speech communication over wireless medium in the presence of eavesdropper never been considered before, it is because both the security mechanisms are in different layers, i.e., the PLS is present in the physical layer and the cryptographic security is present in the application layer.

How the quality of the video will be maintained for the legitimate user in the presence of security threat has not been studied before. It needs a proper source and channel coding at the transmitter side in the presence of eavesdroppers.

Novelty:

Till date, the security of video and speech mainly relies on the cryptographic security key, but with the advanced technology of quantum computing, the key can be obtained in a fraction of time. Thus need an alternative solution for securing video and speech data. In this project, we are looking for proposing advance PLS and cryptographic security mechanism and their combined one to improve the overall security and reduce the complexity of the security mechanism. A complete study on BER performance, source coding and channel coding will be provided in the presence of eavesdropper and considering overall security. A real-time testbed will be developed to see the performance of the proposed double-layer authentication of video and speech data.

9. **IITP9:** Remote Monitoring and Maintenance of Micro Cyber Physical Machine Tools

Technology: In this project, intelligent predictive analytics integrated with communication technologies in conjunction with the physical CNC machines are going to remotely monitor and control the micro machining process for achieving sustainable advancement of the current technology, i.e. achieving smart micromachining system with wide accessibility and more adaptability. The smart micromachining system will be consisting of (1) CNC micro machine (2) data acquisition system (3) advanced connectivity that ensures real-time data streamlining from the physical space to cyber space and feedback from the cyber space (4) Development of maintenance planning methodology and software tool (5) intelligent big data analytic tools that constructs the cyber space (6) smart human machine interfaces.

Novelty: In the recent past, the advancement of Information and Communication Technologies (ICT) has facilitated the implementation of advanced sensors, data collection equipment, wireless communication devices and remote computing solutions. Integration of ICT with the physical machinery, which is called Cyber Physical System (CPS), is transforming the industry into the next level of industrial revolution, frequently noted as industry 4.0 [20]. Such technologies, along with the advances in predictive analytics are capable of remote monitoring and maintenance of globally integrated manufacturing facilities by converting big data into desired information and knowledge to avoid the costly failures and unplanned downtime of machinery [21].

Recently, a CPS architecture consists of Smart Connection, Data-to-info Conversion, Cyber, Cognition and Configuration levels was attempted to remotely monitoring and optimizing CNC sawing machines adaptively [22]. However, the developed architecture is in incipient stage therefore scope of advancement in all levels of the architecture is present. An ICT based approach for human machine interaction was made to trigger test routines of the machine to know about its condition for field service support [23]. A fog computational framework that utilizes wireless sensor networks, cloud computing, and machine learning has been proposed for remote real-time monitoring of tool wear in CNC milling machines [24]. However, significant advancements in the areas of predictive analytics, software portability, computing scalability, infrastructure flexibility, and cyber security are needed to realize cyber manufacturing.

Nowadays, the miniaturization of many consumer products is extending the use of micro-machining operations with high-quality requirements. However, preventing tool failure and reducing the impacts of cutting tool wear on part dimensions and surface integrity for minimum production cost and improvement of product quality are challenging. In fact, industrial practices usually set conservative cutting parameters and early tool replacement policies in order to minimize the impact of tool wear on part quality [25]. However, uncertain tool failure due to size effects in micromachining processes even makes the conventional practice unsuitable and development of a reliable tool wear monitoring is inevitable in such scenario. Different tool wear monitoring techniques using sensors signals, signal processing tools and artificial intelligent tools have been proposed in recent past [26-29]. An on-line adaptive control optimization (ACO) system to adapt the cutting conditions to reach a minimum production cost and considering the real cutting-tool wear state for micro-milling was also proposed [25]. Most of the monitoring and controlling strategies in micromachining reported in the literature are applied locally and are individual machine centric. Aiming to achieving higher level of intelligent remote monitoring and adaptive controlling of in micromachining, i.e., smart micromachining, CNC micro machine tools must be integrated with cyber space through applications of ICT tools. However, remote monitoring and prediction of tool failure and replacement will be challenging considering the high spindle speed, miniature tool and uncertainty of the tool wear behaviour in micromachining.

For long term perspective the behaviour of not only the tool but the machine itself is also important. In [30] a mathematical model is described that takes into account the reliability and maintainability related fraction of the life cycle cost of machine tools in order to compare scenarios in which different maintenance strategies and policies are applied on the components. Both scheduled (preventive maintenance, inspection-based maintenance) and unscheduled (corrective maintenance, condition monitoring) strategies are considered. On the basis of the proposed analytical formulae optimisation is carried out with a view to find the best solution having the lowest life cycle cost.

10. IITP10: Decentralized Real-time Video Analytics for Robotic Swarm Border Surveillance System

Technology: The key phases involved in the proposal are:

(1) Development of Swarm of heterogeneous robots.
(2) Inter-robot and Control Unit Communication
(3) Real-time Video Analytics for Intelligence Gathering
(4) Video-based SLAM and Motion Planning
(5) Blockchain-powered decentralized platform development

(1)Development of Heterogeneous Robotic Swarm

This phase concentrates on the development of Swarm of heterogeneous robots consisting of unmanned ground robots and aerial robots. This involves both hardware and software components:

Hardware:

We plan to develop a physical testing platform comprising of drones and unmanned ground vehicles. For the purpose of the experiments to be conducted in this experiment, four small sized quadrotor-based drones will be used. In addition, small-sized four 4WS wheeled ground robots will also be used. GPS receivers, inertial measurement units, cameras, communication interface, and board computers will be installed on the drones and ground robots. A waypoint tracking controller will be implemented on each hardware platform. The entire robotic swarm will be integrated together on Robot Operating System (ROS) platform.

Software:

We consider Robot Operating System (ROS) for our robotic software implementation. ROS provides common robot-specific services and libraries, such as component communication, hardware abstraction, and low level device control, and allows one to choose from a set of popular programming languages and to customize core libraries to modify architectural parameters such as incoming and outgoing queue sizes, maximum time to wait for incoming messages, rate of publishing, etc.

The development of a swarm of heterogamous robotic systems involves complex hardware devices equipped with sensors and computational units which are often controlled by complex distributed software. Simulation of this system leads to a rapid prototype development and plays an important role for testing various robotic software components, robot behavior and control algorithms in different surrounding environments. To this aim, we leverage various features, such as ROS actions, services, etc, supported by ROS and we shall use Gazebo simulator with ROS plugin.

Even though there exists several software platforms for simulation, the primary reason to use Gazebo is that simulation using it together with ROS's RVis library helps to create simulation model which can directly be deployed to the real robot hardware. In fact, ROS plugin in Gazebo provides necessary interfaces to simulate a robot using ROS messages, services and dynamic reconfigure.

(2)Inter-robot and Control Unit Communication

This phase aims to establish secure light-weght communication between the nodes in the system. Intuitively, the system is distributed in nature and comprises a number of light-weght (heterogeneous) robots and a number of computation-intensive server-units in a peer-to-peer fashion. As our sensor data is in the form of video, the main challenges lie here are (2) real-time collection of this multimedia data through a number of sensors integrated into the robots, (2) Identification of on-board and off-board data processing in order to find a trade-off between communication overhead and light-weight onboard computation requirements, and (3) secure energy-efficient light-wight communication between nodes (robots and servers) in the system. In order to make the system compatible with the underlying blockchain-powered platform, we use asymmetric key cryptographic schemes to achieve security in data transmission.

(3) Real-time Video Analytics and Intelligence Gathering

This is one of the most crucial phases which involves automatic detection of the interesting objects in the monitored area, track their motion and automatically take appropriate action like alerting defense personnel. This is achieved by performing real time processing of video which are captured by the swarm of robots. As we already mentioned that the aerial robots and ground robots work collaboratively and often complement each other's task, we consider the deployment of few computation-intensive robots (we call them master-nodes), along with light weight robotics units, as part of swarm. This forms a decentralised architecture of the system to be well-suited with blockchain technology (discussed later). The light weight robots, while in action, continuously capture videos and communicate them to the nearby master node. The primary tasks of these master-nodes are to perform real-time analysis and to extract high-level information identified with alert situations which would be sent to the defense personnel. Moreover, master-nodes are also responsible to control the movement of light-nodes based on the information extraction, in order to capture complete geographical area of interest. The major challenge here would be: what to extract from video to assist all these decisions. and how efficiently we can do it. In general, the master-nodes analyze the video content by separating the foreground from the background, detecting and tracking the objects and performing high-level analysis. The high-level scenario provides results whether the situation is abnormal or not so as to assist the human supervisor.

(4)Secure Collective Decision Making

Once robot swarms will exit the research labs and operate in real-world missions, they will face situations in which some of the robots in the swarm may behave differently. For example, harsh environmental conditions might cause individual robots to fail, or terrorists might take control of some of the robots and make them behave in misleading ways. Moreover, there is need to achieve consensus in automated decision making process by robotic swarms, without any human intervention.

As a solution, we use blockchain technology which provides a decentralized computation and information sharing platform for robotic swarm, enabling them to cooperate, coordinate and collaborate in a rational decision making process, without trusting each other.

In general, blockchain technology builds a trusted system and enables secure, transparent, and immutable recordkeeping to be maintained by multiple non-trusting members connected via a decentralized peer-to-peer network. Such distributed ledger provides a powerful mean to verify records, without any trusted intermediary such as brokers, agents, etc. The support of smart contracts, a set of turing complete programs, which run on blockchain network is one of the prime reasons behind the major success of the blockchain technology today.

Research on blockchains has gained significant momentum with a wide range of applications spanning cryptocurrencies, supply chain, health care, IoT security, and many others. Recently we are witnessing its footstep in robotic technology as well.

Swarm of robots always form a system which is inherently distributed. This decentralized nature of swarm robotics makes it compatible to combine with blockchain technology and allows it to meet the following attributes crucial to this mission critical system: reliability, availability, safety, and security. Moreover, this also makes the system to scale if necessary.

Novelty:

Drone-based surveillance is a common method, however, it lacks the ability of visibility in the presence of dense vegetation, close inspection, and manipulation. Imparting heterogeneity by augmenting ground robots in a swarm of drones will add the capabilities of close inspection and manipulation. A heterogeneous swarm system can be used for acquiring and processing real time videos using the onboard cameras mounted on the drones and unmanned ground robots, aiming to achieve all-round deep coverage of the area under surveillance. By all-round deep coverage we mean that the birds' eye view will be acquired by the drones while the view beneath the dense canopies of forests will be acquired by the ground robots and subsequently combined together for providing a comprehensive visualization to the defense personnel. In addition, the use of this surveillance system demands to ensure critical systems properties, such as reliability, security, safety and availability. Blockchain Technology has already proved its immense potential to meet all these challenges. Although significant research has been done in other application areas, we observed very few attempts in the context of robotics. To the best of our knowledge, this would be the first attempt to incorporate blockchain technology into heterogeneous swarm systems designed for real time video analytics.

11. **IITP11:** Hyperspectral Video Processing Assisted Automated Segregation of Recyclables from Solid Waste Streams in a Smart City

Technology: We plan to employ hyperspectral imaging for the segregation of recyclable objects from solid waste stream. Conventionally, RGB imaging has been attempted to solve this problem. However, many dissimilar recyclable objects present in the solid waste stream can have similar color rendering them unrecognizable by the RGB camera-based techniques. More recently, thermal imaging technique has been used with some success. However, the presence of semisolid material in solid waste may lead to failure in a thermal imaging segregation system. The spectral range of hyperspectral sensor is 900-1700 nm which can resolve materials with higher accuracy and can address the issue related to the presence of semisolid materials in solid waste stream. Although promising in terms of spectral range and resolution, hyperspectral technique presents computational challenges related to processing of large volume of data in real-time and defining features pertaining to hyperspectral videos. The proposed project will address the technical challenges to incorporate hyperspectral videos processing in waste segregation application.

Novelty:

At present manual segregation for sorting of recyclable objects from source segregated MSW and e-waste is a time consuming and inefficient operation. The novelty of the proposed system lies in the use of a combination of hyperspectral and thermal imaging based technique integrated with robotic manipulation for automated multi-material classification and segregation of recyclables to avoid human drudgery. We plan to develop a classification-cum-sorting framework employing hyperspectral and thermal imaging sensor for multi-material classification of recyclables from MSW into broad categories of materials for recycling.

To the best of our knowledge, there has been no work reported specifically in the area of multi-material classification using thermal imaging for sorting recyclables from MSW stream and e-waste using a robotic system. In the past, a multi-robot system for sorting construction and demolition waste (C&D) was developed that comprised of a manipulator, a conveyor, multiple sensing elements like 3D sensor, metal detector, RGB camera and near-infrared camera (NIR) for sorting C&D waste. Various machine learning approaches such as reinforcement learning, SURF (speeded up robust feature), SIFT (scale invariance feature transform) and PCA-SIFT (principal component analysis-SIFT) applied on RGB images for material classification and grasping have been reported. To separate out metals from waste materials, automated waste segregation has been developed using PLCs [Dudhal et al., 2014] [31]. Another system based on programmable logic controller has been proposed to segregate and recover metals from the scrap [Sharmila et al., 2013] [32]. An intelligent and low-cost programmable robot for color identification and sorting has also been developed based on color recognition algorithm to detect the color of the object and commanding robotic arm to pick and place it to its respective bin [Manjunatha V G., 2014] [33]. Most of the existing works employ visual object sorting. The main limitations of visual object sorting are the sensitivity to the variations in ambient illumination and lack of material-specific information in the RGB images. To the best of the knowledge of investigators of this project, there has been no work reported specifically in the area of multi-material classification using a combination of hyperspectral and thermal imaging for sorting recyclables from solid waste stream using a robotic system in India. The results of the proposed project can provide an efficient tool for automated segregation of recyclable materials from solid waste and thereby benefit municipal authorities and urban local bodies (ULBs), waste management industries, and formal recyclers.

12. **IITP12**: Modern Application of Audio and Visual Sensing in Structural Health Monitoring

Technology: The sensing technologies may involve infrasound acoustic emission measurement from microphone, vibrometer using Doppler effect mounted on unarmed aerial system, video imaging by motion magnification or image processing, optical flow methods, interferometry. It will also involve software for modelling, alarming systems. The total system will be designed based on internet of things. To make the system automated or semi-automated, highly advanced technologies related to artificial intelligence, signal processing, image sensing and for decision making machine learning may also be used.

Novelty:

As per the state- of the art in the field of structural health monitoring, there is very limited application of audio-video sensing network. In last couple of decades, these kind of sensor network system has become very popular in aerospace industry, there is a narrow of almost no application in vibration-based structural damage detection and monitoring. Novelty of the proposal lies in collecting real data from various structures and their analysis in the laboratory environment.

13. IITP13: AI based Water Level Prediction

Technology: We will develop a flood forecasting and early warning system using AI techniques, which will be a web based module and an android app will also be developed for smartphones. Several water level sensors will be deployed on the strategic locations along the river sources. An android app will also be developed so that the information of the water level in the river for the coming two days will be available on an android based smartphone of any user. App would be able to process natural language for various user commands.

Novelty:

Bihar faces an acute flood problem with almost 76% of the population in North Bihar living under zones that are vulnerable to recurrent flood. The 2019 floods were severe and is estimated that it was the worse flood after 1975. Livestock and assets worth millions of dollars along with human lives is lost year after year in Bihar floods.

14. **IITP 14:** Semi-Automated Preliminary Health Assessment of Structures through video processing and deep learning algorithm

Technology: The developed technique and product will facilitate rapid preliminary health assessment for old structures/heritage sites or any structure which has been exposed to natural calamities such as earthquake, floods etc. Also with the aid of semi-automated drones, physically inaccessible sites could be assessed. Further, with the help of deep learning techniques, the exact location of the cracks will be identified, which can be shared with experts for further analysis and categorization of damages. Convolutional Neural Network (CNN), a deep learning algorithm will be used for the processing of images. This method will be used to assign importance to various aspects/objects in the image and be able to differentiate one from the other. Conventionally binary classifiers are used to segregate damage and no-damage images, which consisted of three convolutional layers. The major drawback of this model is that it required a very large database to obtain high predictive accuracy. Even though the model can segregate damage and no-damage images, but it lacked in detecting where the cracks existed in an image. Therefore, a more advanced CNN model will be adopted in this project, comprising of 24 convolutional layers followed by two fully connected layers. Unlike the binary classifier, this model can be used to identify multiple classes of defects. Depending on different types and severity of defects/damages, detailed evaluation or physical inspection if required, can be carried out. Further, the existing models have not been trained on a broad real database and hence not very effective in detecting cracks in images consisting of common obstructions such as trees, wires and other external features. This damage detection solution will be designed for the deployment on an embedded platform of UAVs/drones with a navigation system and uses deep learning algorithms to detect and classify structural cracks on the surfaces.

Novelty:

The novelty lies in providing safer, faster and more reliable inspection of civil engineering structures by using modern drone technology and deep learning. A unique broad real database of videos/images consisting of various types of building facades will be used to train the neural network. Till date, the training of neural network has been conducted on a laboratory dataset of crack and no-crack images. The navigation system of drone will capture the geotagged imagery which is ideal for locating the damage and monitoring the health of the structure. With the developed product it will be possible to quickly, safely and effectively inspect the places where it is otherwise expensive or impossible to reach as a person.

15. **IITP15:** Real-time Anomaly Detection in Traffic Video Streams

Technology: The raw video data will be first filtered for noise. The resulting video will then be segmented into segments which will be classified as normal or abnormal. Features such as dense/ sparse optical flow-based features will be extracted. Those features will comprise the input to classical classification models like support vector machines and hidden Markov models. Moreover, deep learning approaches will be evaluated that combine feature and model learning. The developed techniques will be evaluated on real data with respect to their predictive accuracy, as well as their efficiency (time and memory).

Novelty:
In [34] Kamijo et al. propose a general method for anomaly detection where they consider the behaviour of the vehicle along with its behaviour in relation to other vehicles and the surrounding environment. The authors cluster their trajectory vectors using the popular k-Means approach. In [35] Fu et al. propose a hierarchical framework to cluster vehicle trajectories. They have used a spectral clustering method for grouping similar trajectories. In [36] Yang et al. propose a trajectory analysis method which also uses the spectral clustering method. In this proposal, anomaly is detected based on Indian traffic scenario.

16. **IITP 16:** IoT based Condition Monitoring and Fault Diagnosis of Gearbox

Technology:
In the last decades, lots of research has done and more are going on these five modules. From the literature survey, it has been found that Vibration based fault diagnosis is frequently used methodology in rotary machinery as it carries the signature of faults in the signal. But the noise and vibration associated with raw data often become a major issue in the area of ongoing research. Various Noise reduction method such as Time synchronous averaging (TSA), Angular domain averaging , Order tracking, Hilbert Huang transform, Filter methods etc. have been applied successfully with their limitations. Most of the fault diagnosis technique reported in the literature have taken data at high sampling frequency (above 51ks/sec) and used Time Domain , Frequency domain and Time-Frequency Domainfeatures as input to the data mining algorithm such as genetic algorithm (GA), artificialneural network (ANN), fuzzy logic systems (FLS) etc for pattern recognition. These techniques also require high number of samples for better generalization and high classification accuracy. So, there is a demand of hour in this Industry 4.0 era to develop a new intelligent fault diagnosis expert system, which collect data from gearbox at affordable (low) sampling frequency and have high classification accuracy for minimum number of samples. This willhelp not only in reducing the data saving cost in the IoT Server but also speed up the diagnosis process in online fault diagnosis.

Novelty: The proposed technique will fill the following gap that is not addressed in the literature:

- Multi Fault detection of gearbox at low sampling frequency is not done yet.
- Noise reduction technique based on non- parametric requirement of gearbox is not done yet.
- Development of new intelligent fault diagnosis technique of gearbox based on industry 4.0 for minimum number of samples (taken at low sampling frequency) is not done yet.

17. IITP17: Speech, Video and Data Analytics in Healthcare for TIH

Technology: In this project, we attempt to integrate advanced wearable multimodal sensors, recent powerful digital signal analysis techniques, multi-core embedded processor, digital camera and wireless network technologies with IoT application gateway and interconnect mechanisms to meet demands for improvements in diagnostic accuracy, functionality, power consumption, size and ease of use, as well provide advances in secured data transmission and interoperability. The proposed reconfigurable IoT-driven monitoring system can simultaneously collect and analyze person-specific multiple physiological parameters such as heart rate (HR), respiratory rate (RR), blood pressure, blood oxygen saturation (SpO2), glucose level, body temperature, galvanic skin response (GSR) and physical activity including position/posture, walking, sleeping, climbing and running parameters, blood alcohol concentration level, and securely transfer measured parameters with high-resolution physiological signals to remote monitoring centre for further analysis or diagnosis of potential health risks and specific clinical symptoms.

Novelty:

Limitations of existing neonatal Monitors:

- Existing monitors only provide the video feed of the baby with no contextual health information so it is difficult to analyze that the baby is sleeping or has much serious health problems.
- Not having early health issues and alarming signs.
- Only has contact based health monitoring which may harm the delicate skin of the baby and uncomfortable for the pre-term infants.
- Only availability of contact based sensors in the existing monitors may have cost issues since the wearable devices with electrodes need to be replaced due to damaging, worn and outgrown.

Novelty Points:

Existing Neonatal Monitor	**CSM**	**CLM**	**CDAS**	**BLE/WiFi/3G**	**IoT**
Raybay	√	X	X	X	X
Philips IntelliVue MX800 Patient Monitor	√	X	X	X	X
Cookun Cam	X	√	X	√	√
Oxehealth	X	√	X	X	X
Wellton Healthcare Infant Warmer	√	X	X	X	X
Proposed Monitor	√	√	√	√	√

SM: contact sensor based monitoring; CLM: contactless monitoring; SPD: sleep posture detection; CDAS: crying detection and alert system; BLE: Bluetooth low energy; IoT: internet of things;

18. **IITP18:** Development of Machine Learning based IDS (ML-IDS)

Technology: Machine learning and AI based IDS has been actively used in the recent times for detecting the abnormal behavior of malware. The ML-IDS will monitor the network at different levels of sophistication. In the last few years various deep learning methods like LSTM, RNN, GRU in order to predict the behavior the malware so that malware can be detected at an early stage. We also wish to integrate these techniques as a part of the proposed framework. In the end, a network based device shall be proposed to detect such clandestine malware.

Novelty:

Machine learning has been actively used in the recent times for detecting the abnormal behavior of malware. The ML-IDS will monitor the network at different levels of sophistication. At the lowest level of sophistication we assume that the system utilities like netstat, ps, process lister, network logs, host monitoring tools etc. can be used to monitor the system related activities. The readings from these sensors is considered as untrusted as the clandestine malware can infect these artifacts to hide its activities. At the highest level of sophistication, we analyze the network traffic using a mirrored port via a secured hardware device. This technique can be considered credible. The dissimilarity between the observed traffic observed on the host and the mirrored port can indicate towards the malware that tries to hide traffic from the host based system. The very attempt to hide network traffic indicates the male-fide behavior of the clandestine malware. The system logs collected on the host and traffic will be passed to the various machine learning based classifiers in order to train the model so that such clandestine malwares can be detected at an early stage in future. It can be observed that the clandestine malwares cannot be detected by only host monitoring or by only network monitoring, it requires monitoring of both host as well as network in order to get detected. In the last few years various deep learning methods like LSTM, RNN, GRU in order to predict the behavior the malware so that malware can be detected at an early stage. We also wish to integrate these techniques as a part of the proposed framework.

Further, many such malwares also behave differently under virtualbox and real machine execution. So the behavior observed under VM may not reflect the behavior in real systems. Analysis on a standalone system is expensive since the system needs to be dedicated for analysis. Many malwares modify the stack pointers, alter functional flow of the code in order to execute malicious code. Such techniques are known as system hooking, not all hooks are harmful to the system as a result this technique is prone to false positives.

19. **IITP19:** Secure Monitoring and Data Analysis Management Tool

Technology: At the moment no sharing of video feeds takes place – all video monitoring rely on manual surveillance which is very limited in score. Once notified of an event (intrusion of dumping), officers will screen various cameras with likely recordings of the suspect for evidence. Traditional video analytics providers wouldbuild two separate systems to support each of these use cases, despite their intrinsic similarities.

In contrast, our proposed system will have video analytics modules that can be reused and composed in different ways; in this case, modules for detecting vehicles in a video feed, analyzing vehicle attributes (e.g. type, load, color), estimating driving direction and speed, reading licence plates provide relevant metadata with which video can be annotated, and which can be monitored continuously or queried based on a specific event as in the above examples. Naturally, these same modules can support additional, completely different use cases, e.g. counting trucks passing through a specific road, or determining the favorite car color. In other words, not only video feeds can be shared and accessed by different users, but also analytics capabilities and pertinent video metadata. Wide corporate user and citizen participation is enabled through market mechanisms for supplying camera feeds or analytics modules. The technical, economic, and social research challenges of such an infrastructure are discussed in more detail below.

Novelty:
No such novel framework exists at the moment. We will research the following topics:

- How to cost-effectively process and store video information in the cloud (e.g., should we store the raw video for future needs? or perform the processing in a decentralize way near the source? who pays for it? Optimal decisions may require the formulation of interesting dynamic programming problems.
- How to fairly share costs? The same video information (raw or metadata) could be used multiple times by different customers. How should we share this common cost? We will investigate various fair cost sharing mechanisms and adopt them for our specific needs.
- Designing a specialized portal/market place for video services is a challenging task that requires innovative ideas. Matching demand with supply requires a standardization of interfaces for analytics and metadata information. Specialized brokers (software) will propose how specific queries will be mapped to workflows combining different analytics modules with sources of raw data, using the most economical way to allocate processing resources (decentralization, cloud based processing). Substitute analytics modules will compete in price and quality, and allow customers to choose the right quality-price trade-off.

20. **IITP20**: Edge-AI based Social-Distance Tracker IoT-Camera

Technology: This project aims at using the following technologies for meeting the objectives:

1. **Cloud-IoT-Edge architecture:** In this architecture, computation-intensive tasks (e.g., deep learning model training) should occur at the cloud, and the model will be deployed at the edge for real-time processing. The Camera enabled IoT devices (e.g., surveillance camera) will be used for capturing real-time video/images.
2. **Machine learning and Deep Learning:** The convolutional neural network-based architecture will be used to train the model for counting crowds and tracking. Transfer learning is also useful for building a model. Recommender systems can be made using deep learning or machine learning methods depends on which method gives better performance.
3. **Micro-services:** Docker Containers are useful to deploy models at low resource edge devices or smartphones.
4. **Android/iOS programming:** Building a mobile app android or iOS programming is required.

Novelty:

1. A cloud-IoT-Edge based application for maintaining social distance

2. The app will track and count the crowd in real-time.

3. It is giving decisions based on the crowd pattern that will help the user to understand the current scenario of the market.

4. A recommender system, which will be built on the historical time series data, will recommend a suitable time to go to the shopping complex.

21. IITP 21: Multi-lingual Machine Translation

The project aims at designing, developing and deploying Neural Machine Translation (MT) Systems for English (E) to Indian Languages (X), where X is Hindi, Bengali, and Telugu. Our domains of interest are Judicial, Education and Product reviews. Judiciary and Higher education are the two very important application domains in the mission of TIH. Large volumes of texts are generated daily in the form of reviews about any brand, product, or service via a variety of platforms such as blogs, microblogs, collaborative wikis, multimedia sharing sites, social networking sites, e-commerce platforms etc. Though India is a multilingual country and majority of its population are non-English speakers, at present, almost all the e-commerce platforms provide their services only in English. So, there is a necessity of making the features of e-commerce sites--specifically related to product/services--be available in regional languages. It is difficult for human translators to keep up with dynamic and real time contents, and hence automatic translation is required to make the information available in user's preferred language(s).

Technology: In the face of scarcity of parallel data, we will make **unsupervised NMT (UNMT)** (Artetxe et al 2018; Lample et al 2018) [37-38] as our primary arsenal. Our recently proposed method for Multilingual Unsupervised NMT using Shared Encoder and Language-Specific Decoders (Sen et al, 2019) [39] will be used for translation among multiple Indic languages. This framework evidenced that (i). jointly training multiple languages improves separately trained bilingual models.; and (ii). without training the network for many-to-many translations, the network can translate among all the languages participating in training.

Following are the major components in our proposed project:
***Domain Adaptation and Transfer Learning*:** As a base system, we will use our existing English-Hindi Machine Translation System in Judicial domain. Domain adaptation (Luong et al., 2015)[40] will be used for using the weights and characteristics of our existing English-Hindi judicial domain translation system. By transfer learning (Zoph et al. 2016; Saunders et al., 2019)[41-42], adopting weights from a parent NMT model of English-Hindi, would be a good approach to investigate.

***Data Collection and Parallel (Synthetic) Corpus Creation*:** We shall crawl from the various sources: (i). mix domain data for Hindi, Tamil, Telugu and Kannada from the various sources

(WMT, WAT, OPUS, Wikimatrix, Wikidump); (ii). for English-Hindi, some amount of in-domain parallel corpus will be created through alignment and/or manual translation. For the other Indian languages, we shall mostly depend on the synthetic data prepared automatically through back-translation.

Back Translation- from whatever parallel data is available from (Kunchukuttan et al 2018b)[43] and WAT etc, we will create a crude MT system and apply back translation to monolingual Hindi corpora (Bojar et al 2014) [44]to augment existing corpora.

Subword MT: available parallel corpora is 'treated' to yield character, word, orthographic-syllable (OS) and byte-pair encoded string (BPE) alignments, thereby augmenting the available parallel corpora many times.

Pivot MT: A conventional solution to alleviate the parallel data scarceness is to introduce a "bridge" language (called pivot language) to connect the source and target language though source-pivot and pivot-target translation. Motivated by our recent work on multilingual UNMT, we will keep Hindi as pivot language as there is relatively larger corpus available for English-Hindi.

Combination MT: Each of the approaches mentioned above has its benefits over the approaches depending on the language of choice (that is morphologically rich or poor), parallel data size, vocabulary size of the languages etc. For example, Indian languages (IL) are morphologically rich and the parallel corpora between them are not sufficiently large. So, combination of all these approaches will benefit the MT system.

Novelty:

i. An end to end NMT system will be developed for two important verticals, *viz.* judiciary and education.
ii. An end to end NMT system will be developed for product reviews translation.
iii. Appropriate data augmentation strategies will be developed.
iv. Robust NMT models will be developed for low-resource languages

22. **IITP 22:** Breast Cancer Detection and Classification from Tomosynthesis Dataset

Technology: Artificial intelligence and deep learning techniques can be applied in any field for improved performance without the requirement of handcrafted features. We can use different pre trained CNN structures and also 3D CNN for solving the problem.

Novelty: A globally and locally organised dataset (practical) for classification into benign/malignant.

23. **IITP 23**: Real Time Video Stabilization

Technology: The deep learning can be applied in any field for improved performance without the requirement of handcrafted features. A real time video stabilizer is the need of the hour. Due to large amount of video recording using hand held devices requires a real time video stabilizer.

Novelty: A locally synchronized dataset as ground truth and a deep learning trained network from scratch is aimed. Also we would like to develop a trained model which can take less than a second to stabilize a video containing around 450 frames.

24. IITP24: Combating Misinformation using NLP and Deep Learning

Technology: Our research aims to build NLP models to explore the role of "textual novelty" and "emotions" to identify potential news items/posts for misinformation. The initial investigation would be on English language data. The models we would develop for English would be extended for popular Indic languages, namely Hindi and Bengali.

Novelty/Impact:

In a developing and multilingual nation like India where the cellular network (+low-cost cell phones) has reached almost all corners of the country and penetrated almost all sects of the society (half a billion active users), where people has access to the lowest possible internet tariff, the use of social media platforms is widespread making them an important source of news and place for social and political activity. With limited exposure to technology, many new users are coming online; and due to the diverse socio-political regime, we observe an increasing trend of misinformation in social media and personal messaging apps. The politically-motivated spread of misinformation is reinforcing the less-logical beliefs and creating an atmosphere of polarization be it in religion or politics thus inciting violence and untoward actions. With a strong footing on research in social studies, the current research aims to develop multilingual applications leveraging NLP/ML tasks to address this societal problem with special focus on health, religion, and politics.

The projects which have been reserved for "Open Call for Proposals" are in the broad areas of "Speech, Video and Text Analytics". All the problems have potentials to generate new knowledge, leading to a prototypes, tools and/or products at the end. The key selection criteria would be (i). Novelty of the proposed technique; (ii). Relevance to the mission of i-Hub and the application verticals; (iii). Potential for commercialization.

We have conceptualized all these proposals in consultations with our industry collaborators, namely Accenture, Microsoft, Wipro, TCS Innovation Lab, IBM Research Lab; and foreign as well as national collaborators from academia or PSUs.

10. Section-10: Management

DST will be providing the funding support required to set up and manage the ***"IIT Patna Vishlesan i-Hub Foundation"*** on "Speech, Video and Text Analytics" over the first five years of the project. TIH, IIT Patna shall be the managing body for the disbursement of the expenses incurred under the project. A comprehensive appraisal system will be adopted which shall include, but not limited to the below:

(1) **Hub Governing Body (HGB):** The **Hub Governing Body (HGB)** shall be the apex body for all administrative, legal and financial matters, will be chaired by Director IIT Patna and will have members from Academia, Domain experts, Industry, Investors and Government. The Section-8 Company, TIH will operate under the directives from HGB.

The structure of the approved Hub Governing Body (HGB) is as follows:

Names/Positions	**Designation**
Prof T N Singh, Director, IIT Patna	Chairman
Dr. Asif Ekbal, Associate Professor, Department of Computer Sc and Engg. , IIT Patna	Member
Dr. Rajib Jha, Associate Professor, Department of Electrical Engg., IIT Patna	
Dr. Sriparna Saha, Associate Professor, Department of Computer Sc and Engg. , IIT Patna	Member
1. Dr. Shubhashis Sengupta, MD and Technology Research Director, Accenture 2. Dr. Karthik Sankaranarayanan, IBM Research	Members, Representatives from Industry

3. Dr. Anoop Kunchukuttam, Microsoft 4. Dr. Lipika Dey, TCS Innovation Lab	
1. Prof. Pushpak Bhattacharyya Former Director, IIT Patna Professor, Department of Computer Science and Engg., IIT Bombay 2. Prof. Sanghamitra Bandopadhyay, Director, Indian Statistical Institute 3. Prof. Nilogy Ganguly, Department of Computer Science and Engg., IIT Kharagpur	Members, Representatives from Academia (*External*)
DST, Govt of Bihar has been contacted	Representatives from Local Government

(2) **Project Evaluation Board (PEB) or Working Group (WG):** The board of directors of Innovation hub shall form a Project Evaluation Board (PEB) or Working Group (WG). The roles and responsibilities of the PEB/WG shall be defined for the implementation of different schemes offered by TIH from time to time. The PEB/ WG shall have powers to recommend projects, purchases and expenses of all types within the technical and financial scope of the Mission. The PEB and/or WG will be constituted with the experts drawn from stakeholders, such as Government, academia, research institutions, end-users organizations and industry.

(3). **Project Review Steering Group (PRSG):** For each project, there will be a Project Review and Steering Group (PRSG) that will monitor the progress of the project. PRSG will ensure that project objectives and deliverables are met in a timely manner. This will be constituted with the experts drawn from stakeholders, such as Government, academia, research institutions, end-users organizations and industry

(4) **Project Evaluation Team (PET):** The PET of affiliated Incubation centre may be recognized by TIH or it may also recommend new PET depending as and when deems fit. The PET will take care of admission, progress monitoring, investment and other matters of related to start-ups.

(5) **TIH Administration:** The Innovation hub shall be managed by a professional team comprising CEO, General Managers, Accountants, Junior Assistants, Auditors, and other staffs. The innovation hub shall be registered as a Section 8 company. The organization structure of TIH shall be as follows:

TIH Hub will govern the innovation hub. The TIH is named as ***"IIT Patna Vishlesan i-Hub Foundation"*** . The Section-8 Company of TIH will work under the directives of HGB. There will be a designated Professor-in-Charge as Project Director, who will head the TIH activities. Under this we have three major functional heads, (i). ***IIT Patna Vishlesan CoE*** and (ii). ***IIT Patna Vishlesan TBI*** and *(iii).* ***TIH Administration, Finance and Legal Unit***

CoE, CTO, Managers and Administrative staffs will manage Research and Development, Academic Programs, Centralized Laboratory, and an Industry Relation Unit. They will work under the various project directors and convernors of various committee of the TIH. The ***IITP Vishlesan TBI*** will manage all the startup and incubation related activities. The third unit, i.e. ***TIH Administration, Finance and Legal Unit*** will be common to the CoE and TBI.

The schematic diagram of the management structure is shown in Figure 1[7].

[7] We will study the structure of other IITs which have long experience in running these kinds of facilities.

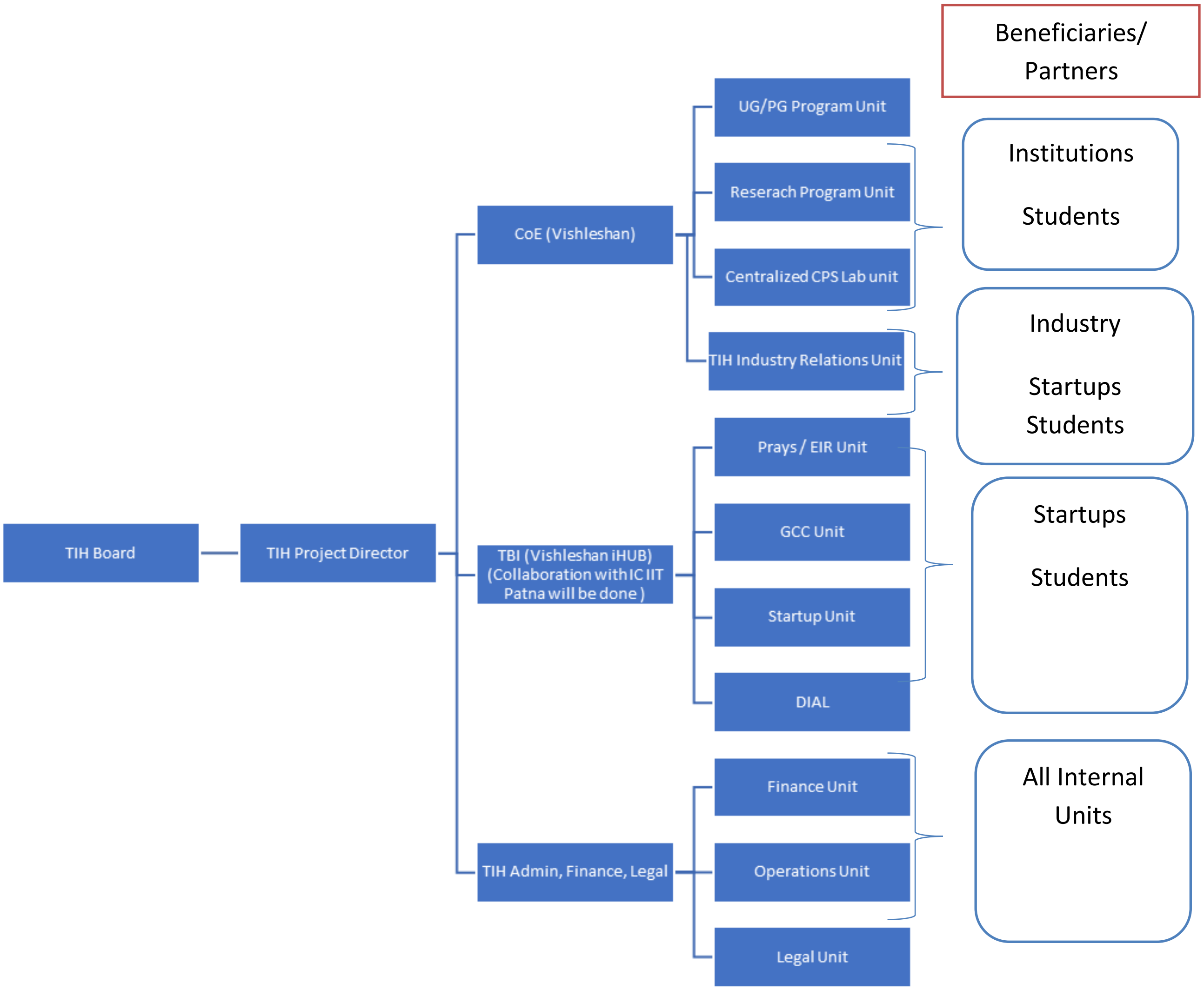

Figure 1: Management Structure of TIH

5. Call for the Proposals and Progress Monitoring

Call for the technical proposals and award: The call for proposals will be advertised by the TIH periodically. Upon the receipt of the proposals, a preliminary scrutiny committee will be set up by the Head of TIH. The shortlisted proposals will be presented by the PIs before the PEB/WG for the final selection. The selected proposals will be placed before the HGB and subsequent approval from the chairman of the Board will be taken.

Call for the Startups and Award depending upon the need: TIH will ask the affiliated incubation centre to advertise the call for proposals periodically. Upon the receipt of the proposals the PIC of Incubation centre in consultation with Head, TIH will set up a preliminary scrutiny committee to shortlist proposals. The shortlisted proposal will be presented before PET for consideration. Finally, the selected companies will be placed before the Head, TIH and subsequently be approved by the chairman of TIH Board.

- ❏ Each beneficiary of different schemes shall be under the observation of TIH, IIT Patna and other professional agencies as deemed necessary. The projects shall be awarded based on a transparent and open processes.
- ❏ The investigators shall be subjected to evaluation periodically. During the evaluation the progress of the projects shall be rated. Any deviation shall be recorded. In case of major deviations which may be a reason of the failure of the project, the support and facility extended to the investigators shall be revoked.
- ❏ Any investigator can avail the benefit of the scheme subject to clearing the evaluation criterion and process.
- ❏ As and when required professional audit firms shall be engaged by TIH for audit works.
- ❏ TIH, IIT Patna shall report the progress of the project from time to time or as and when asked to DST.
- ❏ The progress of the awarded projects shall be monitored by PRSG and reported to the HGB, TIH and DST on regular basis and also as and when called for.
- ❏ The progress of Startups will be monitored by PET, IC IIT Patna and reported to TIH admin; and HGB. TIH, IIT Patna will report to the DST on regular basis. Whenever required, IC, IIT Patna will be asked by TIH to report the updates to the HGB, TIH, IIT Patna and also to the funding agency, i.e. DST.

6. Human Resource Requirement

The TIH will manage the human resources among its functional units. *Please note that depending upon the requirement, the designations may vary, positions could be reduced and/or new positions may be created.*

Sl No.	**Major Group**	**Manpower Requirement**	**Roles and responsibility**

1	TIH	CEO	Secretary to HGB, TIH
2	Knowledge, Technology & Tool Creation	GM/Manager (1) Sr Excutive - Academic Relations (1) Sr Excutive(1) - IP, Tech Jr Excutive (2) Counsellor/ Receptionist (1)	Managing the sub-unit Project management IP, Tech Transfer Management IP, Tech Transfer Management
2	Industrial Programs	GM/Manager (1) Sr Excutive - Industry Relations (1) Sr Excutive - Startup Relations (1) Junior Executives (2)	Managing the sub-unit To develop strong relation with industry To strengthen startup eco system To coordinate with Incubation Centre
3	HR, Finance, Legal and Admin	GM/Manager (1) Sr Accountant (1) Jr Accountant (1) Auditor Sr Admin (1) Jr. Admin (1)	Managing the unit Finance Finance Auditing Routine Admin Activities and Purchases Routine Admin Activities and Purchases

		Executive - Legal / CS Attendant (2) House Keeping (1)	Legal This will cover Managing day-to-day activities House keeping
4	Technical	GM/Manager (1) Scientific officers (1) JTS (2)	Technical Infrastructure IT Management

11. Section-11: Finance:

Attached as Annexure-1

12. Section-12: Time Frame

Attached as Annexure-2

13. Section-13: Cost benefit Analysis

We propose to establish a Technology Innovation Hub (TIH) named as ***"IIT Patna Vishlesan i-Hub Foundation"*** in the area of "Speech, Video and Text Analytics" which aims to create a strong and seamless ecosystem for leveraging the potential and exponential growth of interdisciplinary cyber physical systems (ICPSs). The proposed centre will facilitate and integrate the nationwide efforts of research and development taken in the broad areas of "Speech, Video and Text Analytics" for knowledge generation, innovation, product development, and commercialization.

Component	Sub-components	Cost	Benefits	Remarks

Technology Development	A). Knowledge generation
	B).Tools/Prototype/Product Development
		25 Crores	A).
			(i). Publications: 55;
			(ii). Patents: 25;
			(B). Prototype/ Tool/Product development: 20
				i. Approximately 40-50 projects are planned.
				ii. Each project will generate one research paper, and out of 2 projects there would be 1 patent
				(iii).Deployment of the developed product will be a very important success parameter

HRD and Skill Development	CHANAKYA Scheme for UG Courses	4.75 Crores	(i). 250 UG students will be trained in CPS technologies; (ii). UG program will be aligned to the broad TIH theme	Internships/Fellowships to the UG Students; Project support to the UG students; UG Course up-gradation
	CHANAKYA Scheme for PG Courses	3.5 Crores	(i). 50 PG students (MTech) will be trained with CPS technologies; (ii). 50 PG theses (iii). Rs 100 Lakhs will be received as fees by the institutes hosting those students (50 Students@ Rs 1,00,000 per year for 2 years). TIH is expecting a return of Rs 100 Lakhs from this.	PG students will be awarded fellowships on selection basis; Project support for the PG students; Infrastructure development fund for the projects 50% will be at IIT Patna
	CHANAKYA Doctoral Fellowship	5.00 Crores	(i) 25 PhD scholars specialized in the area of ICPS (ii) 25 doctoral theses in the area of "Speech, Video and Text Analytics". (iii) Additional tools / platforms / products (iv) In addition, Rs 144 Lakhs will be received as fees from PhD scholars through Technology	Each PhD student will be given 19.2 Lacs for 4 years (*The investigators will have to manage 1 year of funding from the other sources*) Out of total scholars, (Chanakya + Through

			projects) by the institutes hosting the technology development projects (~50 PhD scholars; @ Rs 60,000 per year for 4 years). TIH is expecting a return of Rs 60 Lakhs from this.	project), 50-60% we are expecting 35-40 students at IIT Patna
	CHANAKYA Postdoctoral Fellowship	5.00 Crores	(i) 25 Postdoctoral researchers will be specialized in the area of ICPS (ii) 25 doctoral theses in the area of “Speech, Video and Text Analytics”. (iii) Additional tools / platforms / products	Each Postdoc fellow will be given Rs. 19.2 Lacs for 2 years
	CHANAKYA Faculty Fellowship	90 Lakhs	3 Faculty fellows in the areas of “Speech, Video and Text Analytics”	Each Faculty fellowship will be at most 30 Lacs for 3 years
	CHANAKYA Chair Professor	90 Lakhs	3 Chair professors in the areas of “Speech, Video and Text Analytics” Chair professors will be mentoring projects, increasing the efficiency and outcome of projects.	Each Chair professor will be given the compensation of at most Rs. 30 Lacs for period of 3 years
	Professional Skill Development Workshop	30 Lacs	4 professional skill development programs to train 200 professionals in CPS technologies	Each program will train 50 participants
	Advanced Training School	40 Lacs	4 Advanced training programs for BTechs, BSCs, ITI, Polytechnics.	Expected no of participants in each school= 30
	New PG Program	4 Crores	New MTech program in “Speech, Video and Text” Analytics”; Postgraduate students trained= 120 (i). Rs 144 Lakh will be received as fees by the institute (120	Expected no of trained professionals = 30 per year for 4 years

			students @ Rs 120,000 per program). TIH is expecting a return of Rs 72 Lakhs from this.	
Entrepreneurships and Startups	CPS-GCC-Grand Challenge	3.5 Crores	New Ideas, Concepts, Challenges Prototypes in “Speech, Video and Text” Analytics; Expected number of startups = 5 In addition, Prayasees and EIRs are also expected from GCC	5 Startups in new emerging areas of “Speech, Video and Text” Analytics
	CPS-PRAYAS	2.7 Crores	Young Entrepreneurs in “Speech, Video and Text Analytics”; Employment Generation; Increased business	10 Teams to be supported. 40% teams expected to be converted into startups
	CPS-EIR	1.8 Crores	Entrepreneurs and New ventures; Increased no. of start-ups; Commercialized technologies; Employment generation; Increased business in “Speech, Video and Text Analytics”	25 EIRs 25% expected to be converted into either Prayasees or Startup
	CPS-Startup	3 Crores (Cost included in Seed Support system)	Increased no. of student start-ups; Commercialized technologies; Employment generation; Entrepreneurs and new ventures in “Speech, Video and Text Analytics” Around 12 CPS products/applications is expected to be commercialized It is also expected that (i) investment in at least one startup will fetch a return of 10X in 5 to 7 years = 1 Cr (ii) investment in at least 3 startups will fetch a return of 3.3X in 5 to 7 years = 1 Cr	40 Startups in “Speech, Video and Text Analytics”
	CPS-TBI	15 Crores	Commercialization of new technologies; Startups in “Speech, Video and Text Analytics”; Employment generation	Will tie-up with existing IC at IIT Patna

			TBI will facilitate all other schemes under Innovation, Entrepreneurship and Startup Ecosystem	
	DIAL	2 Crores	Speedy commercialization of Technologies; Technology adopted into industry 10 to 12 products in the market; Employment and revenue generation. Around 12 CPS products/applications is expected to be commercialized It is also expected that (i) investment in at least 3 companies will fetch a return of 5X in 2 to 3 years = 4.5 Cr	12 – 14 companies will be accelerated
	CPS-Seed Support System	7.2 Crores	As mentioned under Startup and DIAL. Total companies that will be supported is 40 - 45	3 Cr for Startup 4.2 Cr for 14 DIAL Accelerated companies
	CPS- Strategic Information Services for Entrepreneurship (SISE)	-	Information on CPS related patents Development of product/services based on identified patent Increased Entrepreneurship Identification of new areas for research	
International Collaboration		5 Crores	Collaborating with international experts to explore new research areas; Gaining new knowledge and Experience	Total 1 collaboration

14. **Section-14: Risk Analysis**

We perform the risk analysis in terms of the following parameters.

1. Risk Category: Cost Risk (CR); Time Risk (TR), Scope Risk (SR), Quality Risk (QR), Financial Risk (FR), Legal/Contractual Risk (LR), Regulatory Risk (RR)
2. **Impact:** Defined in the scale of 1 to 5 with Low-1, High =5
3. Probability of Risk: Defined in the scale of 1 to 5 with Low=1 and High=5
4. Risk Response: Corresponds to the measures to be taken to avoid the risks.

A. **Description of Risk:** Inadequacy of the project design in handling the complexity arising out of the magnitude of the project

(A. 1). Risk Category: CR, TR, SR, RR
(A.2). Impact: 4
(A.3). Probability of Risk: 2
(A.4). Risk Response:

1. **Avoidance**:
   - Group reviews and consultative approach, including major stakeholders at the DPR stage: To ensure proper understanding of requirements, well defined project scope and objectives and outcomes, well defined organizational architecture with clear;
   - Human resources are available and capable;
   - Ownership and accountability;
   - Proper mapping of HR skills, competence and experience. Experience of handling big projects is the strength.
2. **Mitigation:**
   - Periodic Review and preventive and corrective actions as necessary with the guidance of Board (that includes representatives from industry and other stakeholders)

4. **Acceptance**:

Minor and inconsequential gaps will be ignored

(B). Description of Risk: Potential non alignment of research initiatives to

- Industry /start-up technology gap in CPS;
- National CPS priority area; and
- Translatable to tools or platforms

affecting translational and commercialization goals

(B. 1). Risk Category: QR, CR, TR
(B.2). Impact: 4
(B.3). Probability of Risk: 3
(B.4). Risk Response:

1. **Avoidance**:

- Creation of explicit guidelines on project selection to ensure proper alignment to project objectives;
- Orientation to selection committee on goals of the projects and strategy adopted to achieve the same

2. **Mitigation:**

- Annual review of the research problems undertaken under the project to ensure ongoing compliance;
- Implementation of preventive and corrective actions based on a route cause analysis in case of observed deviations

(C). Description of Risk: Deviation in the proposed objective and scope because of rapid change in technology

(C. 1). Risk Category: CR, TR, SR, FR, RR
(C.2). Impact: 4
(C.3). Probability of Risk: 2
(C.4). Risk Response: Mitigation

- Drawing the technical expertise from research talent pool of IIT Patna, other institutes, industries and startups to serve the modified objectives.
- Industry partnerships to align to the application layer requirements of technology changes.

(D). Description of Risk: Inadequate legal documentation and legal issues arising out of the same (owing to Large Number of legal documentation of varyingnature to be prepared in very small time span)for defining the role and scope of various stack holders

(D. 1). Risk Category: RR, LR

(D.2). Impact: 5

(D.3). Probability of Risk: 2

(D.4). Risk Response:

1. **Avoidance**:

- Contracting experienced consultants to create the required documentation.
- At least one person from legal/CA/CS background in the team

2. **Mitigation** :

- CA and legal firms will be empanelled to make necessary amendments and to address any issues related to legal and regulatory matters

(E). Description of Risk: Delay in fund release from financial collaborators or budget cuts

(E. 1). Risk Category: QR, SR, LR, FR

(E.2). Impact: 5

(E.3). Probability of Risk: 3

(E.4). Risk Response:

1. **Avoidance**:

- Establish a frequent and regular communication of financial status to stakeholders
- Early projection of financial requirements.

- Demand for fund release raised at 90% utilization of operational/program support expenses

- Setting up of a corpus fund by accepting CSR funding and/or grant and donations from Industry and government body for the same purpose.

2. **Mitigation** :

- Reducing the scope of operations so as to align to the reduced budget in case of a budget cut

- Loans and advances from IIT Patna until funds are available (subject to X months or Y amount) in case of late release of funds

- Loans and advances from Government stakeholders until funds are available (subject to X months or Y amount) in case of late release of funds

- Fund generation by way of liquidating equity in supported company

3. **Acceptance**:

- Beyond mitigation levels, risk will be accepted

- Legal liabilities and obligations will be presented to the board and appropriate remedial measures to be implemented

- Dissolve the company if the operations become completely unviable, with stakeholder consent (DST, IITP, TIH)

(F). Description of Risk Delay in procurement of Essential and Relevant Technical Equipment and Services essential for the success of the innovation hub project

(F. 1). Risk Category: TR, CR, QR,FR
(F.2). Impact: 5
(F.3). Probability of Risk: 1
(F.4). Risk Response:

1**. Avoidance**:

- Setting up of technical and purchase committees competent in procurement activities; OR
- Outsourcing the procurement activities to IIT Patna procurement division for speedy processes
- Adding staff dedicated to technical procurement
- Fool-proof procurement processes will be established to ensure speedy purchase.

2. **Mitigation :**

- If the required equipment/technical service is available in IIT Patna/ nearby institute/organization/ Partner Industry, the same can be taken on rent on temporary basis.

(G). Description of Risk

HR Risks:

- Inadequate number of recruited staff
- Competence of recruited staff
- Poor team dynamics
- High Churn
- Unforeseen long leave of non-technical and technical managerskey people . If some of them get sick, it can delay the project for an indefinite period of time or even derail it.

(G. 1). Risk Category: CR, TR, QRFR
(G.2). Impact: 3
(G.3). Probability of Risk: 3
(G.4). Risk Response:

1. **Avoidance**:

- Collective Review of OD and Org Structure at DPR stage

- Clear definition and alignment of roles, responsibilities, experience levels, skill sets, hierarchical relationships
- Formation of a selection committee with experts from academia, administration and industry
- Team induction and team building programs
- Pay structure and incentives in line with government policies and industry standards.

2. **Mitigation :**

- If staff number found to be inadequate or if there is churn or if long absence of key staff, recruitment will be done and until then (i). services of IIT Patna employee with relevant experience will be taken on part time; (ii). employees can also be hired through outsourcing agencies on temporary basis.
- RCA for high churn, if observed, will be done and necessary preventive and corrective actions will be done.
- Specific training in case of inadequate skills.
- In case of poor team dynamics: Team building activities.
- Review of alignment of roles, responsibility &authority to identify and resolve bottlenecks.

(H). Description of Risk: Contractor failure

(H. 1). Risk Category: CR, TR, FR, LR, QR
(H.2). Impact: 4
(H.3). Probability of Risk: 2
(H.4). Risk Response:

1. **Avoidance**:

- Check references.

- Assess abilities prior to hiring. Provide a scope of work that clearly identifies responsibilities.
- Define remedies clearly in case of non-delivery.
- Create redundancy in critical services that are outsourced.

2. **Mitigation :**

- RCA and corrective and preventive steps. Involve interactive team management to identify issues and act as facilitator to resolve team issues. Implement processes to escalate conflict resolution ton senior management if needed
- Use alternative service providers (where redundancy is available)
- Invoke remedies

(H). Description of Risk: Contractor failure

(H. 1). Risk Category: CR, TR, FR, LR, QR
(H.2). Impact: 4
(H.3). Probability of Risk: 2
(H.4). Risk Response:

1. **Avoidance**:

- Check references.
- Assess abilities prior to hiring. Provide a scope of work that clearly identifies responsibilities.
- Define remedies clearly in case of non-delivery
- Create redundancy in critical services that are outsourced

2. **Mitigation :**

- RCA and corrective and preventive steps.Involve interactive team management to identify issues and act as facilitator to resolve team issues. Implement processes to escalate conflict resolution ton senior management if needed

- Use alternative service providers (where redundancy is available)
- Invoke remedies

(I). Description of Risk: Overly optimistic schedule Legal Action delays; Overly optimistic schedule

(H. 1). Risk Category: CR, TR, FR, LR, QR
(H.2). Impact: 5
(H.3). Probability of Risk: 1
(H.4). Risk Response:

1. **Avoidance**:

- Ensure all contracts signed and MOU in place before starting the sub projects.
- Legal audits to find the consistency of the legal documentation at least once in the project life cycle

2. **Mitigation :**

- Initiate immediate legal review and corrective action.
- RCA to see the root cause and do preventive and corrective actions

(J). Description of Risk: Theft of materials, intellectual property or equipment

(J. 1). Risk Category: FR, RR,LR
(J.2). Impact: 4
(J.3). Probability of Risk: 3
(J.4). Risk Response:

1. **Avoidance**:

- Explicitly define information security and IP handling protocols
- Train teams in the same
- Ensure Non-Disclosure Agreements

- Insurance against theft or fire damage of critical equipment

2. **Mitigation :**

- Initiate corrective actions.
- Secure insurance.
- Ensure all the contracts signed and MoU in place before starting the sub-projects. Follow all the regulatory requirements and complete stakeholder management plan.

15. Section-15: Outcomes

The goal of the mission is to foster the research and development in the broad areas of CPS, especially in “Speech, Video and Text” Analytics. The key outputs indicators would be (i). Promoting translational research in CPS, especially in “Speech, Video and Text Analytics”; (ii). Development of a set of algorithms, technologies, and tools to develop solutions in order to feed into some of the national priorities; (iii). Knowledge generation in terms of publications and patents; (iv). Human Resource Development (HRD) and Skill test for enhancing high-end researchers base by creating next-generation technocrats, Scientists, Engineers, Skilled and Semi-skilled workforce: through Bachelors, Masters, Doctoral, and Postdoctoral programmes; through Faculty Development programmes, Summer and Winter Schools, and Internships; (v). To establish and strengthen the international collaborative research; (vi). To promote research and development in “Speech, Video and Text” Analytics by PPP model; (vii). Engagement with national institutions and PSUs for implementing joint projects; (viii). Enhancing core-competence and innovation in “Speech, Video and Text Analytics” through a very vibrant Start-up ecosystem (setting up Startups, Job creation and Economic growth).

S.No.	Target Area	1stYr	2nd Yr	3rd Yr	4th Yr	5th Yr	Total
1	**Technology Development**						
1.1	No. of Technologies (IP, Licensing, Patents etc.)	0	5	5	5	5	**20**
1.2	Technology Products	0	0	5	5	5	**15**
1.3	Publications, IPR and other intellectual activities	0	10	10	10	15	**45**
1.4	Increase in CPS Research Base	0	10	15	20	25	**70**
2	**Entrepreneurship Development**						
2.1	CPS- Technology Business Incubator (TBI)	0	0	1	0	0	**1**
2.2	CPS- Start-ups & Spin-off companies	0	0	5	15	15	**35**
2.3	CPS-GCC- Grand Challenges and Competitions	0	0	1	0	0	**1**
2.4	CPS-Promotion and Acceleration of Young and Aspiring technology entrepreneurs (CPS-PRAYAS)	0	0	1	0	0	**1**
2.5	PS-Entrepreneur In Residence (CPS-EIR)	0	0	7	7	7	**21**
2.6	CPS – Dedicated Innovation Accelerator (CPS-DIAL)	0	0	1	0	0	**1**
2.7	CPS – Seed Support System (CPS-SSS)	0	0	1	0	0	**1**
2.8	Job Creation	0	2000	2000	2000	2750	**8750**
3	**Human Resource Development**						
3.1	Graduate Fellowships		50	60	70	40	**220**
3.2	Post Graduate Fellowships		8	20	12	2	**42**
3.3	Doctoral Fellowships		15	8	-	-	**23**
3.4	Faculty Fellowships		2	1	-	-	**3**
3.5	Chair Professor		2	1	-	-	**3**
3.6	Skill Development		100	110	100	100	**410**

4	**International Collaboration**						
4.1	International Collaboration		-	1	-	-	**1**

Table 2: Year-wise Targets

Sl no.	**Components**	**Activity**	**Physical Units**
1	**Technology Development**	No of Technologies (IP, Licensing, Patents etc etc)	25
		Technology Products	20
		Publications, IPR and other Intellectual activities	55
2	**HRD and Skill Development**		
		Professional Skill Development + Advanced Skill Development	300+120= 420
		Graduate Fellowships	250
		Post-Graduation Fellowships	50
		Doctoral Fellowships	25
		Post-Doctoral Fellowships	25
		Faculty Fellowships	3
		Chair Professors	3
3	**Centre of Excellences**		1
4	**Innovation and Startup Ecosystem**	CPS-GCC-Grand Challenge and Competition	1
		CPS-Promotion and Acceleration of Young and Aspiring technology entrepreneurs (CPS-PRAYAS)	1
		CPS-Entrepreneur In Residence (CPS-EIR)	25 (25% expected to be converted into either Prayasees or startups)
		CPS-Start-ups & Spin-off companies	40 Startups in "Speech, Video and Text Analytics"
		CPS-Technology Business Incubator (TBI)	1
		CPS-Dedicated Innovation	1 (12 – 14 companies

		Accelerator (CPS-DIAL)	will be accelerated)
		CPS-Seed Support System (CPS-SSS)	1 (Total companies that will be supported is 40 – 45)
5	**Increase CPS Research Base**	PhDs/Post-docs/Researchers in CPS technologies	75
6	**Job Creation**	Through Skill Development, Startups	8750 (approx.)
7	**International Collaboration**		1
8	**Mission Management Unit**		1

Table 3. Outcomes (Physical Units)

S No	Objectives/ Indicators	Expected outputs/ Deliverables	Unit name	Baseline data (National Status)	Measurable Outputs/ Deliverables
1	To promote and foster R&D in Cyber- Physical Systems (CPS), especially in "Speech, Video and Text Analytics" (*Professors, Postdocs, and Researchers*)	Increased core researchers base in advanced and cutting technologies	No of researchers in "Speech, Video and Text" (*all over India*)	25	100
2	To develop technologies, prototypes and demonstrate associated applications pertaining to national priorities.	A set of technologies, tools, algorithms to feed into some of the national priorities	No of technologies	10	45
3	To enhance high-end researchers base, Human Resource Development (HRD) in these emerging areas. (*Doctoral, Postgraduates, Graduates, Trainees*)	Delivery of next-generation technocrats, Scientists, Engineers, Skilled and semi-skilled	No of students	50	470

		workforce.			
4	To establish and strengthen the international collaborative research for cross-fertilization of ideas.	Global standard Collaborative research in the area of “Speech, Video and Text” Analytics	No of collaborations	0	1
5	To enhance core competencies, capacity building and training to nurture innovation and Start-up ecosystem.	Start-up companies, job creation and economic growth	No of start-ups in the broad areas of “Speech, Video and Text Analytics”	10	50
6	To set up world-class interdisciplinary collaboration centers of excellence in several academic institutions around the country, with a substantial amount of funding to enable them to achieve significant breakthroughs.	Centre of Excellence on “Speech, Video and Text Analytics” (*more on the application verticals will be gradually setup*)	No of CoEs	0	1
7	Engagement with Government and Industry R&D labs as partners in the collaboration projects	PPP model demonstration in technology development	No of partnerships developed	10	20-30
8	Mission mode application goals under “Speech, Video and Text Analytics”	Prototypes of “Speech, Video and Text” Analytics	No of prototypes/ testbeds	10	30

Table 4: Outputs in terms of measurable deliverables. Baseline data is computed based on the national scenario (approximately)

16. **Section-16: Grand Challenges**

Below we provide a list of possible grand challenge problems related to the broad areas of Text, Speech and Video Analytics. Some of these will be These have been taken from the industry collaborators.

A). Grand Challenge Set-1: Multimodal AI in Education, Judiciary, Tourism, and E-commerce

Under this, we shall take up a few important and socially relevant problems related to Education, Judiciary and E-commerce. The technologies to be developed will be under the broad theme of speech, video and text analytics.

i. Machine Translation of Judiciary, Educational and Social Media (or, E-commerce) Contents:

India is a multilingual country with 22 officially spoken languages. Majority of the population (almost 80%) do not speak in English, and therefore, developing machine translation system to make these various contents available in different Indian languages will play an important role towards building a digitally literate society. Education, judiciary are two important domains, where a large volume of texts is generated in English. Making this information available in several Indian languages will be beneficial to the society at large to meet the goals of "*Education for All*" and "*Justice for All*" . Social media, on the other hand, is the source that produces enormous amount of information daily, but majorly in English. Translating this information into vernacular languages will facilitate various e-commerce services.

We will take up a few interesting problems on Machine Translation to address the problems of low-resource scenario (as Indian languages are *resource-constrained in nature*): unsupervised neural machine translation under low-resource scenario; domain adaptation and transfer learning involving low-resource languages; domain dictionary creation; parallel corpus filtering etc.

ii. Multilingual QA and Chatbot

Question-answering (QA), Conversational AI system are very important in the judiciary and educational sector. In judiciary, a large volume of texts is generated daily in the form of FIRs, petitions, proceedings etc. The lawyer or judges or end users might be interested to look into some pertinent information such as parties involved, date when the FIR was lodged, section number, verdicts in the case etc. Extracting all this relevant information from the long and complex text by manually reading these is quite infeasible. Question-Answering/ Chatbot can help the stakeholders by providing responses to all these relevant queries.

In Education domain, Chatbot to counsel young students to cope up with peer pressure would be important. Automated question-generation, answer summarization would help the teachers and students. Chabot and QA will assist travelers in selecting tourist places, booking flights, booking hotels etc.

iii. Sentiment Analysis Techniques in Indian languages

Sentiment Analysis has become the holy grail for almost any e-commerce organization, or for political analyst, or for a market surveyor. Although there are dozens of readymade solutions available for English, but there is almost none for Indian languages, except some datasets available for Hindi. The project aims at developing solutions with robust accuracy levels, in the range of 0.7-0.9 F1-score level, for major languages like Hindi, Bengali, Tamil, Telugu, Punjabi, Marathi, Kannada etc. Considering social media cases code-mixing is yet another unseen challenge for Indian subcontinent.

iv. Neural Machine Translation for Extremely Low-resource Languages

Neural machine translation (NMT) has recently shown highly promising results on publicly available benchmark datasets and is being rapidly adopted in various production systems. However, standard NMT systems need huge amount of parallel corpora: hundreds of millions of sentences. Such amount of data are not available for many languages (e.g. Indic languages) and domains (e.g. medical, tourism, judicial, social media and e-commerce contents etc.).

This project aims at developing effective Unsupervised (and Semi-supervised) Neural Machine Translation (bilingual and multilingual) models keeping Indic Languages in focus. This is targeted for Health, Judiciary and Education domains.

v. AI based mis-information detection and prevention system

Rather than pandemics, misinformation spreading kills more people and create social discord world over. There is an urgent need to detect and prevent misinformation spreading through social media through effective AI intervention in Indian context (keeping in view our social, religious and cultural sensitivity). In essence, this will have three primary parts –

- Detect if a particular social network post is fake or un-trustworthy
- Detect virality and spreading potential of the content and the "super-spreaders" in the network
- Analyse veracity / claims in non-reviewed or general publications or blogs.

vi. AI based personalized learning augmentation

AI can be a great enabler for personalized learning through higher adaptiveness, audio-visual interactivity, AI based assessment for individual learning proclivity and path. It is not just for curriculum learning, but for social and behavioral skills as well. A set of projects be initiated on various aspects of AI assisted learning through – AI based course curation, AI based assessment (questionnaire generation, rating and ranking), AI based personalized educational coach, AI based life skill coaching and guidance.

vii. **Hate/Fake Videos Detection**

Indians are highly political and given the nature of the country hate/fake videos are being circulated and become viral soon. Identification of such videos quickly over social media is an essential problem to solve.

viii. **Development of Speech Technologies for Indian Languages**

There are several core speech technologies that need to be developed for Indian languages. Among these, at the first level speech recognition and speech synthesis systems need to be developed. Speech to Speech Machine Translation (SSMT) for translating lectures could be an important step.

B). Grand Challenge-Set2: Multimodal AI for Health Care Systems

Grand challenges will be organized to solve some problems in the healthcare domain which involves the inputs collected from speech, text and videos. The following open problems will be considered for grand challenges and accordingly applications will be invited to solve these problems:

i. Video and Speech Analytics for Elderly Health Care and Teaching especially abled Students: video and speech analytics tools will be designed for elderly health care and teaching especially abled students.

ii. Multi-modal AI for Telehealth: AI based telehealth systems (moving beyond Telemedicine) is destined to be a part of better healthcare delivery mechanism – especially in post COVID context. This is especially visible in areas of telehealth innovations where AI applications are used to support, supplement or develop new remote healthcare models and increase access to millions. According to WHO's eHealth observatory survey, AI in the telemedicine field is directly supplementing innovations in these areas: Tele-radiology,Tele-pathology, Tele-dermatology, and Telepsychiatry.

iii. Multimodal AI for cancer prognosis prediction: Different information collected from patients like tissue-images, clinical information, genetic information, copy-number variations etc. can be combined together to design better disease prognosis/ detection system. Some multimodal breast cancer data sets are freely available. These can be used for designing the grand challenges.

iv. Multimodal AI for translational bioinformatics and drug discovery: information collected from different modalities will be utilized to design drugs for various diseases.

v. Knowledge graph creation for health-care systems: knowledge graphs will be created connecting different diseases with their symptoms. The available biomedical literatures will be utilized for creating the knowledge graph.

vi. Multimedia Lifelog Foodlog: This aims to develop a foodlog website and also application software for calorie identification in order to dietary control which has its social impact for development, and make a system called FoodLog. The aim is to keep Food Record for Health Management using multimedia technology. FoodLog: An Easy Way to Record and Archive What We Eat. An image processing engine analyzes the content of the meals, divide these into different meal category based on calorie value contained. Next is to determine what food types appear in the picture and how they fit into the dietary balance. It then estimates the dietary balance values which helps us to monitor our health.

vii. Novel Technique for Capture, Analysis and Visualisation of Human body movement using distributed camera: Development of next generation distributed video sensing systems for understanding human body movements is the aim of this grand challenge. New models of human body movement and structure movement will be used for modelling the movements of single-joint and whole bodies with applications to animation, biomotion, and gait analysis for diagnosing and treating movement-related disorders. The given research efforts enable novel approaches for realistic animation and the detection of indirect variations in movement, leading to better diagnostic tools and personalized programs for rehabilitation of movement disorders.

C). Grand Challenge Set-3: Multimodal AI for Robotics

Industrial and Social robotics are coming of age and increasingly being adopted, both in large industry, MSME and in household sectors. Several grand challenges will be organized to design various robots to create appropriate AI techniques (such as visual recognitions, spatial reasoning, reinforcement learning) for such robotic firmware to impart improved learning, cognitive and functional capabilities in the fields. The robots will consider inputs collected from speech/video/text.

i) Design of Robots operating in hazardous environment (such as infectious disease treatment, disaster recovery)

ii) Design of Robots operating in remote operations (mining, agriculture)

iii) Design of social robotics (educational assistants, robotic companions for elders).

iv) Conversational agent to assist people with mental health : A large percentage of Indian population is suffering from various mental diseases like depression, ***bipolar disorder***, schizophrenia and other psychoses, dementia, and developmental disorders including autism. Conversational agents can be designed to assist people with mental health. First based on the conversations with the agent, the correct mental health will be detected and then conversational agent can generate some motivational responses based on that. Teams will be invited to work on mental health

D). **Grand Challenge Set-4: Speech, Text and Video Analytics for Security**

As a part of this theme some grand challenges will be organized to apply speech, text and video analytics for solving various problems related to security. We will mainly target the following problem statements:

i) **Online Human Behavior Detection:** Online Human behaviour detection and recognition in untrimmed videos are very challenging computer vision task. Techniques are required to be developed for activity classification and detection.

ii) **Active Authentication on mobile devices:** Security and privacy in mobile devices becomes very important as the loss of a mobile device could compromise personal data of the user. To deal with this problem, Active Authentication (AA) systems needs to be developed which users can continuously monitor the initial access to the mobile device.

iii) **Automatic target verification and identification for air borne surveillance video.**: Nowadays, with the rapid development of consumer Unmanned Aerial Vehicles (UAVs), visual surveillance by utilizing the UAV platform has been very attractive. Most of the research works related to UAV captures visual data, mainly focused on the tasks of object detection and tracking. However, limited attention has been paid to the task of person identification which has been widely studied in ordinary surveillance cameras

iv) **Detecting people looking at each other in videos**: Capturing the 'mutual gaze' of people is essential for understanding and interpreting the social interactions between them. In this grand challenge, we address the problem of detecting people looking at each other in video sequences. We need to focus on some of the important problems related to mutual gaze. (a) Two people talking to each other but not looking each other. (b) Looking at each other with eye occluded and (c) Looking at each other with very close eye.

v) Pose, gait and activity based exact human and their activity detection from Video: The objective of this work is to detect human and their activity from video using different pose, gait and activity. These three descriptions aim to recognize exact human and their activity from a normal and crowded video. The vision-based human detection research is the basis of many applications including video surveillance, health care, and human-computer interaction.

17. Section-17: Evaluation

Systematic and regular evaluation will be an important activity in our proposed i-Hub to ensure that the key objectives are met. The key indicators for evaluating the projects will be qualitative as well as quantitative. These indicators or evaluation metrics will be used to monitor the progress of the TIH with respect to the inputs, activities, outcomes and impacts. i-Hub will adopt a mechanism to regulate an evaluation policy to demonstrate its commitment to the funding agency, beneficiaries and implementing partners by proper utilization of funds, taking appropriate actions and delivering the outputs as agreed. Evaluation will be carried for the

1. Technology Development (R&D),
2. Center of Excellence,
3. HRD & Skill Development,
4. Innovation, Entrepreneurship& Start-ups and
5. International Collaborations.

The i-Hub's monitoring committee will conduct continuous evaluation of the projects undertaken in the hub to ensure that the deliverables are met in time. The continuous monitoring and evaluation will provide the funding agencies, owners and the other stakeholders with the necessary information on the progress relative to the targets.

1. **Accountability**: i-Hub will adopt a mechanism to regulate an evaluation policy to demonstrate its commitment to the funding agency, beneficiaries and implementing partners by proper utilization of funds, taking appropriate actions and delivering the outputs as agreed. The Hub's monitoring committee will conduct continuous evaluation of the projects undertaken in the Hub to ensure that the objectives are met as per the timelines specified.
2. **Operational management/Implementation**: Hub's management committee will co-ordinate the human, financial and physical resources committed.
3. **Strategic management and Capacity building**: Mission HGB along with the experts at the national and international levels will critically analyze the activities of the centre by setting goals to the investigators, staffs, and partners. The committee will closely monitor that the intended deliverables from each project are met in time. An ecosystem will be setup for the commercialization of the products, tools and prototypes

developed as part of the different projects undertaken in the centre.

4. **Evaluation of Project by the Working Groups (WG) of i-Hub**
    a. Hub will create WG to evaluate and recommend the projects/ schemes under the Mission
    b. The WG shall have powers to recommend projects, purchases and expenses of all types within the technical and financial scope of the Mission.
    c. The WG will be constituted with the experts drawn from stakeholders, such as Government, academia, research institutions, end-users organizations and industry.
    d. The WG would meet at least twice in a year.
5. **Evaluation by setting by Project-specific Evaluation Committee**
    a. For each project, there will be a Project Review Steering Group (PRSG) that will monitor the progress of the project.
    b. PRSG will ensure that project objectives and deliverables are met in a timely manner.
    c. The PRSG will hold its evaluation at least twice in a year.
    d. Each project will be evaluated in terms of Quantitative and Qualitative parameters. Every project will be asked to state its very clearly defined evaluation parameters.
    e. Quantitative parameters will include (*but not limited to*): (i).No. of publications from the project; (ii). No. of patents; (iii). No. of technologies transferred; (iv). No of Postdocs/PhDs/MTech/BTechs guided; (v). Project-specific Evaluation metrics
    f. Qualitative parameters will include (*but not limited to*) (i). Project specific Human Evaluation metrics; (ii). Whether the outputs from the projects are beneficial to the community at large; (iii). Whether the products/prototypes/tools developed as part of the project satisfy the stated objectives of national priority missions; (iv). Whether the research undertaken advances the field of CPS in general and "Speech, Video and Text" Analytics in particular.
    g. Every project will be asked to prepare the reports and documentation to submit to the PRSG before the stated evaluation date.
    h. Demonstration of the prototypes and/or tools will be given priority right from the very beginning of the project.
6. "***New Normal***"- In conformity with new normal, we will have the provision for online evaluation of the project.
7. The Chair professors and Faculty fellows under this scheme will closely mentor the various projects; will hold discussion at least once in a month with the investigators, and staffs involved in the individual project; ensure that the knowledge generation,

prototype development, product development, startup and/or industry engagement.

8. Performance Evaluation of Chair Professors and Faculty Fellows will be held by the HGB and External Experts.

18. References

1. Bahdanau, D., Cho, K., and Bengio, Y.Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473 (2014).

2. Adeel, Ahsan, Mandar Gogate, and Amir Hussain. "Contextual deep learning-based audio-visual switching for speech enhancement in real-world environments." Information Fusion 59 (2020): 163-170.

3. Engel, Jesse, et al. "DDSP: Differentiable Digital Signal Processing." arXiv preprint arXiv:2001.04643 (2020).

4. Purwins, Hendrik, et al. "Deep learning for audio signal processing." IEEE Journal of Selected Topics in Signal Processing 13.2 (2019): 206-219.

5. N. Nandan, SudhanMajhi, H.C. Wu, "Secure Beamforming for MIMO-NOMA Based Cognitive Radio Network," IEEE Communication Letters, vol. 22, no. 8, pp. 1708-1711, Aug. 2018.

6. N. Nandan, SudhanMajhi, and H. C. Wu, "Maximizing Secrecy Capacity of Underlay MIMO-CRN through Bi-Directional Zero-Forcing Beamforming," IEEE Transactions on Wireless Communications, vol.17, no. 8, pp. 5327-5337, Aug. 2018.

7. Xiaohang Song, NithinBabu, Wolfgang Rave, Sudhan Majhi, and Gerhard Fettweis, "Two-Level Spatial Multiplexing using Hybrid Beamforming Antenna Arrays for mm Wave Communications," IEEE Transactions on Wireless Communications, vol. 17, no. 7, pp.4830-4844, July 2018.

8. Sudhan Majhi, N. Nandan, "Secrecy Capacity Analysis of MIMO System over Multiple Destinations and Multiple Eavesdroppers," Wireless Personal Communications, Springer , vol. 100, no. 3, pp. 1009-1022, 2018.

9. N. Nandan and SudhanMajhi, "Secrecy Outage Analysis by Applying Bi-directional Beamforming in Underlay MIMO-CRN," Accepted in 14th International Wireless Communications and Mobile Computing Conference, Cyprus, 2018.

10. Pecorella, T.; Brilli, L.; Mucchi, L. The Role of Physical Layer Security in IoT: A Novel Perspective. *Information* **2016**, *7*, 49.

11. Yinian Mao and Min Wu, "A joint signal processing and cryptographic approach to multimedia encryption," in *IEEE Transactions on Image Processing*, vol. 15, no. 7, pp. 2061-2075, July 2006.

12. VasileiosMavroeidis, KamerVishi, Mateusz D. Zych, AudunJøsang , "The Impact of Quantum Computing on Present Cryptography"International Journal of Advanced Computer Science and Applications (IJACSA), 9(3), 405-414, March 2018

13. M. Wang, G. Yang, J. Lin, A. Shamir, S. Zhang, S. Lu, S. Hu,"Deep online video stabilization," arXiv preprint arXiv:1802.08091, 2018.

14. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778, 2016.

15. Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. Science, 359(6380), 1146-1151.

16. Iwanek, K. (2018). WhatsApp, fake news? The internet and risks of misinformation in India. The Diplomat, 30.

17. William Yang Wang. 2017. "Liar, Liar Pants on Fire": A new benchmark dataset for fake news detection. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). http://aclweb.org/anthology/P17-2067.

18. See, A., Liu, P. J., and Manning, C. D.Get to the point: Summarization with pointer-generator networks. arXiv preprint arXiv:1704.04368 (2017).

19. Zhu, J., Li, H., Liu, T., Zhou, Y., Zhang, J., and Zong, C. MSMO: Multimodal summarization with multimodal output. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (2018), pp. 4154–4164

20. L. Monostori et al., Cyber-physical systems in manufacturing, CIRP Annals - Manufacturing Technology, 65 (2016) 621–64.

21. B. Esmaeilian, S. Behdad, B. Wang, The evolution and future of manufacturing: A review, Journal of Manufacturing Systems, 39 (2016) 79–100.

22. J. Lee, H. D. Ardakani, S. Yang, B. Bagheri, Industrial big data analytics and cyber-physical systems for future maintenance & service innovation, Procedia CIRP 38 (2015) 3 – 7.

23. E. Uhlmann, C. Geisert, N. Raue, C. Gabriel, Situation Adapted Field Service Support Using Business Process Models and ICT Based Human-Machine-Interaction, Procedia CIRP 47 (2016) 240 – 245.

24. D. Wua, S. Liub, L. Zhang, J. Terpennya, R. X. Gao, T. Kurfess, J. A. Guzzo, A fog computing-based framework for process monitoring and prognosis in cyber-manufacturing, Journal of Manufacturing Systems 43 (2017) 25–34.

25. R. Coppel, J. V. Abellan-Nebot, H. R. Siller, C. A. Rodriguez, F. Guedea, Adaptive control optimization in micro-milling of hardened steels—evaluation of optimization approaches, Int J AdvManufTechnol (2016) 84:2219–2238.

26. Y-S Hong, H-S Yoon, J-S Moon, Y-M Cho, S-H Ahn,Tool-Wear Monitoring during Micro-End Milling usingWavelet Packet Transform and Fisher's LinearDiscriminant, Int. J. Prec. Eng. Manuf. Vol. 17, No. 7, pp. 845-855.

27. W-H Hsieh, M-C Lu, S-J Chiou, Application of backpropagation neural network for spindle vibration-based tool wear monitoring in micro-milling, Int J AdvManufTechnol (2012) 61:53–61.

28. K. Patra, A.K. Jha, T. Szalay, J. Ranjan, L. Monostori,Artificial neural network based tool condition monitoring in micromechanical peck drilling using thrust force signals, Precision Engineering 48 (2017) 279–291.

29. J.Ranjan, K. Patra, T.Szalay, M. Mia, M. K. Gupta, Q. Song, V.A.Pashnyov, D. Y.Pimenov, Artificial intelligence based hole quality prediction in micro-drilling using multiple sensors, Sensors, 2020, 20(3), 885

30. Alice Reina, ÁdámKocsis, Angelo Merlo, IstvánNémeth, and Francesco Aggogeri: Maintenance decision support for manufacturing systems based on the minimization of the life cycle cost, Procedia CIRP, Volume 57, (2016), 674-679

31. Dudhal, S.M ., Jonwal, B. S., Chaudhari, H. P. 2014. Waste segregation using programmable logic controller. International Journal for Technological Research in Engineering, 1(8), 593-595.

32. Sharmila N., Bansilal, TavreMilind. 2013. A Remote Controlled Metal Sorting and Cleaning System. Proceedings of TEQIP II sponsored National Conference on Wireless Communication, Signal Processing, Embedded Systems-WiSE.

33. Manjunatha V G. 2014. Camera Based Color Identification Robot for Typecasting. International Journal of Engineering Research & Technology (IJERT). 3(3), 32-37.

34. S. Kamijo, M. Harada, and M. Sakauchi, "Incident detection based on semantic hierarchy composed of the spatiotemporal MRF model and statistical reasoning," in Proc. IEEE Int. Conf. Syst., Man, Cybern.Oct. 2004, vol. 1, pp. 415–421.

35. Z.Fu, W.Hu, T.Tan, "Similarity Based Vehicle Trajectory Clustering and Anomaly Detection", in Proc. Intl. Conf. on Image Processing (ICIP'05), vol 2, pp 602-605, 2005.

36. Y. Yang, Z. Cui, J. Wu, G. Zhang, and X. Xian, "Trajectory analysis using spectral clustering and sequence pattern mining," Journal of Computational Information Systems , vol. 8, no. 6, pp. 2637–2645, 2012.

37. M. Artetxe, G. Labaka, E. Agirre, and K. Cho. 2018. Unsupervised neural machine translation. In Proceedings of ICLR 2018

38. Lample, Philipp Koehn, Vishrav Chaudhary, and Marc'Aurelio Ranzato. "Two New Evaluation Datasets for Low-Resource Machine Translation: Nepali-English and Sinhala-English." arXiv preprint arXiv:1902.01382 (2019).

39. Sukanta Sen, Kamal Kumar Gupta, **Asif Ekbal** and Pushpak Bhattacharyya (2019). Multilingual Unsupervised NMT using Shared Encoder and Language-Specific Decoders. In *Proceedings of Association for Computational Linguistics* **(ACL)**, *pp. 3083-3089*.

40. Luong, Minh-Thang and Manning, Christopher D. "Stanford Neural Machine Translation Systems for Spoken Language Domain." International Workshop on Spoken Language Translation, 2015.

41. Zoph, Barret, Deniz Yuret, Jonathan May and Kevin Knight. "Transfer Learning for Low-Resource Neural Machine Translation." *EMNLP* (2016).

42. Saunders et al., 2019] Saunders, Danielle, Felix Stahlberg, Adria de Gispert, and Bill Byrne. "Domain Adaptive Inference for Neural Machine Translation." In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics 2019*.

43. Anoop Kunchukuttan, Pratik Mehta and Pushpak Bhattacharyya, The IIT Bombay English-Hindi Parallel Corpus, LREC 2018.

44. Bojar, O., Diatka, V., Rychlý, P., Stranák, P., Suchomel, V., Tamchyna, A., and Zeman, D. 2014. HindEnCorp- Hindi-English and Hindi-only Corpus for Machine Translation. In Proceedings of the Ninth International Conference on Language Resources and Evaluation, pages 3550–3555, Reykjavik, Iceland.

45. Web: onlinedst.gov.in

Annexure 1 : IIT Patna TIH Project Cost Details

Section 1 : Project Cost Summary

Table No: Fin-1 : Recurring and Non Recurring Year-wise estimated costs(in Rs.Crores)

S No	Budget Head	1st Yr	2nd Yr	3rd Yr	4th Yr	5th Yr	Total
1	Recurring	11.75	20.00	30.00	11.80	5.65	79.20
2	Non-Recurring	7.50	13.00	8.00	1.20	1.10	30.80
	Grand Total in Rs Lakhs	19.25	33.00	38.00	13.00	6.75	**110.00**
	Already Received						
1	Recurring	11.75	0.00	0.00	0.00	0.00	11.75
2	Non-Recurring	7.50	0.00	0.00	0.00	0.00	7.50
	Total Receipts	19.25	0.00	0.00	0.00	0.00	**19.25**

Table No: Fin-2: TIH Cost Analysis - (in Rs. Lakhs)

S No	TIH Key Result Areas	1st Yr	2nd Yr	3rd Yr	4th Yr	5th Yr	Total
	Recurring						
1	Technology Development	600	700	900	0	0	2200
2	Centers of Excellence (Vishleshan)	21	41	144	144	141	491
3	HRD & Skill Development	354	707	769	339	130	2299
4	Innovation, Entrepreneurship, and Start-ups Ecosystem	150	400	903	443	168	2064
5	International collaborations	0	23	155	125	0	303
6	TIH Management Unit	50	129	129	129	126	563
	A. Sub Total (Recurring)	1175	2000	3000	1180	565	7920
	Non-Recurring						
1	Technology Development	50	100	100	0	0	250
2	Centers of Excellence (Vishleshan)	300	523	400	60	60	1343
3	HRD & Skill Development	300	360	6	6	5	677
4	Innovation, Entrepreneurship, and Start-ups Ecosystem	0	247	224	49	40	560
5	International collaborations	0	0	50	0	0	50
6	TIH Management Unit	100	70	20	5	5	200
	B. Sub Total (Non-Recurring)	750	1300	800	120	110	3080
	Grand Total in Rs Lakhs	750	1300	800	120	110	3080

Table No: Fin-3: TIH Cost Analysis - Key Result Areawise (in Rs. Lakhs)

S No	TIH Key Result Areas	Budget Head	1st Yr	2nd Yr	3rd Yr	4th Yr	5th Yr	Total
1	Technology Development	Recurring	600	700	900	0	0	2200
		Non-Recurring	50	100	100	0	0	250
		Capital	0	0	0	0	0	0
		Sub-Total	650	800	1000	0	0	2450
2	Centers of Excellence (Vishleshan)	Recurring	21	41	144	144	141	491
		Non-Recurring	300	423	350	50	50	1173
		Capital	0	100	50	10	10	170
		Sub-Total	321	564	544	204	201	1834
3	HRD & Skill Development	Recurring	354	707	769	339	130	2299
		Non-Recurring	300	360	6	6	5	677
		Capital	0	0	0	0	0	0
		Sub-Total	654	1067	775	345	135	2976
4	Innovation, Entrepreneurship, and Start-ups Ecosystem	Recurring	150	400	903	443	168	2064
		Non-Recurring	0	115	124	29	20	288
		Capital	0	132	100	20	20	272
		Sub-Total	150	647	1127	492	208	2624
5	International collaborations	Recurring	0	23	155	125	0	303
		Non-Recurring	0	0	50	0	0	50
		Capital	0	0	0	0	0	0
		Sub-Total	0	23	205	125	0	353
6	TIH Management Unit	Recurring	50	129	129	129	126	563
		Non-Recurring	75	0	0	0	0	75
		Capital	25	70	20	5	5	125
		Sub-Total	150	199	149	134	131	763
	Total	Recurring	1175	2000	3000	1180	565	7920
		Non-Recurring	725	998	630	85	75	2513
		Capital	25	302	170	35	35	567
	Grand Total in Rs Lakhs		1925	3300	3800	1300	675	11000

Table No: Fin-4: Key Result Area-wise and year-wise estimated costs(in Rs. Lakhs)

S No	TIH Key Result Areas	1st Yr	2nd Yr	3rd Yr	4th Yr	5th Yr	Total	%
1	Technology Development	650	800	1000	0	0	2450	22.27
2	Establishment of CoEs	321	564	544	204	201	1834	16.67
3	HRD & Skill Development	654	1067	775	345	135	2976	27.05
4	Innovation, Entrepreneurship and Start-up ecosystem	150	647	1127	492	208	2624	23.85
5	International collaborations	0	23	205	125	0	353	3.21
6	TIH Management Unit	150	199	149	134	131	763	6.94
	Total cost in Rs Lakhs	**1925**	**3300**	**3800**	**1300**	**675**	**11000**	**100.00**

Table No: Fin-5 : Budget for Permanent Facility for TIH (in Rs. Lakhs)

S No	Budget Head	1st Yr	2nd Yr	3rd Yr	4th Yr	5th Yr	Total
1	Construction of permanent Facility for TIH	100	300	800	100	100	1500
	Grand Total in Rs Lakhs	100	400	800	100	100	**1500**

Notes :

1. Funds are requested to create a permanent facility for TIH at the expense of Rs 15 Cr
2. This is in addition to the budget for key result areas as given in Tables Fin-1, Fin-2, Fin-3 and Fin-4

Section 2 : Key Result Area : Year-wise Targets and Estimated Costs

Note: Tables in this section are numbered by the serial number of corresponding key result area in Table No . Fin-2 in Annexure 1 Section 1

Table No: Fin-3.1 Estimated Expenditure for Key Result Area -Technology Development(in Rs. Lakhs)

S No	Technology Development		Targets						Budget					
		Unit Cost	Yr1	Yr2	Yr3	Yr4	Yr5	Total	Yr1	Yr2	Yr3	Yr4	Yr5	Total
1	Projects A	50	14	16	20	0	0	50.00	650	800	1000	0	0	2450
	Total		14	16	20	0	0	50.00	0	0	0	0	0	2450

Notes :

1. The focus here is on projects - basic research as well as translational research. Steps afterwards is taken care of under the entrepreneurship umbrella
2. Creation of the centralized CPS technical facility is given under CoE budgets
3. Seed fund for product development is given under innovation, entrepreneurship and startup ecosystem budgets
4. First year non-recurring budget also includes purchase of centralized equipments.

Table No: Fin-3.2 Estimated Expenditure for Key Result Area -Setting up of CoE (in Rs. Lakhs)

S No	CoE Setup		Targets						Budget					
		Unit Cost	Yr1	Yr2	Yr3	Yr4	Yr5	Total	Yr1	Yr2	Yr3	Yr4	Yr5	Total
1	CoE Setup and operations	699	0.14	0.21	0.25	0.20	0.20	1.00	107	160	194	154	151	766
2	Centralized Technical Facility (Equipment and Infrastructure)	935	0.20	0.38	0.33	0.05	0.05	1.00	214	404	350	50	50	1068
	Total								321	564	544	204	201	1834

Notes :

1. CoE is the major functional subunit of TIH. It manages all activities related to knowledge creation, HRD and skill development, Centralized CPS facility and international collaborations
2. Centralized technical facility will be accessible to all projects and startups working under TIH support

Table No: Fin-3.3 : Estimated Expenditure for Key Result Area -HRD & Skill Development (in Rs. Lakhs)

S No	HRD & Skill Development		Targets						Budget					
		Unit Cost	Yr1	Yr2	Yr3	Yr4	Yr5	Total	Yr1	Yr2	Yr3	Yr4	Yr5	Total
1	**CHANAKYA - UG**													
1.1	Graduate Internships	1	10	50	70	80	40	250	30	50	70	80	40	270
1.2	Development Fund (For Projects done under Graduate Internships)	1	5	25	35	40	20	125	50	25	35	40	20	170
1.3	CPS Infrastructure development fund (Speech, Video and Text Analytics)	100	0	1	0	0	0	1	0	100	0	0	0	100
2	**CHANAKYA - PG**								0	0	0	0	0	
2.1	Post Graduate Fellowships (M Tech/ MS)	3	10	10	20	15	2	57	50	30	60	45	6	191
2.2	Development Fund (For Projects done under PG Fellowships)	2	3	10	20	15	2	50	50	20	40	30	4	144
2.3	CPS Infrastructure development fund (External program upgrade)	100	0	1	0	0	0	1	0	100	0	0	0	100
3	**CHANAKYA-DF (Doctoral Fellowships)**	20	25	15	12	0	0	52	269	352	240	0	0	861
4	**CHANAKYA-PDF (Post-Doctoral Fellowships)**	20	10	8	10	4	0	32	150	160	200	80	0	590
5	**CHANAKYA-Faculty Fellowship**	30	0	2	1	0	0	3	0	60	30	0	0	90
6	**CHANAKYA-Chair Professor**	30	0	2	1	0	0	3	0	60	30	0	0	90
7	**CPS- PSDW (Professional Skill Development Workshop)**	5	0	1	2	2	1	6	0	5	10	10	5	30
8	**CPS-New PG Programme (Speech, Video and Text Analytics)**	400	0.14	0.49	0.13	0.13	0.13	1	55	195	50	50	50	400
9	**CPS-Advanced Skill Training School (ASTS)**	10	0	1	1	1	1	4	0	10	10	10	10	40
	Total								**654**	**1067**	**775**	**345**	**135**	**2976**

Notes :

1. As part of this budget, one external institute will be supported to upgrade their PG program
2. Advances skill development school activities will be done in tie up with Nielet or a similar entity focussed on skill training
3. Each chair professor is expected to guide and mentor 10 projects
4. The projects (under Technology Development Key Result Area) may be supplemented by HRD funds for PhD/PDF, subject to a cumulative maximum of 2 PhD scholars and one PDF scholar in a project. This provision may be utilized by projects both internal and external to IIT Patna.
5. UG internship and PG fellowships will be selected through open call and can be applied by students all over the country.
6. PhD Fellowship unit cost revised to 20 Lakhs from 17 Lakhs in line with latest guidelines for such fellowships.
7. PDF Fellowship unit cost is revised to 20 Lakhs from 30 Lakhs by limiting the period to 2 years for the fellowship.
8. All other fellowships are estimated as per unit cost guidelines of the mission

Table No: Fin-3.4 : Estimated Expenditure for Key Result Area -Innovation Entrepreneurship and Startup Ecosystem (in Rs. Lakhs)

S No	Innovation, Entrepreneurship, and Start-ups Ecosystem		Targets						Budget					
		Unit Cost	Yr1	Yr2	Yr3	Yr4	Yr5	Total	Yr1	Yr2	Yr3	Yr4	Yr5	Total
1	CPS-Technology Business Incubator (TBI)	917	0.07	0.12	0.45	0.17	0.19	1	57	108	390	148	164	867
2	CPS-GCC - Grand Challenges and Competitions	350	0.13	0.38	0.57	0.00	0.00	1	45	132	200	0	0	350
3	CPS-Promotion and Acceleration of Young and Aspiring technology entrepreneurs (CPS-PRAYAS)	270	0.10	0.34	0.50	0.15	0.09	1	28	93	134	41	24	270
4	CPS-Entrepreneur In Residence (CPS-EIR)	90	0	10	10	5	0	25	0	36	86	18	0	90
5	CPS- Start-up (Fund included in CPS-SSS)	10	2	6	10	10	2	30	0	0	0	0	0	0
6	CPS-Dedicated Innovation Accelerator (DIAL)	200	0.00	0.34	0.34	0.33	0.00	1	0	68	67	65	0	200
7	CPS-Seed Support System (CPS-SSS)	720	0.03	0.29	0.35	0.31	0.03	1	20	210	250	220	20	720
	Total								150	647	1127	492	208	2497

Notes :

1. CPS TBI is the second major component of TIH and it will be managed by collaborating with Incubation Centre IIT Patna, the existing technology incubator to avoid redundancy. The centre has its own space and lab facilities which will be augmented to support CPS pre-incubatees, prayasees and startups.
2. 10 Prayasees, 30 Startups under CPS-startup and 14 startups under CPS-DIAL will be supported during the project period, with minimum number of startups supported not less than 40.
3. Seed fund of Rs 10 Lakh per CPS startups and Rs 30 Lakh per CPS DIAL startups will be supported from CPS -SSS
4. CPS -GCC will conduct 10 events and will also take care of TIH marketing
5. Target for EIR is set to 25

Table No: Fin-3.5 : Estimated Expenditure for Key Result Area -International Collaborations (in Rs. Lakhs)

S No	International Collaborations		Targets						Budget					
		Unit Cost	Yr1	Yr2	Yr3	Yr4	Yr5	Total	Yr1	Yr2	Yr3	Yr4	Yr5	Total
1	International Collaborations	353	0	0	1	0	0	1	0	23	205	125	0	353
	Total								0	23	205	125	0	353

Notes :

1. 1 Collaboration will be targeted with 5 projects.
2. Total of 10 publications and 2 patents are expected

Table No: Fin-3.6 : Estimated Expenditure -TIH Management Unit(in Rs. Lakhs)

S No	TIH Management Unit		Targets						Budget					
		Unit Cost	Yr1	Yr2	Yr3	Yr4	Yr5	Total	Yr1	Yr2	Yr3	Yr4	Yr5	Total
1	TIH Management Unit	763	0.20	0.26	0.20	0.18	0.17	1	150	199	149	134	131	763
	Total								150	199	149	134	131	763

Notes :

1. TIH Management Unit is responsible for setting up and supervising over all operations of TIH. It also provides horizontal services to all sub units such as financial management, accounting, legal, administrative services etc

Section 3 : Key Result Area Components : Year-wise estimated costs

Section Note:

1. Tables in this section are numbered as < Table No of the Key Result Area in Annexure 1>'.'<Serial Number of the component of the Key Result Area that is being detailed>
2. For those components where standard unit cost estimates of mission is followed, a separate detail table is not provided

1. Key Result Area : Technology Development

Table No: Fin-3.1.1 Estimated Expenditure Details for Key Result Area -Technology Development (in Rs. Lakhs)

S No	Budget Head	ESTIMATED COST IN Rs LAKHS					
		Ist Yr	2nd Yr	3rd Yr	4th Yr	5th Yr	Total
A.	Recurring						
	1. Project Staff	400	450	600	0	0	1450
	2. Domestic Travel	60	60	90	0	0	210
	3. Contingencies	30	30	30	0	0	90
	4. Consumables	30	30	50	0	0	110
	5. Miscellaneous	20	30	30	0	0	80
	6. Over Heads	60	100	100	0	0	260
	Sub-Total	600	700	900	0	0	2200
B.	Non-Recurring						0
	1. Equipment	50	100	100	0	0	250
	Sub-Total	50	100	100	0	0	250
C.	Capital	0	0	0	0	0	0
	Grand Total	650	800	1000	0	0	2450

Notes : Technology Development : Project

1. This table gives the budget for 81 technology development projects at unit cost of Rs 50 lakhs.
2. The above budget is a standard outlay and is used for budgeting purposes. However, actual number of project can be upwards of 81 projects and project size may vary between Rs 20 Lakh to Rs 60 Lakhs depending upon the project content, duration etc.
3. At least 25% of total project budget is proposed to be used for external projects
4. CoE will attempt to ensure that 70% of projects are in the area of knowledge creation and 30% projects are on translational research leading to tools and platform creation, resulting in at least 15 products/prototypes.
5. Every project will have an industry collaboration, which will contribute to the project in cash or in kind (facility, mentorship, technology adoption etc). The contribution from the industry will be over and above this budget
6. 50 publications and 20 patent are expected from these projects
7. The project manpower include researchers (PhD, PDF), PG Fellowship holders or other project staff as per CoE project selection committee approval and budgets. At least 25 person increase in CPS research base (PDF+PhD) through these project (remaining 50 increase in CPS research base will be through doctoral and post doctoral fellowships)
8. CoE will release the project budgets as per project timeline and disbursement plan approved at the time of selection and based on project progress evaluation.
9. Once the project is approved by CoE, the budget for the same will be committed from CoE. Hence the complete budget for a project is shown at the year of approval.

Table No: Fin-3.1.1 (Unit Cost) Estimated Expenditure Details per Project under Technology Development (in Rs. Lakhs)

		ESTIMATED COST IN Rs LAKHS			
S No	**Budget Head**	**Ist Yr**	**2nd Yr**	**3rd Yr**	**Total**
A.	Recurring				
	1. Project Staff	10	10	10	30
	2. Domestic Travel	1.5	2	1	4.5
	3. Contingencies	0.5	0.5	0.5	1.5
	4. Consumables	1	0.5	1	2.5
	5. Miscellaneous	0.5	0.5	0.5	1.5
	6. Over Heads	1.5	1.5	2	5
	Sub-Total	15	15	15	45
B.	Non-Recurring				
	1. Equipment	5	0	0	5
	Sub-Total	5	0	0	5
C.	Capital	0	0	0	0
	Grand Total	20	15	15	50

Notes : Technology Development : Project

The table above gives the budget for a standard project under technology development

1. The above budget is a standard outlay and is used for budgeting purposes. However, actual project size may vary between Rs 20 Lakh to Rs 70 Lakhs depending upon the project content, duration etc.

2. The project manpower can include researchers (PhD, PDF), PG Fellowship holders or other project staff as per CoE project selection committee approval and budgets

3. CoE will release the project budgets as per project timeline and disbursement plan approved at the time of selection and based on project progress evaluation.

2. Key Result Area : CoE

Table No: Fin-3.2.1 Estimated Expenditure Details for Key Result Area -CoE - Overall (in Rs. Lakhs)

S No	Budget Head	ESTIMATED COST IN Rs. LAKHS					
		Ist Yr	2nd Yr	3rd Yr	4th Yr	5th Yr	Total
A.	Recurring						
	1. Project Staff	5	20	75	75	75	250
	2. Domestic Travel	1	2	15	15	12	45
	3. Contingencies	1	2	12	12	12	39
	4. Consumables	1	1	6	6	6	20
	5. Miscellaneous	1	1	6	6	6	20
	6. Project review and evaluation expenses including travel and honorarium of experts	8	10	24	24	24	90
	7. Marketing expenses	4	5	6	6	6	27
	Sub-Total	21	41	144	144	141	491
B.	Non-Recurring						
	1. Lab R&D Infrastructure & Equipment	250	423	350	50	50	1123
	Sub-Total	250	423	350	50	50	1123
C.	Capital						
	1. Furnishing, Tables, Chairs, Cubicles, Electrical works and other Capex items	50	100	50	10	10	220
	Sub-Total	**50**	**100**	**50**	**10**	**10**	**220**
	Grand Total	**321**	**564**	**544**	**204**	**201**	**1834**

Notes : TIH CoE (Named as IIT Patna Vishleshan i-Hub Foundation) This unit is responsible for Technology development (Creation of centralized advanced CPS labs, academic programs, basic research, tools/platform research, HRD development, industry relations and liaison with commercialization/startup support unit. This unit reports to the TIH CEO.

The table above gives the budget for CoE management
1. Manpower includes CoE Head, coordination teams for UG/PG programs (UG PG program Upgrade, UG/PG fellowships etc) , Research Programs (Projects, PhD, PDF, Faculty and Chair professor fellowships), technical team for Centralized CPS labs management, Liasoning team for industry and TIH - Runway which is the startup facilitation unit.
2. Travel budget is for CoE officials for administrative and program purposes
3. Project selection and progress review is a major activity and includes multiple committees and external experts. It is expected that there will be a minimum of 15 such meetings annually. Hence budget is separately allocated for this, which also includes travel and honorarium for experts in addition to other expenses.
4. The equipment budget is for the centralized facility. Out of 11.23 Cr, 10 Cr will be spent on Equipment and 1.23 Cr will be for Lab infrastructure
5. Capital budget is given for the office and common facilities of CoE management and team, and for creating seating facilities for the researchers and students

Table No: Fin-3.2.1 B Estimated Expenditure Details for Key Result Area -CoE - Centralized Technical Facility(in Rs. Lakhs)

S No	Budget Head	ESTIMATED COST IN Rs LAKHS					
		Ist Yr	2nd Yr	3rd Yr	4th Yr	5th Yr	Total
A.	Recurring	0	0	0	0	0	0
	Sub-Total	0	0	0	0	0	0
B.	Non-Recurring						0
	1. Equipment	200	368	320	38	38	964
	2.Lab Infrastructure	48	50	25	10	10	143
	3. Miscellaneous	2	5	5	2	2	16
	Sub-Total	250	423	350	50	50	1123
C.	Capital	0	0	0	0	0	0
	Grand Total	250	423	350	50	50	1123

Notes :

1. The above table gives the budget for centralized CPS lab facility that will be created under the CoE. This lab will be used by all projects and researchers under the CoE.
2. UG Program and PG program will have separate labs for regular student use. Access to centralized CPS lab will be on a need basis

Table No: Fin-3.2.1 B.B Estimated Expenditure Details for Equipment in Centralized Technical Facility(in Rs. Lakhs)

Sl No	Equipment	Estimated Cost
1	GPU-enabled Servers	813
2	Eye Tracker and similar devices	100
3	Storage, Robotic Simulator, IoT Kits, CAD etc	100
4	Lab infrastructure	110
	Total	1123

Notes :

1. The above table gives the budget for equipment and infrastructure for centralized CPS lab facility
2. This equipment list is not exhaustive. Equipment may be added/changed later based on research requirements subject TIH management approval.
3. In case of a change, it will be ensured that the total budgets are within the total amount projected.

3. Key Result Area : HRD & Skill Development

Table No: Fin-3.3.1.3 : Estimated Expenditure Details for Chanakya UG : Infrastructure Development Fund(in Rs. Lakhs)

S No	Budget Head	ESTIMATED COST IN Rs LAKHS
A.	Recurring	0
	Sub-Total	0
B.	Non-Recurring	
	1. Equipment	80
	2.Lab Infrastructure	20
	3. Miscellaneous	0
	Sub-Total	100
C.	Capital	0
	Grand Total	100

Notes : HRD and Skill Development : UG : CPS Infrastructure development fund (AI & Data Analytics)

1. The table above gives the budget for the lab set up for the B Tech program in Speech, Video and Text Analytics
2. The amount will be used in the Year 2 of the project, when the BTech program will start.

Table No: Fin-3.3.2.3 : Estimated Expenditure Details for Chanakya PG Upgrade : Infrastructure Development Fund(in Rs. Lakhs)

S No	Budget Head	ESTIMATED COST IN Rs LAKHS
A.	Recurring	0
	Sub-Total	0
B.	Non-Recurring	
	1. Equipment	80
	2.Lab Infrastructure	20
	3. Miscellaneous	0
	Sub-Total	100
C.	Capital	0
	Grand Total	100

Notes : HRD and Skill Development : PG Upgrade : CPS Infrastructure development fund (Speech, Video and Text Analytics)

1. The table above gives the budget for the lab set up for upgrading the PG program in Speech, Video and Text Analytics in one college
2. The amount will be used in the Year 2 of the project.
3. The institute will be selected based on application and evaluation of the experience and competence to conduct the program.
4. The budget only covers Lab and technical infrastructure set up and excludes student fellowships

Table No: Fin-3.3.7 : Estimated Expenditure Details for Professional Skill Development Workshop (in Rs. Lakhs)

S No	Budget Head	ESTIMATED COST IN Rs LAKHS					
A.	Recurring	Ist Yr	2nd Yr	3rd Yr	4th Yr	5th Yr	Total
	1. Contingencies	0.00	1.00	2.00	2.00	1.00	6.00
	2. Travel, honorarium to experts etc	0.00	2.40	4.80	4.80	2.40	14.40
	3. Miscellaneous	0.00	0.60	1.20	1.20	0.60	3.60
	Sub-Total	0.00	4.00	8.00	8.00	4.00	24.00
B.	Non-Recurring						
	1. Teaching Material	0.00	1.00	2.00	2.00	1.00	6.00
	Sub-Total	0.00	1.00	2.00	2.00	1.00	6.00
C.	Capital	0.00	0.00	0.00	0.00	0.00	0.00
	Grand Total	0.00	5.00	10.00	10.00	5.00	30.00

Notes : HRD and Skill Development : CPS Professional Skill Development Workshop

1. The table above gives the budget for professional skill development workshop for 50 people

2. The amount will be used in the Year 2 onwards of the project. There will be 6 such workshops, catering to 300 people

3. The agency to conduct this program will be selected based on experience and competence to conduct the program.

Table No: Fin-3.3.8 : Estimated Expenditure Details for New PG Program in Speech, Video and Text Analytics (in Rs. Lakhs)

S No	Budget Head	ESTIMATED COST IN Rs LAKHS					
A.	Recurring	Ist Yr	2nd Yr	3rd Yr	4th Yr	5th Yr	Total
	Fellowship	0	45	45	45	45	180
	Contingencies	2.5	2.5	2.5	2.5	2.5	12.5
	Miscellaneous	2.5	2.5	2.5	2.5	2.5	12.5
	Sub-Total	5	50	50	50	50	205
B.	Non-Recurring						
	Equipment and lab infrastructure	40	135	0	0	0	175
	Teaching Material	5	5				10
	Books, Journals etc	5	5				10
	Sub-Total	50	145	0	0	0	195
C.	Capital	0	0	0	0	0	0
	Grand Total	55	195	50	50	50	400

Notes : New PG Program : CPS Infrastructure development fund (Speech, Video and Text Analytics)

1. The table above gives the budget to set up and run a new PG program in Speech, Video and Text Analytics at IIT Patna

2. The PG course will admit 30 students per year and will start in Year 2 of the project. Total admission in project duration will be 120.

3. 50% of fellowship (1.5 Lakhs per student) will be met from this budget and the other 50% will be mobilized by IIT Patna

4. An exclusive lab will be set up for this program

Table No: Fin-3.3.9 : Estimated Expenditure Details for Advanced Skill Training School (in Rs. Lakhs)

S No	Budget Head	Amount in Rs Lakhs					
		Year-1	Year-2	Year-3	Year-4	Year-5	Total
A.	Recurring						
	Contingencies	0	2	2	2	2	8
	Travel, honorarium to experts etc	0	2	2	2	2	8
	Miscellaneous	0	2	2	2	2	8
	Sub-Total	0	6	6	6	6	24
B.	Non-Recurring			0		0	
	Equipment	0	3	3	3	3	12
	Teaching Material	0	1	1	1	1	4
	Sub-Total	0	4	4	4	4	16
C.	Capital	0		0		0	0
	Grand Total	0	10	10	10	10	40

Notes : CPS Advanced Skill Training School (Speech, Video and Text Analytics)

1. The table above gives the budget to run four instances of skill development school

2. The school will admit 30 students per instance and will start in Year 2, 3, 4 and 5 of the project, catering to 120 people

3. The budget will be allocated to a specialized skill development facility like Nielet or ITS/Polytechnics

4. CoE will release the funds as per unit disbursement plan (See table below).

5. Once the scheme is approved by CoE, the budget for the same will be committed. Hence the complete budget is shown at the year of approval (2nd, 3rd, 4th and 5th year of the project).

Table No: Fin-3.3.9 (Unit Cost) : Estimated Expenditure Details for Advanced Skill Training School (in Rs. Lakhs)

S No	Budget Head	Amount in Rs Lakhs		
		Year-1	Year-2	Total
A.	Recurring			
	Contingencies	1	1	2
	Travel, honorarium to experts etc	1	1	2
	Miscellaneous	1	1	2
	Sub-Total	3	3	6
B.	Non-Recurring			
	Equipment	3	0	3
	Teaching Material	0.5	0.5	1
	Sub-Total	3.5	0.5	4
C.	Capital	0	0	
	Grand Total	6.5	3.5	10

4. Key Result Area : Innovation, Entrepreneurship and Startup Ecosystem

Table No: Fin-3.4.1 Estimated Expenditure Details for CPS TBI (in Rs. Lakhs)

S No	Budget Head	ESTIMATED COST IN Rs. LAKHS					
		Ist Yr	**2nd Yr**	**3rd Yr**	**4th Yr**	**5th Yr**	**Total**
A	Recurring						
	1. Human Resources**(Core Management Team /Mentors and Tech Support Persons /Business Development Professionals)	10	31	55	55	59	210
	2. Travel (@ Rs. 40,000 pm)	2	2	10	5	10	29
	3. Utility and maintenance	5	10	20	20	20	75
	4. Marketing, networking & publicity	3	4	6	4	6	23
	5. Training Programmes, Events, and Start-up-Resonators	5	6	50	25	35	121
	6. Other Administrative Expenses including consumables, printing, publications, books, journals, etc.	5	6	12	12	12	47
	7. Miscellaneous and Contingencies	1	2	6	6	6	21
	Sub-Total	***31***	***61***	***159***	***127***	***148***	***526***
B	Non-Recurring						
	1. D&D Rooms (Dies & Designs, FAB lab)	4	6	10	10	5	35
	2. Office Equipment including state- of-the art communication network, Video Conferencing Facilities	1	0	25	0	0	26
	3. Contingencies for non-recurring expenditure and other items	1	3	6	6	6	22
	Sub-Total	***6***	***9***	***41***	***16***	***11***	***83***
C	Capital						
	1. Renovation/furnishing of space for CPS-TBI ; (20,000 sf ; @ 600 psf);(Furniture / Test Benches / Installations; Incubation Cubicles and Spaces /Interaction Centers) excluding the cost of land & building	10	10	50	5	5	80
	2. Thrust Area Equipment (Equipment /Machineries; Clean Rooms / Test Rigs / IT Systems; Instruments/Tools & Dies/ Measuring Devices, etc)	10	28	140	0	0	178
	Sub-Total	***20***	***38***	***190***	***5***	***5***	***258***
	Grand Total	**57**	**108**	**390**	**148**	**164**	**867**

Notes : TIH TBI : Runway - iHUB This unit is responsible for all Entrepreneurship and Startup Ecosystem Related activities of TIH. Incubation Centre IIT Patna, the existing technology business incubator in IIT Patna will manage this TBI for TIH. The centre has a dedicated 30,000 Sq ft facility where it can host incubatees and labs. The ESDM focus of the centre is complementary with CPS objectives. The centre has supported companies in image processing & voice recognition. The above table gives the budget for the functioning of TBI, which will run the startup related schemes for TIH.

1. Manpower includes TBI Head, coordination teams for EIR, Prayas and Startup schemes.
2. Travel budget is for TBI officials for administrative and program purposes
3. Team selection and progress review, training programs and event participation are major activities and includes multiple committees, external experts etc. Hence budget is separately allocated for this, which also includes travel and honorarium for experts in addition to other expenses.
4. Very minimal marketing budget is allocated, as it is clubbed with grand challenge budget to keep overall budgets within limits
5. As there are existing lab facilities and office and meeting facilities, minimal budgets are allocated to Non recurring head for minor upgrades
6. Renovation budget will support creation of additional space for incubatees, EIRs, Prayas teams etc
7. The equipment budget is to create minimum dedicated CPS lab facilities for startups. For advanced lab access, TIH centralized facility will be utilized.

Table No: Fin-3.4.2 Estimated Expenditure Details for CPS Grand Challenges and Competitions (in Rs. Lakhs)

S No	Budget Head	ESTIMATED COST IN Rs. LAKHS			
		Ist Yr	2nd Yr	3rd Yr	Total
A.	Recurring				
	I. All India Competitions (Operating Costs for 10 challenges under 1 GCC)				
	1. Human Resources	10	10	20	40
	2. Travel, honorarium to experts etc	10	10	20	40
	3. Miscellaneous	5	1	10	16
	4. Marketing, promotion and publicity	5	1	10	16
	5. Networking and training programmes	5	1	10	16
	6. Other administrative expenses including consumables, printing, publications, books, journals etc	9	2	20	31
	II. Awards				
	1. Reward @ Rs 5.00 lakhs per winner for 5 ideas	0	20	20	
	Sub-Total	44	45	110	199
B.	Non-Recurring				
	I. Prototyping Grant/ Seed Fund @Rs 20.00 Lakhs each for 5 winners	0	80	80	160
	Sub-Total	0	80	80	160
C.	Capital				
	1. Competitions location specific arrangements like furniture, tables, chairs, dash boards, product development and demonstration arrangements etc	1	7	10	18
	Sub-Total	1	7	10	18
	Grand Total	45	132	200	377

Notes : CPS : Grand Challenges and Competitions : To conduct about 10 challenges in 2.5 years

1. Marketing budget also includes the marketing activities for TIH CoE and TBI
2. Most of the events will be conducted in Year 2.
3. Marketing efforts of TIH (Both CoE and TBI) will be clubbed with the challenges

Table No: Fin-3.4.3 Estimated Expenditure Details for CPS PRAYAS (in Rs. Lakhs)

S No	**Budget Head**	**ESTIMATED COST IN Rs. LAKHS**					
		Ist Yr	**2nd Yr**	**3rd Yr**	**4th Yr**	**5th Yr**	**Total**
A.	Recurring						
	1. Prototyping Grant/ Seed Fund @Rs 10.00 Lakhs each for 10 ideas	10	40	80	10	0	140
	2. Travel, honorarium to experts etc	2	2	5	2	1	12
	3. Miscellaneous	1	1	1	1	1	5
	4. Other administrative expenses including consumables, printing, publications, books, journals etc	1	1	1	1	1	5
	Sub-Total	14	44	87	14	3	162
B.	Non-Recurring						
	1. Raw material, Spare parts, consumables etc	2	7	10	6	3	28
	2. Fabrication/ Synthesis charges of working model development or process that includes design engineering/ Consultancy/ Testing/ Experts costs etc	2	7	10	6	3	28
	Sub-Total	4	14	20	12	6	56
C.	Capital						
	1. Establishment of PRAYAS Center, Fabrication LAB, location specific arrangements like furniture, tables, chairs, dash boards, product development and demonstration arrangements etc	5	20	15	5	5	50
	2. Operation and maintenance of Fab lab @ Rs 20.00 lakhs per year for 5 years	5	15	12	10	10	52
	Sub-Total	10	35	27	15	15	102
	Grand Total	**28**	**93**	**134**	**41**	**24**	320

Notes : CPS : Prayas: To support 10 prayas teams in 5 years

1. The above table gives budget for one instance of Prayas, with funding of upto 10 lakhs for 10 teams
2. Minimal amounts are kept for facilities and lab maintenance, as existing facilities will be reused.

Table No: Fin-3.4.6 Estimated Expenditure Details for CPS DIAL (in Rs. Lakhs)

S No	Budget Head	ESTIMATED COST IN Rs. LAKHS					
		Ist Yr	2nd Yr	3rd Yr	4th Yr	5th Yr	Total
A.	Recurring						
	I. 3 Years, 3 cohorts, 5 + 5 + 4						
	1. Human Resources	0	20	20	20	0	60
	2. Travel, honorarium to experts etc	0	6	6	6	0	18
	3. Miscellaneous	0	6	6	6	0	18
	4. Marketing, promotion and publicity	0	12	12	12	0	36
	5. Networking and training programmes	0	5	5	5	0	15
	6. Other administrative expenses including consumables, printing, publications, books, journals etc	0	9	8	8	0	25
	Overheads @2 Lakhs per company for 14 companies	0	10	10	8	0	28
	Sub-Total	0	68	67	65	0	200
B.	Non-Recurring						0
	I. Seed Fund @Rs 35.00 Lakhs for 14 companies	0	0	0	0	0	0
	Sub-Total						0
C.	Capital						0
	Training Facility	0	0	0	0	0	0
	Sub-Total	0	0	0	0	0	0
	Grand Total	0	68	67	65	0	200

Notes : CPS : DIAL: CPS TBI will set up DIAL accelerator and targets to support 14 companies over 3 to 4 years with a seed fund support of Rs 30 Lakhs per company

1. Marketing budget also includes the marketing activities for TIH CoE and TBI
2. The DIAL activities will start in Year 2 and will continue till Year 4.
3. There will be three cohorts of 5, 5 and 4 companies each.
4. Seed fund of Rs 30 Lakh per CPS DIAL startups will be supported from CPS -SSS and budgeted there, hence the budgets are not shown in this table

Table No: Fin-3.4.7 Estimated Expenditure Details for CPS Seed Support System (in Rs. Lakhs)

S No	Budget Head	ESTIMATED COST IN Rs. LAKHS					
		Ist Yr	2nd Yr	3rd Yr	4th Yr	5th Yr	Total
A.	Recurring						
	Sub-Total	0	0	0	0	0	0
B.	Non-Recurring						0
	I. Seed Fund @Rs 30.00 Lakhs for 14 companies under DIAL	0	150	150	120	0	420
	I. Seed Fund @Rs 10.00 Lakhs for 30 companies under startup	20	60	100	100	20	300
	Sub-Total	20	210	250	220	20	720
C.	Capital						
	Training Facility	0	0	0	0	0	0
	Sub-Total	0	0	0	0	0	0
	Grand Total	20	210	250	220	20	720

Notes : CPS : Seed Support System :

1. The scheme will support 30 startups with Rs 10 Lakhs each and 14 accelerated companies with Rs 30 Lakhs each. Target to support at least 40 companies.

2. The DIAL funding will be in Year 2 (5 companies), Year 3 (5 companies) and Year 4 (4 companies).

3. The Startup funding will be in Year 1 (2 companies), Year 2 (6 companies), Year 3 (10 companies) and Year 4 (10 companies) and Year 5 (2 companies).

5. Key Result Area : International Collaboration

Table No: Fin-3.5.1 Estimated Expenditure Details for International Collaborations (in Rs. Lakhs)

S No	Budget Head	ESTIMATED COST IN Rs. LAKHS					
		Ist Yr	2nd Yr	3rd Yr	4th Yr	5th Yr	Total
A.	Recurring						
	1. Project Staff	0	5	70	70	0	140
	2. Contingencies	0	2	7	7	0	14
	3. Consumables	0	2	7	7	0	14
	4. Miscellaneous	0	2	8	8	0	16
	5. International travel/ exchange programmes	0	5	25	25	0	50 0
	6. International workshops/conferences/ meetings	0		30			30 0
	7. Over Heads	0	3	8	8	0	16
	Sub-Total	0	19	155	125	0	299
B.	Non-Recurring	0					
	1. Equipment	0	4	50	0	0	54
	Sub-Total	0	4	50	0	0	54
C.	Capital	0	0	0	0	0	0
	Grand Total	0	23	205	125	0	353

Notes : International Collaboration

1. This table gives the budget for 1 International collaboration project at unit cost of Rs 350 lakhs per collaboration.

2. The collaboration will have 3 projects . The partner will contribute equal amount of funds.

3. A total of 6 publications and 2 patents are expected.

4. Each collaboration will run for 2 years

4. CoE will release the project budgets as per approved detailed DPR for the collaboration.

5. Once the project is approved by CoE, the budget for the same will be committed from CoE. Hence the complete budget for a project is shown at the year of approval.

Table No: Fin-3.5.1(Unit Cost) Estimated Expenditure Details for International Collaborations (in Rs. Lakhs)

S No	Budget Head	ESTIMATED COST IN Rs. LAKHS			
		Ist Yr	2nd Yr	3rd Yr	Total
A.	Recurring				
	1. Project Staff	5	70	70	145
	2. Contingencies	2	7	7	16
	3. Consumables	2	7	7	16
	4. Miscellaneous	2	8	8	18
	5. International travel/ exchange programmes	5	25	25	55
					0
	6. International workshops/conferences/ meetings		30		30
					0
	7. Over Heads	3	8	8	19
	Sub-Total	19	155	125	299
B.	Non-Recurring				
	1. Equipment	4	50	0	54
	Sub-Total	4	50	0	54
C.	Capital	0	0	0	0
	Grand Total	23	205	125	353

Notes : International Collaboration

1. This table gives the budget for 3 technology development projects at unit cost of Rs 100 lakhs each, in an international collaboration.
2. The above budget is a standard outlay and is used for budgeting purposes. However, actual number of project will be determined based on the DPR / MoU for the collaboration based on project content, duration etc.
3. Total of 6 publications and 2 patents are expected
4. The project manpower include researchers (PhD, PDF), PG Fellowship holders or other project staff as per guidelines of the collaboration
5. CoE will release the project budgets as per project timeline and disbursement plan approved at the time of selection and based on project progress evaluation.
6. Once the project is approved by CoE, the budget for the same will be committed from CoE. Hence the complete budget for a project is shown at the year of approval.

6. Key Result Area : TIH Management Unit

Table No: Fin-3.6.1 Estimated Expenditure Details for TIH Management Unit (in Rs. Lakhs)

S No	Budget Head	ESTIMATED COST IN Rs. LAKHS					
		Ist Yr	2nd Yr	3rd Yr	4th Yr	5th Yr	Total
A.	Recurring						
	1. Manpower	44	75	75	75	75	344
	2. Domestic Travel (Team)	1	9	9	9	6	34
	3. Contingencies, Consumables. Miscellaneous	1	18	18	18	18	73
	4. Legal/Audit Compliances/ Board-AGM expenses/ Travel and Sitting fee of directors etc	2	24	24	24	24	98
	5. Website creation and management, PR	2	3	3	3	3	14
	Sub-Total	50	129	129	129	126	563
B.	Non-Recurring						
	1. Lab R&D Infrastructure & Equipment	75	0	0	0	0	75
	Sub-Total	75	0	0	0	0	75
C.	Capital						
	1. Furnishing, Tables, Chairs, Cubicles, Electrical works and other Capex items	25	70	20	5	5	125
	Sub-Total	**25**	**70**	**20**	**5**	**5**	**125**
	Grand Total	**150**	**199**	**149**	**134**	**131**	**763**

Notes : TIH - Management : This unit will manage TIH, and will offer horizontal support services (Admin/Finance/HR/Legal) to TIH sub units (Vishleshan, which is the CoE and Runway iHUB - the Startup and commercialization support unit)

1. Manpower includes TIH CEO and Horizontal Support Unit. Subunit (CoE/Runway iHUB) manpower budgets are separately estimated
2. Travel budget is for TIH officials for administrative and networking purposes (PRSG meeting etc)
3. Budget is separately allocated for legal and financial compliances, board meetings/AGMs, Travel &Sitting fee of Board Members
4. Except for Website, no separate marketing budget is allocated, as it is clubbed with CPS GCC budget to reduce overall cost
5. No Equipment budget is given under TIH directly as the budgets for the same is provided under CoE
6. Capital budget is given for the office and facilities for TIH management and Horizontal support unit, including a board room and meeting rooms

Section 4 : Cost Sharing & Sustainability

Section Note:

1.This section explains the various possibilities explored in sharing the cost or recovering the cost of the projects

1. Cost Shared by partners over and above the estimated cost provided in Section 1 and 2

S No	TIH Key Result Areas	Cost Supplemented or Shared					
		1st Yr	2nd Yr	3rd Yr	4th Yr	5th Yr	Total
1	Technology Development	75	150	150	250	500	1125
2	CoE	36	36	36	36	36	180
3	HRD - New PG Program	0	18	18	18	18	72
4	Innovaiton Entrepreneurship and Startup - TBI	0	255	155	155	155	720
5	International Collaboration	0	0	500	500		1000
	Grand Total	**111**	**459**	**859**	**959**	**709**	**3097**

Notes : Cost Sharing:

1. **Technology development** : in the projects funded by TIH, 15% of the project cost will be added by the industry partner. Subsequently, in Year 4 TIH expects a minimum of 5 fully sponsored project at Rs 50 Lakhs per project. In Year 5, TIH expects 10 fully sponsored projects. This will subsequenlty be maintained or increased in order to ensure sustainability of operations as well

2. **TIH COE** provides 30,000 Sq Ft, cost of the infrastructure at Rs 3 Lakhs per month (Rs 10 per sq ft per month).

3. **HRD** : TIH provides 50% fellowship to students of the new PG program to be started and only 50% is added to the project cost. It comes to Rs 18 Lakh s per year from Year 2 onwards

4. **Startup** : IC shares existing infrastructure, space, labs, access to mentorship etc, leading to a shared cost of around 7 Cr over 5 years

5. **International collaboration** : The international collaborators will add equal amount in project budgeting , above projected cost

2. Cost Recovery and Income Generation possibilities (TIH & Partner Institutes)

S No	TIH Key Result Areas	1st Yr	2nd Yr	3rd Yr	4th Yr	5th Yr	Total
1	**Technology Development**						
	Tech Transfers	0.00	0.00	20.00	40.00	40.00	100.00
	PhD Fees (~ 35 PhD under projects)	0.00	15.00	45.00	45.00	0.00	105.00
2	**Establishment of CoEs Technical Services to industry/startup**	0.00	0.00	5.00	10.00	10.00	25.00
3	**HRD & Skill Development**						
	CHANAKYA - PG						
	Post Graduate Fellowships (M Tech/ MS) - Fees (to Host Institutions & TIH together)	6.00	20.00	40.00	30.00	4.00	100.00
	CHANAKYA-DF (Doctoral Fellowships - 25 PhD Scholars)	12.00	40.00	48.00	0.00	0.00	100.00
	CPS- PSDW (Professional Skill Development Workshop)	0.00	5.00	5.00	5.00	5.00	20.00
	CPS-New PG Programme (Speech, Video and Text Analytics) - Fees (Ongoing income)	0.00	60.00	60.00	60.00	60.00	240.00
4	**Innovation, Entrepreneurship and Start-up ecosystem**						
	Prayas	0.00	2.50	2.50	2.50	2.50	10.00
	Startup Facility usage (Investment Return of Rs 200 lakhs will be only after 5 Years)	0.00	6.00	8.00	8.00	8.00	30.00
	DIAL Accelerated companies investment returns	0.00	0.00	0.00	0.00	450.00	450.00
5	**International collaborations Tech Transfer**	0.00	0.00	0.00	50.00	50.00	100.00
	Total Cost Recovered in Rs Lakhs	**18.00**	**148.50**	**233.50**	**250.50**	**629.50**	1280.00

Notes : Cost Recovery: All cost recovery is for TIH as well for partners.

1. **Technology development** : Approximately 35 PhD Scholars will be working under these projects and they will pay fees. Also IP and tech transfer is a source of cost recovery within Project Period. Subsequently, in Year 4 TIH expects a minimum of 5 fully sponsored project at Rs 50 Lakhs per project. In Year 5, TIH expects 10 fully sponsored projects. This will subsequenlty maintained or increased in order to ensure sustainability of operations as well

2. **TIH COE :** Lab usage fees for external entities and projects/startups will be a source of cost recivery and income generation option over a longer period

3. **HRD** : Fees will be major source of income from all major HRD components. The fees will come to the partner institutions, which will then provide an overhead to TIH for TIH sponsored programs

4. **Startup** : Usage charges for labs and space are primary cost recovery mecahnisms. However, the returns from such initiatives are minimal due to the location factors. Hence investment returns are to be depneded upon for self sustainability. The returns projected are indicative only.

5. **International collaboration** : Technology transfers will be the source of cost recovery for international collaborations. The amounts are indicative only

3. Self Sustainability Options

S No	TIH Key Result Areas	Total Within Project Period	TIH Earning within Project Period	Projected Earning Post Project Years 6 to 8
1	**Technology Development**			
	Tech Transfers / Projects overheads	100.00	50.00	300.00
	PhD Fees (~35 PhD Scholars)	105.00	26.25	50.00
2	**Establishment of CoEs Technical Services to industry/startup**	25.00	25.00	36.00
3	**HRD & Skill Development**			
	CHANAKYA - PG			
	Post Graduate Fellowships (M Tech/ MS) - Fees (to Host Institutions & TIH together)	100.00	25.00	25.00
	CHANAKYA-DF (Doctoral Fellowships)			
	PhD Scholar Fees (3 lakhs per person)	100.00	25.00	25.00
	CPS- PSDW (Professional Skill Development Workshop)	20.00	10.00	10.00
	CPS-New PG Programme (Speech, Video and Text Analytics) - Fees (Ongoing income)	240.00	120.00	120.00
4	**Innovation, Entrepreneurship and Start-up ecosystem**			
	Prayas	10.00	2.50	2.50
	Startup Facility usage (Investment Return of Rs 200 lakhs will be only after 5 Years)	30.00	7.50	100.00
	DIAL Accelerated companies investment returns	450.00	225.00	250.00
5	**International collaborations Tech Transfer**	100.00	25.00	0.00
	Total Income Generated in Rs Lakhs	**1280.00**	**541.25**	**918.50**

Notes : TIH Income Generation and self sustainability options

1. **Technology development** :

a) In the project period, the technology transfers are the major options for income generation. About 10 tech transfers at Rs 10 Lakh per transfer is assumed for the projections in the project period, with TIH getting 25% of the tech transfer fees (75% will be between partners, inventors etc).

b) In the post project period, it is assumed that around Rs 300 Lakh will be earned by TIH (from 30 Tech transfers ; 10 per year for 3 years). The projections are indicative. The number and tech transfer value of projects is expected to increase in the subsequent years once the centre is established as a reliable research hub for industries. Selfsustainability will be primarily driven through projects and startups.

2. **TIH COE :** Lab usage fees for external entities and projects/startups will be a source of income generation. Rs 12 Lakhs per year is projected in the post project period.
3. **HRD** : Fees will be major source of income from all major HRD components. The fees will come to the partner institutions, which will then provide an overhead to TIH for TIH sponsored programs. Total income of Rs 180 Lakhs is project for 3 years in the post project period
4. **Startup** : Usage charges for labs and space are income generation options. However, the returns from such initiatives are minimal due to the location factor. Hence investment returns are to be depneded upon for self sustainability. The projections are given assuming a success rate of about 10 - 15% for Startups and 20% for DIAL accelerated startups
5. **International collaboration** : Technology transfers will be the source of income for international collaborations within project period. No income is projected afterwards, as the collaborations requires a large fixed financial commitment, without assured returns.

Annexure 2 : Project Timelines : Part 1

	FY 21 -22												FY 22 - 23				FY 23 - 24				FY 24 - 25				FY 25 - 26			
Major Activity	1	2	3	4	5	6	7	8	9	10	11	12	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
TIH Set Up	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫√																
Board constitution, Legal Entity Registration		🟦	🟦	🟦	🟦	🟦√																						
Prepartion of Guidelines & Policies, approval	🟦	🟦		🟦	🟦	🟦√				🟦	🟦	🟦√																
Core Team Onboarding & Team expansion			🟦	🟦	🟦	🟦	🟦	🟦	🟦√	🟦	🟦	🟦	🟦	🟦√														
TIH Governance						🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫
AGM (A) / Board Meetings (B)						🟦 A						🟦 **B**		🟦 A		🟦 **B**		🟦 A		🟦 **B**		🟦 A		🟦 **B**		🟦 A		🟦 **B**
Progress Reports (F - FY, M - Mid Term)									🟦 M			🟦 M		🟦 F		🟦 M		🟦 F		🟦 M		🟦 F		🟦 M		🟦 F		🟦 F
Establishment of CoE				🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫√	🟫	🟫	🟫	🟫	🟫√									
Team Onboarding				🟦	🟦	🟦	🟦	🟦	🟦√	🟦	🟦	🟦	🟦	🟦√														
Centralized Lab Facility Set Up				🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦√														
Centralized Lab Expansion (based on projects)														🟦	🟦	🟦	🟦	🟦	🟦√									
Technology Development				🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫
Call for Research Proposals				🟦	🟦	🟦			🟦	🟦	🟦		🟦		🟦		🟦		🟦									
Project Selection Committees (WG)						🟦					🟦			🟦		🟦		🟦		🟦								
Projects Assigning/Sanctioning/ Fund Release							🟦	🟦√				🟦√		🟦√		🟦√		🟦√		🟦√								
Project Monitoring Reviews (PRSG)												🟦		🟦		🟦		🟦		🟦		🟦		🟦		🟦		🟦
IP / Tech Transfer / Startup Spin off															🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦
HRD and Skill Development -					🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫
Chair/Faculty Fellowships					🟧	🟧	🟧	🟧√				🟧	🟧√	🟧	🟧√	🟧	🟧√											
UG Program - Internships & Upgrade							🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧
Call for Proposals for upgrade											🟦	🟦	🟦															
Projects evaluation, Sanctioning and funding														🟦√														
Course commencement														🟦√														
UG Internships / Project development funds							🟦	🟦	🟦√		🟦	🟦	🟦√			🟦	🟦√			🟦	🟦√			🟦	🟦√			🟦
UG Program Reviews																🟦		🟦		🟦		🟦		🟦		🟦		🟦
PG Program - New and Upgrade & Fellowships							🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧
Call for Proposals											🟦	🟦																
Projects evaluation Sanctioning and initial funding												🟦√	🟦√															
Course commencement														🟦√														
PG Fellowships / Project development funds							🟦	🟦√					🟦	🟦√			🟦	🟦√			🟦	🟦√			🟦	🟦√		🟦
PG Program Reviews												🟦				🟦		🟦		🟦		🟦		🟦		🟦		🟦
Research Fellowships (PhD / PDF)						🟧√	🟧	🟧	🟧√	🟧	🟧	🟧	🟧	🟧√	🟧	🟧√	🟧	🟧√	🟧	🟧√	🟧√	🟧	🟧	🟧	🟧	🟧	🟧	🟧
Skill Development											🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧
Professional Skill Development Workshops														🟦√				🟦√				🟦√				🟦√		
Advanced Skill Development School											🟦	🟦	🟦	🟦√	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦√	🟦	🟦	🟦	🟦	🟦	🟦

Note :
1. '√' marks interim milestones or final completion of an activity (eg: Projects evaluation, Sanctioning and funding etc for infrastrucutre projects)
2. '√' marks completion of selection process in case of fellowships, internships, startups, EIR etc that run for longer periods. If selection happens in multiple cycles, each cycle is marked.
3. '√' marks proposed time of the year planned for shorter events such as professional skill development school

Annexure 2 : Project Timelines : Part 2

	FY 21 -22												FY 22 - 23				FY 23 - 24				FY 24 - 25				FY 25 - 26			
Major Activity	1	2	3	4	5	6	7	8	9	10	11	12	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
International Collaboration										🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫
Inviting partners and sign MoU										🟦	🟦	🟦	🟦√	🟦	🟦	🟦√												
Projects Assigning/Sanctioning/ Fund Release													🟦√				🟦√											
Project Reviews														🟦		🟦		🟦		🟦		🟦		🟦		🟦		🟦
IP / Tech Transfer / Startup Spin off																	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦
Innovation, Entrepreneurship and Startup Ecosystem							🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫	🟫
TBI Operations							🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧
MoU between TIH and IC							🟦	🟦√																				
Fund Release and operations commencement									🟦√																			
Lab set up										🟦	🟦	🟦	🟦	🟦	🟦√													
CPS -GCC										🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧								
CPS Prayas									🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧
Lab Upgrade for Prayas										🟦	🟦	🟦	🟦	🟦	🟦	🟦												
Calling for Proposals									🟦	🟦			🟦			🟦					🟦							
Selection of proposals											🟦√		🟦√				🟦√				🟦√							
Progress Review													🟦		🟦		🟦		🟦		🟦		🟦		🟦		🟦	
CPS EIR									🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧
Calling for Proposals									🟦	🟦			🟦		🟦		🟦		🟦		🟦							
Selection of EIRs											🟦√		🟦√		🟦√		🟦√		🟦√		🟦√							
Progress Review													🟦		🟦		🟦		🟦		🟦		🟦		🟦		🟦	
CPS Startup									🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧
Calling for Proposals									🟦	🟦			🟦		🟦		🟦		🟦		🟦		🟦					
Selection of Startups											🟦√		🟦√		🟦√		🟦√		🟦√		🟦√		🟦√					
Progress Review													🟦		🟦		🟦		🟦		🟦		🟦		🟦		🟦	
CPS DIAL										🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧
DIAL Planning and Preparation										🟦	🟦	🟦	🟦															
Calling for Proposals													🟦				🟦				🟦							
Selection of Startups and Program														🟦√	🟦	🟦		🟦√	🟦	🟦		🟦√	🟦	🟦				
Progress Review															🟦				🟦				🟦					
CPS Seed Support												🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	

Note :
1. '√' marks interim milestones or final completion of an activity (eg: Projects evaluation, Sanctioning and funding etc for infrastrucutre projects)
2. '√' marks completion of selection process in case of fellowships, internships, startups, EIR etc that run for longer periods. If selection happens in multiple cycles, each cycle is marked.
3. '√' marks proposed time of the year planned for shorter events such as professional skill development school

This is to state that Wipro Limited hereby consent to partner with Indian Institute of Technology Patna (Host Institute name) in the proposed NM-ICPS Technology Innovation Hub (TIH) in "Speech, Video and Text Analytics". I am aware and agree to the activities mentioned in the proposal under Industry Partnership.

I hereby consent to support the TIH in the terms, that would be defined as per the mutual interests and scope of the work (SoW).

Summary profile of the Industry is given below:

Name of Industry/Organisation	: Wipro Limited
Nature of Business	: Research and Development
Number of Employees	: 1,71,425
Annual Turnover	: 589,060 [2018-2019 - Figures in ₹ Million]

I hereby affirm that my Industry is committed to participate in the proposed TIH "Speech, Video and Text Analytics" as indicated in the proposal.

Amitava Das

Date: 26th May 2020

Wipro AI Research and CTO Office

Place: Bangalore, India

Zimbra **asif@iitp.ac.in**

RE: Request to be a member of Govering Body of Technology Innovation Hub on "Speech, Video & Text Analytics"

From : Sengupta, Shubhashis <shubhashis.sengupta@accenture.com> Wed, May 20, 2020 10:29 AM

Subject : RE: Request to be a member of Govering Body of Technology Innovation Hub on "Speech, Video & Text Analytics"

To : Dr. Sriparna Saha <adean_rnd@iitp.ac.in>

Cc : Pushpak Bhattacharyya <pushpakdiro@gmail.com>, director <director@iitp.ac.in>, asif <asif@iitp.ac.in>

Sriparna:

Please find below a set of proposal abstracts. Many of them may be known already, but all has tremendous industrial and social values. I can get into specific details in each case and corroborate with what we see in the industry.

Please let me know your comments.

Thanks - Shubhashis

Project Proposal Abstracts

1. **Indian Language mixed-code Voice Assistants for functional domains**

Abstract: Many Indian corporate and social enterprises (like Banks, Hospitals and other Health care services, Public Services, Utilities) are looking forward to changing their traditional IVR (Interactive Voice Response) systems to AI-powered chatbots and voice bots. This shift will help them to have better customer interaction, knowing the customer better, better engagement and service. One of the key technical issue in wide-spread adoption of voice bots in Indian context is lack of mature Automated Speech Recognition (ASR) and comprehension and Text to Speech (TTS) models – especially for Indian regional languages and mixed-code (e.g., Hindi + English, Tamil + English) conversations. We propose that R&D effort be spent on creating appropriate thesauri, language models, machine and deep learning models to aid such Ai power virtual assistants in Indian industry context.

2. **AI + Robotics:**

Abstract: Industrial and Social robotics are coming of age and increasingly being adopted, both in large industry, MSME and in household sectors. The applications can be of operating in hazardous environment (such as infectious disease treatment, disaster recovery), remote operations (mining, agriculture), social robotics (educational assistants, robotic companions for elders). The proposal is to create appropriate AI techniques (such as visual recognitions, spatial reasoning, reinforcement learning) for such robotic firmware to impart improved learning, cognitive and functional capabilities in the fields.

3. **Multi-modal AI for Telehealth:**

Abstract: AI based telehealth systems (moving beyond Telemedicine) is destined to be a part of better healthcare delivery mechanism – especially in post COVID context. This is especially visible in areas of telehealth innovations where AI applications are used to support, supplement or develop new remote healthcare models and increase access to millions. According to WHO's eHealth observatory survey, AI in the telemedicine field is directly supplementing innovations in these areas:

- Tele-radiology,
- Tele-pathology,
- Tele-dermatology, and
- Tele-psychiatry.

We propose to create a framework for a multi-modal AI system (Text, image and video, voice and sensor based) to augment the telehealth capability for leading Indian hospitals. This has go beyond remote patient monitoring to provide truly intelligent and interactive healthcare intervention, assistive guidance and alerts.

4. **AI based mis-information detection and prevention system**

Rather than pandemics, misinformation spreading kills more people and create social discord world over. There is an urgent need to detect and prevent misinformation spreading through social media through effective AI intervention in Indian context (keeping in view our social, religious and cultural sensitivity). In essence, this will have three primary parts –

- Detect if a particular social network post is fake or un-trustworthy
- Detect virality and spreading potential of the content and the "super-spreaders" in the network
- Analyse veracity / claims in non-reviewed or general publications or blogs.

5. **AI based personalized learning augmentation**

Ai can be a great enabler for personalized learning through higher adaptiveness, audio-visual interactivity, AI based assessment for individual learning proclivity and path. It is not just for curriculum learning, but for social and behavioral skills as well. We propose that a set of projects be initiated on various aspects of AI assisted learning through – Ai based course curation, AI based assessment (questionnaire generation, rating and ranking), AI based personalized educational coach, AI based life skill coaching and guidance.

Thanks and regards - Shubhashis

From: Dr. Sriparna Saha <adean_rnd@iitp.ac.in>
Sent: Monday, May 18, 2020 11:54 PM
To: Sengupta, Shubhashis <shubhashis.sengupta@accenture.com>
Cc: Pushpak Bhattacharyya <pushpakdiro@gmail.com>; director <director@iitp.ac.in>; asif <asif@iitp.ac.in>
Subject: [External] Re: Request to be a member of Govering Body of Technology Innovation Hub on "Speech, Video & Text Analytics"

A gentle reminder!!

We are in the process of finalizing the DPR which has to be submitted soon in DST. Kindly send us a few topics (with abstract) on which in future there would be an open "call for proposal" from our TIH.

Best regards
Sriparna

From: "Dr. Sriparna Saha" <adean_rnd@iitp.ac.in>
To: "shubhashis sengupta" <shubhashis.sengupta@accenture.com>
Cc: "Pushpak Bhattacharyya" <pushpakdiro@gmail.com>, "director" <director@iitp.ac.in>, "asif" <asif@iitp.ac.in>
Sent: Sunday, May 10, 2020 10:40:52 PM
Subject: Re: Request to be a member of Govering Body of Technology Innovation Hub on "Speech, Video & Text Analytics"

Dear Dr. Shubhashis,

Thanks for agreeing to be a part of our TIH. We are currently working on preparing the Detailed Project Report (DPR) to be submitted to DST.

In this connection, we request the following help from you.

Please note that 50% of the TIH funding has to be utilized for distributing projects to experts of other institutes working on the theme of " Speech, Video & Text Analytics".
In the DPR, we have to mention about few problem statements, on which in future there would be an open "call for proposal" from our TIH.

We request you to kindly suggest a few proposals (short abstract) related to the theme of TIH which can be considered for future "call for proposals".

Looking forward to hearing from you.
Best regards

Sriparna

From: "shubhashis sengupta" <shubhashis.sengupta@accenture.com>
To: "Dr. Sriparna Saha" <adean_rnd@iitp.ac.in>
Cc: "Pushpak Bhattacharyya" <pushpakdiro@gmail.com>, "director" <director@iitp.ac.in>
Sent: Thursday, April 16, 2020 1:42:21 PM
Subject: RE: Request to be a member of Govering Body of Technology Innovation Hub on "Speech, Video & Text Analytics"

Thank you.

Best regards - Shubhashis

From: Dr. Sriparna Saha <adean_rnd@iitp.ac.in>
Sent: Thursday, April 16, 2020 9:59 AM
To: Sengupta, Shubhashis <shubhashis.sengupta@accenture.com>
Cc: Pushpak Bhattacharyya <pushpakdiro@gmail.com>; director <director@iitp.ac.in>
Subject: [External] Re: Request to be a member of Govering Body of Technology Innovation Hub on "Speech, Video & Text Analytics"

Dear Shubhashis,

I have the following information:

Terms of Reference:
(a) The Hub Governing Body shall be the Apex body for overall supervision, control, directions and mid-course correction in the implementation of Hubs at Host Institutes.
(b) Will approve key guidelines for implementation of the Hub.
(c) Governing Bodies of each hub will be the final authority to provide guidelines for implementation and operating the Hubs and all other matters related to them.
Governing Bodies 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 or create a new TBI if there is no TBI in HI.
Support for projects will be based on the requirement, open call and with due scientific diligence and processes.
(d) Hubs Governing Body could co-opt eminent people (India/ abroad) as members.
(e) **The Hubs Governing Body would meet as often as required and at least once in a year.**
(f) Appoint sub-committees from time-to-time and assign and/or mandate them to appropriate technical streams or assign tasks that fall within the scope of such Committees for efficient implementation of Hubs at Host Institutes.

Hubs activities: For the purpose of clearly defining the objectives and the activities of the Hubs, it has been divided into four major streams, namely

(a) Technology Development: Through 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
(b) HRD & Skill Development: Through Fellowship Based UG/ PG, Ph.D., Post- Doctoral and Short Term Training for Faculty
(c) Innovation, Entrepreneurship, and Start-up Ecosystem: To enhance competencies, capacity building, and training to nurture innovation and Start-up
ecosystem.
(d) International Collaborations: To establish and strengthen international collaborative research/ Technology Development.

Best regards,

--
Dr. Sriparna Saha
Associate Dean, Research and Development
Associate Professor, Department of Computer Science and Engineering
IIT Patna, Patna, India-801106
Ph no: +91-612-3028128, +91-612-3028291
Other Email ids: sriparna@iitp.ac.in,sriparna.saha@gmail.com
Web: http://www.iitp.ac.in/~sriparna/

From: "shubhashis sengupta" <shubhashis.sengupta@accenture.com>
To: "Dr. Sriparna Saha" <adean_rnd@iitp.ac.in>

Cc: "Pushpak Bhattacharyya" <pushpakdiro@gmail.com>, "director" <director@iitp.ac.in>
Sent: Thursday, April 16, 2020 9:47:01 AM
Subject: RE: Request to be a member of Govering Body of Technology Innovation Hub on "Speech, Video & Text Analytics"

Dear Sriparna:

Thanks for the invitation to be a member of the Board and I am glad to accept it. Can you please let me know what will be the activities and what will be expected of a member?

Thanks and best regards - Shubhashis

From: Dr. Sriparna Saha <adean_rnd@iitp.ac.in>
Sent: Thursday, April 16, 2020 9:17 AM
To: Sengupta, Shubhashis <shubhashis.sengupta@accenture.com>
Cc: Pushpak Bhattacharyya <pushpakdiro@gmail.com>; director <director@iitp.ac.in>
Subject: [External] Request to be a member of Govering Body of Technology Innovation Hub on "Speech, Video & Text Analytics"

This message is from an EXTERNAL SENDER - be CAUTIOUS, particularly with links and attachments.

Dear Sir,

Greetings from IIT Patna!
Hope this email finds you well.

I am Dr. Sriparna Saha, Associate Dean Research and Development of IIT Patna.

DST has decided to set up a **technology innovation hub** (TIH) at IIT **Patna** as part of its National Mission on Interdisciplinary Cyber-Physical Systems (NM-ICPS) program. The innovation hub will emphasize **speech, video, and text analytics**.

Each Hub will have its Hub Governing Body. This Hub Governing Body shall be the Apex body for overall supervision, control, directions and mid-course correction in the implementation of Hubs at Host Institutes.

As you are an expert in the field of "Speech, Video & Text Analytics", we would be highly obliged if you agree to be a **member of this Hub Governing Body.**

Look forward to hearing from you.
Best regards

Sriparna

--
Dr. Sriparna Saha
Associate Dean, Research and Development
Associate Professor, Department of Computer Science and Engineering
IIT Patna, Patna, India-801106
Ph no: +91-612-3028128, +91-612-3028291
Other Email ids: sriparna@iitp.ac.in,sriparna.saha@gmail.com

Web: http://www.iitp.ac.in/~sriparna/

This message is for the designated recipient only and may contain privileged, proprietary, or otherwise confidential information. If you have received it in error, please notify the sender immediately and delete the original. Any other use of the e-mail by you is prohibited. Where allowed by local law, electronic communications with Accenture and its affiliates, including e-mail and instant messaging (including content), may be scanned by our systems for the purposes of information security and assessment of internal compliance with Accenture policy. Your privacy is important to us. Accenture uses your personal data only in compliance with data protection laws. For further information on how Accenture processes your personal data, please see our privacy statement at https://www.accenture.com/us-en/privacy-policy.

____

http://www.accenture.com

Zimbra **asif@iitp.ac.in**

Fwd: Re: Request for collaboration (IIT Patna)

From : Dr. Preetam Kumar <pkumar@iitp.ac.in> Tue, May 26, 2020 10:06 AM
Subject : Fwd: Re: Request for collaboration (IIT Patna)
To : asif <asif@iitp.ac.in>

Dear Dr Asif,

Please find beliw the Abstract of proposal for open call. Second abstract is expected soon.

Thanks
----- Forwarded Message -----
From: sriganesh rao <sriganesh.rao@calligotech.com>
To: Dr. Preetam Kumar <pkumar@iitp.ac.in>
Cc: Rajaraman Subramanian <rajaraman.subramanian@calligotech.com>
Sent: Tue, 26 May 2020 09:23:28 +0530 (IST)
Subject: Re: Request for collaboration (IIT Patna)

Dear Dr Kumar,

I am giving the abstract of the work for Smart Grid Analytics below. I will provide for other proposal later.

TITLE: AI-BASED SMART GRID DATA ANALYTICS

ABSTRACT:

Smart grid is a vital part of energy because it allows energy providers to draw full value from the Smart grid. It allows for a layer of communication between local actuators, central controllers and logistic units, which enables better response during emergencies and more efficient use of resources.

It is proposed to use a Advanced Data Analytics platform that can capture and analyse data from different endpoints which enables utility companies to distribute resources more efficiently, cut costs and discover better ways to server their customers. The Data Analytics platform would have the capability to collect large volumes of Data either in structured, semi-structured or unstructured format. Different data ingestion mechanisms will be supported like streaming, micro-batches, logs, bulk etc. These data will be analysed in real-time. AI would be used at all layers - data ingestion, data cleansing and data visualization. The Platform will be optimized on High Performance Compute and High-Performance Data Storage drastically reducing data access, model training and inference time.

This project objectives are to develop Predictive analytics giving insights to:

* Demand & Supply; Consumption Patterns
* Forecasting renewal energy production
* Decide on a real-time basis, which energy source to use and in which proportion.
* Early warning system to help in better planning and response of service provider
* Potential cost savings along with uses of different energy sources
* Improved revenue cash flow due to intelligent switchover between available energy sources
* Flexibility to add more renewable solution

Thanks & regards

Sriganesh Rao

On 25-05-2020 19:03, Dr. Preetam Kumar wrote:

> Dear Dr. Rao,
>
> At present, we need abstract of the work/proposal only. Product details will be required in the 2nd or 3rd stage.
> Please mail me the precise problem statement for both project.
>
> Thanks,
> Preetam Kumar
> ----- Original Message -----
> From: sriganesh rao <sriganesh.rao@calligotech.com>
> To: Dr. Preetam Kumar <pkumar@iitp.ac.in>
> Cc: Rajaraman Subramanian <rajaraman.subramanian@calligotech.com>
> Sent: Mon, 25 May 2020 18:59:07 +0530 (IST)
> Subject: Re: Request for collaboration (IIT Patna)
>
> Dear Dr Kumar,
>
> This is further to discussion last night. What was the outcome of the
> meeting with Intel's Ramanathan Sethuraman?
>
> You also wanted additional Use Case Proposals for the proposed CoE using
> our Product. I am furnishing Smart Grid Analytics and Telecom Analytics
> :
>
> A] PROJECT - "ADVANCED DATA ANALYTICS SOLUTIONS TO THE SMART &
> INTEGRATED LOCAL ENERGY SYSTEMS, THROUGH CALLIGO'S INTELLIGENT DATA
> ANALYTICS PLATFORM (CIDAP)"
>
> THE PREDICTIVE ANALYTICS FEATURE OF CIDAP GIVES INSIGHTS TO:
>
> * Demand & Supply ; Consumption Patterns
> * Forecasting renewal energy production in the next 48 hours
> * Potential cost savings along with uses of different energy sources
> * Decide on a real-time basis, which energy source to use and in which
> proportion.
> * Early warning system helps in better planning and response of
> service provider

> * Improved revenue cash flow due to intelligent switchover between
> available energy sources
> * Pilferage Management ; Reduced diesel consumption ;More carbon
> credits
> * Flexibility to add more renewable solution
> * On-demand support
>
> CIDAP'S PREDICTIVE ANALYTICS WOULD WORK BY
>
> * knowing what occurred in past via review of past events
> * Reasoning for why event occurred via data analysis
> * Review and monitor what is occurring via on-line systems
> * Reasoning for why event occurring via dashboards data analysis
> * Predict what event is going to occur via data modes
> * Reason why event would occur via data mining.
>
> B] TELECOM ANALYTICS USING CIDAP
>
> Telecom companies that want to be innovative and maximise their revenue
> potential must have the right solution in place so that they can harness
> the volume, variety and velocity of data coming to their organization
> and leverage on actionable _insights_ from that data.   Telecom
> companies are sitting on terabytes of data that are stored in silos and
> scattered across the organization. For simpler and faster processing of
> only relevant data, telcos need an advanced analytics driven data
> solution that will help them to achieve timely and accurate insights
> using data mining and predictive analytics.
>
> The massive amount of data when captured wisely and analysed
> professionally can reveal powerful insights. Big data and advanced
> analytics provide telcos with the tools and techniques to harness and
> integrate new sources and new types of data in larger volumes and in
> real-time.
>
> Data analytics can help operators enhance the overall value of their
> business in regards to service optimization, customer satisfaction and
> revenues
>
> Using the capabilities of CIDAP, Telcos can turn an enormous structured
> and unstructured data into actionable customer insights. The big data
> that customers generate, with the right analytics, enables telcos to
> DEVELOP ENRICHED 360 CUSTOMER PROFILES, ESTABLISH CUSTOMER-CENTRIC KPIS
> and DEVELOP MORE TARGETED OFFERS.
>
> With CIDAP's advanced data architectures, operators can also store new
> types of data, retain that data longer, and join diverse datasets
> together to gain new insights
>
> CIDAP can provide the following enablement
> □ Enable the client to perform analytics in house
> □ Understand the client's requirement and assist the client team to
> develop, integrate and implement analytics in their environment
>
> DO let me know if this suffices.
>

> Thanks & regards.
>
> Sriganesh
>
> On 23-04-2020 22:31, sriganesh rao wrote:
>
> Dear Dr Kumar,
>
> We thank you for your mail below and congratulate IIT Patna for being selected by DST, GoI, as the Technology Innovation Hub in the area of Speech, Video and Text analytics.
>
> We are happy to collaborate with IIT Patna as the Industry Partner for this initiative, and would like to work with you in the area of Health Research/ Healthcare Analytics as one of the application areas of Speech, Video and Text Analytics.
>
> We are enclosing our Letter of Intent for your kind perusal. We look forward to hearing from you.
>
> Thanks & Regards
>
> SRIGANESH RAO
> Managing Director
> Mobile: +91 9845155800
>
> CALLIGO TECHNOLOGIES PVT LTD
> #55/C, Nandi Mansion,
> 40th Cross, 8th Block, Jayanagar,
> Bangalore - 560070, India
> Ph: (080)-26542726, 26542736
> www.calligotech.com [1] [1 [1]]
>
> On 2020-04-18 08:11, Dr. Preetam Kumar wrote:
>
> Dear Dr. Rao,
>
> Greetings from IIT Patna!
>
> I am happy to inform you that IIT Patna has been selected by DST, GoI , as the Technology Innovation Hub in the area of Speech, Video & Text
> Analytics . An initial grant of RS. 7.25 CRORES has been approved for the TIH. We have to select application area from the following application verticals and I have selected HEALTH RESEARCH to start with:
>
> 1.Ayurveda, Yoga and Naturopathy, Unani, Siddha and Homoeopathy (AYUSH)
> 2. Chemicals and Petro-Chemicals (Rasayan aur Petro-Rasayan)
> 3. Fertilizers
> 4. Pharmaceuticals
> 5. Civil Aviation
> 6. Coal
> 7. Commerce
> 8. Telecommunications
> 9. Posts
> 10. Consumer Affairs

> 11. Food and Public Distribution
> 12. Defence Production
> 13. Defence Research and Development
> 14. Earth Sciences
> 15. Electronics and Information Technology
> 16. Environment, Forest and Climate Change
> 17. CBDT, GST, Aadhar, Fintech, Director & Indirect Taxation
> 18. Investment and Public Asset Management (DIPAM)
> 19. Financial Services
> 20. Fisheries
> 21. Animal Husbandry and Dairying
> 22. Food Processing Industries
> 23. Health and Family Welfare
> 24. Health Research
> 25. Heavy Industry
> 26. Public Enterprises
> 27. Internal Security
> 28. Border Management
> 29. Housing and Urban Affairs
> 30. School Education and Literacy
> 31. Higher Education
> 32. Labour and Employment
> 33.Judiciary
> 34. Micro, Small and Medium Enterprises
> 35. New and Renewable Energy
> 36. Petroleum and Natural Gas
> 37. Power
> 38. Railways
> 39. Road Transport and Highways
> 40. Rural Development
> 41. Land Resources
> 42. Scientific and Industrial Research
> 43. Bio-Technology
> 44. Shipping
> 45. Skill Development and Entrepreneurship
> 46. Empowerment of Persons with Disabilities
> 47. Statistics and Programme Implementation
> 48. Steel
> 49. Textiles
> 50. Tourism
> 51. Sports
> 52. Atomic Energy
> 53. Space
> 54. NITI Aayog
> 55. National Security Council (Rashtriya Suraksha Parishad)
>
> Considering the rich experience of Calligo in the area of Data Analytics, we wish to collaborate with Calligo Technologies Pvt Ltd as an industry partner for this initiative.
>
> Waiting for your kind consent,
> Best Regards
> Preetam Kumar
>
> Web-profile: https://www.iitp.ac.in/index.php/en-us/people-2/faculty/2-

uncategorised/194-view-profile-8

> --
>
> DR. PREETAM KUMAR, SMIEEE
>
> Associate Professor
>
> Department of Electrical Engineering
>
> Indian Institute of Technology Patna, India
>
> Tel: +91-612-3028048

Links:
------
[1] http://www.calligotech.com

Links:
------
[1] http://www.calligotech.com
--
Dr. Preetam Kumar, SMIEEE Associate Professor Department of Electrical Engineering Indian Institute of Technology Patna, IndiaTel: +91-612-3028048

Zimbra **asif@iitp.ac.in**

Re: Request to be a member of Govering Body of Technology Innovation Hub on "Speech, Video & Text Analytics"

From : Lipika Dey <lipikadey@gmail.com> Tue, May 19, 2020 06:06 PM

Subject : Re: Request to be a member of Govering Body of Technology Innovation Hub on "Speech, Video & Text Analytics"

To : Dr. Sriparna Saha <adean_rnd@iitp.ac.in>

Cc : director <director@iitp.ac.in>, Pushpak Bhattacharyya <pushpakdiro@gmail.com>, asif <asif@iitp.ac.in>

Dear Dr. Saha,

I am not sure in what form you want the information. What are you looking for in the paragraph?

Some of the problems that I can think of are

1. Procedure extraction from Technical Text Documents - extracting experimental procedures, manufacturing procedures, treatment procedures as mentioned in text. The need can be to train for a task or compare multiple similar procedures.
These can also be linked to video documentation of similar tasks. There can be several applications that I can think of in different areas.

2. The above systems can be further enhanced with conversation systems. By conversation systems here I mean actually supervisory and guiding agents rather than QA systems.

3. General purpose Speech enabled conversation systems for Teaching Learning Systems

4. Video plus Speech Analytics has so many applications - right from Elderly Health Care to Teaching specially abled students

5. Predictive Analytics with News and Structured Business data - Connecting the dots spread across multiple documents - building knowledge networks - applying predictive techniques on these. Applicable for Supply Chain Reason, better demand forecasting,

Thanks

On Mon, May 18, 2020 at 11:58 PM Dr. Sriparna Saha <adean_rnd@iitp.ac.in> wrote:

> A gentle reminder!!
>
> We are in the process of finalizing the DPR which has to be submitted soon in DST. Kindly send us a few topics (with abstract) on which in future there would be an open "call for proposal" from our TIH.
>
> Best regards

Sriparna

--
Dr. Sriparna Saha
Associate Dean, Research and Development
Associate Professor, Department of Computer Science and Engineering
IIT Patna, Patna, India-801106
Ph no: +91-612-3028128, +91-612-3028291
Other Email ids: sriparna@iitp.ac.in,sriparna.saha@gmail.com
Web: http://www.iitp.ac.in/~sriparna/

From: "Dr. Sriparna Saha" <adean_rnd@iitp.ac.in>
To: "lipikadey" <lipikadey@gmail.com>
Cc: "director" <director@iitp.ac.in>, "Pushpak Bhattacharyya" <pushpakdiro@gmail.com>, "asif" <asif@iitp.ac.in>
Sent: Sunday, May 10, 2020 10:38:35 PM
Subject: Re: Request to be a member of Govering Body of Technology Innovation Hub on "Speech, Video & Text Analytics"

Dear madam,

Thanks for agreeing to be a part of our TIH. We are currently working on preparing the Detailed Project Report (DPR) to be submitted to DST.

In this connection, we request the following help from you.

Please note that 50% of the TIH funding has to be utilized for distributing projects to experts of other institutes working on the theme of " Speech, Video & Text Analytics".
In the DPR, we have to mention about few problem statements, on which in future there would be an open "call for proposal" from our TIH.

We request you to kindly suggest a few proposals (short abstract) related to the theme of TIH which can be considered for future "call for proposals".

Looking forward to hearing from you.
Best regards

Sriparna

From: "lipikadey" <lipikadey@gmail.com>
To: "Dr. Sriparna Saha" <adean_rnd@iitp.ac.in>
Cc: "director" <director@iitp.ac.in>, "Pushpak Bhattacharyya" <pushpakdiro@gmail.com>
Sent: Saturday, April 18, 2020 10:40:01 AM
Subject: Re: Request to be a member of Govering Body of Technology Innovation Hub on "Speech, Video & Text Analytics"

Dear Dr. Sriparna,
Thank you for the invitation.
I will be happy to participate.

> Warm regards
>
> On Sat, Apr 18, 2020 at 8:43 AM Dr. Sriparna Saha <adean_rnd@iitp.ac.in> wrote:
>
>> Dear Madam,
>>
>> Greetings from IIT Patna!
>> Hope this email finds you well.
>>
>> I am Dr. Sriparna Saha, Associate Dean Research and Development of IIT Patna.
>>
>> DST has decided to set up a **technology innovation hub** (TIH) at IIT **Patna** as part of its National Mission on Interdisciplinary Cyber-Physical Systems (NM-ICPS) program. The innovation hub will emphasize **speech, video, and text analytics**.
>>
>> Each Hub will have its Hub Governing Body. This Hub Governing Body shall be the Apex body for overall supervision, control, directions and mid-course correction in the implementation of Hubs at Host Institutes.
>>
>> As you are an expert in the field of "Speech, Video & Text Analytics", we would be highly obliged if you agree to be a **member of this Hub Governing Body.**
>>
>> Look forward to hearing from you.
>> Best regards
>>
>> Sriparna
>>
>> --
>> Dr. Sriparna Saha
>> Associate Dean, Research and Development
>> Associate Professor, Department of Computer Science and Engineering
>> IIT Patna, Patna, India-801106
>> Ph no: +91-612-3028128, +91-612-3028291
>> Other Email ids: sriparna@iitp.ac.in,sriparna.saha@gmail.com
>> Web: http://www.iitp.ac.in/~sriparna/
>
> --
> Dr. Lipika Dey
> Principal Scientist
> Innovation Labs
> Tata Consultancy Services,
> New Delhi, India

--
Dr. Lipika Dey
Principal Scientist
Innovation Labs

Tata Consultancy Services,
New Delhi, India

Zimbra **asif@iitp.ac.in**

RE: Request to be a member of Govering Body of Technology Innovation Hub on "Speech, Video & Text Analytics"

From : Karthik Sankaranarayanan <kartsank@in.ibm.com> Mon, May 11, 2020 12:02 PM

Subject : RE: Request to be a member of Govering Body of Technology Innovation Hub on "Speech, Video & Text Analytics" 1 attachment

To : adean rnd <adean_rnd@iitp.ac.in>

Cc : asif@iitp.ac.in, director@iitp.ac.in, pushpakdiro@gmail.com

Dear Sriparna,

Please find attached a draft set of topics/abstracts. Please let me know your thoughts -- if they're at the right level of detail, the number of abstracts, topics, etc.

Thanks
Karthik

> ----- Original message -----
> From: Karthik Sankaranarayanan/India/IBM
> To: adean_rnd@iitp.ac.in
> Cc: asif@iitp.ac.in, director@iitp.ac.in, pushpakdiro@gmail.com
> Subject: Re: [EXTERNAL] Re: Request to be a member of Govering Body of Technology Innovation Hub on "Speech, Video & Text Analytics"
> Date: Mon, May 11, 2020 10:45 AM
>
> Sure, I'll get back to you on this. Thanks.
>
> ________________________________
>
> **Karthik Sankaranarayanan, PhD**
> STSM, Sr Manager, AI for Interaction Deparment,
> IBM Research
> Bangalore, India
> Email: kartsank@in.ibm.com
> Web: https://researcher.watson.ibm.com/researcher/view.php?person=in-kartsank
>
> > ----- Original message -----
> > From: "Dr. Sriparna Saha" <adean_rnd@iitp.ac.in>
> > To: kartsank <kartsank@in.ibm.com>
> > Cc: director <director@iitp.ac.in>, Pushpak Bhattacharyya <pushpakdiro@gmail.com>, asif <asif@iitp.ac.in>
> > Subject: [EXTERNAL] Re: Request to be a member of Govering Body of Technology Innovation Hub on "Speech, Video & Text Analytics"
> > Date: Sun, May 10, 2020 10:36 PM
> >
> > Dear Dr.Karthik,

Thanks for agreeing to be a part of our TIH. We are currently working on preparing the Detailed Project Report (DPR) to be submitted to DST.

In this connection, we request the following help from you.

Please note that 50% of the TIH funding has to be utilized for distributing projects to experts of other institutes working on the theme of " Speech, Video & Text Analytics". In the DPR, we have to mention about few problem statements, on which in future there would be an open "call for proposal" from our TIH.

We request you to kindly suggest a few proposals (short abstract) related to the theme of TIH which can be considered for future "call for proposals".

Looking forward to hearing from you.
Best regards

Sriparna
--
Dr. Sriparna Saha
Associate Dean, Research and Development
Associate Professor, Department of Computer Science and Engineering
IIT Patna, Patna, India-801106
Ph no: +91-612-3028128, +91-612-3028291
Other Email ids: sriparna@iitp.ac.in,sriparna.saha@gmail.com
Web: http://www.iitp.ac.in/~sriparna/

From: "kartsank" <kartsank@in.ibm.com>
To: "Dr. Sriparna Saha" <adean_rnd@iitp.ac.in>
Cc: "director" <director@iitp.ac.in>, "Pushpak Bhattacharyya" <pushpakdiro@gmail.com>
Sent: Thursday, April 16, 2020 10:21:59 AM
Subject: Re: Request to be a member of Govering Body of Technology Innovation Hub on "Speech, Video & Text Analytics"

Dear Dr. Sriparna Saha,

Thank you for the invitation. It would be my privilege to serve as a member of the Hub Governing Body. I humbly accept.

Look forward to knowing more about the Hub, and its goals and activities.

Regards
Karthik

Karthik Sankaranarayanan, PhD
Senior Research Scientist & Senior Manager, AI for Interaction Dept
IBM Research
Bangalore, India
Email: kartsank@in.ibm.com
Website: https://researcher.watson.ibm.com/researcher/view.php?person=in-kartsank

---

----- Original message -----
From: "Dr. Sriparna Saha" <adean_rnd@iitp.ac.in>
To: kartsank@in.ibm.com
Cc: Pushpak Bhattacharyya <pushpakdiro@gmail.com>, director <director@iitp.ac.in>
Subject: [EXTERNAL] Request to be a member of Govering Body of Technology Innovation Hub on "Speech, Video & Text Analytics"
Date: Thu, Apr 16, 2020 9:19 AM

Dear Sir,

Greetings from IIT Patna!
Hope this email finds you well.

I am Dr. Sriparna Saha, Associate Dean Research and Development of IIT Patna.

DST has decided to set up a *technology innovation hub* (TIH) at IIT *Patna* as part of its National Mission on Interdisciplinary Cyber-Physical Systems (NM-ICPS) program. The innovation hub will emphasize **speech, video, and text analytics**.

Each Hub will have its Hub Governing Body. This Hub Governing Body shall be the Apex body for overall supervision, control, directions and mid-course correction in the implementation of Hubs at Host Institutes.

As you are an expert in the field of "Speech, Video & Text Analytics", we would be highly obliged if you agree to be a **member of this Hub Governing Body.**

Look forward to hearing from you.
Best regards

Sriparna

--
Dr. Sriparna Saha
Associate Dean, Research and Development
Associate Professor, Department of Computer Science and Engineering
IIT Patna, Patna, India-801106
Ph no: +91-612-3028128, +91-612-3028291
Other Email ids: sriparna@iitp.ac.in,sriparna.saha@gmail.com
Web: http://www.iitp.ac.in/~sriparna/

TIH-proposals-Karthik.docx
15 KB

April 18, 2020

To
Dr. Atul Thakur
Associate Professor
Department of Mechanical Engineering
&
Dr. Subrata Hait
Associate Professor
Department of Civil and Environmental Engineering
Indian Institute of Technology Patna
Bihta, Bihar – 801 106

Dear Sir(s),

This letter is to confirm our joint participation and support for the project entitled *"Hyperspectral Video Processing Assisted Automated Segregation of Recyclables from Solid Waste Streams in a Smart City"* under the proposed Technology Innovation Hub (TIH) at Indian Institute of Technology (IIT) Patna under the aegis of the National Mission on Interdisciplinary Cyber Physical Systems (NM-ICPS) of the Science and Engineering Research Board (SERB), DST, Govt. of India.

We confirm that we will be contributing as an industry partner inthe said project. We will be happy to provide any technical support for the successful execution of this project and the development of the proposed technology.

We look forward to working with IIT Patna on this project.Thank you.

Best wishes and regards,

Amit

(Dr. Amit Singh Chauhan)
Founder &Director
NatureSense Technologies P. Ltd.
Kanpur, Uttar Pradesh – 208 016

NATURESENSE TECHNOLOGIES PVT. LTD.

पत्रांक - 811 20April, 2020

To
Dr. Atul Thakur
Associate Professor
Department of Mechanical Engineering
&
Dr. SubrataHait
Associate Professor
Department of Civil and Environmental Engineering
Indian Institute of Technology Patna
Bihta, Bihar – 801 106

Dear Sir(s),

This letter is to confirm our joint participation and support for the project entitled *"Hyperspectral Video Processing Assisted Automated Segregation of Recyclables from Solid Waste Streams in a Smart City"* under the proposed Technology Innovation Hub (TIH) at Indian Institute of Technology (IIT) Patna under the aegis of the National Mission on Interdisciplinary Cyber Physical Systems (NM-ICPS) of the Science and Engineering Research Board (SERB), DST, Govt. of India.

We confirm that we will be contributing as an urban local body (ULB) partner inthe said project. We will be happy to provide any support in the urban framework for the successful execution of this project and the development of the proposed technology.

We look forward to working with IIT Patna on this project.Thank you.

Best wishes and regards,

२०|4
(Sri DhirendraPaswan)
Municipal Commissioner
Ara Municipal Corporation
Ara, Bhojpur – 802 301, Bihar

नगर आयुक्त
आरा नगर निगम, आरा

CALLIGO TECHNOLOGIES
High Performance & Beyond!

The Director,
Indian Institute of Technology Patna
Bihta
Patna - 801103

Dated 22 April 2020

Dear Sir,

REF: Technology Innovation Hub for Speech, Video and Text Analytics.
SUB: Letter of Intent to be Industry partner for IIT Patna for the Technology Innovation Hub in area of Speech, Video and Text Analytics

Kindly accept our congratulations on being selected by DST, GoI, as the Technology Innovation Hub in the area of Speech Video, and Text analytics. We would like to collaborate with your esteemed organisation, as "Industry Partner" for the Technology Innovation Hub in the area of Speech, Video, and Text Analytics.

Further, we would like to work in the area of Health Research/ Healthcare Analytics as one of the application areas of Speech, Video and Text Analytics with Dr Preetam Kumar- Associate Professor, Dept of Electrical Engineering. Kindly consider this as our Letter of Intent for the same.

We are a category defining Data Science and Machine Learning software company focused on helping organizations seeking to realize their full potential to capture new value by leveraging the convergence of High-Performance Computing and Big Data and unleashing the potential of Dark Data using Artificial Intelligence and Machine Learning. We are solving problems in the domains of Healthcare, Workplace Transformation, Telecom, Smart Grid, Smart Cities, Security and Public Safety, with our 'Calligo Intelligent Data Analytics Platform' (CIDAP).

Our solution also uses a deep neural network-based Edge Analytics, Calligo Health Engine, which is capable of working with downloadable Deep Learning models. We use this to perform rapid assessment of avoidable blindness. Our solution has won the "Health10X" competition conducted by George Institute of Global Health, Australia and has been presented and demonstrated at the WHO special 'Focus Group of AI for Healthcare' conference at ICMR, New Delhi. We are now Member of the WHO's special Focus Group of AI for Healthcare.

We look forward to working with IIT-Patna Technology Innovation Hub, in the area of Speech, Video and Text Analytics and eagerly await to hear from you.

Your's sincerely.

SRIGANESH RAO
Managing Director
+91 9845155800
Enclosed: Company and Product brochures.

Vidhya Sangha Technologies Pvt Ltd

Vidhya Sangha Technologies

Education for a Sustainable Future

E: a.t.kishore@ieee.org ; Ph: 9742310003 ; vidhyasanghatech.org

Date : 13-04 -2020

Subject:

Letter of Industry-support for Project title " Edge-AI based Social-Distance Tracker IoT Camera" with Dr Rajiv Misra(IIT Patna) and team- Sourashekhar Banerjee(Research Scholar) reg.

This letter is in support of the proposal titled **“Edge-AI based Social-Distance Tracker IoT Camera”** by PI Dr Rajiv Misra from Indian Institute of Technology Patna (IIT Patna) as part of DPR of ICPS-IH(DST, GoI) of IIT Patna. The project proposed by Dr Misra and his team plans to explore the challenges and opportunities of video analytics on the fly using big data computing powered by Edge Computing, Cloud and AI for tracking Social Distancing in public areas such as Market, Streets etc. The project takes a holistic approach that uses video-data and uses AI, Deep Learning using cloud applications to Edge systems to make the prediction of tracking social distance events using video-analytics .

As a startup company, we would be interested to participate in the project as an Industry Partner..

Thus, we are quite excited about the project and look forward to building a longer-term relationship with the research team.

Please feel free to contact me if you have any questions about this collaboration.

Yours sincerely,

Kisbore A T,

CEO, Vidhya Sangha Technologies

Re: Request to be a member of Govering Body of Technology Innovation Hub on "Speech, Video & Text Analytics"

From : Anoop (അനൂപ്) <anoop.kunchukuttan@gmail.com> Wed, May 20, 2020 08:14 AM

Subject : Re: Request to be a member of Govering Body of Technology Innovation Hub on "Speech, Video & Text Analytics"

To : Dr. Sriparna Saha <adean_rnd@iitp.ac.in>

Cc : Pushpak Bhattacharyya <pushpakdiro@gmail.com>, director <director@iitp.ac.in>, asif <asif@iitp.ac.in>

Hello Dr. Saha,

I will get back to you by early next week. Hope that works.

Regards,
Anoop.

On Tue, May 19, 2020 at 12:00 AM Dr. Sriparna Saha <adean_rnd@iitp.ac.in> wrote:

> A gentle reminder!!
>
> We are in the process of finalizing the DPR which has to be submitted soon in DST. Kindly send us a few topics (with abstract) on which in future there would be an open "call for proposal" from our TIH.
>
> Best regards
> Sriparna
>
> --
> Dr. Sriparna Saha
> Associate Dean, Research and Development
> Associate Professor, Department of Computer Science and Engineering
> IIT Patna, Patna, India-801106
> Ph no: +91-612-3028128, +91-612-3028291
> Other Email ids: sriparna@iitp.ac.in,sriparna.saha@gmail.com
> Web: http://www.iitp.ac.in/~sriparna/
>
> ---
>
> **From:** "Dr. Sriparna Saha" <adean_rnd@iitp.ac.in>
> **To:** "anoop kunchukuttan" <anoop.kunchukuttan@gmail.com>
> **Cc:** "Pushpak Bhattacharyya" <pushpakdiro@gmail.com>, "director" <director@iitp.ac.in>, "asif" <asif@iitp.ac.in>
> **Sent:** Sunday, May 10, 2020 10:34:52 PM
> **Subject:** Re: Request to be a member of Govering Body of Technology Innovation Hub on "Speech, Video & Text Analytics"
>
> Dear Dr.Anoop,

> Thanks for agreeing to be a part of our TIH. We are currently working on preparing the Detailed Project Report (DPR) to be submitted to DST.
>
> In this connection, we request the following help from you.
>
> Please note that 50% of the TIH funding has to be utilized for distributing projects to experts of other institutes working on the theme of " Speech, Video & Text Analytics".
> In the DPR, we have to mention about few problem statements, on which in future there would be an open "call for proposal" from our TIH.
>
> We request you to kindly suggest a few proposals (short abstract) related to the theme of TIH which can be considered for future "call for proposals".
>
> Looking forward to hearing from you.
> Best regards
>
> Sriparna
> --
> Dr. Sriparna Saha
> Associate Dean, Research and Development
> Associate Professor, Department of Computer Science and Engineering
> IIT Patna, Patna, India-801106
> Ph no: +91-612-3028128, +91-612-3028291
> Other Email ids: sriparna@iitp.ac.in,sriparna.saha@gmail.com
> Web: http://www.iitp.ac.in/~sriparna/
>
> ---
>
> **From:** "anoop kunchukuttan" <anoop.kunchukuttan@gmail.com>
> **To:** "Dr. Sriparna Saha" <adean_rnd@iitp.ac.in>
> **Cc:** "Pushpak Bhattacharyya" <pushpakdiro@gmail.com>, "director" <director@iitp.ac.in>
> **Sent:** Monday, April 20, 2020 2:38:14 PM
> **Subject:** Re: Request to be a member of Govering Body of Technology Innovation Hub on "Speech, Video & Text Analytics"
>
> Dear Dr. Sriparna,
> Sorry for the late reply, I got caught up with a few things.
>
> It is great to hear about TIH being set up at IIT Patna focussed on text, speech video analytics. Congratulations on that and I look forward to great innovations from the TIH.
>
> Thanks for your kind invitation to be a member of the Hub Governing Body. It will be an honor for me to provide my inputs in the journey of the TIH. I look forward to this role. It would be great if I can talk to you sometime to better understand what the role entails before I can make a final call.
>
> Thanks and Regards,
> Anoop.
>
> On Thu, Apr 16, 2020 at 9:29 AM Dr. Sriparna Saha <adean_rnd@iitp.ac.in> wrote:
>
>> Dear Sir,
>>
>> Greetings from IIT Patna!

Hope this email finds you well.

I am Dr. Sriparna Saha, Associate Dean Research and Development of IIT Patna.

DST has decided to set up a **technology innovation hub** (TIH) at IIT **Patna** as part of its National Mission on Interdisciplinary Cyber-Physical Systems (NM-ICPS) program. The innovation hub will emphasize **speech, video, and text analytics**.

Each Hub will have its Hub Governing Body. This Hub Governing Body shall be the Apex body for overall supervision, control, directions and mid-course correction in the implementation of Hubs at Host Institutes.

As you are an expert in the field of "Speech, Video & Text Analytics", we would be highly obliged if you agree to be a **member of this Hub Governing Body.**

Look forward to hearing from you.
Best regards

Sriparna

--
Dr. Sriparna Saha
Associate Dean, Research and Development
Associate Professor, Department of Computer Science and Engineering
IIT Patna, Patna, India-801106
Ph no: +91-612-3028128, +91-612-3028291
Other Email ids: sriparna@iitp.ac.in,sriparna.saha@gmail.com
Web: http://www.iitp.ac.in/~sriparna/