I prefer my teams to be smaller: 2 Members.

Under specific circumstances 3 members. I would accept if and only if there is a strong motivation from the project scope and complementing the skills of the team member.

If you are interested email me with the resumes and a 100-word statement outlining why I should offer this project to you.

Do not ask for just more details, all required details are here. If you want, ask SPECIFIC questions.

1. ~~Neural Machine Translation: English to (Tamil / Sinhala)~~

~~This is to build a Neural MT to translate from English to one of the languages mentioned above. I suggest we try to transfer learning or meta-learning for this challenge. Data are available.~~

1. Word embedding Evaluation Tool: This is to build a unified tool to evaluate Wordembeddins produced. Evaluating existing several word embeddings. <https://pdfs.semanticscholar.org/343d/39534682bb7b2eec14f573360877eb80cd59.pdf>
2. Theoretical Research1: Explainable Deep Nets: Read this for more details and write me a proposal. <https://arxiv.org/pdf/1905.08883.pdf> Need strong candidates who aspire to obtain a Ph.D. I will not entertain those who seek industrial placement right after the degree with this project.
3. Theoretical Research2: Embedding knowledge into Deep nets: Read this for more details and write me a proposal. <https://arxiv.org/abs/1905.03540> Need strong candidates who aspire to obtain a Ph.D. I will not entertain those who seek industrial placement right after the degree with this project.
4. ~~Voice Feature Extraction: This research is to explore the latest deep learning techniques towards a strong voice embedding network. I have a couple of such networks researched already. You have to find a way to merge the best of them both and improve upon it.~~
5. Merging Word Embeddings: There have been several proposals on how word embeddings from 2 different embeddings can be merged. I need a strong candidate to propose a novel approach to do this in an effective way. <https://stats.stackexchange.com/questions/315272/merging-two-word-embedding-models>, <https://www.kaggle.com/c/quora-insincere-questions-classification/discussion/71778>
6. An improved OCR for structured Documents. An OCR that works well for structured documents such as, Birth certificate, Medical Report Etc. **This is a startup idea from a team comprised of Amazon engineer, POST Doc and a CEO of another successful startup**
7. ~~Credit Rating for Banks: Data available under NDA. You need to build a credit rating model using data to help grant loans and credits in banks~~
8. ~~Data Governance Framework for Social Media. This is a highly demanded tool by FB, twitter and other social media. Buck for a good bank. You need to build a sequence tagger with other enhancement to auto annotate tweets and FB posts to meet privacy expectations.~~
9. A Bank Assistant: Conversational AI. We would be building a bot to which we can talk to and it will assist us with banking queries. Local language (Sinhala/Tamil)
10. Information extraction in Voice Medical Conversation. We have researched on speaker detection and voice clustering thus far. As a continuation of that project what is beneficial is to extract out the conversation between individual in the context and be able to narrate the conversation in the perspective of a physician in the room. An example scenario is Patient John is at the clinic with Doctor Tom for patient’s diagnosis. The conversation will have back and forth discussion trying to understand why the patient visited the doctor. Once the patient receives the consult the doctor now has to summarize the conversation for record keeping. This project aims at extracting the conversation between the two parties and write a summary narrative in English and make sure the appropriate medical context is preserved. **This is a startup idea from a team comprised of Amazon engineer, POST Doc and a CEO of another successful startup**
11. Supermarket Object Detection: We have a project that identified a problem domain where average retail customers would go buy things for their daily consumptions and use a mobile app to capture their receipts. The receipts will be used to manage their expense and seek for any deals or savings from their purchase for their upcoming shopping. However, the customers don’t often have a method to lookup available savings in a frictionless manner. In order to address the problem we are proposing a method where the user should be able to snap a picture of their shopping cart and be able to identify all the items in the shopping cart for our system to determine any savings instantaneously. This project should at least be able to detect the items in surface level in the shopping cart to determine any savings at all, and the system should be able to correlate the determined items from the shopping cart with that of the receipts after purchase. **This is a startup idea from a team comprised of Amazon engineer, POST Doc and a CEO of another successful startup**

**Pick Me Projects, Some of them are co-supervised by Dr.Sehan.**

**1. Image verification on Driver registration (verify Vehicle Images, Plate details, registration certificate, Driveriving License, Driver image to automatically cross check and validate) This is very similar to Project 7. So I will do it together.**

**2. Regular driver verification, verify the profile image, plates etc on regular intervals + Face Recognition based login**

**3. Route Discovery - Can we discover the best route for specific times of the day for different vehicle types based on actual trip paths**

**4. Shuttle Route Optimization - Infer regular travel patterns and suggest routes where scheduled services might make sense**

The immutability and consistency properties of Blockchains prevent unauthorized manipulation of data. However, they are not designed to detect the integrity and quality of the data added to a blockchain where they exhibit garbage in, garbage out behavior. The inability to represent the physical state in digital form (i.e., parity between physical and digital world) with certain reliability and quality guarantees has been hindering the adoption of blockchains in many domains such as supply chains and Internet of Things (IoT) related applications. While there are high-level proposals on performing data integrity/quality validations at a trusted oracle (a node that feeds data into a blockchain) or a using smart contract to validate them, those solutions are application-specific, static, and lack performance. Moreover, these attempts try to reinvent the wheel, as domains such as Complex Event Processing (CEP) are known to have matured solutions in detecting complex events in diverse applications while having simplified query languages, efficient data structures, and high performance. Therefore, the objective of this project is to develop a CEP engine as a smart contract library that can be called via another smart contract or an oracle to attest the reliability and quality of data added to a blockchain. Moreover, the solution should provide a mechanism to auto-generate smart contracts (caller contract) that call desired functions on CEP engine contract (callee contract) based on the high-level CEP queries given by the user to test the quality of data added to a blockchain at the time of transaction acceptance/inclusion in a block. This project requires an understanding of blockchains, CEP, and data quality. While initially the project can be build around Ethereum and Solidity, it is expected to be an opensource tool/library that can be extended to other blockchain platforms such as Hyperledger and smart contract languages such as Node.js, Go, and Java.  
  
External Supervisors - Dr. Dilum Bandara (Data61) and Dr. Srinath Perera (WSO2)