Proposed Specification of Intelligence Collation & Gathering System Software (I-CAS)

S No		Qualitative Requirements	Trial Directives/ Functionality
1.		**General Specifications**:- Collect Intelligence Inputs from different sources and build a centralized database for analysis & reporting, connects all frontiers and servers information at real time, centralization of disparate solutions for key Data and Seamless Integration of Multi-Party system, Design and build a Central solution across one complete hierarchy chain.	The System should be able to collect data from all frontiers and build a central database which can be shared among all hierarchy.
2.		**Web Application**:-	
	i	The overall vision of this application is to create a system that facilitates the Intelligence team to anticipate and find the relevant information of activities happened, people and organization involved in activities.	The Applications should be user friendly and easy to use so that goal of gathering information and intelligence can be created.
	ii	This web application will be hosted on two servers, the server with-in the network and also on the server outside the network. For security reasons the server outside the network will contain only two days information. A database sync utility will be manually run every day on a specific time (generally on a low traffic period) which will move daily activities information from outside server to inside server.	The application must be a truly client server architecture with two database server, master & slave.
	iii	The intelligence team of SSB Head Quarter will review activities received through application on outside the server. They will do formatting of activities if needed, merging of activities if received from multiple sources on the basis of unique activity identity generated by system so that all the relevant information will be available at one place.	HQ must have this functionality, At HQ Level, gathered data can be manipulated according to relevancy and send to multiple hierarchy accordingly.
	iv	If the information is critical then they will share the information with DIG, IG, ADG and/or DG. All the officers will receive that information on application.	If information tagged with critical flag then it must generate an alert for higher level hierarchy.
	v	The main purpose of this application is to feed the daily activities with details of people and organization involved in that activity and generates reports.	Application must accept daily activity with details.
3		**The modules of web application are:-**	
	i	**Dashboard**	
		a) A page to see all the relevant information at one place	Widgets for marking important must be shown.
		b) This would be customized for different ranks/ designations.	
	ii	**Feed events**	
		a) This module is used to feed the events. Based on type of events there will be different forms to feed event information.	A separate module must be provided to feed event.
	iii	**Person**	
		a) Maintain person profile.	Each profile must be maintain separately
		b) Tag events with person so that we will get all the events related to that person.	Provision for persons can be tagged with events.
	iv	**Organization**	
		a) Maintain Organization profile	Provision for maintaining organization profile.
		b) Tag events with organization so that we will get all the events related to that organization.	Provision for Organization can be tagged with events.

	vi		Application should maintain person profile and Tag events with person so that we will get all the events related to that person	Provision for persons can be tagged with events.
	vii		Application should maintain Organization profile and Tag events with organization so that we will get all the events related to that organization.	Provision for Organization can be tagged with events.
	viii		Application should mange important entities like event type, sector headquarter. Battalion headquarter etc.	Provision for maintain all itineraries in database separately
	ix		Application should generate subjective Reports, Analytical Reports, comparative Reports and Graphical Report.	Provision for creating multiple reports & add on must be provided.
	x		Application should have the functionality to manage users like Add/Modify User, Assign rights to user and Disable or delete a user.	Provision for Add, Delete, Modify, Assign or disable rights to user must be provided
	xi		A database syncing process should be there to move daily activities information from local server to secure server.	The architecture of data must support master slave or master.
7	**SPECIFICATIONS OF AUTOMATED REPORTING &ANALYSIS SOLUTION:-**			
	i		Application should store and process the captured event information like coordinate, location, picture, video and textual information	Provision for capturing information's like coordinate, location.
	ii		Application should generate different reports and provide data mining	Provision to generate automated reports.
	iii		System should capable to input data from different sources	Provision for data inputs from different source.
	iv		Confidential information by the End user can be mined and modelled to create intelligent reports pertaining.	Provision for creating multiple reports & add on must be provided.
		a	Incidents	
		b	People	
		c	Organizations (Crime)	
		d	Area of effect	
8	**General Reports** :-			
	i		Daily Intelligence Report	System / Software should be able to generate reports as per requirement.
	ii		Daily Incident Report	
	iii		Weekly Intelligence Assessment report	
	iv		Monthly Intelligence Requirement Report	
	v		Monthly Intelligence Assessment Report	
	vi		Monthly report on Mosque/Madrasas	
	vii		Monthly details of OPS conducted on the basis of input submitted by 'G' setup	
9	**Reports on Coordination Meetings** :-			
	i		Coordination meeting with Counterpart (Bn& SHQ level)	System / Software should be able to generate reports as per requirement.
	ii		LIA Meeting (SHQ & FTR Level)	
	iii		SMAC Meeting	
	iv		MAC Report	
	v		HS Level Meeting report	
	vi		Border district coordination meeting	
	vii		Report on Coordination meeting between DG, SSB & Chief RGOB & Chief of Nepal Police.	
	Note:- All the parameters specified above should be checked by constituted Board of officer for functionality of software as per requirement.			

Proposed Qualitative requirements / TDs for Machine Learning and Natural Language Processing:-			
S No	**Qualitative requirements**		**Trial Directives / Functionality**
1	The solution should be able to ingest multiple documents, PDF's, text files etc at the same time. The system should accept the documents in a batch form or one at a time. The documents should be visible in a pipeline. The user should be allowed to drop the document as long as it's in the pipeline and hasn't been processed yet. The system should identify and remove duplicates		Solution should be able to upload documents in the PDF, Document & Text files Formats. System should upload the same word document again. The system should also to identify it's a duplicate and discard the document. It should only be accepted once
2	The system should have different workflows for ingesting different type of documents. The workflows should be customized for different sort of documents available with the customer		
3	The system should automatically extract and store metadata against every document type uploaded. The metadata should be used to carry out analysis to link different documents together		System should extract information about the document including Creator, creation date, last modified date etc and store it in database. User should be able to view and all documents from same creator
4	System should be able to run all machine automation algorithms on English, Hindi		Vendor should showcase that the system can accept documents in English as well as Hindi. User should be able to search in both languages.
5	The system should be able to extract the following from un- structured data:-		
	i	People	Vendor should upload a document with all the information in a standard word document. The system should automatically identify and extract the information to an accuracy of 60% or more.
	ii	Organizations	
	iii	Places	
	iv	Events / Incidences	
	v	Addresses	
	vi	Images and Videos	
	vii	Phone Numbers	
	viii	IMEI's / IMSI / Phone Model	
	ix	Vehicle Numbers	
	x	IP Address	
	xi	Credit Card / Debit Card Numbers / Bank Account Numbers	
	xii	Dates and Timelines	
	xiii	Mail ID's	
6	The system should have a pipeline of RPA (Robotic Process Automation) modules to cleanse the data		The system should automatically replace the abbreviations in the text with grammatically correct words to be picked from a dictionary. The system should also pick up names of organizations and locations from the dictionary and match it against the document to ensure that no organizations or locations are missed by Natural language processing

7	The system should have modules to ingest – • Passport Data • Vehicle registration data • License Card Holder data • CDR's • Interrogation reports • Intelligence inputs • FIR's • Travel related information • Crime related information • Other data sources available with the client	Vendor should show that different types of data such as internal documents, images, videos, news articles and OSINT data can be merged in the same system
8	System should be able to run all machine automation algorithms on English, Urdu and Hindi.	Vendor should upload documents in different languages and show that the user can search in different languages
9	System should have multiple Machine Automation models as an integrated framework	Same as above points
10	System should classify the documents into multiple categories using statistical models	Vendor should showcase that the uploaded documents are automatically classified in specific categories
11	System should extract all entities from unstructured data. The entities which need to be extracted have been defined as above.	Same as point 5
12	System should extract and save themes from all unstructured data. The system should be able to form a conceptual and contextual understanding of all data	Vendor should showcase that the uploaded documents are automatically classified in specific categories and themes (heavy weight keywords) extracted from it
13	System should be able to carry out entity disambiguation from large datasets. The threshold to distinguish two entities should be in the hands of the client	Vendor should be able to show that all references to a particular organization or location are clubbed under that one organization / location and multiple entries of the same organization / location are not made in the database
14	System should carry out sentiment analysis on every document uploaded into the system. The system should also carry out sentiment against every entity mentioned in the document	Against every entity in the document, the system should give sentiment for that entity in that document
15	System should be able to automatically build relationships among entities	On the network graph system should show entities connected to each other
16	System should have inbuilt metrics for similarity, regression, correlation and recommendations metrics	Vendor commitment to train the models on user's data. These models come into play
17	System should have options for 'What if' hypothesis	Vendor commitment to train the regression models on users data
18	The system should have capability to stack multiple regression modules into a pipeline – Stack one algorithm over another and subsequently carry out ensemble for predictive intelligence on threat alerts	Same as above two points
19	The system should have inbuilt AI models for Facial Biometrics. The facial biometrics should run across all images to identify individuals who are available in multiple images. The system should	Vendor should upload images and videos and show that the people in the images or in the videos are automatically matched against the images in the database

	automatically detect images from the uploaded images and videos and match it against all the pictures available to the user	
20	The system should have AI library for object identification and should be automatically able to identify objects out of the picture and tag them	Vendor should upload videos / images of weapons and system should identify that there is a weapon in the image
21	The system should have modules and provisions to resolve entity collision and prompt the user for final decision to resolve a collision	Same as point 13
22	The entire source code of application have to submitted to escrow account	Vendor commitment
	Note:- All the parameters specified above should be checked by constituted Board of officer for functionality of software as per requirement.	

Proposed Qualitative requirements / TDs Big Data Analytics :-		
S/No	**Qualitative requirements**	**Trial Directives / Functionality**
1	The solution should be able to analyze TeraBytes of data. It should be able to hold as many documents as can be held in the database. The solution should be built on a Big Data Analysis framework such as Hadoop, MongoDb, GraphDB etc. Any other Big Data analysis framework may be used as well	Vendor should provide details of the database being used and provide evidences from the net that the databases in use are not Relational DB.
2	The system should bring disparate datasets into a single library	User should be able to search or filter for documents belonging to a particular classification.
3	Solution should have support for any structured and unstructured data sets and should build indexes efficiently for easy search, discovery and analysis, using compaction and indexing techniques	System should support full text indexing. The user should be able to search for any keyword available in the text.
4	User should be able to configure alerts for any new update on their key analysis. The system should generate an automatic alert if any new input is available for user defined keywords	The user should be able to configure his dashboard to show alerts. All new information on the topic should automatically be reflected on the dashboard.
5	It should be possible to add new data sources to the existing data repository for increasing the scope of analysis	Vendor's certificate.
6	The system should be able to ingest mail dump in the form of Psd file and automatically extract intelligence from it	Vendor should upload an outlook dump file and show that the system can automatically extract email ID's from it apart from text on which classification will take place
7	It should be possible to carry out a conceptual search across the entire data set	User should be able to search or filter for documents belonging to a particular classification.
8	The solution should classify documents in specific categories. Should automatically relate or link multiple documents whereas files in different formats can be linked to each other	Vendor should be able to showcase different types of searches as defined.
9	The solution should provide modules for a user to define his alerts in.	The user should be able to configure his dashboard to show alerts. All new information on the topic should automatically be reflected on the dashboard.
10	Solution should have multiple searching algorithms including – ▪ Fuzzy search - For matching meta-tags and return a list of most likely correction of given words. ▪ Boolean Search – Uses APCM (Adaptive Probabilistic Concept Modeling) like technique to rank the results that match the Boolean Query. Supported operators are AND, OR, NOT, XOR. ▪ Conceptual Search – Enable searches to be processed and retrieved conceptually on the concepts against the article. ▪ Keyboard Search - Prophecy conceptually matches queries that	Vendor should be able to showcase different types of searches as defined.

	consist of a single keyword. It stems the keyword, and then it finds documents that contain words that have the same stem as the keyword. Support all document formats. ▪ Phrase Occurrence Search - Uses a phrase occurrence search to find documents containing a range of occurrences of a phrase. ▪ Default Phrase search - Uses quotation marks ("..") to treat the string as a phrase and return only documents in which a matching phrase occurs. ▪ Exact Phrase search - Querying with a term or a phrase in quotation marks, it matches them in their exact pre-stemmed form. ▪ Proximity Search – Looks for documents where two or more separately matching terms occur within a specified distance, where distance is the number of intermediate words or characters. Proximity search goes beyond the simple matching of words by adding the constraints of proximity. ▪ GIS Search : Ability to search for geotagged events and entities by drawing a polygon fence on the map.	
11	The system should allow manual and fully automatic linking between related pieces of information, regardless of their format. The concept in document should be linked to those in another file. They can also be linked to related concepts within video or email.	Vendor should be able to showcase different types of searches as defined.
12	System should support full text searching on the entire dataset available	System should support full text indexing. The user should be able to search for any keyword available in the text.
13	System should have a visual link analysis platform to with multiple components – • Relationship graph to study co-relation among people, places and topics • Geospatial analysis • Timeline Analysis • Preconfigured analysis • Charts and Reports	System should have inbuilt capabilities for analysis on Link graph, GIS, Timeline analysis, reports. The user should be able to create his charts on the fly.
14	System should allow the analyst multiple queries to retrieve data – • Simple 360 degree search • Complex Queries for structured filters • Multiple views to search for separate entities	User should be able to search data through multiple ways as defined.
15	User should be allowed to view the entity related document, the search entities should be automatically highlighted. Users	Vendor should be able to showcase that the entities in the ingested documents are already highlighted.

	should be allowed to mark new entities in the document on the fly	
16	Should be able to generate a relevancy graph for each of the various entities including location, name of person, name of organization, keywords etc	Vendor should be able to showcase that the entities in the ingested documents are already highlighted.
17	System should have multiple views for a user to be able to get a 360 degree view on a person, location, organization or an event	User should be able to search for a person or an organization or location and get all information about that entity (from multiple uploaded documents) in one single view
18	Should be able to depict the relation between various extracted entities in a graphical form with a representation of how strongly one is connected to other. The graph should be dynamic as clicking any link should open the relevant content.	Network link analysis between entities. On clicking any entity the relevant content should be shown on the right.
19	Investigators should be able to annotate reports, read and write comments/annotations on reports to aid in collaboration of work during investigative phase	User should be able to create a report from within the software itself by choosing multiple things which need to come into the report
20	Query templates should support entity based search and the time should be selectable on a timeline	Vendor should show timeline analysis.
21	System should have multiple widgets and dashboards. The user should be able to create his own dashboard using any of the multiple widgets available inside the system	User should be able to create his own dashboard.
22	Report generation module should be capable of generating reports based on query / result to include at least the following – • Time / Date based query • Topic Importance / Priorities • Location based • Name of people / organization / group	User should be able to create a report from within the software itself by choosing multiple things which need to come into the report.
23	Information should be displayed using advanced visualization and charts	Vendor should be able to showcase that the entities in the ingested documents are already highlighted.
24	The solution should have support for Association, network, link, temporal and statistical analysis to help build a comprehensive analytical picture, revealing relationships, patterns and trends in data	Vendor should show timeline analysis.
25	The system should allow analyst to carry out partitioning and sharding of data. The user should be able to join different data tables together and carry out excel like functions – filter, sort and pivot on the data set.	User should be able to do pivots or filters on the dataset.
26	User should be able to view any of the data in the Big data repository and dynamically create charts on the data set	User should be able to create charts and reports on the fly.
27	The system should have GIS capability of variety of GIS operations of simple	System should have inbuilt GIS modules with capabilities as defined.

	mapping, indexing and Spatial analysis	
28	The system should have advance overview wizard for spatial analysis, Statistical analysis and suitability analysis.	User should be able to search for events on GIS.
29	The system should have Geospatial search and analysis to leverage operational analytics	User should be able to search for events on GIS.
30	The system should plot data from classified and open source database to track activities.	Vendor's commitment to include databases relevant to the customer. Vendor should be able to showcase some open source databases that have been scrapped and data available
31	The system should have Analytics on – • Clustering • Scatter Plots graphs • Timeline Analysis • Query on map layers	User should be able to cluster events on GIS together.
32	The system should have GIS data editor to use the map to add and update features to populate empty map layers.	User should be able to mark locations on GIS.
33	The system should have map layer to access analytic functions and to display and interact with the layers. It should contain facility to store, capture, query, analyze and display information on geographic layers. Provision should exist to create and save layers that encapsulate all of the GIS aspects necessary for map display, map analysis, data compilation and management. Sharing of map/ layer packages to adopt and share common views	User should be able to cluster events on GIS together.
34	The system should have Provision to create, display and edit military symbols on map and plot movement of unit / formation / group on map	User should be able to mark locations on GIS.
35	Digital terrain and elevation data should be available for advanced analysis. There should be provision to further analyze in detail routes of interest with 3-D view	Vendor's certificate to upload GIS tiles available with the client.
36	The system should have capability to manage multiuser databases that can be used and edited simultaneously by multiple users. This should be scalable to N number of users.	Vendors to provide undertaking in this regard.
37	The entire source code of application have to submitted to escrow account.	Vendors to provide undertaking in this regard.
	Note:- All the parameters specified above should be checked by constituted Board of officer for functionality of software as per requirement.	

Proposed Qualitative requirements / TDs Social Media Analytics:-				
S/No	Qualitative requirements			Trial Directives / Functionality
	Classification	Sub Category	Specifications	
1	Data Investors		Support for both Crawling and Scraping model as well as ability to ingest data from proprietary API's wherever available from the source end	Vendor's certificate
			Ability to Browser Based interactive scraping. This is for instances where there is a lot of Java Script based backend code and multilevel user interaction is mandatory	Vendor should be able to showcase scrapping from face book as a proof of this
			Ability to ingest data from multiple Social Media and Web platforms via their API's	Ingest data from multiple social media platforms
			The system should support for Twitter streaming API for real time (minimal latency) data retrieval	Vendors architecture diagram
			Ability to provide input filters in the form of: • Geographically bound polygon • Multiple keywords • Multiple Social media handles	User should be able to search / retrieve data based on defined filters
			System should have a provision for a user to add his own "twitter", "face book" profiles as Avatars. These profiles are used to get data off private profiles on face book.	Vendors architecture diagram
		Twitter	The system should support for full time stack of Twitter REST API to extract extended information	Vendors architecture diagram
		Facebook	Support for full stack of Facebook Graph API for ingesting available information	Vendors architecture diagram
			Support for Scraping based Facebook data extraction to get access to FB data that is hidden from Graph API	Vendor should show scrapping of a private profile as a demonstration
			Support for handling and creating multiple FB avatars to ensure scraping	Vendors architecture diagram
			Support for scheduling these FB avatars scraping periods and frequencies	Vendors architecture diagram
			Support for doing precision FB crawl by specifying or assigning Avatars to specific profiles or groups	Vendors architecture diagram
			System has a provision to monitor people profiles in Facebook	Vendor should show scrapping of a person's profile as a demonstration
			System has a provision to identify suspicious profiles on Face book based on users likes	Vendor should demonstrate that the system can identify common profiles who have liked multiple pages of terrorist groups
		Instagram	Scrapping of data from Instagram via specific scrappers created for Scrappers	Vendor should show demonstration of scrapping

				data from Instagram
		YouTube	Support for full stack of YouTube API	Vendor should show demonstration of getting data from YouTube
		News and Blogs via RS Feeds	The framework should be able to ingest data on real time basis from multiple RSS feeds. The framework should be intelligent enough to manage the update frequencies to ensure the sanctity of data	Vendor should show demonstration of getting data from Newsfeeds
		Dark Web	The system should be able to extract information from Dark Web marketplaces	Vendor should showcase demonstration of getting data from Dark Web
		WhatsApp	The system should be able to inject and analyze data from WhatsAPP via connecting the phone to the server. System can get data related to Group Name , Group Admin, Group Created On, Group Image, Group Status, Total number of contact ,msg, image, video, document in groups, List of all Participant, All Image of Group, All Video of Group, All Document Uploaded in Group, All Contact share in Group, All Location share in Group	Vendor should be able to connect a phone to the system and show WhatsAPP data on the server
		Google Blogs, Tumbler &Word press	The system should ingest data from Google Blogs, word press and tumbler	Vendor should showcase demonstration of getting data from defined social media sources
2	Big Data Repository		System should be built on Big Data repository to handle large amounts of data	Vendors architecture
			System should have complete provision for proper and optimized indexing mechanisms to ensure fast response to analytical queries	Vendors architecture
			Database should be scalable enough to ensure fast insertion of high volume streaming data	Vendors architecture
			Properly managed to ensure de duplication and optimized storage capacity usage	Vendors architecture
3	Text Analytics		System should have inbuilt NLP capability to carry out entity extraction from unstructured data in the form of – • People • Places • Events • Organizations	Vendor should be able to demonstrate that the system is able to extract people, places and organizations from OSINT data

			System should automatically calculate sentiment against a piece of text. Individual sentiment analytics should be done against entities defined in the text. The sentiment score should be carried out internally and not using a third party library over the internet.	System should do sentiment analysis against the entered text automatically. The user should be able to train the model on his own by changing the sentiments of data
			System should classify every piece of text and extract themes out of it.	System should do classification against the entered text automatically. The user should be able to train the model on his own by changing the sentiments of data
			System should carry out text summarization on the data inside the system	System should do summarization of news articles
4	Data Analytics, Reports and Dashboard		User should be able to create views in the form of – • Geo Fenced data • Keywords or events • Persons	User should be able to search for data in the defines formats
			From Whatsapp data system should do Classification Chart of all msg, Sentiment chart of all msg, Emotional chart of all msg, All msg time line, msg chart per week wise, negative users, positive users etc.	System show sentiment and classification of the whastApp messages
			System should have a custom query builder to carry out Boolean operations	Demonstrate custom queries to search for data
			System should have multiple widgets and allow a user to create his own dashboard using any of the widgets available to him	Demonstration where a user can create his own dashboards
			Dashboard view should provide information in visually rich form factors in terms of Maps, Charts, Tag Clouds, Sort lists etc..	Demonstration of said features
			Dashboard view should collate data from all sources relevant to user's analysis	Demonstration of said features
			Users should have option for a quick access time filter on a day, week and month basis	Demonstration of said features
			All analysis should be filterable by date range	Demonstration of said features
			All analysis should be filterable my multiple text filters	Demonstration of said features
			Filter data based on input keywords using multiple Boolean operations	Demonstration of said features
			Option to use multiple search option in combination with each other	Demonstration of said features
			System should give the user the capability to do Deep Dive analysis into each source separately	Demonstration of said features

			System should have an event calendar	Demonstration of said features
			System should allow users to have multiple views in terms of trends, timelines, viral media, user views etc	Demonstration of said features
			System should have flexibility to do timeline/temporal analysis to understand the flow of events	Demonstration of said features
			System should have a link analysis module to understand the interrelations amongst many entities	Demonstration of said features
			System should have multiple reports for different platforms like user comparison, hashtagvirality, sentiment charts etc	Demonstration of said features
			Multiple analytical containers should be sharable among users	Users analysis should be sharable among different users
			System has a provision to identify suspicious profiles on Facebook based on users likes	Vendor should demonstrate that the system can identify common profiles who have liked multiple pages of terrorist groups
			System can identify people on social network based on mail ID, phone numbers etc	Demonstration of said features
			System should identify trends, key influencers against a particular event	Demonstration of said features
			System identifies Geo Locations (wherever possible) for user checking, pictures, tags, tweets etc	Demonstration of said features
			System supports report generation in terms of graphs, documents, xls, pdf etc	Demonstration of said features
			System supports Fuzzy search, Proximity search, Conceptual search on the gathered	Demonstration of said features
			System supports cascading query results i.e. subsequent queries should be possible	Demonstration of said features
			System has a link analysis module to identify common followers / common following / friends etc of multiple profiles	Demonstration of said features
5	System overview		System should be scalable to add more Sources when available	Vendors commitment
			Should provide a full system and Subsystem health overview, alerting system technicians to servers that are down or to services that are running	System health overview
			System should have a Two factor Authentication system for login access	User should get an OTP for access
6	The entire source code of application have to submitted to escrow account			Vendors commitment
	Note:- All the parameters specified above should be checked by constituted Board of officer for functionality of software as per requirement.			