

Pre Bid Query for Selection of Agencies/Partners for developing Question and Answering (Q&A) Agent – chatbot on UPkSK

SN	RFP Clause no. / Sl. No. / Para no. & Page no.	Description as per RFP	Query	Response
1	Page 5 of 48 - Scope of work	To develop a voice-enabled conversational Q&A agent that uses information from UPkSK to identify the closest/appropriate facilities that offer the services needed by the citizen.	Who will be providing the Infrastructure required for hosting the Agent / Chatbot / Conversational Agent	A) Development + Staging - Preferably, AWS cloud infrastructure will be provided by the selected bidder. Accordingly, it will be the vendor's responsibility to determine and deploy the specifications most appropriate for the project. B) Pre Prod and Production - AWS cloud infrastructure (as of now) will be provided by IHAT/GoUP Note - During the transition to the Government of Uttar Pradesh (GoUP), the selected bidder shall provide full technical and operational support for migration in the event of a change in server infrastructure from AWS Cloud to any other hosting environment such as Microsoft Azure, Google Cloud Platform (GCP), or an on-premises/dedicated data centre.
2	Page 5 of 48 - Scope of work	To develop a voice-enabled conversational Q&A agent that uses information from UPkSK to identify the closest/appropriate facilities that offer the services needed by the citizen.	The Conversational Agent / Agent need to convert voice to text as well as Text to Voice - Please clarify	Conventional agent only needs to convert speech to text. It wont need to convert text to speech.
3	Section 6 – Deliverables & Timelines / Pilot Phase	Pilot implementation	Please confirm that the pilot phase is intended for limited-user evaluation only, and that any functional enhancements or scope additions beyond the defined Scope of Work would be handled through mutual discussion and formal change control. Also clarify whether the number refers to concurrent active users at a given time or average daily active users.	1. The final testing will be conducted by the provider. The chatbot will be tested with approximately 200-500 users. Users who speak Bundeli, Awadhi, and Khadi Boli will be selected for testing. 2. Yes, any changes/additions to Scope of Work will be handled by mutual discussions. 3. The average daily users may be 150 -200
4	Page 5 of 48 - Scope of work	To ensure the sustainability of the chatbot, use nationally recognized open-source LLMs such as Bhashini.	Is it Mandatory to use Bhashini as LLM or we can use other LLM Model or so , Which front-end technology preferred to give prompt to the Agent?	Usage of Bhashini model is mandatory and along with Bhashini, you can use other open-source LLMs. Make sure that the user/GoUP data is not forwarded to any server other than the server dedicatedly bought for this project.
5	Page 5 of 48 - Scope of work		Any data privacy or any compliance to be followed - Please specify as the data privacy is utmost important	All the compliant frameworks designed by GoI/GoUP should be followed such as Digital Personal Data Protection (DPDP) Act 2025, cyber security policies , ABDM etc.
6	Page 5 of 48 - Scope of work		What is the AMC/ATS duration expected meaning expected duration of Support & Maintenance - Please specify	No AMC since this is only PoC.
7	Page 5 of 48 - Scope of work		Any kind of reporting services required meaning any reports or dashboards required to judge the efficiency and effectiveness of the conversational agent ? Please specify	Dashboards are not required. However, detailed report may be produced to showcase the efficacy of chatbot.
8	Page 4 About UPkSK	UPkSK consolidates facility and geography-based information from multiple health data systems such as Manav Sampada, DVDMS, UPHMIS, HMIS, BCPM-MIS, etc., and provides administrator with critical information for decision-making.	In what formats will the source documents be provided? PDF Word Excel HTML Plain Text Database (e.g., SQL, NoSQL, or other formats)	All required documents for preparation of chatbot will be provided at the time of SRS development.
9	Page 4 About UPkSK	UPkSK consolidates facility and geography-based information from multiple health data systems such as Manav Sampada, DVDMS, UPHMIS, HMIS, BCPM-MIS, etc., and provides administrator with critical information for decision-making.	Could you please confirm which database is used across the multiple systems? Additionally, is it required to review and understand the database schema before we proceed, so that we can assess feasibility?	A) All current information is available in UPkSK itself. It uses MS-SQL. B) Historical data will be available on middle layer whose DB is My SQL. C) Complete API documentation will be shared upon signing the contract to understand database schema.
10	Section 5 – Scope of Work (Page 5–6)	Text & voice-enabled conversational Q&A agent	Please confirm that the scope requires only a web-based chatbot interface for pilot users, and does not require development of a mobile application (Android/iOS), as clarified during the pre-bid meeting.	Mobile and Web-based chatbot

11	Section 5 – Scope of Work	Single user ID for all users	Please reconfirm that the use of a single shared user ID and password is acceptable for the pilot phase, as explicitly stated in the RFP, and aligns with the intended evaluation approach.	Single Shared login credentials for all pilot users is acceptable
12	Section 5 – Scope of Work (Voice-enabled chatbot)	“Voice-enabled conversational Q&A agent”	Please confirm that “voice-enabled” refers to mic-based speech input/output within the application, and does not require inbound/outbound telephony, IVR, or dialler integration, as telephony requirements are not specified in the RFP.	No, this project does not require inbound/outbound telephony, IVR, or dialler integration
13	7. INDICATIVE TEAM STRUCTURE		Do we need to provide CVs of the proposed team or only the names is good enough - Please clarify	CVs of resources proposed by the bidder. The cv of every resource has to self attested. In section 4 of page 19 fill in the details.
14	Page 7 of 48 - 10. GENERAL QUALIFYING CRITERIA	The Bidder should have experience in similar nature of work in India of at least 1 year in the last 05 (five) completed Financial Years as on 31st March 2025 and must be in existence at the time of Proposal submission i.e., on Proposal Due Date	This means - Minimum one year of experience in the development and implementation of an Intelligent BOT or AI Agent is required - Please clarify	As defined in the Bid Document " Similar works is Development of AI Chatbot"
15	Section 10 – General Qualifying Criteria		May MSME bidders be considered for relaxation in the criteria?	No relaxation, RFP clause will remain same.
16	Section 10 – General Qualifying Criteria (Page 7-8)	“Projects of similar nature”	Please confirm that “projects of similar nature” refers to AI chatbot / AI-based conversational application development, and that healthcare domain experience may be from government or private sector, as discussed during the meeting.	Experience can be from both Government and Private Sectors.
17	Annexure 3 – Financial Proposal	Single line item project fee	Please confirm that the BOQ is a single-line, lump-sum project cost, and no module-wise or item-wise breakup is required in the financial proposal.	It is a single line item with Lump Sum Project Cost.
18	Annexure 3 – Financial Proposal	“Inclusive of HR Cost, Development Cost, Services and all other cost including taxes”	Please confirm that the quoted price is expected to be a one-time, all-inclusive project fee, with no recurring, subscription-based, or AMC charges, except GST as applicable.	It is one time, all-time inclusive project fee with no recurring, subscription-based, or AMC charges except GST as applicable
19	Section 16 – Payment Terms (Page 12)	Milestone-based payments	Please confirm that all activities including development, deployment, pilot evaluation, and handover are covered within the milestone payments, and no post-project AMC or support fee is envisaged.	A-Yes. B-No AMC is envisaged
20	Section 15 – Selection Criteria (Page 11)	QCBS methodology	Please confirm that the tender follows QCBS methodology (70% Technical : 30% Financial) and that L1 selection is not solely based on price.	The tender follows QCBS selection methodology and weightage as per the RFP
21	Section 4 – Key Dates & Bid Submission (Page 4-5)	Submission of bids	Please confirm that bid submission is strictly offline (hard copy submission) and that no online / email / e-procurement submission is permitted for this tender.	As per the method of submission defined in the RFP
22	Entire RFP	EMD / Bid Security	Please confirm that no Earnest Money Deposit (EMD) / Bid Security is required, as no such clause is specified in the RFP and was also confirmed during the pre-bid meeting.	Confirmed - No EMD
23	Section 19 – SLA & Penalties (Page 13)	Timelines and penalties	Please confirm that SLA penalties apply only to delays attributable to the Agency, and that delays due to IHAT / GoUP dependencies, data availability, approvals, or external factors will result in timeline extensions without penalty.	Confirmed
24	Section 20 – Force Majeure (Page 13-14)	Force majeure provisions	Please confirm that in case of force majeure or external events, timelines will be extended accordingly and no penalties will be levied on the Agency.	Confirmed
25	Section 22 & 25 – Data Ownership & Confidentiality	Data usage post handover	Please confirm that post-handover, IHAT / GoUP shall have the right to operate, modify, and extend the chatbot solution internally or through another agency, without vendor lock-in.	Yes, IHAT / GoUP shall have the right to operate, modify, and extend the chatbot solution internally or through another agency, without vendor lock-in
26	Section 5 – Scope of Work	Suggested follow-up questions	Please confirm that suggested follow-up questions are contextual conversational prompts only, and do not require workflow automation, task execution, or report generation.	Yes, follow-up questions are contextual conversational prompts only.

27	Section 5.2 – Efficiency Evaluation Criteria	Accuracy benchmarks	Please confirm that the 85% accuracy benchmark applies to a predefined pilot evaluation question set, and that accuracy may vary based on data availability, dialect complexity, and user input conditions.	<p>A) The final testing will be conducted by the provider. The chatbot will be tested with approximately 200-500 users. Users who speak Bundeli, Awadhi, and Khadi Boli will be selected for testing.</p> <p>B) The districts/blocks where Bundeli, Awadhi, and Khadi Boli are spoken will be randomly selected for the testing phase.</p> <p>C) An ODK-based tool will be developed by the provider to assess the chatbot's performance at the community level. A predefined set of queries (in some consultation/support with agency) will also be prepared by the provider and shared with citizens, who will be requested to ask these queries to the chatbot.</p> <p>D) Domain experts / research investigators from the provider side will evaluate the chatbot based on the following criteria:</p> <ul style="list-style-type: none"> (i) Correctness of responses (ii) Ability to correctly understand user intent (iii) Ability to understand and respond to the dialects specified in the RFP <p>E) The domain experts will assign a performance score to the chatbot, which will be shared with the developer.</p> <p>F) Kindly note that test dataset queries will not be shared with the developer in prior for any purposes.</p> <p>G) It is expected that the chatbot maintains 85% accuracy as mentioned in the RFP</p>
28	Section 5 – Scope of Work	Data usage & freshness	Please confirm the data refresh frequency for UPkSK APIs (real-time / daily / periodic), and that chatbot responses will be based on the latest available data, with no expectation of real-time transactional accuracy unless explicitly supported by source systems.	<p>A) Some databases are refreshed on monthly basis while some are refreshed on weekly basis.</p> <p>B) The change in data refresh frequency will be analysed by IHAT and can be modified as and when needed.</p>
29	Section 5 – Scope of Work	Nature of responses	Please confirm that the chatbot is intended to provide informational and decision-support responses only, based on UPkSK data and approved documents, and will not be expected to provide clinical diagnosis, medical advice, or treatment recommendations.	The chatbot is not intended to provide clinical diagnosis, medical advice or treatment recommendations. However, it is expected to guide the citizens to reach the facilities of their need by providing follow-up questions.
30	Section 5 – Scope of Work	Location-based queries	Please confirm whether "nearest facility" queries should be computed using existing UPkSK logic / GIS coordinates, and that no custom routing, navigation, or travel-time calculation logic is required as part of this scope.	Nearest facility queries should be answered using UPkSK provided GIS coordinates. There is no custom routing, navigation or travel-time calculation logic is required as part of this scope
31	Section 5.2 – Efficiency Evaluation Criteria	Language & dialect handling	Please confirm that multi-language and dialect support (e.g., Hindi, English, regional dialects) will be evaluated based on functional understanding and relevance, and that minor linguistic or accent variations will not be treated as functional defects during pilot evaluation.	The objective of this project is to make the chatbot understand the dialects well even if the accent varies. Therefore, the performance degradation due to minor linguistic and accent variations will be treated as functional defects during pilot evaluation.
32	Section 5 – Scope of Work / Pilot Phase	Model learning & refinement	Please confirm that feedback, validation, and refinement of chatbot responses during the pilot phase will be a joint activity between IHAT / GoUP and the Agency, and that the Agency will not be solely responsible for domain validation of responses.	<p>A) During the development phase, the agency may like to test the chatbot with certain users at the community to have feedback. This activity is the responsibility of the agency only.</p> <p>B) However, during the pilot testing with the 200-500 users, it will be a joint activity</p>
33	Objective / Contract Term	Contract duration of one year	Please confirm that upon completion of the one-year contract period, the developed chatbot, source code, configurations, and deliverables will remain usable by IHAT / GoUP, and any further support or enhancements would be subject to a separate agreement.	After completion of the contract, no other contract will be executed for any further scope
34	Section 13 – Technical Proposal / Security	Security assessment	Please confirm whether any third-party security audit, VAPT, or certification is required for the chatbot solution, or whether adherence to existing IHAT / GoUP security guidelines is sufficient.	Security Audit point added in addendum
35	Clause 5.1(C) (Scope of Work), Page 5; Clause 5.2 (Efficiency Evaluation Criteria), Page 6	Voice input support for Hindi dialects including Bundeli, Awadhi and Khadi Boli is mandatory, with a minimum accuracy benchmark of 85%, but dataset availability is not specified.	<p>Does IHAT possess existing voice datasets for Bundeli, Awadhi and Khadi Boli, or is the agency expected to collect primary data for model training and fine-tuning?</p> <p>Primary data collection would significantly impact project timelines, cost structure, and delivery risk compared to using existing validated datasets.</p>	IHAT do not have any voice dataset. The agency will collect the primary data or developing the model and the agency is expected to complete the project within the specified timeline as mentioned in RFP.

36	Clause 5.2(4) (Efficiency Evaluation Criteria), Page 6	The efficiency criteria specify an average response time of 20 seconds without defining latency scope.	Does the 20-second response time apply to full end-to-end processing (speech-to-text, processing, database query, text-to-speech) or only LLM inference time? End-to-end latency is a critical constraint for a voice-based system and impacts architecture and infrastructure sizing.	Yes, the 20-second response time applies to full end-to-end processing (speech-to-text, processing, database query, text-to-speech)
37	Clause 2 (Objective), Page 3; Clause 6 (Key Deliverable Timelines), Page 6	The contract duration is defined as one year, while deliverables are listed only up to T+180 days.	What is the expected scope of work for the remaining six months of the one-year contract after T+180 days? The agency needs clarity on whether ongoing support, maintenance, or enhancement activities must be budgeted for the full contract period.	Contract is for one year (12 months) while project is expected to be finished within 6 months. The extra time (6 months after expected project ends) is provided to cover any unexpected delays likely due to natural disasters. If the deliverables are achieved as per schedule, then the contract is deemed to be completed after the full and final payment.
38	Clause 6 (Deliverable 4), Page 6; Clause 16 (Payment Terms), Page 12	Deliverable 4 requires rework based on stakeholder feedback and is linked to a 30% payment milestone, without limits on iterations.	Can IHAT define a cap on the number of rework iterations or limit the scope of changes during the rework phase? Without defined limits, the rework phase carries a high risk of scope creep and payment delays.	There will no capping as such. Iterations will perform till the time of final expected results are not obtained.
39	Clause 18 (Performance Guarantee), Page 12	The RFP mandates deduction of 10% from each invoice as a performance guarantee until project completion.	Can the 10% performance guarantee be submitted in the form of a Bank Guarantee instead of invoice-level deductions? Invoice deductions significantly impact agency cash flow, while a Bank Guarantee is a standard and financially efficient alternative.	RFP clause remains unchanged
40	Technical proposal Pg No-10	Technical proposal	Please confirm whether the 20-slide technical presentation is required to be submitted as part of the hard-copy technical bid at the time of bid submission, or if it will be presented separately during the technical evaluation/presentation stage and submitted only in soft copy at that time.	Present separately
41	Indicative team structure Pg No-6	Indicative team structure	Please confirm that development activities will be carried out at the bidder's location, and that only reviews, demonstrations, and coordination meetings will require interaction with IHAT/GoUP.	A) Development activities can be carried out at the bidder's location B) Reviews, demonstrations, and coordination meetings will be conducted at Lucknow, UP
42	Scope of work Pg No-5	Scope of work	Please specify the expected data volume and historical depth (facilities, indicators, and time range) to be accessed from UPkSK during the pilot phase.	5 GB minimum having all facilities , RMNCHA+ indicators in Last one year to three year
43	Service level agreements (sla) and penalties Pg No-13	Service level agreements (sla) and penalties	Please clarify whether SLA penalties will be applicable only for delays solely attributable to the bidder, excluding delays arising due to non-availability of data, API access, approvals, feedback, or dependencies from IHAT/GoUP or third parties. Additionally, please confirm whether the maximum penalty cap of 10% is cumulative across the entire project and subject to mutual review and acceptance of milestone deliverables before invocation.	The penalties are applicable only for delays attributable to the bidder, which excludes the delay caused due to the failure in Authority's obligations. The maximum penalty cap of 10% is cumulative across the entire project and it is not subjected to mutual review
44	Clause 11, Page 8	Clause 11, CRITERIA FOR EVALUATION OF TECHNICAL PROPOSAL	It has been mentioned in the point 5 Supporting documents that Agency has to provide a Certificate from the department of Human Resource Department mentioning the name, nature of employment, years of association with the firm, and designation on the letter of the bidder. So this list should mention all resources working in the Organization or only those resources who would be deployed to work on this requirement?	The bidder shall give the list of 25 permanent staff, including Team leader, software developers, AI ML Developers, Server Admin, Database admins.
45	Generic Query	Objectives & KPIs	Will there be a labelled benchmark dataset for evaluation, or should we design and annotate this dataset as part of the scope?	There is no labelled dataset available for evaluation. Developers will design and annotate as required
46	Generic Query	Workload & Performance	Are there specific concurrency requirements (e.g., minimum number of concurrent WebSocket or HTTP/2 connections the system must support)?	To ensure the high performance delivery of the chatbot to the citizens, the selected bidder should design appropriate server infrastructure and protocols
47	Generic Query	Workload & Performance	Are there limits on maximum conversation length (tokens/turns) that impact context-window and memory design for the LLM backend?	To ensure the high performance delivery of the chatbot to the citizens, the selected bidder must identify appropriate conversational length and memory design.

48	Generic Query	Architecture & Hosting	Are there preferences for deployment style (Kubernetes, Docker Swarm, VMs only) and mandatory tools (e.g., Rancher, OpenShift, AKS/GKE/EKS- equivalent on gov cloud)?	This is upto selected bidder. However, docker/Kubernetes based deployment is preferred.
49	Generic Query	Architecture & Hosting	Do you require multi- AZ or multi-region deployment for HA/DR, and if yes, what RPO/RTO targets should we design for?	AZ should be within Indian region and multi-AZ is not needed
50	Generic Query	Architecture & Hosting	Should LLMs and vector indices be deployed in an isolated subnet/VPC separate from web frontends and APIs for security segmentation?	Yes, it is recommended
51	Generic Query	Architecture & Hosting	Are there existing ingress controllers / API gateways (e.g., NGINX, Kong, Istio) that we must integrate with instead of deploying our own?	This will be discussed with selected bidder at the time of SRS documentation.
52	Generic Query	Data & Integration	Are geospatial services (distance matrix, travel time) already available as APIs, or should we host our own geospatial engine (e.g., PostGIS, OSRM)?	A) UPkSK will provide only Latitude and Longitude of the health facilities B) For other geo locations, selected bidder need to use their own geo spatial engine.
53	Generic Query	Data & Integration	Will APIs provide consistent schemas and versioning, or do we need to build schema-adapter layers to handle heterogeneous/legacy interfaces?	APIs need to develop based on schema adapter layer.
54	Generic Query	Data & Integration	Are there quality constraints on incoming data (null handling, inconsistent codes, missing geo- coordinates) that we need to sanitize via a data-cleaning pipeline?	Mostly sanitized data will be provided by IHAT but in certain cases selected bidder need to use Data cleaning Pipeline.
55	Generic Query	Data & Integration	Do we need a dedicated feature store or data warehouse layer for analytics queries from the chatbot (e.g., district- wise trends)?	The AI model server can be used to store and analyse queries
56	Generic Query	NLP / LLM / ASR-TTS	Are there hard constraints on using external APIs (e.g., open cloud LLM endpoints) versus fully self-hosted models due to data residency/security?	The data is sensitive and it belongs to GoUP. Therefore, the external APIs can be used only when it doesn't leak any data outside.
57	Generic Query	NLP / LLM / ASR-TTS	Do you want a pure RAG architecture over UPkSK data, or a hybrid architecture with: intent classifier, slot- filling NLU, rule-based engine, and LLM as NLG layer?	The bidder must suggest the suitable architecture design after careful analysis of the requirements. It will be mutually agreed up at the time of contract signing.
58	Generic Query	NLP / LLM / ASR-TTS	What is the maximum allowed context length (in tokens) per query for the LLM?	The maximum token length should be identified by collecting a few sample queries from selective citizens.
59	Generic Query	NLP / LLM / ASR-TTS	Do you require language/dialect auto-detection at runtime, or will the user/channel specify language explicitly?	Language/dialect to be auto-detected at runtime by the chatbot.
60	Generic Query	Security & Compliance	What is the retention period for raw vs anonymized data?	All raw and anonymized data will be stored during the development phase. During deployment, it all will be transferred to GoUP server. It might be used by GoUP when the chatbot is scaled.
61	Generic Query	Security & Compliance	Do you require separate key management (KMS/HSM) for encryption keys, and should keys be managed by the state/NIC rather than the vendor?	The selected bidder is responsible for handling the keys for encryption and decryption
62	Generic Query	Security & Compliance	Is SSO required for internal users (planners/admins), and if yes, Is there any preferred identity provider (ADFS, LDAP, OAuth2, SAML) should we integrate with?	All user credentials will be created within the application only. No external integration is required for internal users.
63	Generic Query	Monitoring & DevOps	Which observability stack is preferred or already mandated (e.g., Prometheus + Grafana, ELK/EFK, Open Telemetry) for metrics, logs, and traces?	A) Selected Bidder should consider appropriate server monitoring tool while designing the server infrastructure in consent with IHAT B) Alarms should be intimated to IHAT/GoUP officials if in case any criticality found over servers like high usage of CPU, etc.
64	Generic Query	Monitoring & DevOps	Do you require automated alerting with on-call rotation (SMS/Email/ChatOps), and what escalation paths should be configured?	Not required
65	Generic Query	Monitoring & DevOps	Are there existing CI/CD tools and workflows (e.g., Jenkins, GitLab CI, Azure DevOps) we should plug into, or can we propose a new pipeline?	A) Code repository of chatbot will be handled by IHAT GitHub account. B) Will be finalised at the time of SRS preparation.
66	Generic Query	Monitoring & DevOps	Do you require blue-green / canary deployments for backend services and models to minimize downtime during rollouts?	Will be finalised at the time of SRS preparation.
67	Generic Query	UX & Channel Integration	How long should conversation state be preserved for returning users?	For a PoC, the focus is on testing feasibility and functionality. Therefore, not on long-term retention or personalization is required
68	Generic Query	Healthcare Logic & Safety	Are there clinical safety rules that must be enforced at the orchestration layer (e.g., certain red-flag symptoms must always trigger an emergency referral message)?	not required

69	Generic Query	Healthcare Logic & Safety	Should medical content be validated against a specific knowledge base (e.g., state SOPs, national guidelines), and can we get machine-readable versions?	not required
70	Generic Query	Healthcare Logic & Safety	Do you require a rules engine for eligibility/service routing (e.g., age, gender, service type) separate from the LLM layer for determinism and auditability?	Any citizen who has smart phones can use the application
71	Generic Query	Healthcare Logic & Safety	How should the system behave when model confidence is low (thresholds for escalating to predefined messages, human review, or alternate flows)?	The bidder must identify the methodology to improve the system when the confidence is low. The final development model should achieve the accuracy mentioned in the RFP
72	Generic Query	Operations & Handover	What are the expected SLAs for incident response and resolution (severity-wise), and should they be codified in a runbook?	A) A detailed runbook shall be maintained, covering incident classification, resolution steps, escalation procedures, monitoring, and communication protocols. B) Incident type (critical, moderate and lower issues) along with turn-around time to resolve the issues are added to addendum
73	Generic Query	Operations & Handover	What level of access should be granted to your internal tech team for future operation (full cluster access, read-only monitoring, ability to redeploy models)?	Full access to all systems and information
74	Generic Query	Operations & Handover	What artefacts must be handed over at project closure (source code, IaC scripts, docker images, model weights, annotated datasets, runbooks)?	Everything inclusive of source code, scripts, API codes, etc. generated during this PoC must be handed over.
75	Clause 3, Page 3	Clause 3. About UPkSK	Will we get Location Coordinates of users also from GIS systems ?	Chatbot, before initiating the conversation, will ask for permission to capture the user's current location.
76	Clause 3, Page 3	Clause 3. About UPkSK	Which all UPSK indicators will be available for the Chatbot App?	All indicators available in UPkSK will be available for the Chatbot app
77	Clause 3, Page 4	Clause 3. About UPkSK	Is the supply chain of drugs and medicines also reported in HMIS portal.	All these information are already available in UPKSK.
78	Clause 3, Page 4	Clause 3. About UPkSK	What are the key information that can be extracted from HMIS.	No need of contacting HMIS since all information are available in UPKSK
79	Clause 3, Page 4	Clause 3. About UPkSK	At what frequency does Nodal officer updates data at HMIS portal.	No need of contacting HMIS since all information are available in UPKSK
80	Clause 3, Page 4	Clause 3. About UPkSK	Will the APIs of HMIS be provided to us ?	No need of contacting HMIS since all information are available in UPKSK
81	Generic Query		Do we need to give Area / District wise recommendation using NIKSHAY App.	No application other than UPkSK will be integrated with the chatbot.
82	Clause 5.1, Page 7	Clause 5.1	What languages are to be supported? RFP mentions English, Hindi and Hinglish. Are there any other expected languages in the scope of this RFP?	No change
83		General Query	What is the total size of the dataset which needs to be trained (in MBs, GBs, etc)?	5 GB approximately.
84		General Query	How many APIs need to be integrated, if any?	A) UPkSK has Facility, Program, HR, Equipment, and Drug data. The APIs will be designed within this periphery. B) The details will be discussed upon signing the contract.
85		General Query	Are the functions going to be read-only or read-write?	Will be finalised at the time of SRS preparation.
86		General Query	Total number of scheduled languages required to be supported?	Text: 2 Scheduled languages: Hindi and English and 1 other language: Hinglish needs to be supported. Speech: English as well as Hindi (Dialects: Awadhi, Khadi Boli and Bundeli) needs to be supported.
87		General Query	Is there a provision for extension of the tender after completion of the initial one-year term?	No, there will be no extension beyond the 1 year as mentioned in RFP.
88		General Query	What is the approximate budget assigned for this RFP project.	Confidential Information
89	Section 5.1 (A)	Conversational Q&A agent that uses information from UPkSK to identify the closest/appropriate facilities	Please clarify is what format, and from where the information of UPkSK will be received? 1. Is the information stored in Database? Then what kind of Database it is? (MySQL, MongoDB, or some other DB) 2. Is the information fetched via API? Please confirm if the format is JSON, or SOAP or XML.	1. MS-SQL and MySQL 2. Restful API

90	Section 5.2 – Point 4	Average response time to queries – Expected to answer in 20 secs	Please specify the minimum data speed during which we should determine the response time, since response time will vary depending on the data connectivity speed	A) At the community level, low data connectivity speeds are to be expected. B) Some LLM models are specifically designed to handle low-bandwidth environments. C) We request that such models be identified and utilized so that users can receive responses within 20 seconds. D) Delayed responses are considered a failure of the system and negatively impact users' satisfaction and trust
91	5. SCOPE OF WORK 8. ROLES AND RESPONSIBILITY OF IHAT / Page #7	Support Necessary approvals.	1.What specific approvals are envisaged (data access, security clearance, hosting, pilot approval, etc.)? 2.Will IHAT/UPTSU be the single point of coordination for all such approvals?	IHAT shall obtain all necessary approvals.
92	10. GENERAL QUALIFYING CRITERIA / Page#8	2. Minimum Annual turnover of INR 2 Cr per year, from providing IT services, from India operations in the last 03 (three) completed Financial Years ending 31st March 2025. A copy of CA certificate confirming the same should be enclosed	1. Considering the state level implementation and importance of the project, the required Minimum Annual Turnover of INR 2 Cr per year appears to be significantly low. Therefore, we request you to increase the minimum annual turnover requirement to ₹50 Crores to appropriately define the qualifying criteria for bidders. We kindly request you to consider this	RFP clause remains unchanged
93	11. CRITERIA FOR EVALUATION OF TECHNICAL PROPOSAL / Page #8	Table 1: The scoring criteria to be used for Technical Bid evaluation shall be as follows 4. Average annual turnover- (Maximum Marks - 10) Average Annual Turnover from providing IT services, from Indian operations, for the preceding 03 (three) Financial Years (i.e., 2022-23, 2023-24 and 2024-25): INR 2 to 10 crores- 5 mark INR more than 10 Crore-10 Marks	1. We request you to amend this scoring criteria of Technical Bid evaluation as outlined below, in alignment with the suggested revisions to the Bidder's General Qualifying Criteria: <i>"Average Annual Turnover from providing IT services, from Indian operations, for the preceding 03 (three) Financial Years (i.e., 2022-23, 2023-24 and 2024-25): "More than INR 50 Crores and up to INR 75 Crores – 5 Marks More than INR 75 Crores and up to INR 100 Crores – 7 Marks More than INR 100 Crores – 10 Marks"</i> Please consider	RFP clause remains unchanged
94	19. SERVICE LEVEL AGREEMENTS (SLA) AND PENALTIES / Page #13	Timelines and Penalties, whenever applicable, to be levied by the Nodal/Payment Authority as follows and to be adjusted against the fee payable.	1. The timeline specified for "Initial Model Development" appears to be inadequate. Therefore, we request you to kindly consider extending the timeline to T + 90 days to ensure proper model design, training, validation, and quality outcomes.	No change
95	13. TECHNICAL PROPOSAL / Page #10	Documents for evaluation of each criterion mentioned under Section 11 (technical Evaluation), including a brief technical proposal, not more than 20 slides (proposal which exceed the word/slide limit may NOT BE EVALUATED) , comprising a clear vision, understanding of Background & Context, Potential Solution & Rationale, Technical Approach, Methodology, Projected Impact along with Work Plan, Time lines, and HR personnel required.	1. Please confirm whether the Technical Proposal is required to be submitted in PDF format converted from a Word document or from a PowerPoint (PPT) file.	The presentation PPT converted to PDF shall be submitted in hard copy as a part of the Technical proposal. The soft copy (.PPT) shall be asked at the time of presentation

96	16. PAYMENT TERMS / Page #12	<p>1. 10 % of the total project fee On Signing of contract and Kick-off</p> <p>2. 15% of the total project fee On completion of initial model</p> <p>3. 20 % of the total project fee On successful completion of Model improvement and production ready model tested with limited citizens</p> <p>4. 30% of the total project fee On completion of demonstration to stakeholders and reworked on the model to accomplish the suggestions requested by stakeholders</p> <p>5. 25% of the total project fee On completion of the deployment and maintenance support</p>	<p>The payment terms are very stringent and unfavourable to service provider as almost all exposure of servicer provider is done upto Go-live and handover of the solution (i.e. 180 days). Therefore we request you to amend the payment terms as per below to maintain the cashflow of servicer provider.</p> <p><i>"1. 10 % of the total project fee On Signing of contract and Kick-off</i></p> <p><i>2. 30% of the total project fee On completion of initial model</i></p> <p><i>3. 30 % of the total project fee On successful completion of Model improvement and production ready model tested with limited citizens</i></p> <p><i>4. 15% of the total project fee On completion of demonstration to stakeholders and reworked on the model to accomplish the suggestions requested by stakeholders</i></p> <p><i>5. 10% of the total project fee On completion of the deployment</i></p> <p><i>6. 5% of the total project fee On completion of maintenance support"</i></p> <p>Please consider.</p>	No Change, This is a standard payment matrix
97	Training	GENERAL	<p>1. Is there any training required to internal users ? If yes, please provide details :</p> <p>1.1 No. of training session required</p> <p>1.2 Whether training will be online/Offsite.</p> <p>1.3 User type wise number of users for training</p>	<p>A) Yes, training will be required for the research investigators who will be testing the chatbot in the field.</p> <p>B) Approximately 3-4 batches of training may be required.</p> <p>C) The training may be conducted online or offline, depending on the selection and availability of the research investigators.</p> <p>D) The final number of research investigators will be decided at a later stage.</p>
98	General	Bid Submission time	1. We request you to provide at-least 12 working days time to submit the bid after the replies to bidders queries are published.	No change, bid submission date will remain same.
99	Section 6	Deliverable timelines	Are the timelines calendar days or working days from date of contract signing?	Calendar Days
100	Section 7	Indicative Team Structure	Is the team structure mandatory or indicative? Can roles be combined if justified by experience?	roles be combined if justified by experience
101	Section 10 (Eligibility)	Similar Work definition	Will AI chatbot projects outside the health sector but similar in complexity be considered as "Similar Works"?	Yes
102	Section 11 – Technical Evaluation	Certification marks	If the bidder holds equivalent international certifications, will they be considered for scoring?	No
103	Section 12	Pre-Bid Meeting	Will the responses to pre-bid queries be shared publicly with all bidders via corrigendum/email?	Both
104	Section 14	Financial Proposal	Should the quoted project fee include cloud hosting, STT/TTS APIs, and LLM inference costs, or will these be provided by IHAT?	YES
105	Section 14	Taxes	Please confirm whether TDS will be deducted from payments and applicable percentage.	Yes TDS will deduct @ 10%
106	Section 16	Payment Terms	Is milestone payment release subject to formal sign-off/acceptance certificate from IHAT?	After the signing of contract from both the party - either on soft copy or on hard copy
107	Section 18	Performance Guarantee	Will the 10% performance guarantee be deducted invoice-wise or can a Bank Guarantee be submitted instead?	We will deduct 10% as performance guarantee on each invoice
108	Section 19	SLA & Penalties	Is there a grace period before penalties are applied for milestone delays?	No
109	Section 4.2	Appendix C – Team Deployment,	Kindly clarify how many team members are deployed and also kindly provide us the Appendix C – Team Deployment format (Format is Missing)	Appendix C shall be provided by the Onboarded Agency during signing of the contract
110	Section 5.1 (B)	Appendix E – Duties of the Client and the Agency	Kindly provide the Format. (Format is Missing)	Appendix E shall be provided by the Onboarded Agency during signing of the contract

111	Section 5.1 (E)	Application deployed to limited number of users for pilot testing	Kindly specify the expected number of pilot users, pilot duration, and geographical coverage (district/block-wise). This is required for sizing infrastructure and performance testing.	<p>A) The chatbot will be tested with approximately 200-500 users. Users who speak Bundeli, Awadhi, and Khadi Boli will be selected for testing.</p> <p>B) The districts/blocks where Bundeli, Awadhi, and Khadi Boli are spoken will be randomly selected for the testing phase.</p> <p>C) The testing phase is expected to be completed within 30 working days. Any issues, gaps, or limitations identified during testing will be addressed by the developer.</p>
112	Section 5.2 – Point 1	% of citizens' queries answered compared to those identified in scope	Please clarify how "queries identified in scope" will be defined and documented. Will a predefined query taxonomy be shared, or is the agency expected to derive it?	Agency is expected to derive the queries sets
113	Roles & Responsibility of Agency	Establish technical integration with UPkSK	Kindly confirm whether UPkSK integration support, sandbox environment, API documentation, and test credentials will be provided before development commencement. This is essential for timely delivery.	YES
114		Assuming that the solution will be primary targeted for the citizens. Please list out the other user roles who will be accessing the chatbot system.		This chatbot is expected to answer citizens queries. Therefore any user including the government officials can access the chatbot during the testing phase. However, the chatbot is not expected to answer administrative related queries to the administrators.
115		Any LLM that is going to be used in the solution, should it be used as is or training of the LLM needs to be considered.		In order to achieve high accuracy, it is a general practice to enhance the baseline LLM models with new data.
116		While implementing the solution, is there any restrictions on using the Hugging Face or the cloud services as part of the solution.		If you are planning to reuse the open-source LLM models under the hugging face such as BERT, then there will be no issue. It must be used within the cloud dedicated for this project. Make sure that the data is not forwarded to the hugging face servers for any purposes.
117		Is there any unstructured data like documents etc. being stored in UPkSK from which the information to be extracted for generating the response.		No