AI-Powered Smart Survey Tool for Survey Data Collection

1. **Track:**

Data Collection

2. Description:

The National Sample Survey under MoSPI collects comprehensive socio-economic data from households and enterprises. Government data collection continues to rely heavily on conventional survey tools that enumerator driven. They are also difficulty in conducting surveys due to the India's language diversity. With the increased penetration of smartphones, AI, and messaging platforms like WhatsApp, there's a strong opportunity to redesign how official surveys are built and deployed. The problem is to develop a smart, AI-enabled tool that can make data collection more efficient, inclusive, and intelligent serving the evolving needs of national statistical systems.

3. **Expected Outcomes/Solutions**

Participants should aim to deliver the following:

- A working prototype of an AI-powered smart survey tool (web, mobile, or hybrid).
- No-code/prompt-based survey builder, with question bank integration.
- Support for multi-channel survey delivery (e.g., WhatsApp, phone call, AI avatar).
- Prepopulation using identifiers; multilingual and voice-based interfaces.
- AI-driven customised questioning and real-time validation.
- Paradata capture (time, location, device, mode) and quality flagging.
- Auto-coding of responses and structured data storage.
- Dashboard for monitoring enumerator performance and data quality.
- Code, documentation, and brief implementation summary or pitch.

4. **Relevance to National Priorities or Ongoing MoSPI Initiatives**

This challenge aligns with MoSPI's digital transformation agenda under its Data Innovation Lab. MoSPI is actively working to modernize data collection systems through automation, integration with new technologies, and adoption of AI. The initiative is directly relevant to upcoming large-scale surveys and the Economic Census. Creating AI-based survey tools will contribute to the vision of building a national Statistics-as-a-Service

(StaaS) architecture and supporting India's Digital Public Infrastructure (DPI) for data.

5. Background Resources or Datasets (if available)

Participants may access these resources which are available on the web:

- Sample survey questionnaires from past NSS rounds.
- LGD codes for location mapping and geo-coding.
- Open datasets for prepopulation logic (e.g., Census, SECC, Socio-economic indicators).
- MoSPI metadata standards and classification codes (e.g., NCO, NIC, ISIC).

6. Key Features Required

- Survey Creation
    - No-code interface for users to build surveys.
    - Option to generate survey forms using natural language prompts.
    - Ability to add questions from a standardized question bank or library.
    - Support for conditional logic, loops, and skip patterns.
- Data Prepopulation
    - Use of unique identifiers (e.g., household ID, Aadhaar, phone number) to pre-fill known data fields.
- AI-Driven Adaptive Questioning
    - Automatically infer respondent traits and personalize follow-up questions.
    - Use of lightweight LLMs or rules-based classifiers for dynamic routing.
- Multilingual and Multi-Modal Interface
    - Language selection at survey start with automated translation.
    - Support for WhatsApp, IVR/phone-based surveys, and AI avatar chat interfaces.
- Real-Time Validation and Auto-Coding
    - Built-in rules and AI to detect inconsistent responses.
    - Auto-coding of text/categorical responses using standard classifications.
    - Storage in structured databases (SQL/NoSQL).

- Monitoring and Quality Dashboard
    - Dashboards for supervisors to monitor progress, quality, and performance.
- Data Privacy, Security, and Consent
    - Consent Management: The tool must include a clear, auditable mechanism for capturing and managing respondent consent at the beginning of the survey.

7. Paradata & Quality Assurance

The tool must collect and utilize paradata for ensuring high-quality data collection, including:

- Time taken per question/module
- Start/end timestamps and pauses
- GPS coordinates and survey location
- Device type and mode of interview
- Patterns in enumerator behavior
- AI-based flags for poor quality or suspicious entries

    - Data Security: All data, both in transit and at rest, must be encrypted. The solution should demonstrate secure data handling practices.
    - Privacy by Design: The system should be built on principles of data minimization and purpose limitation.

8. Bonus Features / Future Scope

Teams are encouraged to demonstrate innovation beyond the core requirements. Bonus consideration will be given for features like:

- An integrated chatbot for self-service enumeration
- OCR for document scanning
- Lightweight real-time analytics and visualizations
- Biometric or facial recognition-based respondent verification
- AI-based summarization or classification of open-ended responses
- Offline Design

9. Impact Potential

This tool will serve as a foundational element for AI-driven data collection platforms across multiple sectors, including health, education, labour,

agriculture, and governance. It can transform how official data is collected, validated, and utilized in real time, while also making the process inclusive and respondent-friendly. It also contributes to India's vision of building scalable, trusted, and intelligent Digital Public Infrastructure for data.