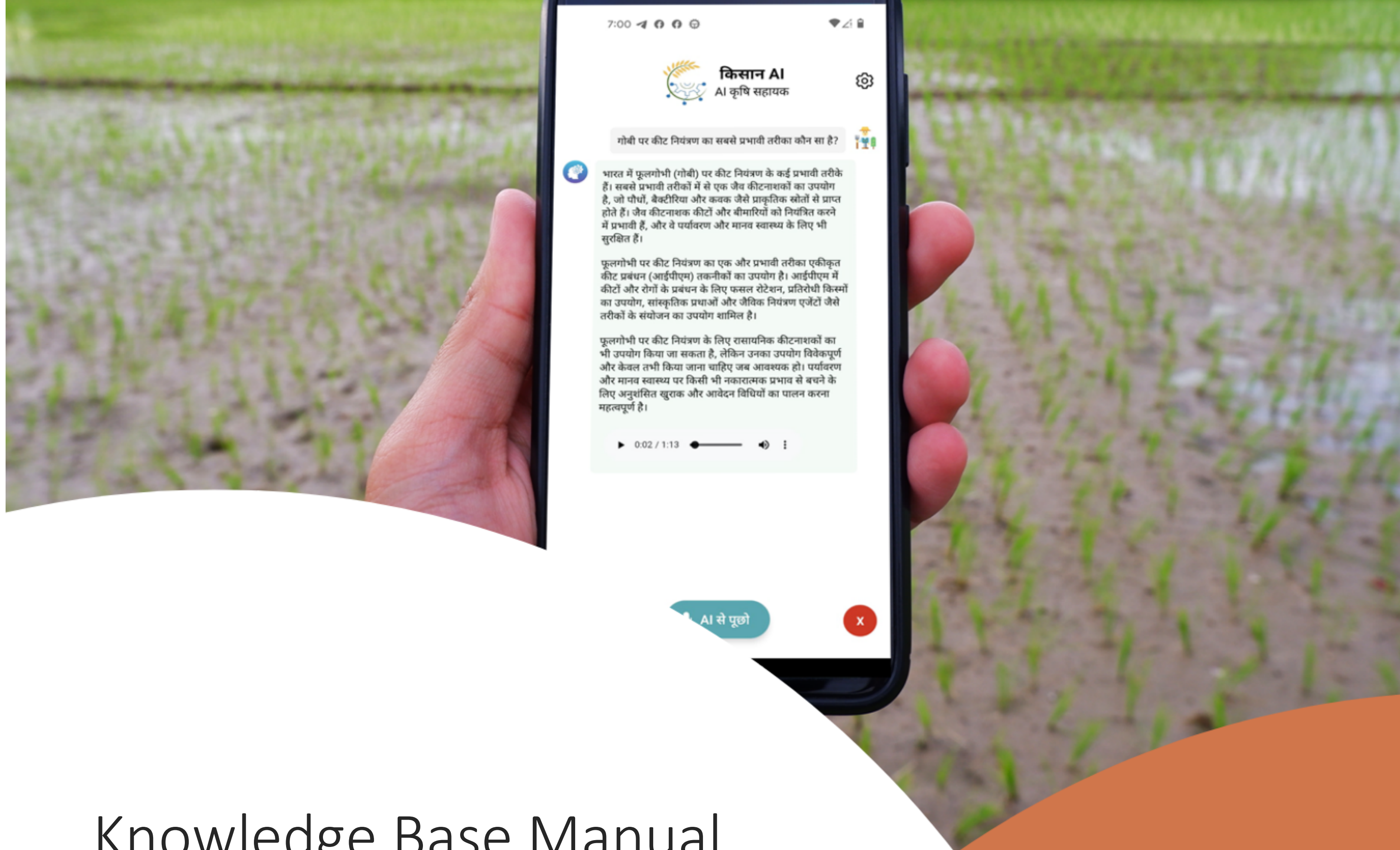

गोबी पर कीट नियंत्रण का सबसे प्रभावी तरीका कौन सा है?

भारत में फूलगोभी (गोबी) पर कीट नियंत्रण के कई प्रभावी तरीके हैं। सबसे प्रभावी तरीकों में से एक जैव कीटनाशकों का उपयोग है, जो पौधों, बैक्टीरिया और कवक जैसे प्राकृतिक स्रोतों से प्राप्त होते हैं। जैव कीटनाशक कीटों और बीमारियों को नियंत्रित करने में प्रभावी हैं, और वे पर्यावरण और मानव स्वास्थ्य के लिए भी सुरक्षित हैं।

फूलगोभी पर कीट नियंत्रण का एक और प्रभावी तरीका एकीकृत कीट प्रबंधन (आईपीएम) तकनीकों का उपयोग है। आईपीएम में कीटों और रोगों के प्रबंधन के लिए फसल रोटेशन, प्रतिरोधी किस्मों का उपयोग, सांस्कृतिक प्रथाओं और जैविक नियंत्रण एजेंटों जैसे तरीकों के संयोजन का उपयोग शामिल है।

फूलगोभी पर कीट नियंत्रण के लिए रासायनिक कीटनाशकों का भी उपयोग किया जा सकता है, लेकिन उनका उपयोग विवेकपूर्ण और केवल तभी किया जाना चाहिए जब आवश्यक हो। पर्यावरण और मानव स्वास्थ्य पर किसी भी नकारात्मक प्रभाव से बचने के लिए अनुशंसित खुराक और आवेदन विधियों का पालन करना महत्वपूर्ण है।

# Knowledge Base Manual

## AI-Powered Climate Advisory Bot for Multilingual Farmers in India

Titodi

KissanAI

UNDP

# 1. Introduction

Welcome to the Knowledge Base Curation Manual, a pivotal resource developed by KissanAIi's Data Curation Team for curating a rich repository of data focused on advancing sustainable agricultural practices. This manual serves as an index and guide to a well-structured compendium of resources tailored to support climate-smart agriculture initiatives, particularly within the diverse agroecological contexts of Jharkhand and Madhya Pradesh in India.

## 1.1 Purpose of the Manual

This document delineates the systematic process of compiling, standardizing, and verifying a vast array of data. It guides readers through the elements of curated CSV files, the sourcing of data, the content focal areas, and the comprehensive data curation methodology employed by the project.

## 1.2 Scope of the Knowledge Base

The knowledge base is an accessible, dynamic platform that brings together data on crop diversification, soil health, water conservation, agroforestry, and livelihood enhancements. Structured to facilitate easy navigation, it offers agricultural professionals, researchers, and policymakers a pathway to a wealth of information instrumental in confronting the challenges of climate change and enhancing rural livelihoods.

## 1.3 Intended Users of the Manual

Aimed at empowering stakeholders with vested interests in sustainable agriculture, this manual caters to a broad audience that spans researchers, policymakers, and agricultural practitioners. It serves as a navigational aid in leveraging curated datasets and resources to promote practices that bolster climate resilience and biodiversity, with an eye towards sustainable development in rural areas.

## 1.4 Collaborative Curation and Community Involvement

With the knowledge base's presence on GitHub, it opens up to continuous and collaborative updates. This document invites and guides domain experts globally to contribute, expand, and refine the content, adhering to open-source principles that encourage community participation and collective intelligence. Contributions are meticulously reviewed to maintain the highest standards of data integrity and relevance. This collaborative approach not only enriches the knowledge base but also fosters a sense of shared purpose in the global quest for sustainable agriculture solutions.

# 2. Strategy

The strategy underpinning the curation of the Knowledge Base is predicated on meticulous selection, processing, and refinement of data. This ensures the provision of accurate, up-to-date, and actionable content for our Retrieval Augmentation Generation (RAG) based system, supporting the delivery of precise agricultural guidance.

## 2.1 Source Selection

The Knowledge Base is populated with data from two principal types of sources:

1. **Documentary Sources**: These include scientific research papers, reports, and guidelines from esteemed agricultural research institutions, universities, and organizations. The emphasis is placed on the reputation and authority of these entities to ensure the reliability of the data.
2. **Multimedia Sources**: Reputable agricultural YouTube channels have been leveraged to diversify the content of the Knowledge Base. These sources provide practical insights and visual demonstrations of sustainable agricultural practices.

## 2.2 Identification and Manual Curation

A dedicated team is tasked with:

- Sifting through the selected documents and multimedia content to identify relevant Climate Resilient Agriculture (CRA) data.
- Performing manual curation and labeling of the data, which involves extracting and categorizing information in a manner that aligns with the requirements of the RAG system. This includes creating descriptive labels that facilitate accurate retrieval and generation of responses by the chatbot.

## 2.3 Automation and Pipeline Development

To enhance efficiency and consistency, the strategy involves the use of:

- **Python-Based Automation**: A pipeline of Python scripts has been developed to automate various aspects of the curation process. This includes:
  - Preprocessing and formatting of textual content.
  - Extraction of relevant information from multimedia sources.
  - Initial labeling based on predefined categories and criteria.

- **Continual Refinement**: The pipeline is subject to ongoing review and refinement to improve its accuracy and to adapt to new data sources or changes in agricultural practices and technologies.

## 2.4 Ensuring Relevance and Usability

The overarching aim of the curation strategy is to ensure that:

- The Knowledge Base remains a current and authoritative resource for CRA information.
- The data is structured in a way that is readily usable by the RAG system, enabling the chatbot to provide contextually appropriate responses to user queries.

## 2.5 Collaborative Enhancement

In alignment with the principles of open-source development and collaborative work:

- The Knowledge Base is designed to support contributions from external experts and stakeholders via GitHub.
- An established review process is in place to oversee contributions and maintain the Knowledge Base's integrity and relevance.

## 3. Sources and Content

This section outlines the diverse and reputable sources that contribute to the Knowledge Base (KB), ensuring a rich and varied collection of data that underpins the information provided to end-users. The KB's content has been meticulously selected to include an extensive range of data types that are pertinent to climate-resilient agricultural practices.

### 3.1 Sources from Organizations

| Organization |
|---|
| Birsa Agricultural University, Kanke, Ranchi, Jharkhand |
| Indian Institute of Pulses Research, Kanpur |
| Central Research Institutes for Dry Land Agriculture (CRIDA), Hyderabad |
| United Nations Development Project |
| Indian Council of Agricultural Research |
| Watershed Organisation Trust (WOTR) |
| Mongabay India |

The KB comprises authoritative information from well-established agricultural organizations and research institutes. Each source has been chosen for its unique contribution to the field of sustainable agriculture:

- **Birsa Agricultural University, Ranchi, Jharkhand**: The official database offers PDF documents detailing crop varieties, soil suitability, and cultivation recommendations specific to Jharkhand, alongside methods to enhance soil quality.
- **Indian Council of Agricultural Research (ICAR) and CRIDA, Hyderabad**: Data from these institutions provides in-depth insights into dryland farming, including crop characteristics, optimal planting times, and localized advice on fertilizers, pest management, and soil and water conservation strategies.
- **Mongabay India**: This platform offers strategies to adapt farming practices to climate change, assisting farmers in navigating environmental challenges.
- **Watershed Organisation Trust (WOTR)**: The focus here is on climate-resilient agriculture in Madhya Pradesh, providing research on managing farming activities amid seasonal variations.
- **United Nations (UN)**: The UN’s contribution centers on Climate Change Adaptation within the Agricultural Value Chain, with methodologies, case studies, and synopses that reflect the adaptive strategies in farming practices.

### 3.2 Sources from YouTube Channels

| Video Channel | Total Videos | CRA Videos |
|---|---|---|
| Digital Greens | 7540 | 70 |
| Discover Agriculture | 2235 | 1 |
| Agriculture India | 1795 | 1 |
| SPK (Subhash Palekar Natural Farming) | 707 | 43 |
| Down to Earth | 1530 | 6 |
| Krishi Jagaran | 149 | 27 |
| Pathfinder by Unacademy | 5 | 5 |
| Madhya Pradesh Deendayal Antyodaya Yojana State Rural livelihoods Mission | 79 | 4 |

The team meticulously reviewed all available videos from the chosen YouTube channels, selecting only those that focus on CRA practices. A key consideration in this selection was the video length; priority was given to videos shorter than 20 minutes to ensure the content is easily digestible for end-users.

Video content has been curated from YouTube, with each channel offering a unique perspective on agricultural practices:

- **Digital Green**: Shares the use of technology and data in agriculture, aiming to improve rural agricultural practices and livelihoods.
- **Discover Agriculture**: Provides education on modern farming technologies and sustainable practices.
- **Agriculture India (@AgriGoI)**: Focuses on the evolution of agriculture in India, highlighting natural farming techniques and agricultural policies.
- **SPK (Subhash Palekar Natural Farming)**: Promotes natural farming practices through workshops and success stories.
- **Down to Earth**: Delivers news and opinions on environmental issues, with a focus on climate change, biodiversity, and renewable energy.
- **Krishi Jagaran**: Offers news and educational content on agricultural innovation and rural development in India.
- **Pathfinder by Unacademy**: While not exclusively agricultural, it occasionally includes agricultural topics in its broader educational content.

### 3.3 Types of Data Collected

The KB contains a comprehensive array of data types, including but not limited to:

- **Agricultural Best Practices**: Techniques and strategies for crop diversification, soil health management, and water conservation.
- **Climate-Resilient Practices**: Adaptive methods for farming in response to environmental and climatic shifts.
- **Innovative Farming Techniques**: Modern approaches to sustainable agriculture, leveraging technology and data-driven insights.

- **Case Studies and Research Synopses**: Detailed analyses and summaries of projects and their outcomes within the realm of sustainable farming.

# 4. Curation Process

The curation process is meticulously crafted to transform raw data into a reliable and insightful Knowledge Base, serving as the bedrock for the chatbot's intelligent responses. The following subsections elaborate on the key stages:

## 4.1 Data Collection and Identification

- **CRA Data Identification**: This foundational step involves sifting through the sources to pinpoint relevant Climate Resilient Agriculture (CRA) data.
- **Manual Curation and Labeling**: Subject matter experts manually curate and label the identified data, ensuring its relevance and applicability to the RAG-based system.

## 4.2 Automation and Pipeline Code

- **Python Automation Scripts**: Routine tasks are automated through Python scripts, promoting efficiency in data scraping, preliminary cleaning, and sorting.
- **Pipeline Evolution**: Regular updates to the automation scripts ensure they remain effective as data types and standards evolve.

## 4.3 Conversion to Markdown and CSV

- **Simplification to Markdown**: Data from HTML and PDF sources is converted to Markdown to facilitate easier handling and processing.
- **Structuring into CSV**: After simplification, data is structured into CSV files, enabling seamless integration and analysis within the Knowledge Base.

## 4.4 Data Cleaning and Standardization

- **Uniformity and Precision**: Through the curation pipeline, raw data is cleaned and standardized to ensure consistency across the Knowledge Base.
- **Refined Datasets**: The end product of this stage is a collection of CSV datasets that are thoroughly standardized and ready for use.

## 4.5 Video Curation Pipeline

- **Video to Audio**: Videos are processed to extract audio tracks, which are then transcribed to capture the spoken content accurately.
- **Transcription and Translation**: The transcribed audio is translated as needed, ensuring that language barriers do not impede the curation process.
- **LLM Fact Extraction**: Leveraging Language Learning Models, we transform transcriptions into structured facts, closely adhering to the ground truth presented in the videos.

- **Thematic Extraction and Labeling**: In parallel, themes and labels are extracted to provide a categorization framework for the video content.
- **Content Integration**: These facts and themes are meticulously incorporated into the Knowledge Base, enriching it with verified and actionable content derived from multimedia sources.

## 4.6 Elements of Curated Data

This crucial stage involves detailing the data attributes to be included in the Knowledge Base:

- **Organization**: Source entity information.
- **Label**: Categorization details, including subtitles and tags.
- **Title**: Concise description or headline of the data point.
- **Description**: In-depth details about the data.
- **Source**: Citation or origin of the data, including specific document names or URLs.
- **Additional Metadata**: Any supplementary details that enhance the data's context or utility.

## 4.7 Manual Cross-Verification

- **Expert Review**: All curated datasets undergo an additional layer of scrutiny by our in-house team to guarantee their accuracy and quality.
- **Validation and Reliability**: This critical phase ensures the data's integrity before its integration into the Knowledge Base.

## 4.8 Finalization

- **Inclusion Criteria**: Completion of the manual and automated curation stages qualifies the data for inclusion in the Knowledge Base.
- **Ongoing Updates**: The Knowledge Base is not static; it is continually updated to reflect new insights and maintain its relevance and accuracy.

# 5. Collaborative Curation and Open Contribution

A core tenet of our Knowledge Base (KB) is its collaborative and open-source nature, which will be prominently reflected throughout this manual. The KB is hosted on GitHub, a platform that enables continuous updates and collaborative contributions from a global community of experts.

## 5.1 Embracing Open-Source Principles

The manual will underscore the importance of open-source principles, advocating transparency, collaboration, and shared stewardship of the knowledge that our KB contains. By making the KB available on GitHub, we invite contributions that can expand, refine, and validate the agricultural knowledge crucial for farmers.

## 5.2 Continuous Evolution and Enrichment

We will detail the process by which updates and contributions are made to the KB, highlighting the review mechanisms that ensure the accuracy and reliability of the information. This process involves:

- **Community Contributions**: Encouraging experts from around the world to contribute to the KB, thereby leveraging collective wisdom and expertise.
- **Expert Review**: Implementing a rigorous review process by domain experts to ensure that every contribution meets our high standards for data integrity and relevance.
- **Version Control**: Utilizing GitHub's version control features to track changes, manage contributions, and maintain a historical record of updates.

## 5.3 Guidelines for Contributors

To foster a collaborative and inclusive environment for the development of the Knowledge Base (KB), we provide a structured pathway for contributions. Adhering to these guidelines will ensure that the KB evolves into a comprehensive and authoritative resource.

### Submitting Contributions via GitHub Pull Requests

- **Fork the Repository**: Start by forking the KB repository to your GitHub account.
- **Create a New Branch**: Make a new branch in your forked repository for each distinct change or addition you propose.
- **Make Your Changes**: Add or update the content in your branch, adhering to the provided content standards and formats.
- **Pull Request**: Once you're ready to submit, create a pull request (PR) against the main branch of the KB repository. Clearly describe the changes you are proposing in the PR description.

## Standards and Formats for Data Submission

- **Data Formats**: Submissions should be in the formats previously outlined in the manual, primarily Markdown and CSV, to ensure compatibility with the KB structure.
- **Adherence to Templates**: Utilize the provided templates and structures for data entries, ensuring consistency across the KB.
- **Documentation**: Include detailed documentation with your data submission, explaining the source, relevance, and any other pertinent details that will aid in the review process.

## Review and Acceptance Criteria for Contributions

- **Quality Control**: Contributions will be reviewed for their accuracy, relevance, and adherence to the established themes and categories of the KB.
- **Compliance with Standards**: Ensure that your submission complies with the data and content standards set out in the manual.
- **Community Review**: Contributions may be open to review by the wider community before acceptance. This peer-review process can provide additional validation and feedback.
- **Acceptance**: The final decision to integrate a contribution will be made by the repository maintainers, taking into account feedback from the community and alignment with the KB's objectives.

By embracing this model of collaborative curation, we invite the community to take an active role in the continual growth and refinement of the KB. Contributors' expertise and insights are invaluable to maintaining the KB as a premier resource that supports the dissemination of up-to-date and impactful climate-resilient agricultural practices.