




Gurukul

Science, Technology and Research

Generated: **Thu Jul 17 2025** Dataset Type: **Structured** Source: **IITGN**

Parameter	Value
Short Description What this dataset is about	Educational Data for Indian Education System
Long Description / About Dataset Brief detailed insight into dataset content/use	Key Features <ul style="list-style-type: none"> • Curriculum-Aligned: Derived directly from NCERT textbooks with chapter-level annotations. • Bilingual: Available in both English and Hindi. • Manually Reviewed: Entries curated with filtering using Claude Sonnet API and human review. • Instructional Style: Questions are cleaned, de-duplicated, and rewritten to be self-contained. • Use Cases: Useful for building and evaluating educational LLMs, especially for: <ul style="list-style-type: none"> ◦ Generative QA ◦ Retrieval-augmented systems ◦ MCQ generation ◦ Exam-oriented tutoring
Purpose Why this dataset was created	N.A.
Author Dataset creator and maintainer	Isha Narang, Sneha Gosai, Mayank Singh
Dataset Type Format and structure of dataset provided	Structured
Year Range Time period covered by the dataset	N.A.
Primary Key / Indicator Unique identifier for dataset records	N.A.
Hosted / Redirected Platform where the dataset is stored	Hosted

Parameter	Value
Sector Industry domain and application area	Science, Technology and Research
License Usage rights and restrictions	Attribution-ShareAlike 4.0 International (CC BY-SA 4.0)
Source Organization Original data provider organization	IITGN
Frequency How often the dataset is refreshed	Daily
Date & Time Last modification timestamp	Mon Dec 15 2025
Category Data organization and format type	
If redirection which source Original data provider organization	N.A.
Geographical Coverage Geographic scope of the dataset	Global
Uploaded by Uploader of this dataset file	Lingo Research Group
Time Granularity Data's time based detail level	NA
Visibility Who can view this dataset	Open
Data Collection Method How, when and from where the data was gathered	N.A.
Data Quality Score Machine learning readiness assessment	★★★★★

 **Gurukul Dataset Link:**
https://aikosh.indiaai.gov.in/home/datasets/details/gurukul_2.html