**PEDIATRIC PULMONOLOGY CHATBOT**

## Sources Used for data gathering

* Mayo Clinic — [https://www.mayoclinic.org](https://www.mayoclinic.org/)
* PubMed Central (PMC) — [https://pmc.ncbi.nlm.nih.gov](https://pmc.ncbi.nlm.nih.gov/)
* National Institutes of Health (NIH)
* Peer-reviewed Pediatrics and Pulmonology Journals
* Clinical guidelines and educational materials from pediatric pulmonology societies
* Additional reputable medical platforms including Medscape and Cleveland Clinic

## 2. Methodology

* Used medical search engines and databases such as PubMed, PubMed Central, and the Mayo Clinic website.
* Typed keywords related to pediatric pulmonology conditions, for example:
  + *asthma*
  + *bronchiolitis*
  + *cystic fibrosis*
  + *pneumonia*
  + *sleep apnea in children*
  + *wheezing in infants*
* Selected recent review articles, meta-analyses, clinical practice guidelines, and patient education resources focusing on respiratory diseases in children.
* Cross-checked information across multiple sources to verify accuracy, consistency, and comprehensiveness.
* Extracted detailed data about disease definitions, symptoms, causes, risk factors, diagnosis, treatment, prevention, and frequently asked questions to populate the chatbot knowledge database.

## 4. Date Accessed

* Data gathered between July 08, 2025, and July 14, 2025.

## . Description of Contents

* Comprehensive overview of pediatric pulmonology conditions including causes, symptoms, differential diagnoses, treatment options, surgical procedures, and rehabilitation timelines.
* Included doctor-patient conversation examples for clear communication strategies.
* Summaries of recent advances in management and clinical guidelines.

## 6. File Format and Size

* Sources accessed were primarily in Microsoft document which was 4.7MB

**Disease list of pediatric pulmonology**

**Common and Chronic Respiratory Diseases**

* Asthma
* Reactive Airway Disease (RAD)
* Bronchitis (acute and chronic)
* Bronchiolitis (commonly RSV-related)
* Pneumonia (bacterial, viral, fungal)
* Cystic Fibrosis
* Bronchopulmonary Dysplasia (BPD) / Chronic Lung Disease of Infancy
* Chronic Respiratory Failure / Insufficiency
* Wheezing disorders
* Chronic cough

## **Congenital and Structural Abnormalities**

* Congenital lung abnormalities (e.g., congenital cystic adenomatoid malformation)
* Congenital airway anomalies (tracheomalacia, bronchomalacia, laryngomalacia)
* Chest wall deformities
* Airway malacia and stenosis

## **Rare and Diffuse Lung Diseases**

* Children’s Interstitial Lung Disease (chILD)
* Bronchiolitis Obliterans
* Pulmonary fibrosis
* Neuroendocrine cell hyperplasia of infancy (NEHI)
* Alveolar hemorrhage syndromes

## **Infectious Diseases**

* Respiratory Syncytial Virus (RSV) infection
* Pertussis (Whooping cough)
* Tuberculosis (TB)
* Other viral and bacterial lower respiratory tract infections

## **Neuromuscular and Functional Disorders**

* Neuromuscular respiratory diseases (e.g., muscular dystrophy, spinal muscular atrophy)
* Muscle weakness affecting breathing and cough
* Sleep-disordered breathing (obstructive and central sleep apnea)
* Apnea of prematurity and infant apnea

## **Airway and Aerodigestive Disorders**

* Pulmonary aspiration syndromes (due to swallowing dysfunction or gastroesophageal reflux)
* Aerodigestive disorders involving combined airway and digestive tract issues

## **Vascular and Hematologic-Related Pulmonary Diseases**

* Pulmonary hypertension (including secondary to congenital heart disease or chronic lung disease)
* Hereditary hemorrhagic telangiectasia (HHT)
* Pulmonary complications of sickle cell disease

## **Other Important Conditions**

* Primary ciliary dyskinesia (PCD)
* Non-cystic fibrosis bronchiectasis
* Respiratory failure requiring ventilator support or tracheostomy
* Subglottic stenosis
* Acute respiratory distress syndrome
* Laryngeal web
* Esophageal atresia
* Tracheoesophageal fistula
* Alpha 1 antitrypsin deficiency
* Alveolar capillary dysplasia
* Asbestosis
* Atrioventricular septal defects
* Bronchopulmonary sequestration
* Chylothorax
* Congenital diaphragmatic hernia
* Congenital lobar emphysema
* Diaphragmatic paralysis
* Hemoptysis
* Hypoplastic left heart syndrome
* Parapneumonic effusion and empyema
* Pectus carinatum
* Pectus excavatum
* Truncus arteriosus
* Pleural effusion
* Pneumococcal pneumonia
* Pulmonary contusion
* Pulmonary stenosis
* Pulmonary stenosis
* Spontaneous pneumomediastinum
* Thoracic trauma
* Sudden infant death syndrome
* Transient tachypnea of newborn
* Total anomalous pulmonary venous return
* Sarcoidosis
* Aspirin exacerbated respiratory disease
* Community acquired pneumonia
* Bronchial atresia

Followed by their CPT codes and ICD 10 codes.

We converted the doc.x file to txt file before building the model for the chatbot

**BUILDING OF THE MODEL USING PYTHON ENVIRONMENT**

Step 1: Define the Objective

Patients or caregivers often have questions related to pediatric pulmonology—such as symptoms, medication side effects, care instructions, or general respiratory health concerns. These questions frequently arise outside clinical hours or when access to a healthcare professional is limited. This can lead to anxiety, misinformation, delayed care, or unnecessary clinic visits.

The problem is to **provide timely, reliable, and non-critical advice to pediatric patients or their caregivers based on their inquiries about lung and respiratory health**, without replacing professional medical consultation but supporting informed and responsible health management.

**Why this solution?**

* **Accessibility:** Many caregivers need immediate reassurance or preliminary guidance, which is not always feasible with limited healthcare resources.
* **Efficiency:** Reduces the workload on healthcare professionals by automating responses to routine, non-critical questions.
* **Patient Engagement:** Encourages proactive health management by delivering easy-to-understand advice.
* **Safety:** Limits risk by carefully focusing on non-critical advice, emphasizing that this doesn’t replace professional diagnosis or treatment.
* **Scalability:** Can be deployed via chatbots, mobile apps, or patient portals, making advice available 24/7.

**How does this solution solve the problem?**

* Use Natural Language Processing (NLP) to understand the patient’s or caregiver’s query expressed in everyday language.
* Leverage a specialized dataset (Pediatric Pulmonology.txt) that contains verified medical knowledge relevant to pediatric lung health.
* Build a machine learning model to classify queries and generate appropriate, medically accurate, non-critical advice.
* Integrate safety checks and disclaimers, ensuring users know when to seek professional care.
* Provide immediate responses to common inquiries, reducing wait times and misinformation.

## **Step 2: Data Loading and Exploration**

We are using the dataset titled **"Pediatric Pulmonology.txt"** which contains pediatric respiratory health-related data. This dataset includes patient inquiries, medical notes, or clinical advice snippets focused specifically on pediatric pulmonology topics.

* It is **domain-specific**, focused exclusively on pediatric lung health, making it highly relevant for training an NLP advice model in this niche.
* It contains real-world examples of patient/caregiver queries and expert advice, which helps the model learn meaningful, contextual responses.
* Other datasets are either too general (like generic medical FAQs) or not focused on pediatrics or pulmonology, which would reduce model accuracy and relevance.

**Why this approach to loading the data?**

* The dataset is in a raw text file format (.txt), requiring flexible parsing to convert it into structured tabular form for machine learning.
* We emphasize robust error handling and path management to ensure smooth loading in various environments and prevent failures due to file path issues.

**Challenges faced during data gathering and with the dataset itself:**

* The dataset was manually curated from multiple sources (clinical notes, FAQs, expert advice), which might lead to inconsistent formatting and varying data quality.
* The .txt format is unstructured and may require preprocessing and cleaning before ML tasks.
* Ensuring the data is relevant and correctly labeled for non-critical advice classification requires careful inspection and validation.

**How the data was chosen:**

* Selected for its relevance to the pediatric pulmonology domain and the inclusion of real inquiry-response pairs.
* Validated by domain experts to ensure medical accuracy and suitability for non-critical advice generation.

## **Step 3: Exploratory Data Analysis (EDA)**

To thoroughly understand the pediatric pulmonology inquiry dataset by analyzing its structure, content, and quality. This involves exploring the text data and associated labels to prepare for effective NLP model development focused on providing non-critical patient advice.

**Why is this step necessary?**

* To verify the integrity and completeness of the dataset specific to pediatric pulmonology inquiries.
* To identify any inconsistencies, missing values, or noisy text that could negatively impact model training.
* To understand the distribution of patient inquiry types or advice categories, ensuring balanced representation.
* To reveal characteristics of the textual data, such as average length or common terms, guiding preprocessing and feature extraction choices.
* This step minimizes risks of model underperformance and ensures the advice generated aligns with the dataset’s domain-specific nuances.

**How did we go about the process?**

* Assessed dataset size and structure including number of records and feature columns.
* Checked for missing or duplicate entries that may distort model learning.
* Sampled patient inquiries and advice texts to gain qualitative insights into language style and content complexity.
* Evaluated class distribution if categorical labels are present, detecting any imbalances in advice categories.
* Measured text length statistics (e.g., average number of words) to inform tokenization and model input design.
* Optionally, created visualizations such as bar charts or word clouds to highlight frequent terms and category frequencies.

## Step 5: Data Cleaning

Cleaning the raw text data involves removing noise, normalizing text, and preparing it for downstream NLP tasks like tokenization and modeling.

### Why?

* Raw clinical text often contains inconsistencies such as punctuation, special characters, varying capitalization, and irrelevant sections.
* Cleaning improves model accuracy by focusing on meaningful text and reducing irrelevant noise.
* Normalized text helps create consistent tokenization and feature extraction.

### How?

* Lowercase all text for uniformity.
* Remove punctuation, special characters, and extra whitespace.
* Remove stop words if applicable.
* Optionally, apply stemming or lemmatization to reduce words to their root forms.
* Handle domain-specific terms carefully (e.g., “cystic fibrosis” should stay intact).
* Remove or handle headers and non-informative sections detected earlier.

Step 6

Model building

The dataset was used to train textcatbow model and BioGPT large language model.

Step 7 deployment

Key points from research and deployments support this focus:

* Deliver real-time, accurate, and accessible medical guidance tailored to pediatric conditions, improving caregiver health literacy and confidence.
* Help assess severity of symptoms and suggest appropriate care pathways, supporting better health-seeking behaviors and potentially lowering ED overcrowding.
* Maintain a user-centered approach prioritizing empathy, personalization, and trust to foster engagement and adherence.
* Address privacy and compliance with pediatric healthcare regulations.
* Enable scalable two-way communication via SMS using APIs like Twilio integrated with NLP platforms (e.g., Dialogflow) [User context].

## Suggested refined chatbot goals for your pediatric pulmonology SMS bot:

* Symptom triage and guidance: Help parents identify the seriousness of lung-related symptoms and recommend next steps.
* Answer frequently asked questions: Provide validated answers about diseases, treatments, risk factors, and prevention.
* Support patient-provider communication: Simulate doctor-patient dialogue for common concerns, clarifying confusion and anxiety.
* Promote appropriate care navigation: Guide users to schedule appointments or emergency care as needed.
* Offer understandable pediatric health education: Communicate complex medical info in clear, empathic language suited for caregivers.

## Chatbot’s core use cases based on this goal — for example:

* Parent texts "My child has wheezing and cough, what should I do?" → chatbot triages severity & advises action
* Parent asks "How can I prevent asthma attacks?" → chatbot replies with prevention tips and common triggers
* Parent requests "Can you explain what sleep apnea is?" → chatbot provides child-friendly disease description
* Parent reports "My child has a fever and runny nose." → chatbot suggests home care or visit based on symptom severity

**Team members**

Jennifer Imogie

- Leslie El

- Abubakar Barakat Oyiza