ChatGPT

Objective Definition

Pediatric growth assessment involves tracking a child's **height/length**, **weight**, **BMI**, **head circumference**, and **developmental milestones** over time. These metrics are typically plotted on age- and sex-specific growth charts (e.g. WHO or CDC charts [1]). Height-for-age and weight-for-age percentiles diagnose chronic and acute malnutrition, respectively [2] [3]. BMI-for-age percentiles screen for underweight or overweight/obesity (AAP recommends annual BMI screening from age 2 [4]). Head circumference (especially in <2 yr) reflects brain growth [5]. Gross motor and other developmental milestones (sitting, walking, language) are tracked at well-child visits [6]; missed milestones can signal neurodevelopmental issues. Overall, **child growth metrics are critical indicators of nutritional, endocrine, and neurodevelopmental health** [7] [5] and guide interventions (e.g. nutrition support, specialist referrals).

- **Key metrics:** Length/height-for-age, weight-for-age, weight-for-length (or BMI)-for-age, head circumference-for-age, and age-specific developmental milestones [1] [6].
- **Clinical significance:** Deviations (e.g. height-for-age <–2 SD = stunting, weight-for-height <–2 SD = wasting) indicate undernutrition and chronic illness [2] [3]. High BMI-for-age (≥95th percentile) predicts obesity and related comorbidities [8] [4]. Microcephaly or macrocephaly (head circumference anomalies) suggest neurological disorders [5]. Tracking milestones ensures early detection of developmental delay [6].

Dataset Identification

We identified multiple public and institutional datasets relevant to pediatric growth assessment:

- **CDC Pediatric Growth Charts (USA):** Official percentile data (height, weight, BMI, head circumference) for U.S. children (0–20 yrs) [9]. These are *reference charts* derived from historical NHANES and other survey data. *Data type:* LMS parameters and percentile curves (no raw patient records). *Access:* Public via CDC/HealthData.gov [9]. *Use:* Standard clinical tool (with CDC publications).

- **WHO Child Growth Standards (Global, 0–5 yrs) and Growth Reference (5–19 yrs):** International growth charts (length/height, weight, BMI, head circumference, and motor milestones) based on the WHO MGRS multicenter study [1]. *Source:* WHO, combining healthy cohorts from 6 countries. *Data:* Percentile curves; raw data not public. *Access:* Public via WHO site [1]. *Publications:* WHO MGRS group (Lancet 2006 etc).

- **NHANES (USA):** Ongoing U.S. national surveys with direct anthropometry. *Source:* CDC/NCHS. *Data:* Measured height, weight, and other body measurements for children (0–19 yrs); cycles (e.g. 2007–2018). *Size:* ~5,000 individuals per 2-yr cycle. *Demographics:* Nationally representative U.S. (includes all races/ethnicities). *Access:* Public (NHANES website). *Use:* Underpins CDC growth references; available in CDC Vital and Health Statistics reports [9].

- **National Survey of Children's Health (NSCH, USA):** Annual parent-survey of ~30,000 U.S. children (0–17 yrs). Collects parent-reported height/weight (for BMI), plus broad health/demographics [10]. *Size:* ~80,000 pooled across 2016–2023. *Access:* Public via Census Bureau. *Use:* Population health, BMI prevalence and correlates.

- **All of Us Research Program (USA):** NIH cohort with rich EHR and survey data. Includes *legacy pediatric records* for ~19,729 participants (measurements, diagnoses, visits since ~1980) [11]. *Data:* Structured EHR data (height, weight, BMI, diagnoses, etc.); predominantly U.S. adults now, but with archived childhood data. *Demographics:* Diverse U.S. (emphasis on minorities). *Access:* Controlled (researcher application required). *Publications:* Jamia Open demonstration of pediatric obesity trends [11].

- **PEDSnet (USA):** National pediatric learning health system (consortium of children's hospitals). Aggregated *EHR data (demographics, vitals, labs, diagnoses, anthropometry)* for ~7.5 million children (2009–2021) [12] [13]. *Size:* Over 5.3 M children with ≥1 height/weight (pre-cleaning); final curated dataset ~3.5 M children and 19.4 M weight-height pairs [14]. *Demographics:* U.S. from 7 states (CO, DE, FL, IL, IN, KY, MO, NJ, OH, PA, WA) [15]. *Labels:* No explicit labels, but obesity can be derived. *Access:* Limited to PEDSnet sites/partners. *Publications:* JAMIA introduction to PEDSnet (anthropometrics and diagnoses data model) [16].

- **Global Nutrition Surveys (DHS/MICS):** USAID Demographic and Health Surveys and UNICEF MICS collect child anthropometrics in many countries. *Data:* Height, weight (and MUAC) for under-5 children, with demographic context. *Size:* Hundreds of thousands of children (across many country surveys). *Access:* Public via DHS/MICS websites (requires account). *Use:* Malnutrition (stunting, wasting) prevalence, trend analysis.

- **Fels Longitudinal Study (USA):** Century-long birth cohort from Cincinnati (1929–present). *Data:* Longitudinal anthropometry (length, height, weight) across lifespans and generations. *Size:* Thousands of subjects (including descendants). *Access:* Research collaboration by request (not publicly open). *Relevance:* Source of early CDC growth charts (ages 0–3, 1978–2000) [17]. *Publications:* Numerous papers on growth and maturation.

- **childdevdata (Global Developmental Milestones):** Open dataset ("D-score") compiling developmental milestone assessments from 10 studies. Contains **1,116,061 assessments** on **10,831 children** (28,465 visits) across 21 assessment instruments [18]. *Data:* Age, sex, milestone attainment (cognitive/motor language) scores. *Access:* Public via CRAN/GitHub (R package). *Use:* Modeling milestone achievement and developmental norming.

- **PIC (Paediatric Intensive Care) Database (China):** Critical-care EHR from a Chinese children's hospital. *Data:* Time-series vitals, medications, labs, fluid balance, diagnoses, demographics (including age/weight). *Size:* >12,000 ICU admissions (v1.0). *Access:* Public (PhysioNet, with credentialing and DUA) [19]. *Use:* Research in pediatrics and ICU, including growth-related analyses (e.g. fluid balance vs weight).

LLM-Relevant Features

Most pediatric growth datasets are **structured, numeric data** (anthropometric measurements), not natural language corpora. For example, CDC/WHO growth charts are tables of percentiles [9], and NHANES/PEDSnet provide height/weight entries. Thus:

- **Structured vs. Textual:** The data are predominantly numeric. Few contain raw clinical notes or narrative. (EHR systems like PEDSnet/AllofUs may have free-text fields, but these are typically not included in public data extractions [16].) This limits direct LLM training on raw text.
- **Longitudinal Data:** Several sources are longitudinal (repeat measures per child). Fels, PEDSnet and AllofUs have *longitudinal* growth records. childdevdata tracks each child over multiple visits. In contrast, growth chart references and surveys (NHANES, DHS) are cross-sectional. Longitudinal data enable modeling growth trajectories (more powerful for prediction).
- **Labels/Annotations:** Explicit labels (e.g. "failure to thrive", endocrine disorder) are generally absent. Datasets can yield labels by deriving conditions (e.g. BMI ≥ 95th percentile = "obesity"). childdevdata provides milestone attainment outcomes. EHR data include diagnostic codes (e.g. ICD codes for growth disorders) which could serve as labels if accessible.
- **Multimodal Content:** No dataset combines growth data with images or rich text. (Growth charts and body measurements are non-visual data.) Pediatric radiographic or genetic data are outside this scope. Thus LLM applications would rely on textual interpretation rather than multimodal fusion.
- **Suitability for LLMs:** The numeric nature means these datasets are better suited to structured-data models or to *retrieval tools* for LLMs, rather than direct LLM training. However, accompanying documentation, medical literature, and clinical notes (if available) could be used in Retrieval-Augmented Generation (RAG) approaches.

Gaps and Challenges

- **Data Quality:** Even large datasets contain errors. For example, a Brazilian pediatric dataset (50M records, 10.7M children) required outlier detection, with ~5–6% of height/weight values flagged as impossible [20]. EHR data require extensive cleaning: PEDSnet uses ≈1000 automated checks for completeness and plausibility [13]. Inaccurate or missing measurements can bias model training.
- **Privacy and Access:** Pediatric data are highly sensitive. Regulations (HIPAA in the US, GDPR in Europe) strictly limit sharing child health data [21]. Most detailed datasets (EHRs, surveys) are not fully public; researchers often need IRB approval, data use agreements or synthetic surrogates. Secure computing environments and de-identification are mandatory.
- **Representativeness/Bias:** Many data sources are not population-representative. For example, PEDSnet's 3.5M children come from specific hospital networks in certain U.S. states [15], so may underrepresent rural/minority populations. Healthcare-based cohorts may overrepresent disease. Similarly, WHO standards are based on optimal-growth cohorts, not reflecting malnourished communities. This can bias an LLM if it learns from non-representative inputs.
- **Limited Textual Context:** The lack of native clinical text or detailed labels makes supervised LLM training difficult. There's no standard benchmark "question-answer" or report narrative dataset for pediatric growth. Developers might need to create synthetic or annotation datasets.
- **Ethical Considerations:** Using child data requires special consent and ethical oversight. In many jurisdictions, informed consent must be obtained from guardians, and minors' data have extra protections. Engaging with vulnerable populations (children) demands careful governance.

Healthcare Context and LLM Applications

LLMs could support pediatric growth assessment in several ways:

- **Chart Interpretation & Summarization:** An LLM could convert numeric growth data into a narrative summary (e.g. “Height is consistently at the 10th percentile; growth velocity has slowed recently”). For example, a prototype study integrated an LLM into a SMART-on-FHIR growth-chart app. It analyzed a child’s growth curve and history, suggesting specialist referrals and growth disorder evaluations [22].
- **Predictive Alerts:** By analyzing longitudinal growth patterns, an LLM-based tool could predict risk of disorders (e.g. “likely failure-to-thrive” or “prediabetes risk”) before clinical diagnosis, prompting early intervention. (Generic ML research supports this: comprehensive datasets and algorithms can predict growth outcomes with “remarkable precision” to give actionable insights [23].)
- **Patient/Family Education:** An LLM could generate easy-to-understand explanations of growth charts or feeding recommendations for parents (e.g. simplifying CDC guidelines). This requires careful tuning to medical guidelines.
- **Clinical Decision Support:** A conversational LLM interface could answer a clinician’s questions (“What follow-up for this BMI trajectory?”) or flag deviations from normal. In the SMART-on-FHIR example, clinicians found the AI tab useful as a “second opinion” in complex cases, although they noted the model sometimes gave incomplete or incorrect advice [24]. This underscores the need for explainable, validated outputs.

Explainability and Validation: All LLM outputs must be transparent and clinically vetted. In the growth-chart app trial, the LLM’s recommendations were scored by pediatricians; the model made correct referrals in some cases but also produced “hallucinations” (fabricated details) in others [24]. Any clinical LLM must thus include confidence estimates, cite evidence (e.g. highlighting relevant guideline text), and undergo rigorous evaluation. In summary, LLMs offer promise (automating chart narratives, risk prediction, report generation) [22] [23], but safe use requires alignment with pediatric expertise.

Recommendations

- **Ideal Dataset Criteria:** A benchmark dataset for growth assessment should be **large**, **diverse**, and **longitudinal**. It would combine structured measurements (height, weight, etc.) with corresponding patient context (age, sex, ethnicity, medical history). Including *textual elements* (clinical notes or narrative reports) and *ground-truth labels* (e.g. diagnosed growth disorders) would enable richer LLM tasks. Multi-center data with varied demographics minimizes bias. Privacy-preserving design (de-identified or synthetic proxies) is essential.

- **Data Creation/Augmentation:** To build such datasets, stakeholders can:

- **Collaborate across hospitals** (like PEDSnet) to pool de-identified pediatric EHR, ensuring scale and heterogeneity.
- **Use synthetic data generation**, e.g. Generative models to create realistic growth trajectories and clinical notes (building on tools like Synthea). This can augment scarce real data while preserving privacy.

- **Leverage federated learning:** Train LLMs on-site at hospitals without sharing raw data, aggregating model updates for generalization.
- **Annotate existing data:** Add labels (e.g. "obesity", "growth hormone deficiency") and textual summaries to structured records, either manually by experts or via semi-automated tools.
- **International datasets:** Augment with global surveys (DHS/MICS) to cover diverse nutrition profiles, and with developmental data (e.g. childdevdata) for milestone tracking.

- **Evaluation Metrics for LLMs:** Depending on task:

- *Classification/Prediction:* Accuracy, precision/recall/F1 (e.g. predicting obesity or growth-failure). ROC-AUC and calibration (to measure reliability of risk scores).
- *Regression:* Mean Absolute Error (MAE) or RMSE for predicting continuous growth measures (height-for-age z-scores).
- *Summarization/Generation:* BLEU and ROUGE scores compare generated text to reference summaries [25]. However, these capture only surface overlap; clinical correctness must be assessed via domain-specific error metrics or expert review (e.g. checking factual consistency and potential for harm) [25].
- *Clinically-focused metrics:* Expert-rated "accuracy of diagnosis" or "usefulness to clinician" can be scored. Unique pediatric metrics might include correct identification of stunting/wasting (percentile thresholds).

Sources and Validation

This analysis draws on a wide literature and official sources. We searched academic databases (PubMed, Google Scholar) and repositories (e.g. HealthData.gov, WHO) for pediatric growth datasets and ML studies, and cross-referenced them with clinical guidelines. For example, the WHO and CDC growth chart webpages [1] [9] were used to verify which measurements are standard and their clinical use. Published studies (e.g. the All of Us JAMIA Open paper [11], PEDSnet data papers [12] [13], and the CDC obesity screening guidelines [4]) provided data specifics and sizes. All claims here are supported by cited sources or by well-established pediatric practice.

Ethical and Regulatory Considerations

- **Privacy Protections:** Child health data are legally sensitive. In the U.S., HIPAA requires de-identification or strict access controls for pediatric records (and state laws add further limits). In the EU, GDPR grants special protections to minors' data. Any dataset development must **encrypt data**, remove identifiers, and secure IRB/consent approvals [21]. Secondary use of pediatric data typically mandates parental consent or an ethics waiver.
- **Informed Consent:** Research on children's data requires guardians' informed consent, and assent from older children. For large-scale datasets, broad consent or population-based waivers may be needed, with transparency about AI use.
- **Bias and Fairness:** Algorithmic bias is a concern if the training data is unbalanced (e.g. underrepresents certain ethnic groups or socioeconomic strata). Developers must audit models for fairness (e.g. equal accuracy across demographic groups) and correct biases.
- **Clinical Safety and Accountability:** AI tools for pediatric care must undergo rigorous validation. The FDA's proposed framework for AI/ML-based medical software mandates that any device

(including LLM-based decision support) demonstrate analytical and clinical validity [26]. Models should be explainable (e.g. providing justification or references) so clinicians can verify outputs.
- **Ethical Use:** Predictions or summaries about a child's growth can influence treatment decisions. Thus, outputs must be clinically accurate and should be framed as recommendations, not deterministic diagnoses. Ethical deployment also means continuous monitoring for harm (e.g. flagging when a model's advice contradicts standard of care).

Sources: All information above is drawn from peer-reviewed articles, official public data portals, and clinical guidelines (see citations). Wherever possible, statements are supported by published data (see references like [9] [1] [22]).

---

[1] Child growth standards
https://www.who.int/tools/child-growth-standards

[2] [3] [7] [8] Malnutrition in children
https://www.who.int/data/nutrition/nlis/info/malnutrition-in-children

[4] Screening for Child Obesity | Obesity | CDC
https://www.cdc.gov/obesity/child-obesity-screening/index.html

[5] How to Measure Head Growth | Medical School
https://med.umn.edu/adoption/resources/how-measure-head-growth

[6] Developmental Monitoring and Screening | CDC
https://www.cdc.gov/ncbddd/actearly/screening.html

[9] CDC Child Growth Charts - Catalog
https://catalog.data.gov/dataset/cdc-child-growth-charts

[10] National Survey of Children's Health (NSCH) | MCHB
https://mchb.hrsa.gov/data-research/national-survey-childrens-health

[11] Pediatric data from the All of Us research program: Demonstration of pediatric obesity over time - Johns Hopkins University
https://pure.johnshopkins.edu/en/publications/pediatric-data-from-the-all-of-us-research-program-demonstration-

[12] [13] [14] [15] [16] Measuring BMI change among children and adolescents - PMC
https://pmc.ncbi.nlm.nih.gov/articles/PMC11135243/

[17] Fels Longitudinal Study Collection | Boonshoft School of Medicine
https://medicine.wright.edu/population-and-public-health-sciences/fels-longitudinal-study-collection

[18] Child Development Data • childdevdata
https://d-score.org/childdevdata/

[19] Paediatric Intensive Care database v1.0.0
https://physionet.org/content/picdb/1.0.0/

[20] Identifying biologically implausible values in big longitudinal data: an example applied to child growth data from the Brazilian food and nutrition surveillance system - PubMed
https://pubmed.ncbi.nlm.nih.gov/38360575/

[21] [23] Advancing Pediatric Growth Assessment with Machine Learning: Overcoming Challenges in Early Diagnosis and Monitoring
https://www.mdpi.com/2227-9067/12/3/317

[22] [24] GitHub - amva13/growth-llm: @inproceedings{velezextending, title={Extending a Clinical Pediatric Growth Chart App Using a Large Language Model}, author={Velez-Arce, Alejandro and Anaya, Jesus Caraballo}, booktitle={AI for Children: Healthcare, Psychology, Education} }.
https://github.com/amva13/growth-llm

[25] Evaluating large language models on medical evidence summarization - PMC
https://pmc.ncbi.nlm.nih.gov/articles/PMC10449915/

[26] [PDF] US FDA Artificial Intelligence and Machine Learning Discussion Paper
https://www.fda.gov/files/medical%20devices/published/US-FDA-Artificial-Intelligence-and-Machine-Learning-Discussion-Paper.pdf