

Article

Not peer-reviewed version

LLM-Assisted Interpretation of Kinematic Gait Data in Children with Cerebral Palsy: An Expert Agreement Study on Gait Deviation Detection and Surgical Group Recommendations

[Mehrdad Davoudi](#)^{*}, Jacqueline Romkes, [Michèle Widmer](#), Chris Easthope Awai, [Elke Viehweger](#)

Posted Date: 1 June 2026

doi: 10.20944/preprints202606.0019.v1

Keywords: cerebral palsy; gait analysis; large language model; clinical decision support



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC, OpenAlex.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

LLM-Assisted Interpretation of Kinematic Gait Data in Children with Cerebral Palsy: An Expert Agreement Study on Gait Deviation Detection and Surgical Group Recommendations

Mehrdad Davoudi ^{1,2,*}, Jacqueline Romkes ¹, Michèle Widmer ¹, Chris Easthope Awai ^{3,4} and Elke Viehweger ^{1,5}

¹ Neuroorthopaedics and Centre for Clinical Motion Analysis, Department of Orthopaedics, University Children's Hospital Basel, Basel, Switzerland

² Clinic for Orthopaedics, Heidelberg University Hospital, Schlierbacher Landstr. 200a, 69118 Heidelberg, Germany

³ Lake Lucerne Institute, Data Analytics and Rehabilitation Technology DART, Vitznau, Switzerland

⁴ ETH Zurich, Department of Health Sciences & Technology, Zurich, Switzerland

⁵ Department of Biomedical Engineering (DBE), University of Basel, Basel, Switzerland

* Correspondence: mehrdad.davoudi@ukbb.ch

Abstract

Three-dimensional instrumented gait analysis (3D-IGA) is widely used to guide surgical decision-making in children with cerebral palsy (CP), but its interpretation is time-consuming and prone to inter-rater variability. We investigated whether a generative large language model (LLM) could consistently generate gait deviation findings and surgical procedure suggestions that align with expert judgement. Kinematic features for lower-limb joints across the gait cycle, stance, and swing were extracted from eight children with unilateral CP using the open-source GaitSharing Toolkit and a structured prompt, then submitted three times per patient to OpenAI's GPT-5.5 model. The model assessed 28 kinematic deviations and 12 surgical procedure groups using majority voting. One gait analyst rated the deviations, and two paediatric orthopaedic surgeons independently rated the procedures on a 0–2 ordinal scale, blinded to all clinical information beyond the kinematic curves and diagnosis. Agreement with the gait expert averaged 84.7% in the sagittal plane and was lowest at the knee (62.6%). Surgeons' agreement with the LLM reached 83.9% and 73.5%, with the tibialis anterior procedure showing the lowest concordance. Inter-surgeon agreement was 79.2%. The LLM showed high self-consistency (>90% across runs). This work demonstrates the potential of generative LLMs as assistive tools in clinical gait analysis for deviation detection and future treatment planning.

Keywords: cerebral palsy; gait analysis; large language model; clinical decision support

1. Introduction

Cerebral palsy (CP) is the most common cause of physical disability in childhood, with a substantial impact on overall development, particularly on walking ability [1,2]. The global prevalence of CP is estimated to range from 1.6 to 3.3 per 1,000 live births [3]. Although no curative treatment currently exists, a combination of surgical and conservative interventions is used to optimize physical function [1]. In ambulatory children with CP, three-dimensional instrumented gait analysis (3D-IGA) plays a central role in clinical decision-making, especially when surgical intervention is considered [4]. A single 3D-IGA assessment yields hundreds of kinematic, kinetic, electromyographic, and spatiotemporal variables, which must be integrated with clinical findings, imaging results, and patient history. This synthesis imposes a considerable cognitive burden, and

interpretation is widely recognized to depend strongly on the experience and clinical judgment of the treating clinician. The inter-observer reliability of this interpretive process has been extensively investigated in the literature [5–7] suggesting that while the interpretation of gait analysis outcomes is reproducible in principle, clinically meaningful variability persists across institutions, individual clinicians, and over time. In the context of gait with CP, particularly in the presence of multi-level involvement and complex severity, this process may lead to inconsistent clinical judgment and less predictable outcomes [8].

In recent years, machine learning approaches have increasingly been applied to gait analysis data, including electromyography (EMG) and clinical examination parameters, to develop objective artificial intelligence (AI)-driven tools that support more informed clinical decision-making in children with CP [9–11]. Despite these advances, such evidence-informed decision-making frameworks have not been widely integrated into routine clinical practice, potentially due to the lack of accessible platforms capable of consolidating and operationalizing these tools. To address this challenge, Schwartz et al. introduced the Evidence-Based Gait Analysis Interpretation Tools (EB-GAIT) framework [8], which employs Bayesian models trained to estimate the probability of commonly performed surgical procedures. However, current clinical AI tools, including EB-GAIT, rely on predefined feature sets and highly structured input pipelines and do not generate the narrative-style clinical reasoning.

Large language models (LLMs), including Chat Generative Pre-trained Transformer (ChatGPT), have recently demonstrated substantial potential for complex clinical reasoning across a wide range of medical domains [12–15]. While the use of LLMs for gait analysis, particularly for gait pattern classification across different pathologies, has already been explored [16–19], these models have not yet been systematically assessed for interpreting kinematic gait data and decision-making in CP. Whether such models can reliably replicate expert clinical reasoning remains an open and clinically important question, which this study aims to investigate.

To investigate this, we evaluated whether a general-purpose LLM provided exclusively with structured kinematic features can detect individual gait deviations from reference norms and reason about corresponding surgical procedure recommendations. The decision space of the LLM was constrained to a subset of clinically informed deviations and interventions derived from literature [6–8]. We hypothesised that such an approach would allow the LLM to serve as an assistive tool for gait deviation detection and treatment planning, with its outputs receiving high agreement from experts while supporting, not replacing, clinical decision-making (first hypothesis). More specifically, we hypothesized that model performance may vary across planes of motion, anatomical levels, and surgical procedure groups (second hypothesis). Moreover, we examined the inter-rater agreement on LLM-generated treatment recommendations, hypothesizing that clinical factors may influence their judgment, even within a single-centre setting (third hypothesis one). Eventually, we assessed the internal replication confidence (self-consistency) of LLM, expecting variability across anatomical joints and patients (forth hypothesis).

2. Materials and Methods

2.1. Ethics Statement

This study was approved by the local ethics committee (“Ethikkommission Nordwest- und Zentralschweiz, EKNZ Nr. 2025-01762”). Written informed consent was obtained from all participants or, in the case of minors under the age of 14 years, from their parents or legal guardians. To ensure data privacy, only derived kinematic features and a simplified diagnostic descriptor were transmitted to the external LLM service. No patient-identifiable information, including name, surname, age, sex, clinical history, or disease severity, was shared. Furthermore, full kinematic waveforms, spatiotemporal parameters, kinetic data, EMG, video recordings, medical imaging, or clinical examination notes were not transmitted at any stage of the pipeline. The output generated in this study was not used for clinical decision-making and did not influence the treatment recommendations of the participating surgeons. The surgeons were blinded to patient identity

throughout the evaluation process. Access to the study data (C3D files) was obtained on May 7, 2026, by selecting eligible patients (referred to also as “cases” throughout the manuscript) based on anonymised gait laboratory identifiers. No prior knowledge of the patients’ underlying gait abnormality was involved in the selection process.

2.2. Participants

This study included clinical gait data from eight patients with unilateral CP and 20 typically developing (TD) individuals, all aged between 7 and 17 years. Only the affected side was included in subsequent analyses. Inclusion criteria for all participants were the availability of complete kinematic data and the ability to walk barefoot without assistive devices. Only patients with written consent were included. The TD cohort has been described in our previously published works [20,21]. Individuals with any known neurological or orthopaedic impairments, a history of major lower limb surgery, or a leg length discrepancy greater than 1% of body height were excluded. Demographic characteristics of both CP and TD participants are summarized in Table 1.

Table 1. Group details. Uni-CP: children with unilateral cerebral palsy, TD: typically developing children, M: male, F: female, R: right, L: left, NA: not applicable.

Group	Parameter					
	N	Sex	Age (years)	Height (m)	Weight (kg)	Affected side
		M/F	Mean [range]	Mean [SD]	Mean [SD]	R/L
Uni-CP	8	6 M/2 F	10.3 [7.2-14.8]	1.39 [0.15]	36.4 [19.9]	4 R/4 L
TD	20	10 M/10 F	11.4 [7.6-16.7]	1.47 [0.18]	37.3 [12.0]	NA

2.3. Experimental Protocol

Data collection was conducted in the movement laboratory at the University Children's Hospital Basel (UKBB, Switzerland). Participants walked barefoot at a self-selected speed over a minimum distance of 6 m, with one practice trial allowed prior to recording. Kinematic data were collected using a 12-camera motion capture system (Vicon, Oxford Metrics Limited, Oxford, UK) at 150 Hz, with reflective skin-mounted markers placed according to the protocol described by Kadaba et al. [22] and the Plug-in-Gait (PiG) model used for analysis. Seven valid trials were obtained per participant; trials containing tracking errors, missing markers, or non-steady walking patterns were excluded. Gait cycles were defined by visually identifying foot strike and toe-off events, and marker trajectories were filtered using the built-in Woltring filter in Vicon Nexus. Joint kinematics were calculated using the Conventional Gait Model (CGM 1.0, Plug-in-Gait, Vicon Nexus, UK) [23], with body segment inertial parameters based on Dumas et al. [24]. Joint angles for the pelvis, hip, knee, and ankle were extracted in the sagittal, frontal, and transverse planes, and the foot progression angle was computed as an indicator of in-toeing or out-toeing gait. All data were stored in C3D format. For the TD cohort, 39 trials comprising 155 strides were used to establish reference kinematic values. For the CP group, 53 trials from eight patients yielding 131 strides were analyzed, with interpretation restricted to the clinically affected side.

2.4. GaitSharing Toolkit

All processing steps, from raw C3D files to the final LLM-generated report, were performed using GaitSharing, a freely available open-source toolkit previously developed in Python (version 3.13) by the authors. The toolkit is publicly accessible at [25,26]. It implements a fully integrated, end-to-end pipeline composed of the modules described below, each of which is accessible via a graphical user interface. The C3D Extractor module imports preprocessed and modeled kinematics data using the ezc3d library [27], and segments strides for both right and left sides using heel-strike and toe-off

events recorded in the C3D. The strides were time-normalised to 101 data points using linear interpolation (interp1d function). The stance phase proportion was calculated to enable subsequent labeling of stance and swing phases. This was followed by a feature extraction stage to transform each kinematic curve into a compact tabulated representation as an input to the LLM. For each joint (pelvis, hip, knee, ankle, and foot progression angle) and plane of motion (sagittal, frontal, transverse), six summary features were extracted: max, min, their timing within the gait cycle, mean, and range, recommended by [9,10,28]. These are calculated over the gait cycle, as well as for the stance, and swing phases on both sides, first at the stride level and then averaged across all valid strides for each patient. A dataset of TD individuals (described in Section 2.2) was processed through the same pipeline to establish reference values. All the features exported to LLM were paired with a corresponding TD value to further facilitate identification of deviations.

2.5. LLM Interpretation

The LLM Interpreter module of GaitSharing communicates with the OpenAI Application Programming Interface (API) using a user-provided API key. The model employed in this study was gpt-5.5-2026-04-23. All API calls were performed within a single session, with each assessment requiring less than five minutes on a standard gait laboratory workstation (Intel Core i7-8700, 16 GB RAM). For traceability and reproducibility, the toolkit automatically logged the model version, execution timestamp, patient identifier, and diagnostic context for each run in a local text file. Prior to submission, all input files were anonymised to remove patient-identifiable information. File names were replaced with a generic descriptor (“patient_kinematics.txt”), and the diagnostic context provided to the model was simplified to its essential form: “unilateral cerebral palsy with affected right/left side.” A standardized prompt was used for all patients to ensure consistency in reasoning and output structure. The prompt instructed the model to assume the role of a paediatric orthopaedic surgeon and gait analyst specializing in cerebral palsy, with reasoning strictly limited to the kinematic features and diagnostic descriptor.

Phase 1- Gait deviation detection: The model assesses a predefined list of 28 kinematic gait deviations on the affected side only (Table 2) [6–8]. For each deviation, the model responds YES, NO, NMI (need more information) or INC (inconclusive). When responding YES, it must record the exact variable name and value from the input file as supporting evidence.

Table 2. Predefined kinematic gait deviations assessed by the LLM in Phase 1. ROM: range of motion.

Joint	Deviation
Pelvis	Increased anterior pelvic tilt
	Reduced pelvic sagittal ROM
	Pelvic obliquity in stance
	Pelvic rotation asymmetry
	Excessive pelvic rotation
Hip	Excessive hip flexion in stance
	Reduced peak hip extension in stance
	Reduced hip sagittal ROM
	Excessive hip adduction in stance
	Excessive hip adduction in swing
	Excessive hip abduction in stance
	Internal hip rotation in stance
	External hip rotation in stance
Knee	Excessive knee flexion in stance
	Knee hyperextension in stance
	Reduced peak knee flexion in swing
	Delayed timing of peak knee flexion in swing
	Reduced knee sagittal ROM

	Internal knee rotation in stance
	External knee rotation in stance
Ankle	Reduced peak ankle dorsiflexion in stance
	Excessive peak ankle dorsiflexion in stance
	Excessive ankle plantarflexion in swing
	Reduced ankle sagittal ROM
	Ankle valgus in stance
	Ankle varus in stance
Foot	Foot internal rotation
	Foot external rotation

Phase 2- Surgical procedure suggestions: Based exclusively on the deviations marked YES in Phase 1, the model evaluates twelve standard procedure groups (Table 3) [6–8], and responds with the same types as deviations. When responding YES or NMI, it must list the driving deviation numbers from Phase 1 together with a short reasoning for decision. Considering NMI as a possible response for both phases 1 and 2, provided us with this possibility to further evaluate the LLM's reasoning from a clinical point of view. The prompt clearly specified that NMI could be assigned when data were missing or considered ambiguous (Phase 1) or when a relevant deviation was identified but further examination data were needed to confirm the decision (Phase 2). An example of an NMI rationale for calf muscle lengthening extracted from LLM is as follows (ROM: range of motion):

“Deviations 23, 24 – Swing plantarflexion with reduced ankle sagittal ROM may reflect dynamic equinus or dorsiflexor weakness, requiring passive ankle examination and selective motor assessment.”

The deviation and procedure lists were assigned from the existing literature particularly from [6–8]. The full prompt text is provided in the Supplementary Material (StudyPrompt.txt).

Table 3. Predefined surgical procedure groups assessed by the LLM in Phase 2.

No.	Procedure group
1	Calf muscle lengthening
2	Tibialis anterior procedure
3	Tibialis posterior procedure
4	Foot bony reconstruction
5	Hamstring lengthening
6	Rectus femoris transfer
7	Distal knee surgery
8	Derotation osteotomy
9	Hip adductor lengthening
10	Psoas procedure
11	Proximal femoral osteotomy
12	Selective dorsal rhizotomy

2.6. Expert Review

Three independent clinical experts reviewed the anonymised gait curves. The 28 kinematic deviations (Phase 1) were rated by an experienced gait analyst with more than 25 years of experience in clinical gait analysis (J.R.). The 12 surgical procedure groups (Phase 2) were rated independently by two paediatric orthopaedic surgeons: a senior surgeon with over 25 years of experience in neuro-orthopaedic surgery and gait analysis (E.V.), and an experienced paediatric orthopaedic surgeon with five years of clinical practice (M.W.). All reviewers received only the kinematic curves and the diagnostic descriptor (unilateral CP with the affected side).

Reviewers were first asked to predict independently, then looking at each LLM output, rate the decision on a three-level ordinal scale: 2 appropriate (correct decision with valid reasoning); 1

partially appropriate (correct direction but flawed reasoning, or borderline); 0 inappropriate (wrong decision or evidence). The reviewers were also asked to complete a free-text commentary summarizing their assessment. For Phase 1, the mean agreement percentage was computed across anatomical levels (pelvis, hip, knee, ankle, and foot) and planes of motion (sagittal, frontal, and transverse). If all LLM decisions were rated as 2, agreement for that specific variable was 100%. Otherwise, agreement was calculated as the percentage of the summed scores (e.g., across cases) divided by the maximum possible score (i.e., all ratings equal to 2). Similarly, it was reported for each procedure group (Phase 2) across all patients. Agreement between the two surgeons was also calculated per patient and expressed as the inter-surgeon agreement percentage. If both surgeons assigned the same rating (0, 1, or 2) to an LLM decision, inter-surgeon agreement was considered 100%; otherwise, it was 0%. Agreement was then averaged across decisions for each patient.

2.7. LLM Self-Consistency

To investigate internal replication confidence of LLMs outputs, each patient's feature file was assessed three times under identical conditions by OpenAI. Each run produced an independent set of YES/NO/NMI decisions covering all 28 deviations and all 12 procedure groups. When at least two of the three runs returned the same decision (2/3 or 3/3 agreement), that decision was considered as the final LLM output; for example, a YES:2/3, NO:1/3 distribution yielded a final decision of YES. When all three runs generated different decisions (one YES, one NO, one NMI), the item was flagged as inconclusive (INC). For each patient, the proportion of items with complete agreement (3/3=100%) and with majority agreement (2/3=66%) was computed separately for deviations and procedures. In total, in this study the LLM generated 320 decisions across the cohort (28 deviations plus 12 procedures × 8 cases), each derived from three independent runs on the same input data.

3. Results

Figure 1 presents detailed information on the gait deviations identified by the LLM and their corresponding agreement with the expert across all patients. The LLM outputs were categorized as YES, NO, and, in one single decision, INC. A summary of these findings is provided in Table 4. Overall, there was relatively high agreement between the gait expert and the LLM-detected gait deviations in the sagittal plane across most joints, with values of 87.5%, 91.7%, 68.8%, and 90.7% for the pelvis, hip, knee, and ankle, respectively. In contrast, the lowest agreement from the expert was found for the deviations observed at knee int/external rotation (56.3%), as well as pelvic obliquity and ankle varus/valgus (both 62.5%). However, across all evaluated items, the overall mean agreement reached 81.4%. The primary concerns raised by the gait expert regarding the LLM outputs and the underlying evaluation process are summarized in Table 5. These comments closely align with the discrepancies observed in Table 4. A key recommendation was to consider a threshold for defining a clinically meaningful difference from TD in deviation assessment. This was particularly relevant for the assessment of knee int/external rotation. Additionally, the expert highlighted the need to include a bilateral assessment of pelvic obliquity and to revise the terminology used for ankle varus/valgus and foot rotation.

Level	Deviation	LLM								Gait Expert							
		Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8
Pelvic	Increased anterior pelvic tilt	YES	YES	YES	NO	YES	NO	YES	NO	0	2	0	2	2	2	2	2
	Reduced pelvic sagittal ROM	NO	NO	NO	NO	NO	NO	NO	NO	2	2	2	2	2	2	2	2
	Pelvic obliquity in stance	NO	YES	NO	NO	YES	NO	NO	NO	0	2	0	2	0	2	2	2
	Pelvic rotation asymmetry	YES	YES	YES	YES	YES	YES	YES	YES	2	2	2	2	2	2	2	2
	Excessive pelvic rotation	YES	YES	YES	YES	NO	NO	YES	YES	2	2	2	2	2	2	2	2
Hip	Excessive hip flexion in stance	NO	YES	NO	NO	YES	NO	NO	NO	2	2	2	2	0	2	2	2
	Reduced peak hip extension in stance	NO	YES	NO	NO	NO	NO	NO	YES	2	2	2	2	2	2	2	0
	Reduced hip sagittal ROM	NO	YES	NO	NO	NO	NO	NO	YES	2	2	2	2	2	2	2	2
	Excessive hip adduction in stance	NO	NO	NO	NO	NO	NO	NO	NO	2	2	2	2	2	2	2	2
	Excessive hip adduction in swing	NO	NO	NO	NO	NO	NO	NO	NO	2	2	2	2	2	2	2	2
	Excessive hip abduction in stance	NO	NO	YES	YES	NO	NO	NO	YES	0	2	2	2	2	2	2	0
	Internal hip rotation in stance	NO	NO	YES	YES	YES	NO	NO	NO	2	2	2	2	2	2	2	2
	External hip rotation in stance	YES	YES	NO	NO	NO	YES	NO	YES	0	2	2	2	2	0	2	2
Knee	Excessive knee flexion in stance	NO	YES	NO	NO	NO	NO	NO	YES	0	2	0	0	0	2	2	2
	Knee hyperextension in stance	YES	NO	YES	YES	NO	NO	YES	NO	0	2	2	0	2	2	0	2
	Reduced peak knee flexion in swing	NO	NO	NO	NO	NO	NO	NO	YES	0	2	2	2	2	2	2	0
	Delayed timing of peak knee flexion in swing	NO	NO	YES	YES	YES	NO	NO	YES	2	2	2	2	2	0	2	2
	Reduced knee sagittal ROM	NO	YES	NO	INC	YES	NO	NO	YES	2	2	2	0	0	2	2	1
	Internal knee rotation in stance	NO	NO	NO	NO	NO	NO	NO	NO	2	0	2	2	2	2	2	2
Ankle	External knee rotation in stance	YES	NO	YES	YES	YES	YES	YES	YES	0	2	0	0	0	0	0	2
	Reduced peak ankle dorsiflexion in stance	NO	NO	YES	YES	YES	NO	YES	NO	2	2	2	2	2	2	2	2
	Excessive peak ankle dorsiflexion in stance	NO	NO	NO	NO	NO	NO	NO	NO	2	2	2	2	2	2	2	2
	Excessive ankle plantarflexion in swing	YES	NO	YES	YES	YES	YES	YES	YES	2	2	2	2	2	2	2	2
	Reduced ankle sagittal ROM	YES	YES	YES	YES	YES	YES	NO	YES	0	2	1	2	1	2	2	0
	Ankle valgus in stance	NO	NO	NO	NO	NO	NO	NO	NO	2	2	0	2	0	2	2	2
Foot	Ankle varus in stance	YES	YES	NO	YES	NO	NO	NO	NO	2	0	2	0	2	0	2	0
	Foot internal rotation	NO	NO	NO	YES	YES	NO	NO	NO	2	2	2	2	2	2	2	2
	Foot external rotation	YES	NO	NO	NO	NO	YES	NO	YES	2	2	2	2	2	2	2	2

Figure 1. Detailed overview of gait deviations identified by the LLM and their corresponding evaluation by the expert. LLM outputs are categorized as YES (green), NO (red), NMI (need more information), and INC (inconclusive, no colour). Evaluation scores LLM are defined as 2 (appropriate), 1 (partially appropriate), and 0 (inappropriate), represented by a colour gradient from dark to light blue.

Table 4. Percentage agreement (%) between the gait expert and the LLM outputs, stratified by anatomical joint (pelvis, hip, knee, ankle, and foot progression angle) and plane of motion (sagittal, frontal, transverse). Agreements below 75% are highlighted in red. NA: not applicable.

	Sagittal	Frontal	Transverse	Average
Pelvic	87.5	62.5	100	83.4
Hip	91.7	87.5	87.5	88.9
Knee	68.8	NA	56.3	62.6
Ankle and Foot	90.7	62.5	100	84.4
Average	84.7	70.9	86	

Table 5. Overview of gait expert feedback outlining the observed concerns regarding the LLM outputs and the methodology. ROM: range of motion, IC: initial contact.

Concern	Gait expert feedback
Clinically relevant	The model should be informed of a clinically relevant difference threshold (e.g., 1.5 standard deviation from healthy reference data) to distinguish pathological deviations from normal variability.
Phases ROM	Range of motion should be computed over the whole gait cycle only; reporting ROM for stance and swing phases separately is unnecessary and may confuse interpretation. Whole-cycle ROM is most relevant for the knee (stiff knee detection) and pelvic (increased ROM).
Pelvic ROM	Pelvic ROM is typically stable across healthy; therefore, an increased ROM should also be considered a deviation, not a reduced ROM.
Pelvic obliquity	Pelvic obliquity and rotation asymmetry requires comparison of both left and right gait cycles to distinguish anatomical or functional leg length discrepancy from a true pelvic deviation. Assessing the affected side alone may be insufficient.
Knee flexion at IC	Knee flexion angle at IC is an important indicator for downstream decision-making, particularly regarding orthotic prescription, and was not captured.
Ankle angle at IC	Ankle angle at IC is clinically important for identifying flat/fore-foot contact patterns (initial flat foot or forefoot contact), which was not included in the deviation list.
Ankle valgus/varus	Valgus and varus as measured during gait represent the alignment between the tibial segment and the global laboratory axis, not ankle joint angle. This should be acknowledged when interpreting these items. More accurate information can come from clinical examinations and clinical images.
Foot rotation	“Foot internal/external rotation” in the prompt should be renamed to “internal/external foot progression angle,” as the foot rotation signal from marker-based gait analysis is not a reliable measure of true foot rotation.

To further evaluate the performance of the LLM across procedure groups, detailed outputs and their corresponding agreement rates with both surgeons are presented in Figure 2. Notably, the LLM did not produce any “YES” outputs. This suggests that the model appropriately identified “NMI” as a valid outcome when the available data were insufficient, thereby avoiding definitive decisions based on incomplete information. The overall agreement between the surgeons and the LLM was 78.7% (83.9% for Surgeon A and 73.5% for Surgeon B; Table 6). The lowest levels of agreement were observed for the tibialis anterior procedure (31.3% for Surgeon A and 25% for Surgeon B, Table 6) and proximal femoral osteotomy (62.5% for Surgeon A and 0% for Surgeon B, Table 6).

Procedure	LLM							
	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8
Calf muscle lengthening	NMI	NO	NMI	NMI	NMI	NMI	NMI	NMI
Tibialis anterior procedure	NMI	NMI	NO	NMI	NO	NO	NO	NMI
Tibialis posterior procedure	NMI	NMI	NO	NMI	NO	NO	NO	NO
Foot bony reconstruction	NMI	NMI	NO	NMI	NO	NO	NO	NO
Hamstring lengthening	NO	NMI	NO	NO	NO	NO	NO	NMI
Rectus femoris transfer	NO	NO	NMI	NMI	NMI	NO	NO	NMI
Distal knee surgery	NO	NMI	NO	NO	NO	NO	NO	NMI
Derotation osteotomy	NMI	NMI	NMI	NMI	NMI	NMI	NMI	NMI
Hip adductor lengthening	NO	NO	NO	NO	NO	NO	NO	NO
Psoas procedure	NMI	NMI	NO	NO	NMI	NO	NMI	NMI
Proximal femoral osteotomy	NO	NO	NO	NO	NO	NO	NO	NO
Selective dorsal rhizotomy	NO	NO	NO	NO	NO	NO	NO	NO

Procedure	Surgeon A								Surgeon B							
	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8
Calf muscle lengthening	2	2	2	2	2	0	2	0	2	0	2	2	2	2	2	0
Tibialis anterior procedure	2	0	0	2	0	0	1	0	2	0	0	2	0	0	0	0
Tibialis posterior procedure	0	2	2	2	2	2	2	2	0	0	2	2	2	2	2	2
Foot bony reconstruction	2	0	2	2	2	0	2	2	2	0	2	2	2	0	2	1
Hamstring lengthening	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	0
Rectus femoris transfer	2	2	2	2	2	2	2	2	2	2	0	1	1	2	2	1
Distal knee surgery	2	2	2	2	2	2	2	1	2	2	2	2	2	2	2	2
Derotation osteotomy	2	2	1	2	2	1	2	2	1	2	2	2	2	0	2	2
Hip adductor lengthening	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
Psoas procedure	2	2	2	2	2	2	2	2	0	2	2	2	0	2	0	0
Proximal femoral osteotomy	2	2	0	0	0	2	2	2	2	2	0	0	0	2	0	2
Selective dorsal rhizotomy	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2

Figure 2. Overview of procedure groups suggested by the LLM and their corresponding evaluation by two orthopaedic surgeons. LLM outputs are categorized as YES, NO (red), NMI (need more information, yellow), and INC (inconclusive). Evaluation scores are defined as 2 (appropriate), 1 (partially appropriate), and 0 (inappropriate), represented by a colour gradient from dark to light blue.

Despite these discrepancies, inter-surgeon agreement remained relatively high, averaging 79.2% across cases, except for patient number 8, where agreement dropped to 58.4% (Table 7). Both surgeons emphasized the need for additional clinical examination data to support precise decision-making (Table 8). Furthermore, Surgeon A more frequently considered combined procedures such as tendo-Achilles lengthening in combination with tibialis anterior tendon shortening (TAL-TATS) for several patients, whereas Surgeon B highlighted the importance of patient age and the potential compensatory role of the contralateral limb in guiding treatment decisions.

Across all cases, the language model demonstrated an average self-consistency of more than 90% (Figure 3) in detecting deviations and recommending procedures. However, consistency was comparatively lower for foot-level deviations in two cases, where it dropped to 66%.

Table 6. Average percentage agreement (%) between each surgeon and the LLM across all cases, stratified by procedure group. Agreements below 75% are highlighted in red.

Procedure group	Agreement of surgeon A on LLM across all cases (%)	Agreement of surgeon B on LLM across all cases (%)
Calf muscle lengthening	75	75
Tibialis anterior procedure	31.3	25
Tibialis posterior procedure	87.5	75
Foot bony reconstruction	75	68.8
Hamstring lengthening	93.8	87.5
Rectus femoris transfer	100	68.8
Distal knee surgery	93.8	100
Derotation osteotomy	87.5	81.3
Hip adductor lengthening	100	100
Psoas procedure	100	50
Proximal femoral osteotomy	62.5	50
Selective dorsal rhizotomy	100	100
Average	83.9	73.5

Table 7. Average percentage agreement (%) between the two surgeons for each patient, defined as inter-surgeon agreement. Agreements below 75% are highlighted in red.

Patients	Inter-surgeon agreement (%)
Case 1	83.4
Case 2	75
Case 3	83.4
Case 4	91.7
Case 5	83.4
Case 6	83.4
Case 7	75
Case 8	58.4
Average	79.2

Table 8. Case-wise qualitative comments on the LLM-generated outputs for surgery procedures, provided by surgeons A and B. TAL-TATS: Tendo-Achilles lengthening with tibialis anterior tendon shortening, EMG: electromyography.

Case	Surgeon A	Surgeon B
1	This patient is a candidate for the TAL-TATS procedure. An additional assessment for possible bony foot correction is recommended.	Would proceed with calf lengthening if a contracture is confirmed and further consider tibialis anterior shortening based on EMG results. If assessments indicate flatfoot, a bony correction may be needed. Psoas lengthening is not recommended due to the patient's good function and risk of muscle weakening.
2	A knee extension osteotomy should be considered. Hip flexor contractures should be assessed to determine whether lengthening is needed.	Need additional information on the patient's muscle strength, EMG findings, contractures, and foot deformity before planning for surgical treatment.
3	candidate for TAL-TATS. Torsional alignment should be assessed through rotational range and femoral anteversion. A derotation osteotomy is indicated when internal rotation predominates with limited external rotation and increased anteversion; if external rotation is good, the issue is likely muscle weakness, and no bony correction is needed.	Depending on the clinical data, including clinical assessment, video analysis, and the torsional profile, it is likely to consider tibialis anterior shortening, calf lengthening, and hip derotation. The patient's age may influence the decision regarding knee valgus correction; in boys between approximately 12–14 years and girls between about 10–12 years, such correction may be considered.
4	Distal correction is recommended, including calf muscle lengthening and procedures involving the tibialis anterior and posterior. Torsional malalignment should also be evaluated.	Would likely proceed with calf lengthening and tibialis anterior shortening, based on clinical assessment and EMG findings. Hip derotation with foot correction may also be considered depending on torsional values, foot imaging, and kinetic data.
5	Equinus correction should be performed together with proximal bony realignment. Rectus femoris intervention should also be considered; botulinum toxin injection may serve as an alternative.	Calf lengthening, tibialis anterior shortening, and hip derotation appear to be the most plausible options, but additional clinical data is required before making a final decision.
6	This patient requires drop foot correction, with a possible tibialis anterior procedure.	A foot deformity seems more likely in this case, and we need additional information at this level to make an informed surgical recommendation. This includes foot imaging and clinical assessment of the calf muscles to better understand muscle function and potential contracture.
7	Equinus correction is indicated together with a tibialis anterior procedure. Tibial torsional alignment should be verified through clinical examination.	The patient appears to be well-compensated proximally, but at the distal level we may need to consider equinus correction, including tibialis anterior shortening, depending on clinical/EMG.
8	Clinical evaluation of knee and hip function is recommended prior to any surgical planning. No distal surgical intervention is indicated at this stage. An assessment of the torsional profile is required.	The patient shows good distal function, but more details on foot alignment and ankle range of motion are needed. The significance of observed hip and knee external rotation remains unclear. The unaffected side demonstrates compensatory gait patterns, and all these factors must be clarified before making decision.

	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8
Pelvic	100	100	100	100	93.2	100	100	93.2
Hip	100	100	100	95.8	100	100	100	100
Knee	100	95.1	100	90.4	95.1	95.2	100	95.2
Ankle	83	88.6	100	100	100	100	100	94.4
Foot	100	66	100	66	100	100	100	100
Deviations	96.4	94	100	93.9	97.6	98.7	100	96.3
Procedures	97.2	91.5	94.4	100	100	97.2	94.4	94.3

Figure 3. LLM self-consistency (%) across all cases at different segmental levels, averaged over all gait deviations and procedure recommendations. The colour scale, ranging from dark green to yellow, represents the highest (100%) to lowest levels of consistency.

4. Discussion

This study demonstrates that an LLM-based pipeline, driven solely by kinematic features and implemented through an open-source platform can generate gait deviation assessments and surgical procedure recommendations in children with CP that show substantial agreement with expert judgement (Table 4 and Table 6, first hypothesis). The LLM's performance varied across the planes of motion, showing consistently high agreement ratio for sagittal plane measures, and relatively low agreement for knee joint deviations (Table 4) and recommendations for tibialis anterior procedure (Table 6) approving the second hypothesis of the study. Furthermore, while inter-surgeon agreement was on average 79.2% it varied across patients and dropped to 58.4% for one patient (Table 7) confirming the third hypothesis regarding the influence of clinical factors, including experience, on the surgical decision-making. Finally, our results also showed more than 90% LLM self-consistency, with partial inconsistencies observed in foot-level decisions (Figure 3, fourth hypothesis).

4.1. Gait Deviations

In the sagittal plane, knee flexion and ankle plantarflexion angles at initial contact are two clinically relevant parameters associated with hamstring spasticity [29] and equinus gait pathology [30], respectively; however, they were not included in this study. This omission resulted in partial disagreement (Table 5), suggesting that a more detailed segmentation of the gait cycle beyond the stance and swing phases particularly in the sagittal plane would improve the ability of LLMs to detect and interpret clinically meaningful deviations. The relatively lower agreement observed for knee kinematics in the transverse plane (Table 4) is consistent with known limitations of 3D-IGA in measuring knee rotational angles [31]. Furthermore, the analysis was restricted to the affected side, and the model was unable to evaluate bilateral features such as pelvic obliquity in a comparative context. This limitation made it difficult to distinguish true pelvic deviations from anatomical or functional leg length discrepancies (Table 5), resulting in a reduced agreement in the frontal plane. In addition, as noted by the gait expert (Table 5), ankle varus/valgus is not directly measured in standard gait analysis at our center; instead, the reported values represent the angle between the tibial segment and the global laboratory coordinate system. This discrepancy led to the observed disagreements for this measure (Table 4). Knee varus/valgus deviations were also not explicitly linked to the listed procedures in the literature and were not included in the analysis.

As suggested by the gait expert, defining clinically meaningful thresholds relative to reference data could further enhance interpretation. For instance, Baker et al. [32] proposed that deviations exceeding ± 1 standard deviation from TD may represent a minimal clinically important difference. Incorporating such thresholds into future implementations could improve the clinical relevance of LLM-based interpretations. Additionally, while reduced knee flexion ROM over the entire gait cycle may be indicative of stiff-knee gait pathology [33], assessing such features across all subphases of gait may introduce redundant or non-informative deviations. The features used in this study have been applied in large-scale machine learning investigations [9,10], yet further research is warranted to optimize their role in interpretability and LLM-based reasoning. Integrating global gait quality indices such as Gait Deviation Index (GDI) [34], may further enhance decision-support performance.

4.2. Surgical Procedures

Agreement of the surgical reviewers with LLM's reasoning and treatment recommendations varied substantially across the twelve surgical procedure groups (Table 6). While hip adductor lengthening and selective dorsal rhizotomy achieved 100% agreement with both surgeons, agreement for the tibialis anterior procedure was relatively low (31.3% with Surgeon A and 25% with Surgeon B, Table 6). Proximal femoral osteotomy represented a second area of relatively low agreement (62.5% and 50%, respectively, Table 6).

The low level of agreement regarding the tibialis anterior procedure is consistent with the recognized complexity of this clinical decision in literature. In our centre, this intervention most commonly performed as tibialis anterior tendon shortening combined with Achilles tendon lengthening (TATS-TAL). It is typically indicated for persistent swing-phase drop foot in unilateral CP, with the additional requirement of EMG activity of the tibialis anterior during the initial swing phase [35–37]. Consequently, identifying appropriate candidates requires multimodal clinical information and cannot be reliably inferred from kinematic data alone. This limitation was independently highlighted by both surgeons in our study, who requested additional EMG and clinical examination data prior to confirming their decisions. The observed disagreement therefore maybe reflects the inherently complex and multimodal nature of this surgical indication. Furthermore, both surgeons reported ambiguity in the presentation of derotation osteotomy and proximal femoral osteotomy as distinct procedure categories. Although derotation osteotomy primarily refers to rotational corrections (e.g., tibial derotation) and is driven by transverse plane abnormalities, while proximal femoral osteotomy addresses varus/valgus deformities associated with frontal plane deviations, these distinctions were not sufficiently clarified in the review process. This lack of explicit differentiation likely contributed to confusion and represents an additional source of disagreement in the procedure selection.

The inclusion of “need more information” (NMI) as a response category provides valuable insight into the underlying reasoning process beyond binary decision-making. Comprehensive clinical gait assessment typically integrates multiple sources of information, including passive joint range of motion, muscle function, selectivity, spasticity, patient history, and muscle activity, as well as kinetic data. In particular, decisions regarding bony procedures often require imaging data, without which precise recommendations are difficult to establish. Notably, the absence of “YES” outputs for certain procedures combined with the presence of NMI responses supported by appropriate reasoning suggests that LLM was able to appropriately identify gaps in the available clinical information. Nevertheless, it remains to be determined whether the availability of more complete datasets would enable the model to arrive at accurate and clinically sound decisions.

Additionally, inter-surgeon agreement on procedure recommendations averaged 79.2% across the eight cases (range: 58.4-91.7%, Table 7). Wang et al. [6] reported agreement rates of 84-90% within a single institution, whereas Rethlefsen et al. [7] observed agreement ranging from 66% to 90% between surgeons from different institutions. In both studies, interpreters had access to comprehensive clinical information, whereas in the current study only kinematic curves were available for decision-making. This limitation, in addition to experience-related differences, may

explain why agreement levels were comparable to those reported between institutions, despite both surgeons working at the same centre.

4.3. LLM Self-Consistency

The model demonstrated substantial internal consistency across three independent runs, with average agreement exceeding 90% for both gait deviation detection and procedure recommendation across all eight cases (Figure 3). Only one single decision was classified as INC among 320 total decisions, suggesting that three repeated runs may be sufficient to obtain stable output from generic LLMs. However, it is important to emphasize that high self-consistency does not necessarily imply correctness; a model may produce consistently incorrect outputs, and such consistency can foster a misleading perception of reliability. The employed feature-based approach in this study likely contributed to the high level of self-consistency observed. Nevertheless, the impact of using full gait patterns, rather than extracted features, as input to LLMs warrants further investigation in future work.

It should also be noted that, the total processing time from reading gait data from C3D files through GaitSharing, feature extraction, and API requests to the LLM and output generation was approximately 10-15 minutes per patient, which is shorter than the duration typically reported for report generation in conventional clinical gait analysis [38]. In addition, our approach may potentially reduce the time required for interpretation and decision-making by providing a detailed report on observed deviations and relevant surgical interventions; however, this variable was not explicitly measured and compared in the current study.

4.4. Limitations and Future Directions

The primary limitation of this study is the relatively small sample size, which comprises only patients with unilateral CP. As a result, certain CP-related gait pathologies and deformities were not represented, potentially introducing bias in the agreement outcomes for specific gait deviations or recommended procedures. Future work should extend this study design to larger and more diverse patient populations, including individuals with other neurological disorders such as stroke and Parkinson's disease, with corresponding adaptation of the procedural framework.

Moreover, CP represents a heterogeneous clinical condition, and its management is not confined to surgical interventions alone. Non-surgical approaches, including physiotherapy, occupational therapy, and the use of orthotics and assistive devices, play a crucial role in treatment planning. Therefore, future investigations should aim to evaluate LLM-based decision-support systems in more comprehensive clinical contexts. However, in the current study, other clinical signals (e.g., EMG and ground reaction forces), these were excluded from the analysis due to their inherent redundancy. Nevertheless, it is noteworthy that even in the presence of kinematics only, surgeons were nearly as likely to agree with the LLM (78.7%) as with each other (79.2%).

The influence of clinical experience on agreement with the LLM was not a research question in this study, and the review process was kept anonymised accordingly. This factor could, however, be investigated in future work by involving a larger and more diverse cohort of surgeons. Similarly, the gait deviation assessment could benefit from independent rating by two or more gait analysis experts to strengthen the reliability analysis; this was beyond the scope of the present study, where the primary focus was on the final surgical recommendation level. Furthermore, as noted in the surgeons' comments, age is an important factor in decision-making for CP. Given the limited sample size of our study, future work could systematically investigate the impact of incorporating patient age into the LLM and its effect on the model's reasoning.

Finally, we believe that the integration of probabilistic decision-support tools such as EB-GAIT [8], which are developed using local clinical data with natural language reasoning systems trained on publicly available knowledge and scientific literature represents a promising direction for the next generation of clinical gait analysis decision-support systems. Furthermore, the development of time-

series foundation language models for gait analysis could be an important advancement in this direction.

5. Conclusions

In conclusion, this study highlights the potential of generative LLMs as supportive tools in clinical gait analysis, particularly for detecting gait deviations and informing potential treatment planning decisions. We introduced a clinically feasible approach, accompanied by the open-source and freely available GaitSharing toolkit [25,26], to encourage further research on the role of LLMs in clinical decision support. Despite the limited sample size and the exclusive use of kinematic data, the proposed framework achieved a high level of agreement (>75%) with expert assessments for both gait deviation interpretation and surgical procedure recommendations. The strong self-consistency of the LLM (>90%) further supports its potential for the use in clinical setting. Overall, this approach may help accelerate the conventional interpretation of clinical gait analysis while reducing subjectivity by integrating published evidence and broadly accessible knowledge into the reasoning process.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/..., Supplementary Materials StudyPrompt.txt>: The full prompt text used in this study.

Author Contributions: M.D.: writing—original draft, developing the toolkit, data analysis; J.R.: review and editing, methodology; M.W.: review and editing, methodology; C.E.A.: review and editing; E.V.: review and editing, methodology, conceptualization, project administration. All authors have read and agreed to the published version of the manuscript.

Funding: Neuroorthopaedics and Center for Clinical Motion Analysis Unit, Department of Orthopaedics, University Children's Hospital Basel, Basel, Switzerland.

Institutional Review Board Statement: This study was conducted in accordance with the Declaration of Helsinki, and approved by the local ethics committee ("Ethikkommission Nordwest- und Zentralschweiz, EKNZ Nr. 2025-01762").

Informed Consent Statement: Written informed consent was obtained from all participants or, in the case of minors under the age of 14 years, from their parents or legal guardians. Only personnel that had regular legal access to the medical records retrieved patient data. To ensure data privacy, only derived kinematic features and a simplified diagnostic descriptor were transmitted to the external LLM service. No patient-identifiable information, including name, surname, age, sex, clinical history, or disease severity, was shared. Furthermore, raw gait data (e.g., full kinematic waveforms), spatiotemporal parameters, kinetic data, EMG, video recordings, medical imaging, or clinical examination notes were not transmitted at any stage of the pipeline. The outputs generated in this study were not used for clinical decision-making and did not influence the treatment recommendations of the participating surgeons. The surgeons were blinded to patient identity throughout the evaluation process. Access to the study data (C3D files) was obtained on May 7, 2026, by selecting eligible cases based on anonymised gait laboratory identifiers.

Data Availability Statement: The full GaitSharing toolkit, including the LLM Interpreter module, the structured prompt used in this study, and all processing scripts, is openly available under MIT License at [25,26]. The de-identified feature files and the corresponding LLM reports analyzed in this study are available from the corresponding author upon reasonable request, subject to the institutional data-sharing policy of the University Children's Hospital Basel.

Acknowledgments: The authors gratefully acknowledge Prof. Dr. Sebastian I. Wolf, Prof. Dr. Babak Moradi, Dr. Morgan Sanguex, and Dr. Katrin Bracht for their valuable guidance and support throughout this work. Microsoft Copilot (Microsoft, Redmond, WA, USA) was used to assist with English language editing during manuscript preparation; the authors reviewed and revised the output and take full responsibility for the published content.

Conflicts of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as potential conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CP	Cerebral palsy
3D-IGA	Three-dimensional instrumented gait analysis
LLM	Large language model
AI	Artificial intelligence
EMG	Electromyography
ROM	Range of motion
EB-GAIT	Evidence-Based Gait Analysis Interpretation Tools
ChatGPT	Chat Generative Pre-trained Transformer
API	Application Programming Interface
NMI	Need more information
INC	Inconclusive
TAL-TATS	Tendo-Achilles lengthening with tibialis anterior tendon shortening
IC	Initial contact
GDI	Gait Deviation Index

References

1. Rosenbaum, P.; Paneth, N.; Leviton, A.; Goldstein, M.; Bax, M.; Damiano, D.; Dan, B.; Jacobsson, B. A Report: The Definition and Classification of Cerebral Palsy April 2006. *Dev Med Child Neurol Suppl* **2007**, *109*, 8–14.
2. Graham, H.K.; Thomason, P.; Willoughby, K.; Hastings-Ison, T.; Stralen, R.V.; Dala-Ali, B.; Wong, P.; Rutz, E. Musculoskeletal Pathology in Cerebral Palsy: A Classification System and Reliability Study. *Children* **2021**, *8*, 252, doi:10.3390/children8030252.
3. McIntyre, S.; Goldsmith, S.; Webb, A.; Ehlinger, V.; Hollung, S.J.; McConnell, K.; Arnaud, C.; Smithers-Sheedy, H.; Oskoui, M.; Khandaker, G. Global Prevalence of Cerebral Palsy: A Systematic Analysis. *Developmental Medicine & Child Neurology* **2022**, *64*, 1494–1506, doi:10.1111/dmcn.15346.
4. Schwartz, M.H.; Ries, A.J.; Georgiadis, A.G.; Kainz, H. Demonstrating the Utility of Instrumented Gait Analysis in the Treatment of Children with Cerebral Palsy. *PLoS One* **2024**, *19*, e0301230, doi:10.1371/journal.pone.0301230.
5. Skaggs, D.L.; Rethlefsen, S.A.; Kay, R.M.; Dennis, S.W.; Reynolds, R.A.; Tolo, V.T. Variability in Gait Analysis Interpretation. *Journal of Pediatric Orthopaedics* **2000**, *20*, 759–764.
6. Wang, K.K.; Stout, J.L.; Ries, A.J.; Novacheck, T.F. Interobserver Reliability in the Interpretation of Three-dimensional Gait Analysis in Children with Gait Disorders. *Developmental Medicine & Child Neurology* **2019**, *61*, 710–716, doi:10.1111/dmcn.14051.
7. Rethlefsen, S.A.; Hanson, A.; Ciccodicola, E.; Hara, R.; Kay, R.M.; Chambers, H.; Wren, T.A. Update on the Reliability of Gait Analysis Interpretation in Cerebral Palsy: Inter-Institution Agreement. *Gait & Posture* **2024**, *109*, 109–114, doi:10.1016/j.gaitpost.2024.01.031.
8. Schwartz, M.H.; Georgiadis, A.G. Evidence Based Gait Analysis Interpretation Tools (EB-GAIT) Treatment Recommendation and Outcome Prediction Models to Support Decision-Making Based on Clinical Gait Analysis Data. *PloS one* **2025**, *20*, e0328036, doi:10.1371/journal.pone.0328036.

9. Davoudi, M.; Salami, F.; Reising, R.; Patikas, D.A.; Wolf, S.I. Rectus Femoris Electromyography Signal Clustering: Data-Driven Management of Crouch Gait in Patients with Cerebral Palsy. *PLoS one* **2024**, *19*, e0298945, doi:10.1371/journal.pone.0298945.
10. Davoudi, M.; Salami, F.; Reising, R.; Gather, K.S.; Wolf, S.I. Gluteus Medius Muscle Activation Patterns during Gait with Cerebral Palsy (CP): A Hierarchical Clustering Analysis. *PLoS one* **2025**, *20*, e0309582, doi:10.1371/journal.pone.0309582.
11. Reinbolt, J.A.; Fox, M.D.; Schwartz, M.H.; Delp, S.L. Predicting Outcomes of Rectus Femoris Transfer Surgery. *Gait & posture* **2009**, *30*, 100–105, doi:10.1016/j.gaitpost.2009.03.008.
12. Pohlmann, P.F.; Glienke, M.; Sandkamp, R.; Gratzke, C.; Schmal, H.; Schoeb, D.S.; Fuchs, A. Assessing the Efficacy of Ortho GPT: A Comparative Study with Medical Students and General LLMs on Orthopedic Examination Questions. *Bioengineering* **2025**, *12*, 1290, doi:10.3390/bioengineering12121290.
13. Nori, H.; King, N.; McKinney, S.M.; Carignan, D.; Horvitz, E. Capabilities of Gpt-4 on Medical Challenge Problems. *arXiv preprint arXiv:2303.13375* **2023**.
14. Lee, P.; Bubeck, S.; Petro, J. Benefits, Limits, and Risks of GPT-4 as an AI Chatbot for Medicine. *New England Journal of Medicine* **2023**, *388*, 1233–1239, doi:10.1056/NEJMs2214184.
15. Singhal, K.; Azizi, S.; Tu, T.; Mahdavi, S.S.; Wei, J.; Chung, H.W.; Scales, N.; Tanwani, A.; Cole-Lewis, H.; Pfohl, S. Large Language Models Encode Clinical Knowledge. *Nature* **2023**, *620*, 172–180, doi:10.1038/s41586-023-06291-2.
16. Dindorf, C.; Dully, J.; Keilhauer, R.; Lorenz, M.; Fröhlich, M. Evaluating Large Language Models for Gait Classification Using Text-Encoded Kinematic Waveforms. *arXiv preprint arXiv:2603.13317* **2026**.
17. Celik, M.; Birsal, S.E.; Demirci, E.; Dogan, O.; Inan, M. Evaluating ChatGPT-4o's Performance in Gait Pattern Classification Using Sagittal Kinematic Data and Video-Based Inputs in Children with Cerebral Palsy. *Gait & Posture* **2025**, *121*, 43–44, doi:10.1016/j.gaitpost.2025.07.051.
18. Wang, D.; Bobenrieth, C.; Seo, H. Agir: Assessing 3d Gait Impairment with Reasoning Based on LLMs. *arXiv preprint arXiv:2503.18141* **2025**.
19. Keilhauer, R.; Lorenz, M.; Dindorf, C.; Ernst, S.; Wang, C.-Y.; Messer, P.; Stricker, D. Exploring Large Language Models for Automated Gait Analysis.; IEEE, 2025; pp. 136–142.
20. Romkes, J.; Freslier, M.; Rutz, E.; Bracht-Schweizer, K. Walking on Uneven Ground: How Do Patients with Unilateral Cerebral Palsy Adapt? *Clinical biomechanics* **2020**, *74*, 8–13, doi:10.1016/j.clinbiomech.2020.02.001.
21. Sangeux, M.; Viehweger, E.; Romkes, J.; Bracht-Schweizer, K. On the Clinical Interpretation of Overground Gait Stability Indices in Children with Cerebral Palsy. *Scientific Reports* **2024**, *14*, 26363, doi:10.1038/s41598-024-76598-1.
22. Kadaba, M.P.; Ramakrishnan, H.; Wootten, M. Measurement of Lower Extremity Kinematics during Level Walking. *Journal of orthopaedic research* **1990**, *8*, 383–392, doi:10.1002/jor.1100080310.
23. Baker, R.; Leboeuf, F.; Reay, J.; Sangeux, M. The Conventional Gait Model-Success and Limitations. *Handbook of human motion* **2018**, *1*, 489–508, doi:10.1007/978-3-319-14418-4_25.

24. Dumas, R.; Wojtusich, J. Estimation of the Body Segment Inertial Parameters for the Rigid Body Biomechanical Models Used in Motion Analysis. In: Müller B., Wolf S.(Eds) Handbook of Human Motion. *Estimation of the Body Segment Inertial Parameters for the Rigid Body Biomechanical Models Used in Motion Analysis. In: Müller B., Wolf S.(eds) Handbook of Human Motion.* **2018**, pp-47.
25. Gait Sharing: A Toolkit for Clinical Gait Analysis; Available online: <https://github.com/MehrdadDavoudiii/GaitSharingToolkit> (accessed on 25 May 2026).
26. Documentation for GaitSharingToolkit; Available online: <https://mehrdaddavoudiii.github.io/GaitSharingToolkit/> (accessed on 25 May 2026).
27. Michaud, B.; Begon, M. Ezc3d: An Easy C3D File I/O Cross-Platform Solution for C++, Python and MATLAB. *Journal of Open Source Software* **2021**, *6*, 2911, doi:10.21105/joss.02911.
28. Wolf, S.; Loose, T.; Schablowski, M.; Döderlein, L.; Rupp, R.; Gerner, H.J.; Bretthauer, G.; Mikut, R. Automated Feature Assessment in Instrumented Gait Analysis. *Gait & posture* **2006**, *23*, 331–338, doi:10.1016/j.gaitpost.2005.04.004.
29. Rha, D.; Cahill-Rowley, K.; Young, J.; Torburn, L.; Stephenson, K.; Rose, J. Biomechanical and Clinical Correlates of Stance-Phase Knee Flexion in Persons with Spastic Cerebral Palsy. *PM&R* **2016**, *8*, 11–18, doi:10.1016/j.pmrj.2015.06.003.
30. Schweizer, K.; Romkes, J.; Brunner, R. The Association between Premature Plantarflexor Muscle Activity, Muscle Strength, and Equinus Gait in Patients with Various Pathologies. *Research in developmental disabilities* **2013**, *34*, 2676–2683, doi:10.1016/j.ridd.2013.05.025.
31. Osis, S.T.; Hettinga, B.A.; Macdonald, S.; Ferber, R. Effects of Simulated Marker Placement Deviations on Running Kinematics and Evaluation of a Morphometric-Based Placement Feedback Method. *PloS one* **2016**, *11*, e0147111, doi:10.1371/journal.pone.0147111.
32. Baker, R.; McGinley, J.L.; Schwartz, M.; Thomason, P.; Rodda, J.; Graham, H.K. The Minimal Clinically Important Difference for the Gait Profile Score. *Gait & posture* **2012**, *35*, 612–615, doi:10.1016/j.gaitpost.2011.12.008.
33. Sutherland, D.H.; Davids, J.R. Common Gait Abnormalities of the Knee in Cerebral Palsy. *Clinical Orthopaedics and Related Research (1976-2007)* **1993**, *288*, 139–147.
34. Schwartz, M.H.; Rozumalski, A. The Gait Deviation Index: A New Comprehensive Index of Gait Pathology. *Gait & posture* **2008**, *28*, 351–357, doi:10.1016/j.gaitpost.2008.05.001.
35. Widmer, M.; Staganello, M.; Sangeux, M.; Odorizzi, M.; Brunner, R.; Viehweger, E. Single Procedure Tibialis Anterior Tendon Shortening in Combination with Achilles Tendon Lengthening in Unilateral Cerebral Palsy Improves Swing Phase Dorsiflexion in Gait. *Journal of Children's Orthopaedics* **2024**, *18*, 441–449, doi:10.1177/18632521241244624.
36. Rutz, E.; Baker, R.; Tirosh, O.; Romkes, J.; Haase, C.; Brunner, R. Tibialis Anterior Tendon Shortening in Combination with Achilles Tendon Lengthening in Spastic Equinus in Cerebral Palsy. *Gait & posture* **2011**, *33*, 152–157, doi:10.1016/j.gaitpost.2010.11.002.

37. Kläusler, M.; Speth, B.M.; Brunner, R.; Tirosh, O.; Camathias, C.; Rutz, E. Long-Term Follow-up after Tibialis Anterior Tendon Shortening in Combination with Achilles Tendon Lengthening in Spastic Equinus in Cerebral Palsy. *Gait & posture* **2017**, *58*, 457–462, doi:10.1016/j.gaitpost.2017.08.028.
38. Armand, S.; Sawacha, Z.; Goudriaan, M.; Horsak, B.; van der Krogt, M.; Huenaerts, C.; Daly, C.; Kranzl, A.; Boehm, H.; Petrarca, M. Current Practices in Clinical Gait Analysis in Europe: A Comprehensive Survey-Based Study from the European Society for Movement Analysis in Adults and Children (ESMAC) Standard Initiative. *Gait & posture* **2024**, *111*, 65–74, doi:10.1016/j.gaitpost.2024.04.014.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.