

Leveraging Home Video Analysis with Large Language Models to Identify Developmental Performance Insights Beyond Standardized Assessments

Submitter Anzalee Khan

Affiliation The Nathan S. Kline Institute for Psychiatric Research

SUBMISSION DETAILS

I agree to provide poster pdf for attendee download. Yes

I have used the poster abstract template to develop my abstract. Yes

Methodological Issue Being Addressed To what extent can large language model-based analysis of ecologically valid home video data extract quantifiable developmental indicators, and how do these LLM-derived behavioral features statistically correspond with, diverge from, or augment standardized norm-referenced developmental assessment scores? Additionally, can integration of these multimodal data sources improve sensitivity to real-world functional abilities over time?

Introduction Standardized developmental assessments such as Vineland-3, Bayley-4, and CGI scales are essential tools for evaluating early childhood development but may not fully capture a child's functional abilities in naturalistic settings. Home video recordings offer ecologically valid behavioral samples that may capture subtle or context-dependent functional skills not elicited during structured assessment sessions. This study examines: (1) the concordance between structured developmental scores and LLM-derived behavioral observations from home videos; (2) the added explanatory value of naturalistic video-derived features; and (3) potential integration pathways for multimodal developmental monitoring.

Methods 41 participant home videos were reviewed from 2 pediatric/adolescent clinical trials with standardized developmental scores and home video recordings were included. Extracted standardized developmental scores (Vineland-3, BSID-4) and CGI-Improvement scale at baseline and 12 weeks post-intervention. Home videos were qualitatively coded for developmental behaviors across motor, language, cognitive, social-emotional, and sensory domains using LLMs to summarize narrative observations. The model comprehends the context and nuances of the home videos, identifying subtle developmental cues and summarizing complex behaviors into structured, domain-specific insights. This includes recognizing patterns like cause-effect understanding, imitation, social engagement, and motor coordination. The model then categorizes observed behaviors into standardized developmental domains, facilitating comparison with clinical assessment scores. Descriptive statistics and Pearson correlation coefficients were calculated between Vineland-3 and BSID-4 Growth Scale Value (GSV) change scores, and concordance assessment (agreement/disagreement patterns) between standardized scores on clinician-administered assessment and LLM-derived home video observations, and clinician developmental notes on home videos and LLM-derived home video observations. Qualitative video

categories were mapped to developmental domains for comparison.

Results Pearson correlation between Vineland-3 and BSID-4 GSV change scores was strong ($r = 0.95$, $p < 0.05$), indicating agreement between measures. Expressive communication improvements (Vineland-3 GSV +17, BSID-4 GSV +1) aligned with video observations of vocalizations, word approximations, and social engagement. Receptive communication showed declines in standardized scores (Vineland-3 GSV -20, BSID-4 GSV -22), while video narratives indicated emerging receptive skills such as following instructions, suggesting context-dependent abilities. Gross motor improvements were consistent with video evidence of walking, climbing, and coordinated play, though quantitative data was incomplete. Fine motor scores declined (-5 Vineland-3, -6 BSID-4), yet videos demonstrated fine motor skills like grasping and stacking, indicating possible assessment sensitivity issues. Social interaction improvements (Vineland-3 GSV +10) corresponded with video observations of social smiling and engagement, though some videos noted limited responsiveness. Across all mapped domains, 87% of cases showed full or partial agreement between video-derived behaviors and clinician observation/description of directionality of change. Full agreement occurred when both video observations and clinician observation/description reflected comparable improvement or decline. Partial agreement was defined as convergence in at least one subdomain (e.g., expressive but not receptive communication). Discrepant cases (23%) occurred primarily in receptive language and fine motor domains, where standardized scores declined but video behavior suggested emerging or context-specific skills. (Note: A sample size of approximately 75 will be available at the time of the conference (FEB2026)).

Conclusion LLM-based home video analysis yields contextually grounded behavioral insights that may reveal functional abilities not consistently captured by standardized, clinic-based assessments. While strong correlations between Vineland-3 and BSID-4 support their validity within structured testing contexts, home videos frequently captured subtle, emerging, or context-dependent skills that expanded developmental interpretation. Integrating automated naturalistic video analysis with standardized assessment data may enhance developmental monitoring, improve ecological validity, and strengthen sensitivity to real-world performance change. Future work will incorporate multimodal computational features (e.g., pose estimation, speech-to-text, emotion analysis) to advance scalability, objectivity, and measurement precision.

Co-Authors

Anzalee Khan¹, Stacey Eckert², Mary Seddo³, Christian Yavorsky²,
James Lefkowitz⁴, Sheraz Hussain⁵

¹ The Nathan S. Kline Institute for Psychiatric Research

² Valis Biosciences Inc

³ Long Island University

⁴ Wandering Monk Productions

⁵ Ark Intelligence

Keywords

Keywords
LLM
Home Video
Rare Disease

Guidelines I have read and understand the Poster Guidelines

Disclosures Employee: Manhattan Psychiatric Center and Nathan S. Kline Institute

Advisor: Valis Biosciences

Consulting: Clario, WCG, Sprimm