Supplementary Table S2. Option-specific exact agreement between AI models and specialty-specific expert-based reference decisions in common evaluable cases.

Option	ChatGPT	Gemini	Claude
A. O1b — Orthodontic treatment-plan modifications (19 common cases)			
1. Limitation of the amount of planned tooth movement	16/19 (84.2%)	13/19 (68.4%)	7/19 (36.8%)
2. Modification of the direction of planned tooth movement	17/19 (89.5%)	13/19 (68.4%)	15/19 (78.9%)
3. Avoidance of specific orthodontic movements	19/19 (100.0%)	19/19 (100.0%)	2/19 (10.5%)
4. Modification of force levels and/or biomechanics	11/19 (57.9%)	13/19 (68.4%)	6/19 (31.6%)
5. Modification of anchorage approach	18/19 (94.7%)	14/19 (73.7%)	9/19 (47.4%)
6. Modification or limitation of orthodontic treatment objectives	19/19 (100.0%)	15/19 (78.9%)	13/19 (68.4%)
7. Modification of extraction/tooth-retention plan	19/19 (100.0%)	19/19 (100.0%)	19/19 (100.0%)
8. Other	19/19 (100.0%)	18/19 (94.7%)	19/19 (100.0%)
B. P2b — Periodontal/mucogingival interventions (5 common cases)			
1. Free gingival graft	2/5 (40.0%)	2/5 (40.0%)	2/5 (40.0%)
2. Connective tissue graft / augmentation	4/4 (100.0%)	4/4 (100.0%)	4/4 (100.0%)
3. Root coverage	3/5 (60.0%)	5/5 (100.0%)	5/5 (100.0%)
4. Frenotomy / frenectomy	5/5 (100.0%)	5/5 (100.0%)	5/5 (100.0%)
5. Vestibuloplasty	5/5 (100.0%)	5/5 (100.0%)	5/5 (100.0%)
6. Other	5/5 (100.0%)	5/5 (100.0%)	5/5 (100.0%)

*Values represent option-specific exact agreement (n/N, %) between each AI model and the corresponding specialty-specific expert reference. Analyses were restricted to cases in which O1b or P2b was applicable (O1a = 1 or P2a = 1, respectively) for the expert reference and all three AI models. Expert reference decisions were determined by the majority of valid specialist responses, and comparisons without an expert majority were excluded. Agreement included both selection and non-selection of an option. For P2b option 2, one comparison was excluded because no expert majority was available (N = 4). Results are descriptive and should be interpreted cautiously, particularly for P2b because of the small number of evaluable cases.