Ablation Study Results: Teamwork Components Across Medical Datasets

Table 1. Accuracy comparison (%) across 8 medical datasets for baseline system and 6 teamwork configurations. Model: Gemma3 4B, n=20 questions per dataset.

Configuration (multi-agent)	MedQA (text)	MedBullets (text)	DDXPlus (text)	MMLU-Pro (text)	PMC-VQA (image)	PubMedQA (text)	Path-VQA (image)	MedMCQA (text)
Baseline	**28**	**48**	**44**	**44**	**56**	**40**	**60**	**44**
1. Shared Mental Model	30	55	40	30	55	35	65	50
2. Leadership	45	55	60	40	45	30	45	45
3. Team Orientation	40	65	75	40	50	40	50	55
4. Trust Network	25	60	65	40	45	35	50	40
5. Mutual Monitoring	40	40	40	35	45	40	60	45
6. All Components	35	60	63	37	45	40	50	50