

Supplementary Table 4: Confusion matrix of clinical expert predictions on the Visual Turing test. Each block contains four cells that correspond to the predictions of each of the four experts. The first value in the cell represents the number of images that were correctly classified. The second value in parenthesis represents the image count normalized over the true label row. The third value in parenthesis is the number of images normalized over the answered images only (excluding the images with predicted label = Unsure).

		Predicted Label from Expert			
		Real	Synthetic	Unsure	
True Label	Real	E1	75 (54%)(60%)	50 (36%)(40%)	13 (9%)(0%)
		E2	120 (87%)(91%)	12 (9%)(9%)	6 (4%)(0%)
		E3	61 (44%)(56%)	47 (34%)(44%)	30 (22%)(0%)
		E4	94 (68%)(68%)	44 (32%)(32%)	0 (0%)(0%)
	Synthetic	E1	48 (32%)(34%)	95 (63%)(66%)	7 (5%)(0%)
		E2	88 (59%)(64%)	50 (33%)(36%)	12 (8%)(0%)
		E3	49 (33%)(48%)	53 (35%)(52%)	48 (32%)(0%)
		E4	66 (44%)(44%)	84 (56%)(56%)	0 (0%)(0%)
	Unsure	E1	-	-	-
		E2	-	-	-
		E3	-	-	-
		E4	-	-	-