

one score per model (using aligned human-model cases)

