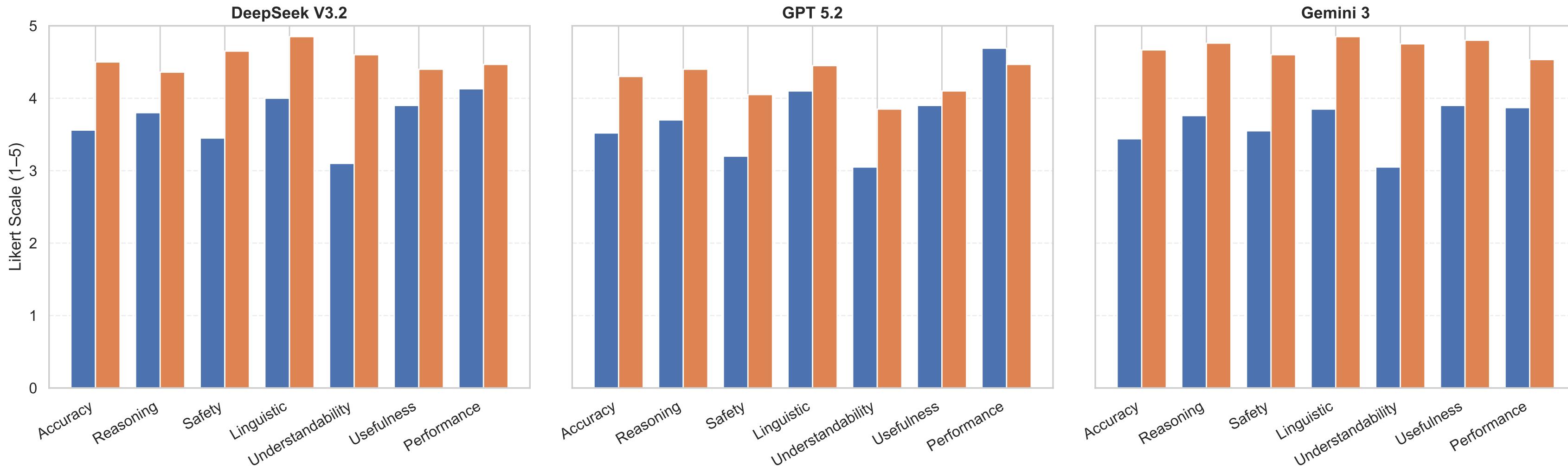Paired Validation (Model × Domain): MELMA-W vs Human Evaluators