F1 Scores by Model and Prompt Type (StandardLLM)