



Figure 1: The Pass@1 accuracy across different difficulty levels, *Easy*, *Medium*, and *Hard* (part (b)) on the HumanEval benchmark when leveraging GPT-3.5 as the LLM backbone.