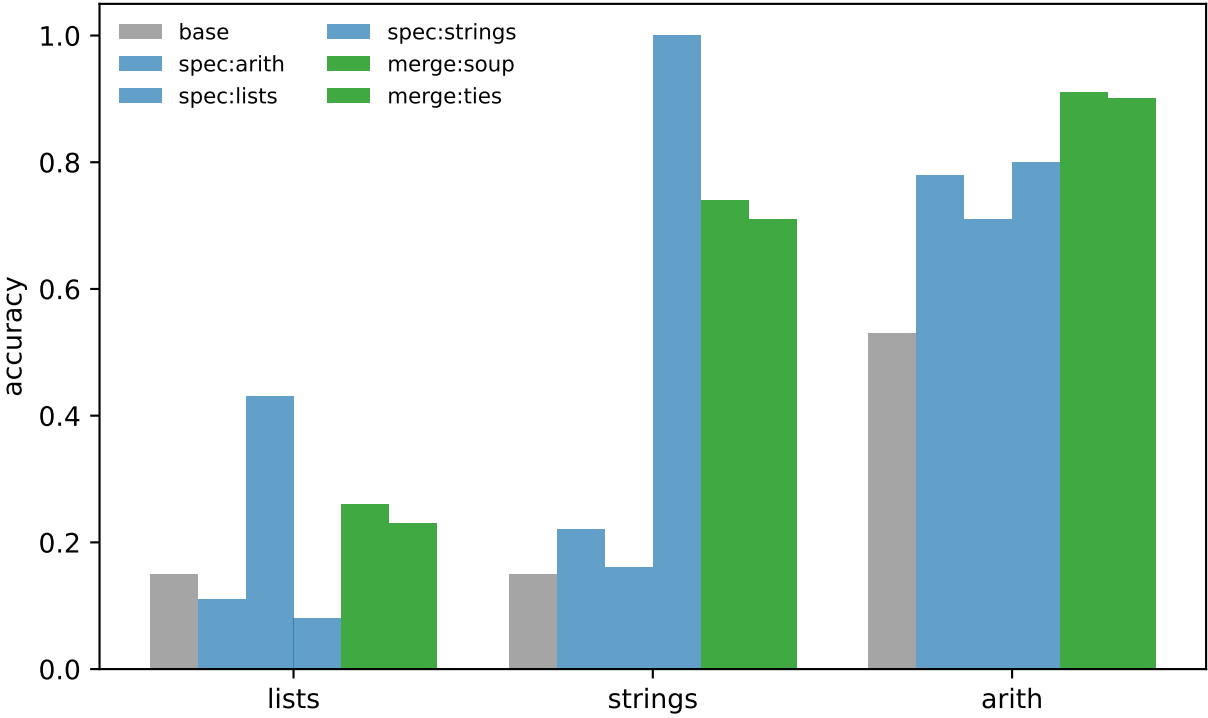


llm_merge — recombining decorrelated specialist LLMs gives the only model competent across all families (balance); overall parity (Qwen/Qwen2.5-0.5B-Instruct)

Per-family: each specialist spikes on its own family; the merges (green) are competent everywhere (but averaging dilutes some peaks)



Overall (solid) vs worst-family (hatched): the merge clearly wins worst-family (balance); overall it matches the best specialist

