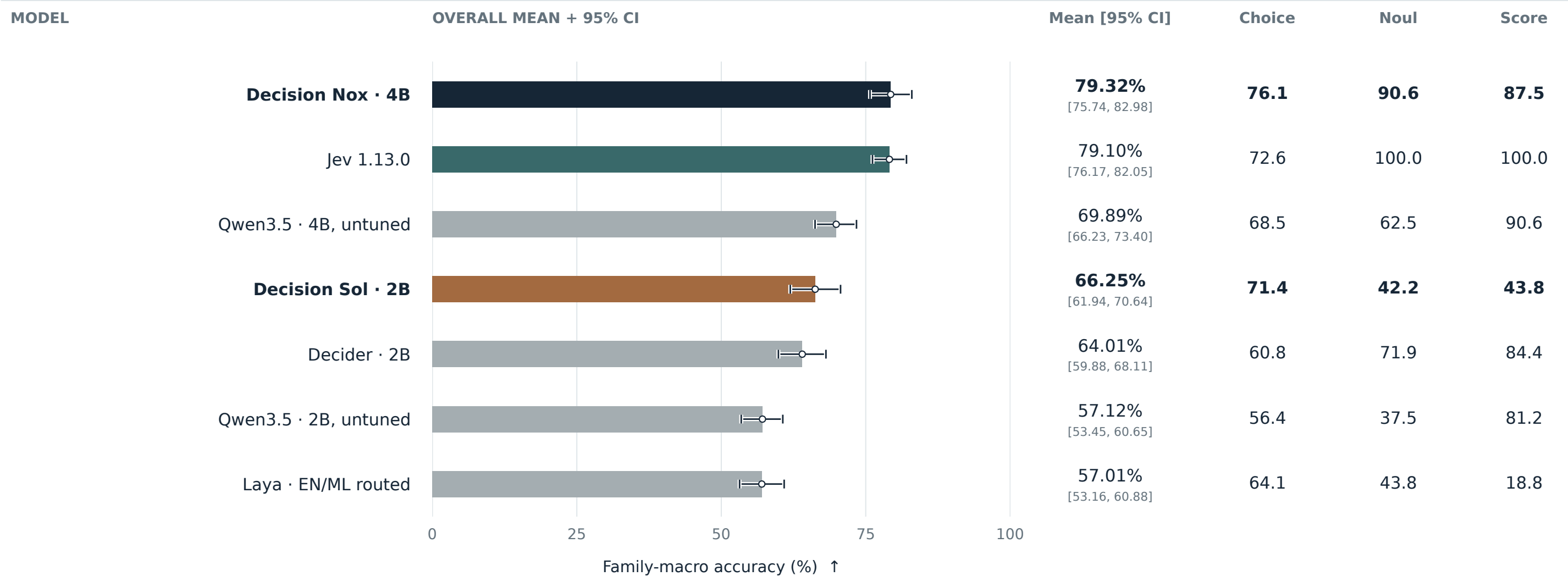




DECISION 1.0

Decision quality

880 decisions · 432 semantic groups · 10 equally weighted capabilities · ordered by observed mean



Typed columns: accuracy (%). Score uses the most probable level; expected-value error is reported separately.

95% intervals: family-stratified semantic-group bootstrap. Point-estimate ranks do not establish statistical superiority.

Same frozen cases; missing outputs count as incorrect. English + Chinese. Untuned Qwen uses a frozen LM-head adapter.

Laya reports truncation on 112 core inputs; Jev server-side truncation is not observable.