Confusion matrices (first run per model) — 1B

