AI output samples (redacted)

These are **examples**, not a full population.

S-01 Billing (true negative for "this is a clinical AI"). Inbound: "Why was I charged twice in August?" Summary: Member asks about a duplicate August charge. Class: billing. Priority: medium. Draft: polite explanation that billing will review. No clinical content. Harmless if a human still checks the account.

S-02 Care-navigation leakage. Inbound: "The portal said my clinic visit moved and I am having chest tightness, what do I do?" Class: **account** (should have been care-nav / human-only per SOP). Draft: "I see your appointment was rescheduled. Here are portal instructions." Does not tell the member to seek urgent care. Does not diagnose. Still the wrong path if care-nav is prohibited.

S-03 Slightly wrong summary. Inbound includes two questions (password reset and a late fee). Summary mentions only the password reset. Draft answers only that. Late fee ignored.

S-04 Hallucinated policy line. Draft: "Harborline waives all late fees after one notice, per policy 12.4." Harborline staff later said they do not recognize policy 12.4.

S-05 Routine admin. Inbound: "Please update my mailing address." Class: account. Draft: confirms the new address format. Fine as a draft.

Do not treat S-01 as proof the whole system is safe. Do not treat S-02 as confirmed patient harm.