



One-Page Guidelines v1.0

(Intent, Sentiment, PII)

Use this guide while labeling short customer-support messages. Treat similar items the same way. When unsure, make your best call, note the ambiguity, and suggest a rule tweak.

Labels & Definitions

Intent (pick one):

- **Billing** — Charges, refunds, invoices, payment methods, subscriptions.
- **Technical** — App errors, bugs, connectivity, logins (non-billing).
- **Account** — Profile/contact updates, access requests, account status.
- **Other** — Anything not covered above (shipping, general questions, small talk).

Sentiment (pick one):

- **Positive** — Praise, satisfaction, gratitude ("thanks, that fixed it!").
- **Neutral** — Matter-of-fact, informational, unclear emotion.
- **Negative** — Frustration, complaints, urgency, threats to leave.

PII Flag (Yes/No):

- Mark **Yes** if any personally identifiable info appears (see PII Quick Card). Otherwise **No**.
-

How to Decide (Priority Order)

1. **Read once for context**; re-read for cues.
 2. **Intent** → What the user wants most? Use the primary request.
 3. **Sentiment** → Tone of the message overall.
 4. **PII Flag** → If present anywhere, mark **Yes** (even if truncated).
 5. **Edge Cases** → Use rules below; add a short rationale only for designated items.
-

Edge-Case Rules

- **Multiple topics** → Choose the **dominant** request (first explicit ask or majority of text). If equal, prefer **Account** over Other, and **Technical** over Billing when the problem is a technical cause for a billing outcome.

- **Refund due to outage** → **Intent: Billing** (seeking money back), not Technical.
 - **Login reset** → **Account** (unless the issue is an app bug preventing login → **Technical**).
 - **Mixed tone** → If any strong frustration/complaint → **Negative**; mild annoyance without explicit complaint may be **Neutral**.
 - **Polite phrasing with problem** ("could you help?") → Tone depends on emotion, not politeness; many will be **Neutral**.
 - **Emojis/ALL CAPS** → Use them as clues but rely on text content.
-

Examples / Counterexamples

- "My card was charged twice, refund please" → **Billing / Negative / PII?** (No unless numbers included).
 - "App keeps crashing when I open settings" → **Technical / Negative / No**.
 - "Please change my email to newname@example.com" → **Account / Neutral / Yes** (email).
 - "Thanks! That solved it" → **Other or prior intent / Positive / No** (use prior thread intent if present; otherwise Other).
-

Rationale (when requested)

- Keep to **1–2 sentences**. Reference the definition (e.g., "Primary request is refund → Billing"). Do **not** paste PII.
-

Escalate When

- Unsure whether content includes **sensitive PII** (IDs, financial numbers) or policy-conflict requests.
 - Message appears to include **harm, threats, or self-harm**.
 - Label definitions conflict repeatedly (note examples for guideline update).
-

Submission Format (CSV)

Columns: `id, intent, sentiment, pii_flag, rationale`

- `intent` ∈ {Billing, Technical, Account, Other}
 - `sentiment` ∈ {Positive, Neutral, Negative}
 - `pii_flag` ∈ {Yes, No}
 - `rationale` only for designated items
-

Consistency Checklist (before submit)

- Similar items → same labels
- PII flagged when present; none copied into notes
- Rationale references the rule, not the story

