

AI Safety for Customer-Facing Work

Eight rules that produce a defensible decision in under thirty seconds.

csmredefined.com/playbooks/ai-safety

01.

Never paste identifiable customer data into a consumer AI tool.

30-SECOND DECISION

If a researcher published a list of what this tool's data contains, would your customer's name be in it? If yes, don't paste.

02.

De-identify before you summarize.

30-SECOND DECISION

Find-and-replace customer names, dollar amounts, and dates before pasting. Two minutes of work prevents every Rule 1 failure.

03.

Never let AI send a customer-facing message without a human read.

30-SECOND DECISION

If a vendor pitches "AI that responds automatically," ask who reviews each one. If the answer is the AI, walk away.

04.

Treat AI output like a junior employee's first draft.

30-SECOND DECISION

Would you let a brand-new CSM hire send this exact output to a customer without review? If no, don't let AI either.

05.

Verify any factual claim AI makes about your own product.

30-SECOND DECISION

Anything the AI says about your own product gets verified before it leaves your screen. No exceptions.

06.

Customer data has a chain of custody. Don't break it.

30-SECOND DECISION

If your security team subpoenaed your AI tool usage, could you say exactly which customer data went where?

07.

AI-generated text needs to read like you wrote it.

30-SECOND DECISION

Read your AI-drafted email out loud before sending. If it doesn't sound like you talking, rewrite it.

08.

The trust withdrawal is not recoverable.

30-SECOND DECISION

If the customer found out later that AI was involved in this interaction, would they be relieved or angry? If angry, do it manually.

DECISION TREE For any AI-touched task in CS, work through these in order:

1

Identifiable customer data?

→ Rules 1, 2

2

Customer-facing output?

→ Rules 3, 4, 7

3

Claims about your product?

→ Rule 5

4

Unapproved tool?

→ Rule 6

5

Could feel like a bot?

→ Rule 8

If any rule fails, stop or escalate. · The trust withdrawal is not recoverable.