THE BUILDER'S STACK · THE PRACTICE · HUMAN FACTORS

The Human Factors Audit

A one hour review you can run on any AI product: yours, a competitor's, or one you are about to buy. Pick one real user job in the product, walk the six stations in order, about ten minutes each, and write down named gaps as you go. The output is a list of gaps ranked by user harm, not a score.

The science behind every station is taught, with sources, in the Human Factors chapters of The Builder's Stack at ainativeproductmanager.com.

PRODUCT UNDER REVIEW

REVIEWED BY

DATE

1 Perception: Is the thing that matters impossible to miss?

The eye sorts a screen in about a fifth of a second, before conscious reading, and it catches only a few attributes: color, size, motion, position.

- [ ] Glance at the riskiest output for a fifth of a second. The warning, confidence, or source pops first, not decoration.
- [ ] The most important signal differs from its background in a strong attribute, not just in wording.
- [ ] No warning is rendered in the same style as the body text around it.
- [ ] Long output has visual breaks where the important things sit.
- [ ] At most two things compete for the first glance, and the right one wins.

Failing looks like: The model flagged its own uncertainty, in gray, mid paragraph, and nobody saw it.

NAMED GAPS AT THIS STATION

2 Working memory: Who is holding the session?

Working memory holds a handful of items. A long session quietly assumes the user's memory is as large as the model's context window, and it is not.

- [ ] After a ten step session, the user can see in one place what the system knows and what it changed.
- [ ] The plan and the working set stay visible without re-reading the transcript.
- [ ] What the model is unsure of is flagged where the user will see it.
- [ ] Undo is one step from anywhere.
- [ ] A user asked to recall the session from memory misses little, because the screen carries it.

Failing looks like: To know what the agent did, you re-read the whole transcript.

NAMED GAPS AT THIS STATION

3 Anxiety: What happens at the sharp moment?

Attention and anxiety draw from the same limited pool. The higher the stakes, the less mind the user has left, exactly when they need it most.

- [ ] Every irreversible action (sent, deleted, spent, published) is identified and listed.
- [ ] Each irreversible action has a preview, a confirm, and a way back, or a stated reason why not.
- [ ] The user can see what will happen before it happens: a plan, a dry run, a diff.
- [ ] A regular user can predict what the system will do next.
- [ ] The single most stressful moment in the product has something that lowers the stakes at that exact point.

Failing looks like: The agent sent the email while the user was still deciding.

NAMED GAPS AT THIS STATION

4 Mental models: Does the product say what it can do?

People understand new tools through the models old tools left behind. With AI the stored model is usually wrong, and a blank input shows nothing about what is possible.

- [ ] The first screen tells a brand new user what the system can do.
- [ ] The capabilities you most want found appear where a new user starts.
- [ ] Real seeded examples replace the empty box.
- [ ] Something states what the system cannot do, or where it is weak.
- [ ] The product never silently does something an expert in the domain would never do.

Failing looks like: Users ask it for the one thing it cannot do, fail, and conclude the product is bad at everything.

NAMED GAPS AT THIS STATION

5 Metacognition: Can the user tell when it is wrong?

Judging an answer takes knowledge the asker may not have, and the people most likely to rely on the answer are the ones least able to check it.

- [ ] The three most consequential outputs can each be verified in under a minute.
- [ ] The product shows what it used (sources, files read, data touched) next to what it produced.
- [ ] Confident answers and shaky answers look different.
- [ ] Checking is cheap: a source link, a diff, a one click comparison.
- [ ] Something tells the user when they are past the edge of what they can judge.

Failing looks like: Every answer looks equally sure, and the wrong ones are fluent.

NAMED GAPS AT THIS STATION

6 Supervision: Who is in charge of the agent?

Nothing supervises itself from the inside. A model cannot reliably detect its own failures, so the check has to come from outside: a human, or an explicit guardrail.

- [ ] For each autonomous behavior, someone or something can see the plan before it runs.
- [ ] Someone or something can watch the run while it happens.
- [ ] Anyone can stop the run in one step.
- [ ] External checks exist that do not depend on the model grading itself.
- [ ] The user can adjust how much autonomy they grant, rather than all or nothing.

Failing looks like: The agent ran a bad plan to completion because no one could see it running.

NAMED GAPS AT THIS STATION

The gap list, ranked by user harm

Carry the gaps from the six stations into one ranked list, the most harmful first, each named concretely. Take the top of the list into the next planning cycle, and re-run the audit after it ships.

1

2

3

4

5

A good audit produces work items, and a re-run should find the list shorter and harder each time, which means the product is absorbing the load instead of the user.