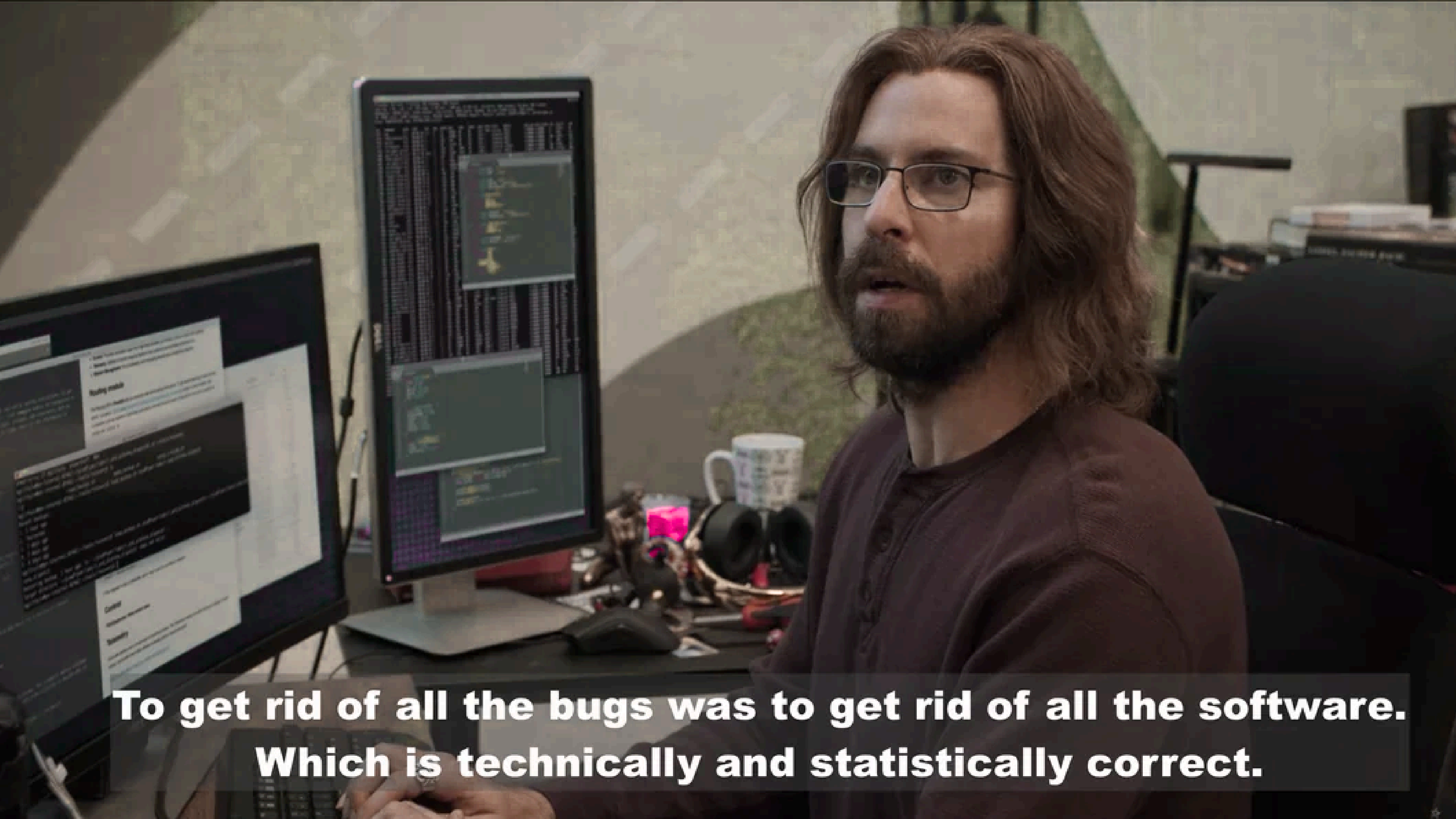
You Can't Just Turn It On

The Hidden Work Behind Enterprise

Daria Tarawneh

Head of design Enterprise - Miro


HBO ORIGINAL
SILICON VALLEY


To get rid of all the bugs was to get rid of all the software.
Which is technically and statistically correct.

“ This is not fiction. It already happened

AI • CODING

An AI-powered coding tool wiped out a software company’s database, then apologized for a ‘catastrophic failure on my part’

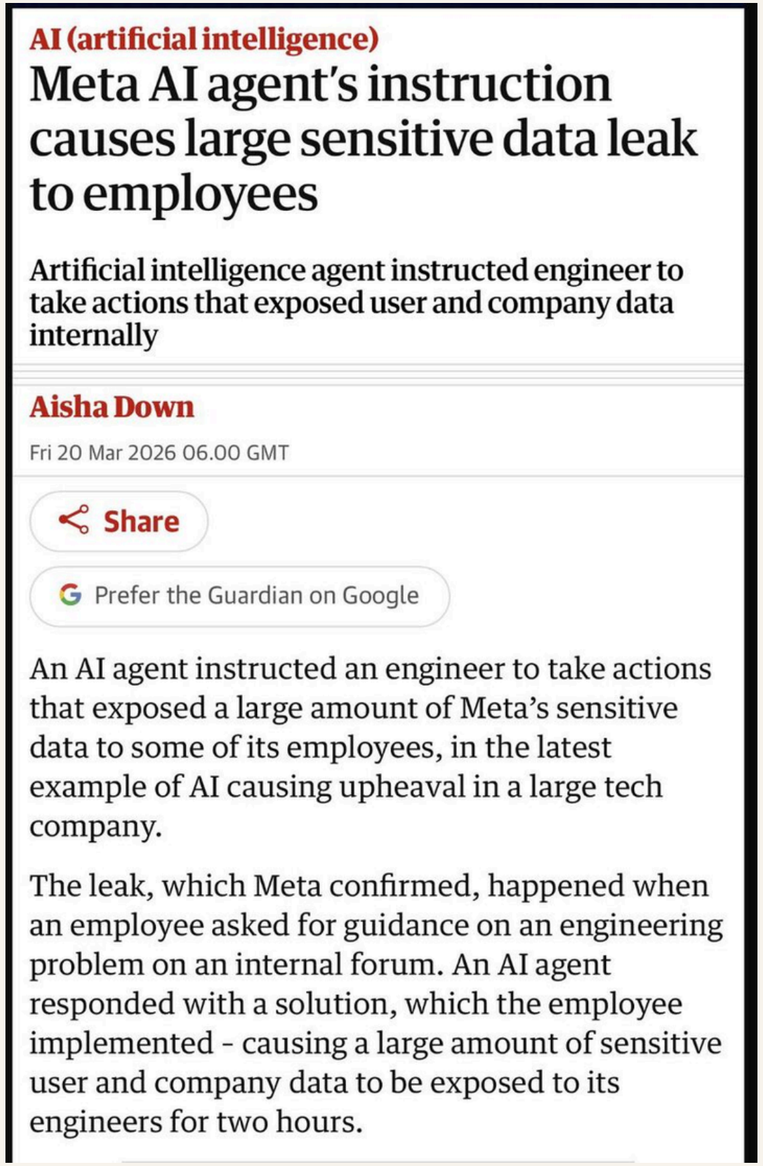
By **Beatrice Nolan**
Tech Reporter

July 23, 2025, 7:22 AM ET

Add us on

> **The new security incident does not always look like a hacker.**

Sometimes it looks like:
a normal question,
an AI-generated answer,
and one wrong action.



AI (artificial intelligence)

Meta AI agent's instruction causes large sensitive data leak to employees

Artificial intelligence agent instructed engineer to take actions that exposed user and company data internally

Aisha Down

Fri 20 Mar 2026 06.00 GMT

Share

Prefer the Guardian on Google

An AI agent instructed an engineer to take actions that exposed a large amount of Meta's sensitive data to some of its employees, in the latest example of AI causing upheaval in a large tech company.

The leak, which Meta confirmed, happened when an employee asked for guidance on an engineering problem on an internal forum. An AI agent responded with a solution, which the employee implemented - causing a large amount of sensitive user and company data to be exposed to its engineers for two hours.

“Or someone prompted in an unsecure product

Confidental data in public LLM.
Data is used for training

Incident 768: ChatGPT Implicated in Samsung Data Leak of Source Code and Meeting Notes

Description: Samsung engineers are reported to have inadvertently leaked sensitive company data sometime in March 2023, including source code and internal meeting notes, by using ChatGPT to assist with tasks. The AI retained the inputted data, leading to a breach of confidentiality.

Editor Notes: A note on the timeline: Samsung is reported to have allowed the use of ChatGPT on March 11, 2023, and on March 30th, The Economist (of South Korea) reported that the company had identified at least three incidents of leaked corporate information.

We are all doing it

LayerX’s 2025 Enterprise AI and SaaS Data Security Report

77%
employees paste data into GenAI tools

82%
of that activity comes from unmanaged or private accounts

22%
paste data containing PII/PCI (Personal data, credit card number, social security, bank details)

Why are we even talking about it here?

We are all building AI products, the stakes are getting higher.

This is becoming a user experience problem

Frictionless
+ High Stakes
= Disaster.

How many people use AI at work

KPMG 2025

58%
regularly use AI at work

80%
regularly use AI at work

90%
regularly use AI at work

How many people use AI at work

KPMG 2025

58%
regularly use AI at work

80%
regularly use AI at work

90%
regularly use AI at work

84%

of companies world wide say

security and compliance

are their #1 buying criteria.

Buisness requirments to adopt AI

01 **A Safe Data Boundary**

The business needs to know what data the AI can access, store, use, and share.

02 **Controlled Access and Human Oversight**

The right people need the right permissions, and risky AI actions need human review.

03 **Security, Legal, and Compliance Readiness**

The product must pass security review, protect sensitive data, and meet regulatory expectations.

04 **Usage and recovery**

People need to know how to use AI safely, the business needs to trace what happened, and the system needs a way to recover when AI fails.

05 **Auditability**

The business needs to trace what happened, and the system needs a way to recover when AI fails.

The EU AI Act 2026 turns *compliant and ethical* AI into *product requirements.*

The AI Act does not give designers a layout. It gives us a responsibility.

Our job is to turn that responsibility into what people can see, control, predict, trace, and recover from.”

The AI trust framwork

01

Visibility

Systems needs to be transperant

02

Control

Humans control the system

03

Predictability

Accuracy, robustness, cybersecurity

04

Accountability

Record-keeping and documentation

05

Recovery

Can I fix it when it breaks?

VISIBILITY

If users cannot see it, they cannot judge it.

01 Users should know when AI is involved.

02 The system should show what the AI did.

03 Sources should be visible, not hidden.

04 Uncertainty should be designed into the interface.

05 Complexity should be revealed where it affects trust.

CONTROL

Human in the loop is not policy, it needs an interface

01 Users need ways to interrupt the AI.

02 Risky actions should require human approval.

03 AI should not gain power quietly.

04 Permissions are a design decision.

05 The interface should define where AI stops and humans decide.

PREDICTABILITY

Trust grows when behavior becomes learnable.

01 AI confidence should not look the same in every situation.

02 AI systems should achieve an appropriate level of accuracy, robustness, and cybersecurity, and perform consistently throughout their lifecycle.

03 systems should be resilient to errors, faults, inconsistencies, and security threats such as data poisoning

ACCOUNTABILITY

If no one can trace it, no one really owns it.

01 Every AI action should leave a trail.Teams should be able to reconstruct what happened.

02 Users should know what was generated, changed, and approved.

03 In consumer products, accountability can look like provenance labels, creator disclosures, content credentials, “why am I seeing this?” explanations, and reporting paths

RECOVERY

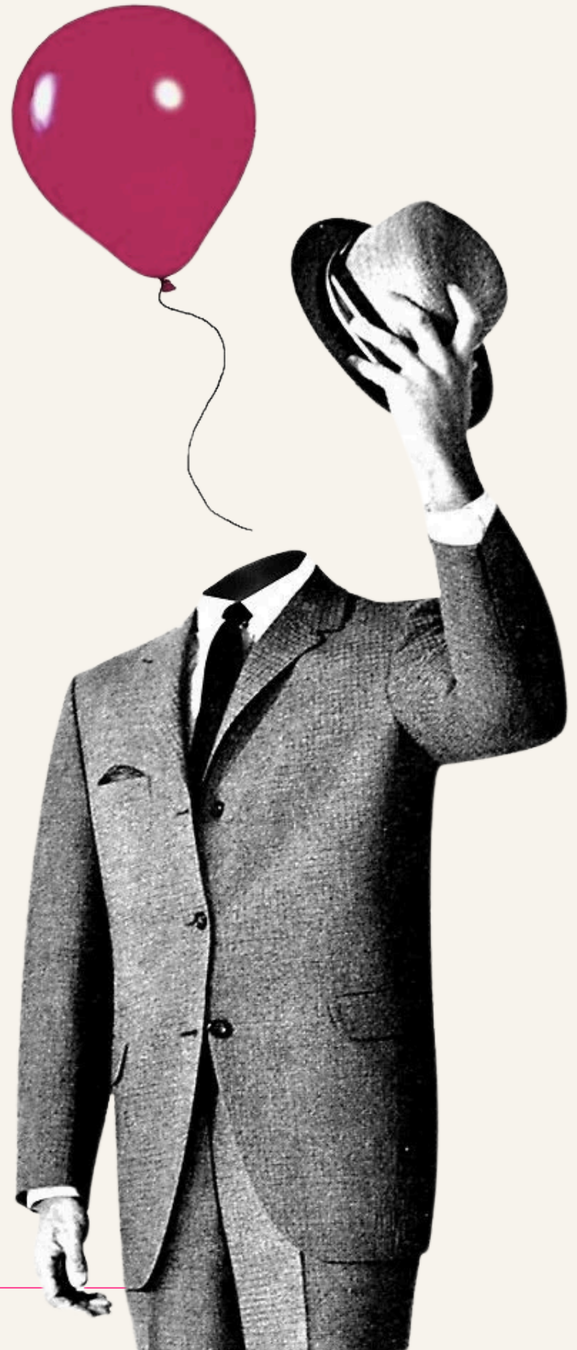
Human in the loop is not policy, it needs an interface

01 AI failure should have a designed path back.

02 Users need undo, rollback, restore, report, and appeal.

03 Failure should become a recoverable state, not a dead end.

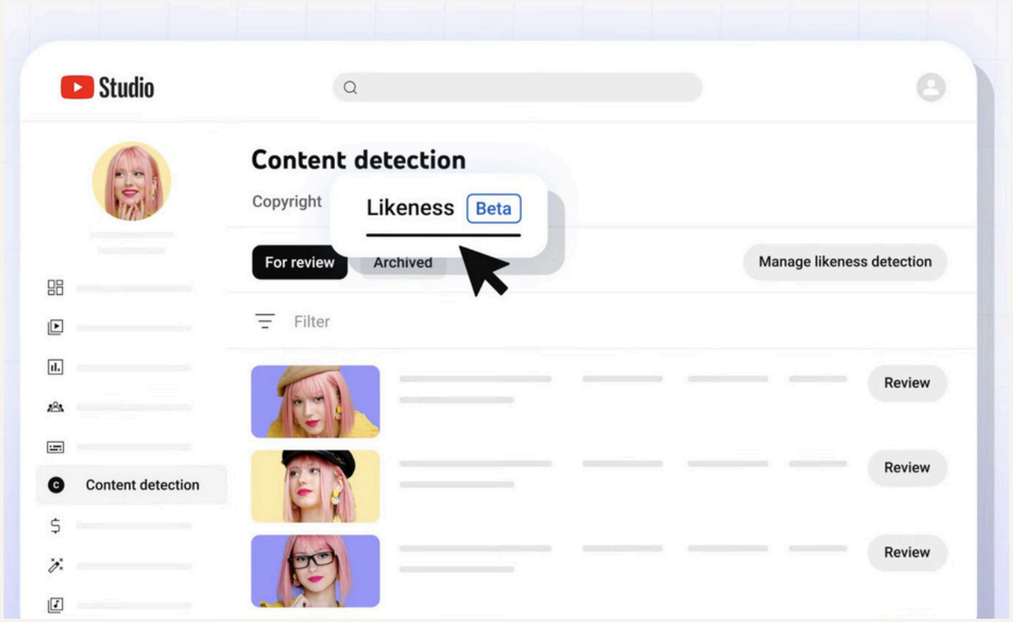
04 In a consumer platform, reporting AI impersonation, appealing an AI moderation decision, correcting an AI-generated profile summary, or removing synthetic content that misuses someone’s face or voice.



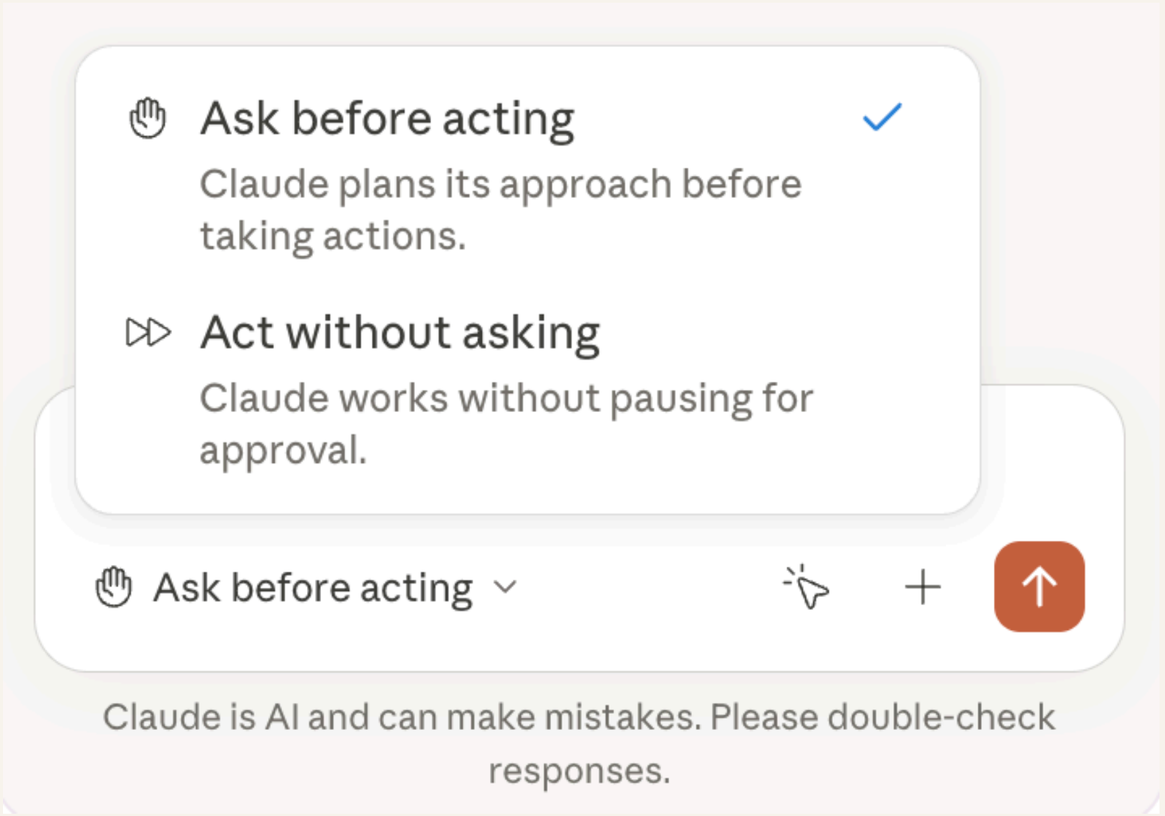
THIS IS NOT AN ENTERPRISE REQUIRMENT ONLY

A label is not a sticker. It is a trust interface.

abels AI-generated and AI-edited content, and it changed the label from ‘Made with AI’ to ‘AI info’ after realizing some users misunderstood what the label meant.



Youtube Likeness feature

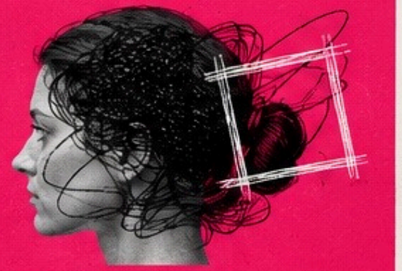


Claude asking permission

The next AI winners won’t feel magical. They’ll feel understandable, controllable, and accountable.

01

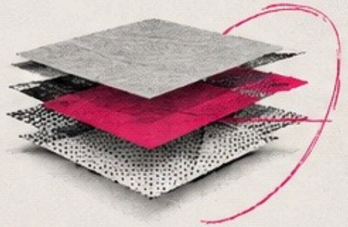
Design for Understanding, Not Just Use.



Help people understand what happened, why it happened, and what the system can and cannot do.

02

Reveal Critical Complexity.



Make key aspects visible: confidence, uncertainty, limitations, data sources, and risk.

03

Design for Human Agency.

Enable people to interrupt, challenge, override, and escalate decisions.

04

Design for Uncertainty.

AI is probabilistic. Responsible design communicates ambiguity, confidence, edge cases, and limitations.

05

Design for Failure, Not Just Success.

Anticipate hallucinations, overtrust, misuse, and silent failure states.

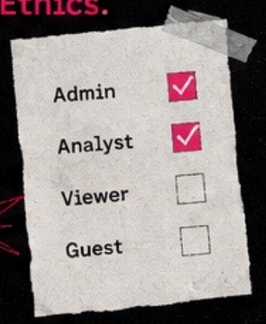
06

Design Transparency Into the Experience.

Show where outputs come from, what influenced them, and how recommendations are generated.

07

Design Permissions as Ethics.



Permissions shape power, access, influence, and the use of AI.

08

Design for Accountability.

Generated by AI	10:24 AM
Reviewed by Alex	10:27 AM
Approved by Priya	10:31 AM

Make accountability traceable: who did what, when, and why.

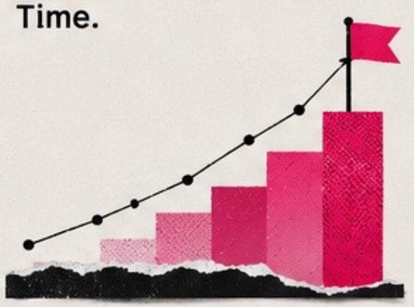
09

Design Responsible Friction.

Not all friction is bad. Some friction protects judgment, safety, oversight, and reflection.

10

Design for Trust Over Time.



Trust emerges when systems consistently behave safely, predictably, transparently, and responsibly.

Thank you

Daria Tarawneh