CASE STUDY / TIER 1 / PERSONAL RESEARCH

Project Heer.

What I learned shipping an AI persona — and why I killed half of it.

AUTHOR
Simrat Bath
Founder, NS Studio LLC

WINDOW
July – September 2025
Pre-FLUX.2, pre-Nano Banana

STACK
DALL-E · Gemini · Flux
ElevenLabs · Replicate

A six-page field report on what happens when you actually deploy an AI persona in public. For F500 leaders deciding whether to ship one.

01

WHAT I BUILT

Two Instagram channels, custom-trained models, public deployment.

Between July and September 2025 I designed and ran Project Heer: a self-funded study of AI persona deployment, trust dynamics, and creator safety. I trained custom image and voice models, built two distinct Instagram personas, posted them publicly, and watched what happened.

The two tracks

Track A — Fashion brand marketing channel. A digitally-fabricated model for a stylized fashion label. The premise: can an AI persona function as a low-cost marketing channel for a real product company?

Track B — College-life persona channel. A second persona running parallel, set in a college-life context, with a different audience and a different disclosure posture. The premise: how do audiences engage when AI content competes with the lifestyle-creator economy on its own terms?

The stack

Image generation: DALL-E, Gemini, and Flux, with custom-trained checkpoints on Replicate to enforce identity consistency across hundreds of frames. Voice and audio: ElevenLabs for cloned and synthetic voiceover. Compositing and continuity logic: hand-built, because nothing off-the-shelf preserved character identity at the fidelity I needed.

This work predates the November 2025 release of FLUX.2, the consumer launch of Nano Banana, and Midjourney's web interface. Every output required substantially more prompt engineering, scaffolding, and hand-correction than the same work would today. The dates are the credibility signal — they are publicly verifiable.

02

WHAT IT TAUGHT ME

Three findings buyers should care about.

I went in expecting an engineering experiment. I came out with a research finding every brand and platform team will eventually have to answer for.

Finding 1 — AI disclosure changes audience behavior.

When viewers became aware (or strongly suspected) that an account was AI-generated, engagement patterns shifted measurably. The disclosure variable — explicit, implicit, or absent — turned out to be one of the most consequential design decisions in the entire system. It is not a marketing-copy detail; it is the primary lever that determines whether AI content reads as creative work or as deception. Most brand teams currently treat disclosure as a legal afterthought. It should be treated as a product decision with a product owner.

Finding 2 — Gender dynamics surface within days, not months.

The feminine-presenting persona received a different inbound DM pattern than the neutral channel within a week of launch — including content that no platform moderation pipeline currently filters at the inbound message layer. This is a creator-safety gap that platforms have not addressed and that brands are structurally unprepared for. If a F500 brand deploys a feminine AI spokesperson, the team running it will absorb the same harassment patterns human creators already report — but with no creator-safety protocol on the books.

Finding 3 — Solo operators have no tooling to defend themselves.

Once the channels were live, I had no good way to triage incoming risk. Platform tools assume either a corporate moderation team or an individual user; nothing exists for the solo operator running an AI persona at scale. This is a gap that will show up the moment a single brand-side employee is asked to manage an AI creator account end-to-end. It is also a gap that will be filled by tooling built in the next twelve months — the question is whether your team builds, buys, or imports the assumptions of whatever they buy.

03

WHAT I'D TELL A TEAM LAUNCHING AN AI PERSONA TODAY

Five decisions I'd make on day one.

If a F500 marketing or product team handed me a brief tomorrow that said "we want to ship an AI spokesperson," here is the order I would force the conversation in.

01 — Decide the disclosure posture before you generate one frame.

Explicit, implicit, or undisclosed. Each posture has a different audience, a different legal surface, and a different failure mode. Picking it after launch is the most expensive way to make this decision.

02 — Assume harassment is part of the operating cost.

If your persona presents as feminine, budget for moderator hours and a defined escalation path before launch. Treat the inbound channel as a hostile environment by default. The team you assign to manage the account should not be the team that absorbs the harassment unsupported.

03 — Define what you would not let the persona do.

Not what it can do — what it must refuse. Endorse a financial product? Take a political position? Respond to a specific user category? Write the refusal list before you write the brand voice guide. If you can't write the refusal list, the brand isn't ready.

04 — Pick the kill criteria up front.

Decide, in advance, what numbers or events would cause you to take the channel down. Engagement floor, complaint volume, reputational incident, regulatory shift. Killing a live AI channel under crisis is ten times harder than killing it against a pre-agreed threshold.

05 — Build the oversight before the persona scales.

Who reviews output? On what cadence? With what authority to pause? An AI spokesperson is a system you do not fully control — that is what makes it useful and what makes it risky. The oversight layer is not optional. It is the product.

04

WHAT THIS MEANS FOR AI OVERSIGHT AT SCALE

The pattern repeats across every AI deployment, not just personas.

Heer is a small experiment. The pattern it surfaces is not. Every F500 team I talk to is being asked to deploy AI systems whose behavior they cannot fully predict, into contexts they cannot fully control, on a timeline shorter than their compliance and trust functions were designed for. The same three pressures that showed up in a two-month solo experiment show up — amplified — in any real-world AI deployment:

Disclosure decisions become product decisions. Whether the system says "I am AI" — and how, and when — stops being a copy choice and starts being a core design parameter. Teams that don't own this explicitly inherit the default their model vendor chose, which was optimized for a different audience than yours.

Hostile inputs arrive faster than your moderation pipeline. Whatever harassment, jailbreak, or adversarial pattern exists in the wild reaches your deployment within days of launch. Plan for it as a baseline cost, not as an incident.

Tooling lags deployment by 12–18 months. The infrastructure you need — monitor, audit, escalate, pause — does not yet exist. Either you build it, buy a partial solution and harden it, or absorb the gap. Pretending the gap isn't there is the most expensive choice.

> *"The disclosure variable turned out to be the most consequential design decision in the entire system. Brands treat it as legal copy. It should be a product decision with a product owner."*

This is the work I do at NS Studio. I help F500 teams answer three questions: which AI products to build, which to kill, and how to ship the rest. If your team is deciding whether to deploy an AI persona, agent, or any AI-facing system into a context your moderation stack wasn't built for, this is a thirty-minute conversation worth having before the first sprint.

BOOK A CALL

Simrat Bath · 30-minute intro
calendly.com/simratbath/30min

READ MORE

NS Studio LLC · advisory
nsstudiollc.com · simratbath.com