PD-with/for-AI

Framework and Lessons for Responsible Use of AI-Generated Synthetic Personas

HAXVIG · D'ANDREA · TELI

UNIVERSITY OF TRENTO · AALBORG UNIVERSITY

A Critical Crossroad

Commercial tools now offer LLM-generated synthetic personas — user research *"without the user."* This paper synthesizes **four studies** to ask: how can Participatory Design responsibly govern these tools?

PD → AI

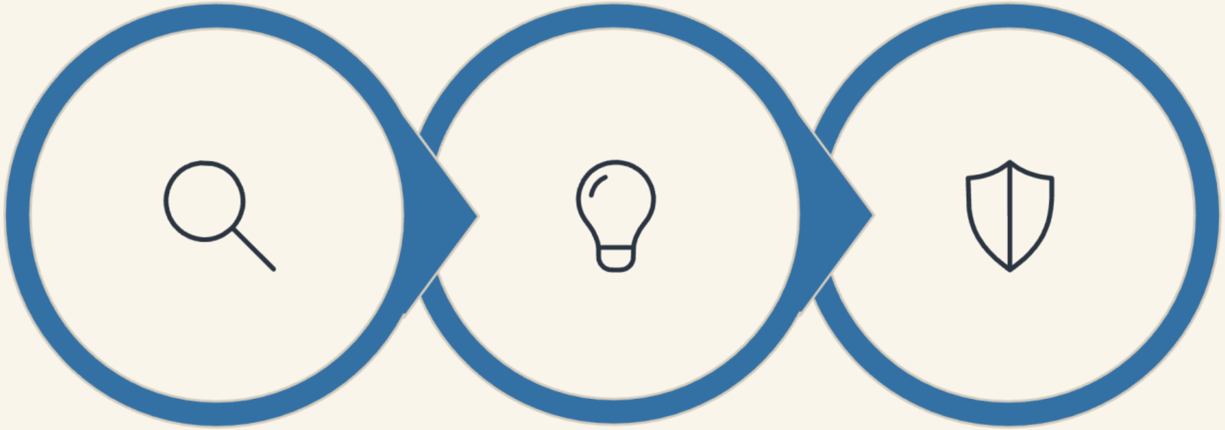
PD evaluates AI through lived experience, surfacing bias that metrics miss

AI → PD

GenAI augments co-creation — but only if PD designers act as stewards

The Coupled Path

The authors argue for a **reflexive, recursive design loop** between PD and GenAI — not a simple feedback cycle, but a relationship governed with care.



PD surfaces bias

Insights inform tools

PD validates tools

PD positions itself as both **shaper and steward** of generative AI in design.

STUDIES 1 & 2

LLMs Default to Bias

Across all four LLMs tested, personas defaulted to **traditional gender roles**:

Male personas

Assigned technical/leadership roles; described as "confident," "ambitious," "strong"

Female personas

Placed in caregiving/creative roles; described as "kind," "friendly," "empathetic"

Subtle framing

Same interest (e.g. sustainability), different framing: women act through "public demonstrations," men through "scientific achievement"

STUDIES 3 & 4

Speculative Power: Personas as Provocations

Co-Created Queer Personas

LGBTQ+ participants validated synthetic personas. These were then used by a CEO — enacted live by ChatGPT — to challenge a technological interface's design for inclusivity.

"Thinking Partners"

Interactive personas created brief **"pauses" for reflection**, helping participants notice overlooked perspectives and articulate inclusion concerns concretely.

Queer, Fictional & Historical

When testing three different persona groups (developed differently) across five tech scenarios, the . Community-validated personas were perceived as more credible and actionable than off-the-shelf characters.

The Believability Hazard

> Treating a convincing simulation as semantic truth is a **"category error"** that risks silencing the very marginalized voices PD aims to empower.

Fluency ≠ Authority

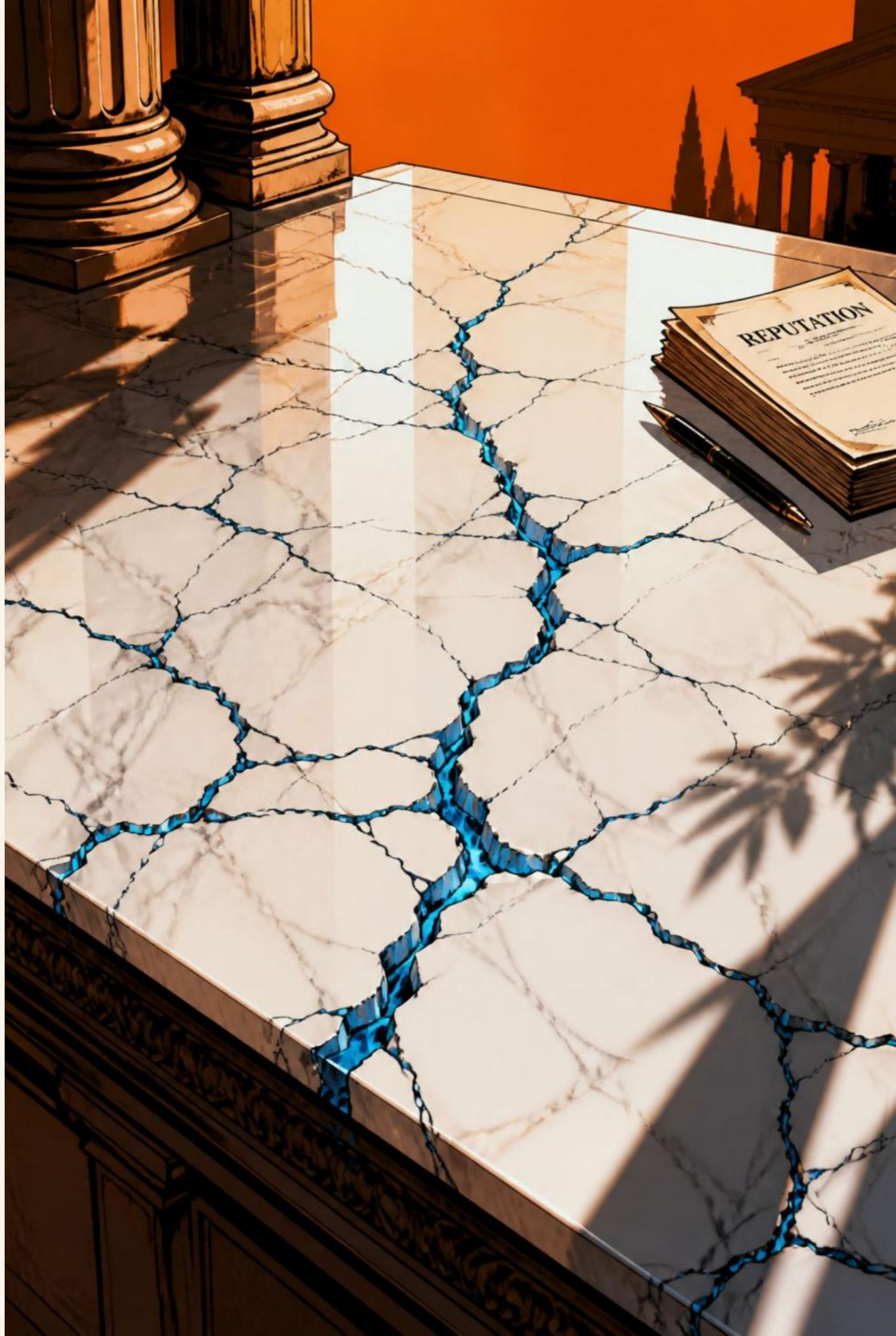
LLMs produce polished prose through statistical patterning, not situated understanding (Bender et al., "stochastic parrots")

Normalization Risk

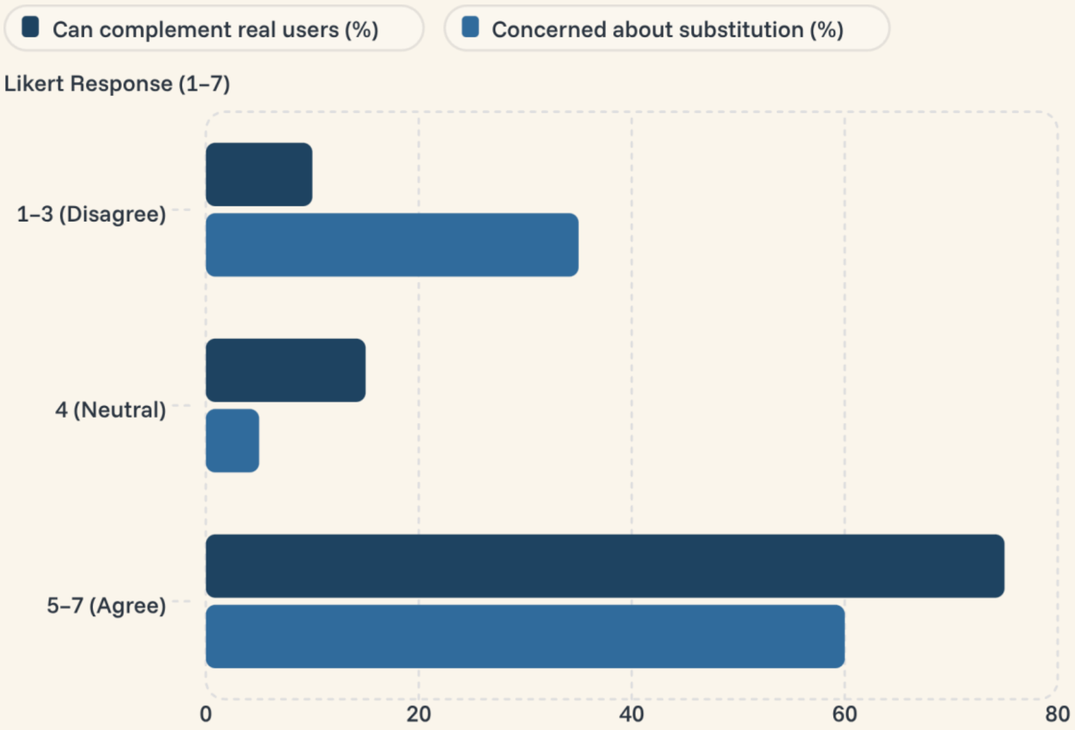
Convincing narratives can flatten differences and normalize dominant framings, especially for marginalized positions

Mandatory Governance

Where a persona came from, how it was constructed, and who reviewed it **must be explicit**



Non-Substitution Principle



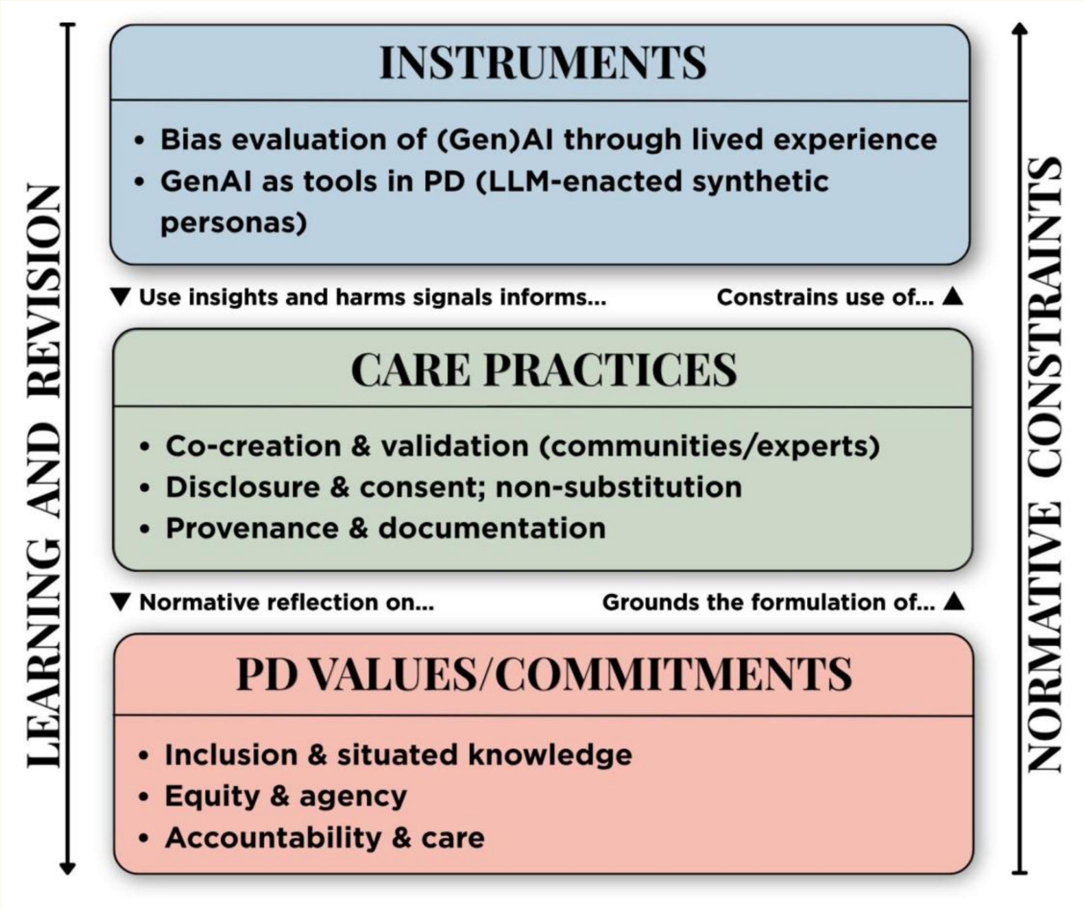
Study 4 participants (n=20) could imagine personas as a **complement**, but expressed **strong concern about substitution**.

Non-Substitution Principle: Synthetic personas are adjuncts that open dialogue and prepare inquiry; they do *not* justify design decisions nor replace engagement with affected people.

Where participation is hindered, personas may scaffold forthcoming engagement — *only* if co-created and locally validated.

The Bidirectional Layered Model

PD values ground the work, care practices operationalize those commitments, and instruments (like synthetic personas) are constrained by both. **Learning flows upward;** normative constraints flow downward.



Care Practices in Practice

→ **Participant Care**

Warn about potential exposure to offensive content; ensure voluntariness with real right to decline, skip, or withdraw

→ **Boundary Conditions**

In high-stakes settings, use personas only to open dialogue — never to justify decisions. Run local validation; document provenance

→ **Model Agnosticism**

Transferability lies in the *process*, not specific prompts. Document model, version, prompts, and validation steps

→ **Human-Only Fallback**

If any care practice cannot be met — use a human-only approach

People, Not Proxies

> Synthetic personas should be considered **governed adjuncts**: they can open dialogue and rehearse workshops, but can never justify a design decision.

Key Contribution

A bidirectional model positioning PD as both **shaper and steward** of GenAI

Core Hazards

Bias and believability require embedded care practices, not optional add-ons

Future Work

Longitudinal deployments and mixed evaluations keeping legitimacy anchored in those affected