

Appearing Legitimate is Not Enough: Interrogating Synthetic Agents in Representational Processes through a Participatory Design Lens

Aditya Nayak^{*1}, Aditi Vashistha^{*2}, Alissa Centivany², Aakash Gautam¹

¹University of Pittsburgh

²FIMS, Western University

aditya.nayak@pitt.edu, avashi4@uwo.ca, acentiva@uwo.ca, aakash@pitt.edu

Abstract

Synthetic agents built atop LLM-based foundation models are gaining popularity as substitutes for human participants across research contexts, including user-testing, market-research, computational social science, surveys, and qualitative research. We are also witnessing an extension of synthetic agents into experimental implementations of policy consultation, jury deliberation, humanitarian diplomacy, and similar contexts where human participation and representation are central to the perceived legitimacy of the institutional processes. The value of participation extends beyond informational contributions and consensus generation; participation is a necessary, legitimizing condition for democratic political institutions and processes. Treating synthetic agents as human substitutes raises serious political, representational, and ethical concerns. Participatory Design's modes of engagement — probing, priming, understanding, and generating — offer helpful tools for engaging with representational questions of personhood. We apply the lens to three case studies of synthetic agents substituting for personhood at varying representational scales: local policy, enterprise jury deliberation, and global diplomacy. We argue that legitimacy and personhood are integral and mutually constitutive while identifying the ethical, representational, and methodological risks of using synthetic agents in representational processes. We conclude by proposing soft and hard boundaries for designing oversight on LLMs and synthetic agents in representational processes.

Introduction

Large Language Models and agents built on top of the foundational models are increasingly being deployed as substitutes for human participants in domains like user-testing (Yang 2026), market-research (Simile.ai 2026), computational social science (Wang, Morgenstern, and Dickerson 2025), surveys (Ball et al. 2025), and qualitative research (Nguyen and Welch 2026). Synthetic agents, that is, LLM-based systems trained or augmented with domain-specific data to mimic context-bound personas, are now being extended into experimental implementations of policy consultation (Reshef Kera et al. 2024), jury deliberation (Ersoz 2026), and humanitarian diplomacy (Albrecht 2025). Their

proponents claim these systems can perceive contextual environments, make decisions, and take actions in ways that approximate human decision-making.

The use of synthetic agents is often justified as a solution to the difficulty of recruiting suitable participant groups. One could argue that in low-risk domains, like market research, the stakes of this substitution are limited to participant representativeness and data quality (Maden 2026; Su 2026) for building better products. While that in itself is contentious, the use of synthetic agents in domains like policy making, diplomatic consultation, and legal juries is alarming. These are deliberative and representational processes where participation and representation extend beyond informational contribution and consensus generation. We use representational process to mean a deliberative or decision-making process whose legitimacy is procedural. It derives from the inclusion and participation of those affected by the decision instead of the correctness of the outcome (Habermas 1996; Young 2000). Participation is a necessary and legitimizing condition of the democratic process, and importantly, it is fundamental to the institutional legitimacy these processes claim. Substituting synthetic agents for human participants in these contexts raises political, representational, and ethical concerns, including the exclusion of the lived experiences of the marginalized groups who would otherwise have been participants.

The concept of personhood has been made and remade whenever societies need new vocabulary for new circumstances (Leibo et al. 2025). But most conceptions of personhood in moral and legal domains rest on consciousness, intentionality, the capacity for reciprocal moral relations, and responsibility and accountability for the consequences of one's actions (Baeyaert 2025; Bronner-Martin 2025). Synthetic agents do not possess these properties, and prior work argues that this limitation cannot be overcome through technical fixes (Bronner-Martin 2025). Despite this, synthetic agents are being deployed in roles that presuppose them.

In this paper, we investigate how synthetic agents come to appear legitimate as substitutes for human participants in representational processes. We approach this through three case studies at different scales, namely local policy (Ana, a community-voice chatbot), enterprise legal deliberation (Synthetic Juror), and global diplomacy (Ask Amina and Ask Abdalla). To analyze these cases, we draw on Sanders

^{*}These authors contributed equally.

Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

et al.’s framework of Participatory Design (PD), which identifies four modes through which a participant is produced from a person, namely, *probing*, *priming*, *understanding*, and *generating* (Sanders, Brandt, and Binder 2010). These modes give us process-oriented perspectives to evaluate what representational work the synthetic agents are doing, and what they leave out. We ask two research questions: **RQ1**: How do synthetic agents in representational processes construct an appearance of legitimate participation? **RQ2**: What boundaries should guide the design and deployment of synthetic agents in representational processes?

Across the three cases, we trace a four-step process through which personhood is manufactured. First, through the framing of the problem. Then, through the institutionalized curation of data. This is followed by the design of user encounters with the persona. Finally, through the technical evaluation of validity. We argue that these four steps collectively produce *manufactured personhood*, an appearance of legitimate participation that bypasses the representational processes from which institutional legitimacy is derived. We contrast our position against existing evaluation frameworks for synthetic agents, such as *believability* (Xiao et al. 2024), *faithfulness* (Maden 2026; Agarwal, Tanneru, and Lakkaraju 2024), *plausibility* (Agarwal, Tanneru, and Lakkaraju 2024), and *algorithmic fidelity* (Argyle et al. 2023), all of which cannot detect this bypass because they measure similarity of outputs rather than the integrity of the underlying processes.

With this paper, we make three contributions. First, by examining three case studies, we offer a four-step account of how synthetic agents manufacture personhood across representational scales. Second, we use the PD modes of engagement to develop a process-oriented critique of synthetic agents, complementing the outcome-centric evaluations that dominate the existing literature. Third, we propose desiderata for design consisting of *soft boundaries* that separate analytic and indexing uses of LLMs from anthropomorphized interfaces, and *hard boundaries* that categorically separate anthropomorphized interfaces from systems that substitute for human participants in representational processes. We make these proposals at a moment when synthetic agents in representational contexts remain experimental, and the boundaries are still tractable.

Related Works

Synthetic Agents as Research Participants

The use of synthetic agents as substitutes for human participants raises a set of interrelated concerns. These include positionality, validity, epistemic and ethical risks, and the boundaries of knowledge production (Alvarado Garcia et al. 2025; Wang, Morgenstern, and Dickerson 2025; Nguyen and Welch 2026; Agnew et al. 2026; Jowsey et al. 2025; Maden 2026). Scholars have cautioned against using synthetic agents as wholesale replacements for human participants (Arawjo 2026), and a recent ACM workshop (Agnew et al. 2026) has begun to consolidate standards and documentation practices for LLM use as simulated research participants (Su 2026; Kuric, Demcak, and Krajcovic 2026; Kim, Yun, and Hyun 2026; Sampson, Lurie, and Metaxa

2026; Maden 2026; Yang 2026).

A recurring critique concerns how synthetic agents handle social identity. To represent humans, LLMs would need to capture the influence of positionality, including the relevance of social identities like race and gender (Wang, Morgenstern, and Dickerson 2025). In practice, they fail, for example, with misportrayal and category error (Wang, Morgenstern, and Dickerson 2025; Nguyen and Welch 2026), group flattening, and identity essentialization (Su 2026). Since these systems are trained on data dominated by majority populations, they gravitate toward median experiences and reject tail cases, marginalizing the populations who already are disproportionately and negatively impacted by technology (Su 2026; Smith et al. 2025), and producing cultural scarcity (Alvarado Garcia et al. 2025). Models reduce complex demographic dynamics, such as systemic inequality produced through discriminatory policies, historical neglect, or economic exclusion, to mere demographic proxies (Maden 2026). Simulation results can appear believable on the surface while the underlying logic, that of treating identity as a collection of inherent traits, is inaccurate (Maden 2026; Su 2026).

A second strand of critique concerns embodiment. Synthetic agents lack the embodied lived experiences that human participants bring to research (Agnew et al. 2026), and in spatial usability evaluation they perform poorly on embodied experience criteria (Kim, Yun, and Hyun 2026). Their implementations often treat surface-level markers of human realism as evidence of behavioral or experiential validity (Su 2026).

Problems in evaluation follow naturally from these concerns. Human participation evaluation centers on identity verification grounded in lived experience. LLMs have no lived experience to verify, and the task shifts from verification to characterization, such as ‘does this simulated entity faithfully reflect properties of the population the researcher intends to study?’ With human participants, data quality is assessed through depth, richness, and experiential specificity, none of which can be automatically measured (Maden 2026). In this space, scholars have proposed alternative evaluation frameworks: *believability*, defined through metrics of consistency with human-like output and robustness under disturbance (Xiao et al. 2024); *faithfulness* (Maden 2026; Agarwal, Tanneru, and Lakkaraju 2024); *plausibility* (Agarwal, Tanneru, and Lakkaraju 2024); and *algorithmic fidelity* (Argyle et al. 2023). However, these frameworks are all output-centric. They do not examine the processes that lead to those outputs. Beyond evaluation, prior work identifies ethical concerns around retrospective consent and the moral alignment of models (Agnew et al. 2026), and epistemic risks including category error, misattribution, the *oracle effect*, and anthropomorphic fallacies (Nguyen and Welch 2026; Stark 2024; Maeda and Stark 2025; Maeda and Quan-Haase 2024). While some work examine legal implications of agents in e-commerce (Riedl and Desai 2025), and socio-political implications of “agency-washing” (Timaite and Castelle 2025), empirical studies of application of synthetic agents as participants in representational processes remain underexplored, and importantly, the high stakes of pol-

icy, juries, and diplomacy make this gap consequential.

Personhood and Legitimacy

The English word “*person*” is derived from the Latin word *persona* which refers to the mask worn by an actor portraying a character within the context of a play (Gunkel and Wales 2021). There is no fixed concept of personhood; it is made and remade whenever societies need vocabulary for new circumstances (Leibo et al. 2025).

Legal literature distinguishes natural and legal persons. A natural person has intrinsic qualities like consciousness (Leibo et al. 2025) and is natural by virtue of being born. This personhood follows from birth, not from legal decree. Corporations and environmental entities are legal persons by decree, artificial and non-natural (Gunkel and Wales 2021; Leibo et al. 2025; Baeyaert 2025). Legal personhood rests on recognition as a subject of law, possessing both rights and responsibilities (Gunkel and Wales 2021), and is historically a pragmatic instrument of governance rather than a recognition of moral worth. Its extension to non-human entities is justified on three grounds (Baeyaert 2025): a *rights-based* model grounded in moral subjectivity (self-awareness, sentience, the capacity to experience harm, the ability to participate in reciprocal moral relations); a *functional* model where legal status is extended as a juridical tool for coherence, welfare, or governance efficacy; and an *agency-based* model grounded in intentionality, goal formation, and bearing responsibility for the consequences of action.

Across all three grounds, AI fits poorly. Even artificial entities like corporations are embedded in networks of human control, whereas AI systems can produce outcomes unforeseen by their developers (Mayne et al. 2025). Moral action, in particular, requires consciousness, intentionality, and the capacity to experience and take responsibility for the consequences. Attempts to evaluate LLMs in domains requiring moral reasoning have found their moral judgment shallow and misaligned with human reasoning (Jiang et al. 2022; Chakraborty, Wang, and Jurgens 2025). Prior work argues that the question of whether AI can be made more ethical is categorically misplaced since the inability of AI to be a moral actor is a philosophical limit and cannot be overcome by technical fixes (Bronner-Martin 2025). The capacity for reciprocal moral relations follows Strawson’s account of the reactive attitudes that constitute a moral community (Strawson 1962). Synthetic agents can be objects of such attitudes but never subjects of them. Moreover, the technical conceptions of agents do not map to socio-legal conceptions of agency (Riedl and Desai 2025). Conceptions from sociology and linguistic anthropology identify agency as ontologically relational, and inseparable from social accountability (Timaite and Castelle 2025). A ‘moral actor’ in AI ethics is defined as a person who has a ‘will to action’ (Laflamme, College, and Bruneault 2023), and a legitimate actor is one whose agency, intentionality, and capacity for reciprocal moral relationships make them bound by moral responsibility and its consequences (Baeyaert 2025). Synthetic agents do not meet this bar.

Personhood in Participatory Design

Representation of personhood is the central problem in participatory design (PD). Sanders et al.’s (Sanders, Brandt, and Binder 2010) framework categorizes the purpose of design engagement into four categories: probing, priming (immersion in the domain of interest), understanding (of current experience) and generating (creating and exploring future scenarios). Each mode aligns with other aspects of participation, including form (making/telling/enacting) and context (stakeholder relationships, online or in-person modalities, venue, group size, and composition) (Sanders, Brandt, and Binder 2010). Crucially, participatory modes of engagement produce a *participant* from a person. The participant is constituted in the process of participation, but never represents the person as a whole. The production of a participant is mediated through the modes of engagement, each in different capacities.

Probing: The foundational work, ‘Cultural Probes’ (Gaver, Dunne, and Pacenti 1999) introduced probing as a participatory mode. Probing explores ‘functions’, ‘experiences’, and ‘cultural placement’ outside the norm to discover new pleasures, forms of sociability, and cultural forms (Gaver, Dunne, and Pacenti 1999; Burrows, Mitchell, and Nicolle 2015; Madden et al. 2014). It uses provocations directed towards ‘inspirational data’ rather than ‘information’ with the purpose of stimulating imagination and asks the question ‘What inspires a person’s experience of a phenomenon of interest?’ (e.g., ‘What inspires fear of vaccines in a population?’), effectively producing the *a priori* of a participant in research.

Priming: Priming is commonly used in behavioral psychology to investigate human judgments and decision-making. It affects perceptions and behaviors by increasing cognitive accessibility of specific mental content. In the context of design, priming is achieved by using stimuli, such as drawings, videos, neutral or hostile sentences (She et al. 2018). Building on memory recall in psychological sciences (Redelmeier, Katz, and Kahneman 2003), priming seeks to produce an alterity in a participant’s experience and sense-making by introducing a shift in the process of recall. Priming is premised on the idea that a participant’s memory is not an exact replay of the complete experience but a recall of selected moments. HCI has adapted priming to support idea generation in designing creativity support tools (Lewis, Dontcheva, and Gerber 2011).

Understanding: Understanding is a precursor to generative engagement. Orr (Orr 1996) looks at technical training of technicians and their diagnosis of machine breakdown. Diagnosing the breakdown requires narrative construction and negotiation between fragments of information from the customer and the condition of the machine. Suchman (Suchman 1987) problematizes the notion embedded in the ‘expert help system’ for machine use, which embodies a conception of human action shared by designers, behavioral sciences, and common sense that what people say or do is best understood as the reflection of underlying plans. Suchman shows that this confuses *plans* with *situated action*. Plans are not rigid scripts dictating every move; they are vague guides used to explain action while navigating unpredictable situ-

ations, designed to accommodate the unforeseeable contingencies of actual situations of action. Understanding, then, engages a person to produce a participant who is in the process of sensemaking.

Generating: In early PD works, generating developed as a mode of engagement in the UTOPIA project (Bødker et al. 1987). It presents a ‘tool perspective’ where a computing system is designed as a collection of tools for skilled workers to use by building upon their tacit skills as a basis for analysis and design. In order to develop their technical imagination the users have to gain insight into technical possibilities as well. Thus, generating engages a person to produce a participant who becomes familiar with technical affordances in the design process.

The modes of engagement in PD produce a participant from a person in different aspects of processes. PD produces an *a priori* to the participant (probing); a participant in the process of memory recall (priming); a participant in the process of sensemaking (understanding); and a participant who is familiar with technical affordances (generating). There are *degrees* of participation and representation afforded by the different modes of engagement but never representation of personhood as a whole.

Methodology

This research is a comparative case study of three synthetic agents deployed in representational processes. The three cases were selected to provide variation along two dimensions. First, is the scope of representation where the cases span local (Ana, a community-voice chatbot in Pittsburgh), enterprise (Synthetic Juror, a legal trial-preparation tool), and global scales (Ask Amina and Ask Abdalla, two AI avatars developed by the United Nations University Centre for Policy Research). Second, populations represented span community members in local policy consultation, jurors in legal deliberation, and refugees and combatant leaders in humanitarian diplomacy. We restricted ourselves to cases with publicly accessible documentation sufficient for close reading. The three cases also vary in technical architecture (Retrieval Augmented Generation (RAG)-based personas in Ana and Amina/Abdalla; a hybrid neuro-symbolic system in Synthetic Juror) and in interface (chatbot, AI avatar, and Slack integration), but variation along these dimensions was not a selection criterion.

Despite the range in scope, relative to other technological systems and infrastructures, synthetic agents in representational processes are an emerging domain, and the cases we examine are pilots, enterprise trials, and prototypes rather than mature deployments. We see this as the tractable moment for articulating boundaries, before institutional inertia and sunk investments make them harder to draw.

Analytical Approach

The analysis proceeded through close reading of the publicly available documentation for each case. This included project reports, blog posts, marketing materials, founder writings, and secondary coverage where available. For Ask Amina and Ask Abdalla, our primary source was the UNU-CPR

working paper (Albrecht 2025), supplemented by investigative journalism (Gault 2025). For Ana, we drew on the New Sun Rising organizational materials and the Community Voice project documentation (Horn 2025; Rising 2025). For Synthetic Juror, we drew on the company’s published blog posts and product documentation (Abbey 2024a,b; Hoogerhuis 2024; Juror 2025).

Rather than conducting a formal thematic analysis, we worked through iterative paired discussion. The first and second authors collected and read the documentation, and developed initial codes for each case. The four authors then discussed the cases in pairs, comparing codes across cases and refining them through successive rounds. We iterated until our codes were stable across cases, focusing on how each synthetic agent framed its problem context, curated its data, designed its user encounters, and evaluated its outputs. These four dimensions emerged as analytically useful through the iterative comparison and now structure the Findings section.

While discussing the codes, the limited focus on processes prompted our examination and subsequent application of PD. In particular, Sanders, Brandt, and Binder (2010) helped us gain a process-oriented vocabulary to set against the outcome-centric evaluations that dominate the existing literature.

Positionality

We are four researchers based in Global North, three of us originally come from the Global South. The first and second authors are graduate students working at the intersection of Science and Technology Studies and Human-Computer Interaction (HCI). The third author is a faculty member with expertise in information science, technology ethics and law who leads action-oriented research projects. The fourth author is a faculty member in HCI who has conducted community-based participatory research over the last decade in both the Global South and the Global North.

Our shared concern in pursuing this work is the overreach of technology into representational contexts, particularly where the populations being represented are already structurally marginalized. We are aware of the tension that we critique tools that are framed as offering support to populations with limited resources, often from our location in well-resourced Global North institutions. We hold onto this tension. Our critique is not that synthetic agents serve these populations poorly (indeed they may speak for these communities well), but that the legitimacy and value of any approach comes from people having the agency and power to speak and act for themselves (i.e., representation). We have tried to maintain this distinction throughout the analysis.

We also acknowledge that our materials are drawn predominantly from documentation published by the organizations themselves. Much of this material is targeted at potential clients or funders, and some claims may be more aspirational. We treat the documentation as evidence of how the systems are envisioned and justified.

Case Studies

The three synthetic agents selected for this research are situated in the context of policy, diplomacy, and legal jury. Collectively, they constitute synthetic agents in political and representational contexts at three different scales: local (Ana), enterprise (Synthetic Juror), and global (Ask Abdalla/Ask Amina).

Ask Abdalla/Ask Amina

Ask Amina and Ask Abdalla are two AI avatars that were created as an experimental tool by the United Nations University Centre for Policy Research (UNU-CPR) for exploring use of AI in Conflict Prevention. Both are RAG-based implementations with curated knowledge bases targeted at specific populations, presented through voice-conversational avatars. Amina is positioned as a digital representation of a refugee living in a camp in Chad, and Abdalla as a digital replica of a combatant leader in the Rapid Support Forces (RSF), a group active in the southeastern part of Sudan from which many refugees are fleeing.

The use-case of these avatars is justified by citing the difficulties in recruiting participants for diplomatic consultation and deliberation. The documentation argues that current methods, including surveys, focus groups, and questionnaires, are time- and resource-intensive, and that participation can be limited by time constraints, geographical remoteness, or unwillingness to participate. It also notes that respondents may have *“an incentive to provide false or incomplete information when replying, especially if respondents believe that a certain type of answer may benefit them in some way”* (Albrecht 2025, p. 4).

The stated goals for the two are operational, which the document states as *“if Amina works, ‘her’ rapid responses could be of great value. For example, they could be used to quickly make a case to donors (often in very different locations and with very little time) on what population needs to be prioritized when earmarking aid to the region. If Abdalla works, ‘his’ responses could help negotiators and mediators prepare for more targeted real-world engagement”* (Albrecht 2025, p. 8).

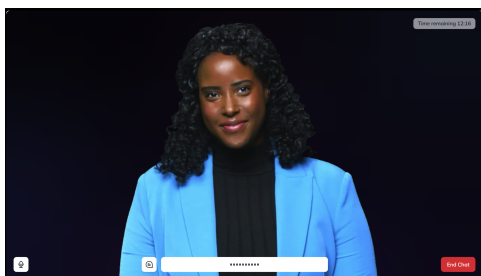


Figure 1: Screenshot of the AI avatar Amina (Gault 2025).

The development team also expresses an intention to extend the synthetic agents to group deliberation, integrating AI agent-generated personas of stakeholders, including community leaders, government officials, and military leaders, into diplomatic negotiation and mediation simulations

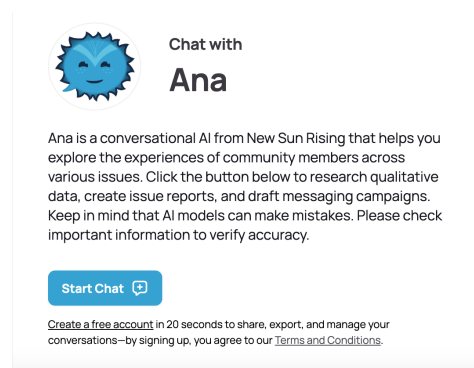


Figure 2: Screenshot of the AI chatbot Ana (New Sun Rising Pittsburgh 2025).

(Albrecht 2025, pp. 4). The two avatars were briefly online for public use but are no longer accessible.

Ana (New Sun Rising)

Ana is a chatbot created by New Sun Rising, a Pittsburgh (Pennsylvania, USA)-based organization, as part of their virtual support center. Like Amina and Abdalla, Ana is a RAG-based implementation, but rather than mimicking a regional demographic persona, it surfaces quoted excerpts and synthesized insights from a community dataset. The underlying dataset is the organization’s Community Voice project, which gathers feedback and testimonials from community leaders and members on the present strengths and challenges of their community, as well as their hopes, fears, and expectations for the future (Rising 2025).

The organization coded the dataset using its Impact Framework (mapped to the UN Sustainable Development Goals), a Resource Framework (Community Capitals), and sentiment analysis (Rising 2025). Ana *“...specializes in data analysis and communication, assisting users in exploring the Community Voice qualitative data-set to enhance grant proposals, communications, and advocacy campaigns”* (Horn 2025). Ana¹ is targeted towards non-profits and local policy makers who require inputs from community members in order to make better informed decisions. It acts as an interface for exploring and discovering existing data relevant to the query.

Synthetic Juror

Synthetic Juror is a legal technology platform that utilizes a hybrid neuro-symbolic architecture to simulate juror behavior, reasoning, and decision-making. The LLM layer handles the “human” elements of communication, processing case files and evidence to understand natural language, and patterns in emotional nuances (Juror 2025).

The system aims to extract deep patterns from the psychographic and demographic data provided by its integrated workbook, ‘Lairm’, to create synthetic juror profiles. Legal teams can use these to replace traditional mock trials with simulations that include geo-specific modeling, adjusting to

¹ Accessible at: <https://vibrancy.newsunrising.org/>

regional attitudes, and narrative testing to identify resonating emotional appeals. The company emphasizes the geo-specific dimension as a distinguishing feature, claiming that the platform “understands the cultural, demographic, and regional nuances that shape juror perspectives” and lets attorneys tailor strategies to address biases and emotional triggers specific to the jurisdiction (Juror 2025).

The platform supports running thousands of simulations and is integrated into the Slack interface of the enterprise client, where synthetic agents interact as individual personas across Slack channels. Synthetic Juror is an enterprise software available for trial upon request sent to the company.

Findings

Across the three cases we trace a four-step process (framing the problem, curating data, designing the encounter, evaluating validity) through which synthetic agents become a representational entity.

Manufacturing the Problem for Synthetic Agents

Establishing the need for a synthetic agent product is the first stage that enables conceptualization of a field of intervention in the representational process. While synthetic agents are generally proposed as a solution to the difficulty in recruiting participants for consultation, each of them speaks to a different problem.

For Ask Amina and Ask Abdalla, the stated objective is to enable humanitarian and relief workers to ask questions about refugee experiences, needs, and sentiments, and to receive responses that closely mirror real refugee perspectives (Albrecht 2025, p. 3). Amina is positioned to enhance the feasibility of needs assessments for UN staff in determining the type of intervention (Albrecht 2025), placing the synthetic agent in a role that makes allocative choices about whose interests come first. Ask Abdalla, on the other hand, is designed to let diplomats, negotiators, and mediators practice their skills with a persona that responds in ways consistent with a real combatant’s known behavioural patterns (Albrecht 2025, p. 3).

Collectively, probing and understanding the field is acknowledged as central to the domain enquiry of diplomacy. However, the complexities of probing are reduced to a feasibility concern in synthetic agents. The framing focuses on manufacturing a “person-who-probes” instead of the process of probing. These two AI avatars are, in fact, proposed as “Anthropologist Agents” that can “act as ‘investigators’ dedicated to studying a person and their context ... they must become similar to anthropologists” (Albrecht 2025, pp. 6). They further list four skill sets for these agents, covering literature review, fieldwork, culturally responsive analysis, and continuous knowledge base updates (Albrecht 2025, p. 6). This frames the synthetic agent as an expert on the lived experiences of the community whom it claims to represent, and importantly, generates a tension between the two incompatible positionalities, as an embodied refugee or combatant on the one hand, and the external observer and investigator on the other.

Ask Amina/Abdalla seeks to configure a mechanized

probe tool which can be attributed the legitimacy of an anthropologist expert. In contrast, Ana draws out synthesized insights and quoted interview excerpts from the Community Voice dataset (Horn 2025). In doing so, it makes moral choices on behalf of its users by prioritizing some voices over others on grounds of relevance, even as it does not claim to embody a persona.

Synthetic Juror intervenes in legal trial preparation for lawyers as an enterprise solution addressing the feasibility challenges of conducting a mock trial. The company frames traditional trial preparation as relying on mock trials, focus groups, and intuition, and positions itself as bringing “a new level of precision by simulating thousands of jurors’ reactions using AI.” (Abbey 2024a). The system claims to model “how real jurors think, incorporating their biases, emotional triggers, and regional quirks ... These aren’t just static profiles. They’re active participants ... You can present any evidence to them (depositions, expert witness, potential outcomes, interrogatories, motions, arguments and the list goes on) ... they respond ... with nuanced feedback that mimics how real jurors think” (Abbey 2024a). The appeal to legitimacy is the ability to narrativize and imitate the case context like a person sitting in the jury. This claim invites interrogation: can synthetic jurors be probed for bias and impartiality, which is the established procedure of *voir dire* in legal jury deliberations?

Across the cases, synthetic agents address an identified feasibility gap by targeting needs like quick surveys, data-driven insights, and variety of mock-simulation of personas. While the promise of quick, real-time qualitative insights are the core offering, it could have been delivered as reports or dashboards. The personas appear as a separate layer that anthropomorphizes the interaction through narrativization.

Manufacturing Persona through Data Access

The synthetic agents across representational contexts use qualitative datasets, surveys, and reports in their RAG implementation. For Amina and Abdalla, the knowledge base contains surveys, reports, articles, and issue briefs from UN agencies, multilateral organizations, think tanks, and NGOs working with refugees and combatants in Chad and Sudan (Albrecht 2025, p. 8). A separate persona-curation dataset draws from cultural artifacts of the community being mimicked, including “... seminal works of literature, folkloric traditions and key religious texts, but also items from the culture’s contemporary digital footprint, including local news media, social media activity and online diaspora community discussion forums.” (Albrecht 2025, p. 8). The voice-avatar also attempts to mimic the tone and filler-sounds between sentence pauses. An investigative article found that the contemporary digital footprint likely included voice and video recordings of community members (Gault 2025).

Persona curation is also central to Synthetic Juror, given its intention to simulate a diverse mix of jury positionalities. As the document states, it “uses advanced patent-pending technology to simulate juror responses based on a broad range of data” (Hoogerhuis 2024). Another document indicates “the synthetic character samples are built using a combination of public data, subscription databases, your firm’s

proprietary information, and client-specific data, ensuring a high degree of realism and relevance" (Abbey 2024b).

Both Amina/Abdalla and Synthetic Juror manufacture personas through a specifically curated dataset that is separate from their knowledge base. Ana, in contrast, draws its data from the Community Voice project curated by the same organization to *"gather feedback and testimonials both digitally and in person"*, in which people were *"invited to share their perspectives on the present strengths and challenges of their community, as well as their hopes, fears, and expectations for the future"* (Rising 2025). The community data was coded by the organization before using it to train the chatbot model (Rising 2025). Unlike Amina/Abdalla and Synthetic Juror, the chatbot Ana does not appear to mimic a region-specific demographic persona.

This data work bypasses the participatory mode of understanding, where sensemaking is constructed through narrative negotiation between participant, researcher, and context (Suchman 1987; Orr 1996). Synthetic agents collapse sensemaking into static dataset curation. Datasets drawn from institutional sources reflect the plans and blueprints of the institutions that collected them, leading to an overrepresentation of institutionally prioritized data. Paradoxically, the embedded presence of institutional blueprint sabotages the very grounds that necessitate ethnographic inquiry for understanding to complement and evaluate the institutional paradigms. There is also a clear demarcation between the datasets that generate insights, and the datasets for persona-curation which focuses on drawing positionality from demographic data and the ability to mimic personhood in the context of the prompt.

Manufacturing Encounters for User Interaction

The three cases offer widely varying affordances to target users. The documentation on Amina and Abdalla justifies the AI avatar persona by contrasting it with NLP-based sentiment analysis, where NLP can only examine existing opinions, personas can engage in dialogue about hypothetical scenarios *"offering predictive insights into how people might respond to future activities and actions"* (Albrecht 2025, p. 4). This establishes a clear distinction between analysis of existing data and predicted or hypothetical data, and importantly, it positions synthetic agents as interaction surfaces for predicted scenarios. While probing and priming could potentially reveal asymmetric power relations, such modes of inquiry are no longer viable for hypothetical and predicted scenarios. Priming is thereby bypassed since there is no recall of lived experience to shift.

Synthetic Juror extends this affordance to legal practitioners through integration with Slack. *"All of this happens right in Slack. Lawyers can @mention a simulated juror just like they would a colleague. They can create sub-channels for different aspects of the case... use Slack's huddle feature to do mock cross-examinations with AI witnesses..."* (Abbey 2024a). It further claims that the *"Lairm Workbench includes a sophisticated character creation tool that allows you to build incredibly detailed synthetic individuals for your case simulations"* (Abbey 2024b).

This affordance directly bypasses priming as it operates in

actual jury deliberation. Priming in juries takes the form of implicit bias instructions designed to make jurors reflect on their own biases, and the *Sommers effect* (Sommers 2006) shows that racial diversity among jurors primes individuals to exchange a wider range of information. By offering the ability to calibrate specific demographic positionalities at will, Synthetic Juror removes the moment of reflexivity that is integral to priming and replaces it with essentialized and flattened identities.

Ana's interaction is a chat interface. That leaves a wide-open possibility for interaction and risks confusing users. The organization recognized this risk and in response, they delivered an information session that included sample prompt templates (New Sun Rising Pittsburgh 2025). A representative template asks Ana to identify the primary themes in the Community Voice data relevant to a specified organization, issue, and audience, and to draft a 500-word summary including quotes from community members (New Sun Rising Pittsburgh 2025). Since templates scaffold the user toward indexing and summarization instead of encounter with a person, there is no moment of recall to prime.

We see these design affordances as different approaches of engaging the user in speculative inquiry and visualization. The synthetic agents either lead the user into visualizing an actual embodied individual (e.g., a refugee from Chad), the opinions of an individual (e.g., Slack messages of a jury member), or the reported insight of an individual (e.g., an insight reported by Ana). In all cases, the synthetic agents give an impression of personhood while mechanically drawing data and inferences from their existing knowledge base.

Manufacturing Validity with Technical Measures

We note that the validity of the synthetic agents is justified by manufacturing the problem context. However, a parallel move occurs at the level of outcome evaluation.

Synthetic Juror claims *"99% accuracy rate in predicting case outcomes. That's not just a marketing claim. It's based on comparing the system's predictions to actual trial results"* (Abbey 2024a). The documentation on Amina describes a study where Amina was fed 20 questions drawn from four surveys not included in her knowledge base, with the delta between her responses and the actual human responses measured as the indicator of accuracy; *"she correctly answered 16 out of 20 questions, achieving an 80 per cent accuracy rate"* (Albrecht 2025, p. 7). In contrast, *"Abdalla's responses, [were] qualitatively assessed. A brief conversation between the two [will] also be qualitatively assessed"* (Albrecht 2025, p. 7).

The available documentation does not describe any evaluation method for Ana. They measure the similarity of synthetic outputs to human outputs (a delta, a percentage accuracy, a qualitative comparison) and treat that similarity as the warrant for representational validity. This evaluation logic bypasses the participatory mode of generating, where the process of setting up a deliberative mechanism is as integral to legitimacy as the artifact it produces. PD tools for generating design through group deliberation reveal that outcomes are not automatic products but are closely tied to the underlying process (Simonsen, Karasti, and Hertzum 2020).

Outcome-centric evaluation makes the process invisible, and importantly it makes the process un-evaluable.

Discussion

The synthetic agents in our three cases are early-stage and experimental, and they reflect the broader nature of representational synthetic agents as an emerging domain. This is the moment when boundaries are most tractable, before institutional inertia and sunk investments make them harder to articulate. In this section, we make explicit what is at stake in the four-step manufacture we traced in Findings, and articulate what synthetic agents categorically cannot do.

Bypassing Representational Process

As the Findings showed, the need for synthetic agents is manufactured by emphasizing the difficulty of recruiting participants as the central bottleneck of representational processes, for which a sophisticated mechanical process is offered as the solution. This amounts to a “machinal bypass” (Kaplan, Palitsky, and Raison 2025) of a multidimensional process where representation is integral to any claims of legitimacy for an institution.

The central function of representation in participatory processes is the fidelity to the participant population (Premananda 2026). From this view, the limited access to participants is not a mere problem to solve with a single solution. It is a symptom of deeper structural failures of representation, and PD scholarship has long questioned whether slow prototyping or feasibility are bottlenecks at all (Elavsky 2026).

Existing evaluation frameworks for synthetic agents (*believability* (Xiao et al. 2024), *plausibility* (Agarwal, Tanneru, and Lakkaraju 2024), *faithfulness* (Maden 2026), *algorithmic fidelity* (Argyle et al. 2023)) all measure similarity between human and synthetic responses, treating outcome resemblance as the warrant for representational validity. The table in Fig. 3 contrasts this with PD modes of engagement applied to the same three domains. PD treats representation as a process whereas synthetic agents reduce it to an outcome. Let us consider the strongest version of the case these frameworks make, a synthetic agent whose responses are statistically indistinguishable from those of the population it stands in for. Even at this limit of perfect fidelity, the agent delivers the resemblance of participation and not participation itself, and, as established above, it is the participation that legitimacy requires. The bypass thus becomes structural since existing outcome-centric evaluation makes emphasis and orientation on the processes impossible.

Reckoning Replaces Judgment The distinction between *reckoning* and *judgment* clarifies why this bypass is consequential (Weizenbaum 1976; Smith 2019). Reckoning is estimation or calculation that can be sophisticated but operates within a well-defined space governed by simple logical rules. For instance, a game of chess is reckoning. Judgment is a different form of intelligence requiring dispassionate and deliberate thought guided by ethical commitments and awareness of consequences, producing actions appropriate to an undefined situation (Weizenbaum 1976). In our case studies, jury deliberation is paradigmatically a process

of judgment. The jury constructs a story from evidence to arrive at a verdict, and importantly, the evidence interpretation cannot be reduced to the verdict itself.

LLMs are reckoning machines due to their epistemic method of reproducing the cosine distances present in their vector-embeddings in the high-dimensional space. These vector-embeddings are abstracted from the past training data and hence are constrained by them. Related literature on simulation of judgment in LLMs argues that the very notion of judgment is operationalized when decisions are made by LLMs, where similarity of outputs ‘appear as alignment’ but may conceal deeper epistemic shift where normative reasoning is replaced by statistical approximation. The authors ask what heuristics are encoded in this process of delegation (Loru et al. 2025)? Such a system cannot orient itself towards the concerns of the future or the consequences of its actions. As we saw in the Findings section, synthetic agents mimic human personhood by imitating human responses and even have a separate dataset curated just for the purpose of imitation. The deliberate use of anthropomorphic terminology, such as ‘reasoning’, ‘thinking’, ‘hallucinations’, indicate an attempt to conflate the human and machine processes. Importantly, it stretches the logic of context representation to claim that reckoning machines are capable of judgment.

This has deep consequences beyond poorly made moral choice. Allocative choices, such as whose needs to prioritize when earmarking aid (Amina/Abdalla), and choices about which voices to prioritize (Ana), are political judgments. In fact, when they are represented as output of a calculation engine, they are often removed from the domain of contestable political judgment and relabeled as technical fact. It launders judgment, conferring on the synthetic agent the functional status of a moral actor while remaining categorically incapable of intentionality or of entering into reciprocal relations as a subject of rights and duties. This relocation of political choice into the seemingly neutral output of a machine is what later allows those in authority to defer accountability for it, a deferral we take up in the next section. Substituting a synthetic agent for a human participant is, in the end, the application of a reckoning mechanism in a context that categorically requires judgment.

Reflexivity Cannot Be Mechanized A related bypass occurs at the level of reflexivity. Synthetic agents often draw their imitation data from “seed-interviews” (Narasimhan 2026, p. 6) which collect opinions and associate them with their demographic data. In our case studies, Ana draws on the Community Voice project, and Amina/Abdalla draw on social media data and surveys. Beyond identity flattening and essentialization (Su 2026; Wang, Morgenstern, and Dickerson 2025), this epistemic method amounts to a claim of mechanically reproducing reflexivity.

Participation in representational processes requires reflexivity. When a citizen responds to a policy consultation or a juror weighs evidence, they reflect on their positionality (as a citizen, as a juror) before generating a response. This response is shaped in the act of reflection. Synthetic agents bypass this moment and replace it with a probabilistic algo-

	PERSONHOOD				SYNTHETIC AGENTS
	PROBING	PRIMING	UNDERSTANDING	GENERATING	
POLICY	Deliberative inquiry, and cultural probes	Awareness of nudging and choice architecture	Citizen ethnography	Generative co-design and participatory budgeting	Opinions of the agent closely match the opinions of the target population
DIPLOMACY	Probing the nature of conflicts	Symbolic gestures in picking negotiation venues	Ethnography and field enquiry	Consulting affected communities for negotiation	The agent's knowledge of the conflict situation and culture matches the target population
JURY DELIBERATION	"Voir Dire" - probing for bias in the jury	"Implicit bias instructions"- make the jury check their own bias; "Sommers effect" – presence of other jury members primes individuals to act in a certain way	"Evidence interpretation"- building a story to make sense of evidence	"Deliberation"- construction of verdict through arguments and voting among the jury	Agents can enact responses similar to those of target jury demography
PROCESS ORIENTED					OUTCOME CENTRIC

Figure 3: Synthetic agents seek to replace a complex representational process.

rithm of mechanical context estimation. The extreme case in our findings was the proposed *anthropologist agent* (Albrecht 2025, p. 6), which conflates the positionality of the embodied refugee or combatant with that of the external observer studying them.

Synthetic Juror represents the inverse extreme. The platform allows users to curate demographic identities for synthetic jurors and claims the resulting agents respond as people from those positionalities would. An individual is reduced to a tuple of social and systemic categories. The training data of foundational models and the RAG implementations built on them are constrained by the computational requirement to converge on the most probable output, which means generated responses identify with the median of the training data and exclude marginalized voices that exist in the tails (Su 2026). Essentialization is, thus, structurally produced by these epistemic methods.

Personhood is constituted relationally through political, legal, and moral dimensions that together determine the relationships a person holds with other persons and institutions. Participation, representation, and legitimacy are entangled in this relational notion. But synthetic agent reproduces the appearance of participation while removing its substance, the process, the judgment, and the reflexivity that make participation legitimate. That is, it “*manufactures personhood*”. When synthetic agents manufacture personhood, they reduce these relational dimensions to essentialized identities and produce an appearance of legitimate participation without the underlying process. *Nobody* participates.

Boundaries for Design

The manufacture of personhood enables a further deferral. When LLMs substitute for human participants, those in authority can claim that representational choices were algorithmically determined by a calculation engine that cannot itself be evaluated. This problem has been examined in the context of e-commerce as the principal-agent-third-party problem of trust and verification (Riedl and Desai 2025), but it has not been conceptualized for representational decision-making. Related work has named the broader pattern an *accountability shield* (Timaite and Castelle 2025) and *algorithmic governmentality* (Rouvroy and Berns 2013; Nayak 2021).

Representational contexts intensify the stakes because the choices are political. Drawing boundaries becomes critical to ensure we can center people’s power and agency in the political processes.

Drawing on the manufactured personhood we traced above, we propose two boundaries for researchers and designers working with LLM-based systems in representational contexts. The first is “soft” and context-dependent. The second is “hard” and categorical. We build this distinction because hedged versions of the second boundary erode under pressure from precisely the actors most likely to scale or transfer these systems beyond their experimental scope (Larson et al. 2016).

The soft boundary separates LLMs functioning as analytic, indexing, or scaffolding tools from LLMs that present themselves as person-like interlocutors. This boundary is contextual. Anthropomorphization is sometimes useful. For example, it can make complex data more navigable and thus accessible (Tan et al. 2025). Other times, it is harmful when it naturalizes claims a system cannot make (Lifshitz and Hung 2024).

The hard boundary separates anthropomorphized interfaces, which are mere “evocative animations” (Stark 2024), from systems that substitute for human participants in representational processes. This boundary is categorical and absolute in the sense that no degree of fidelity, believability, or algorithmic sophistication moves a system across it.

We draw this categorical boundary based on the fact that representational processes (e.g., policy consultation, jury deliberation, diplomatic negotiation) derive their legitimacy not from arriving at the right outcome but from being constituted by the participation of persons. The participation is in the service of legitimacy. For example, a jury verdict in the US is legitimate because twelve persons from society deliberated and not because a tribunal predicted what they would have said. Similarly, a policy consultation rests its legitimacy on having citizens’ input.

On the Hardness of Hard Boundaries Synthetic agents categorically lack the properties that participation requires, which are consciousness, intentionality, embodied experience, the capacity to bear consequences, and the capacity for reciprocal moral relations. These are conditions of per-

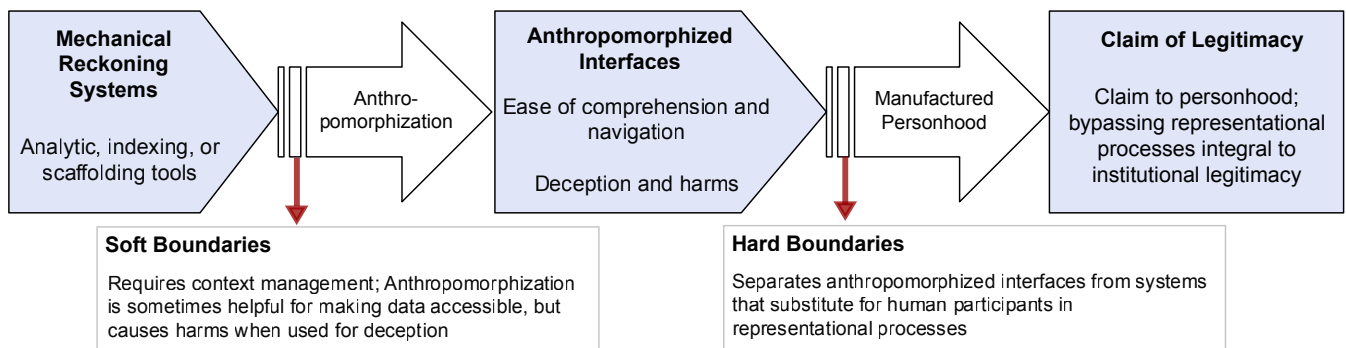


Figure 4: Soft boundaries and hard boundaries to develop oversight for synthetic agents.

sonhood that constitute representational processes.

We understand that this may rule out reasonable middle-ground positions that are argued in the literature and popular media. *Community-consented synthetic representation* assumes a community can authorize a synthetic agent to be its representation in deliberative settings. However, participation is non-transferable. No one can consent on another's behalf to having that other's participation substituted, because the participation is the thing legitimacy requires (Agnew et al. 2026). *Augmentation-as-supplement* assumes a synthetic agent can speak alongside human participants without displacing them. But the synthetic voice in a representational process is doing representational work, howsoever it may be framed. *Training simulations for real deployment* assume practice on a simulation transfers to competence in the real thing. But what is practiced is engagement with a text-prediction system, and importantly, not engagement with a refugee, a juror, or a combatant leader and all the contingencies and uncertainties that are at play in those contexts.

From our perspective, the single remaining use case is education or training that is explicitly scaffolded as non-transferable, where the discontinuity between the simulated encounter and the real one can be a lesson. A diplomat-in-training can encounter a system like Ask Abdalla provided the encounter is framed as engagement with a text-prediction system, and importantly, not as preparation for engaging with actual combatant leaders. A law student can practice argumentation against a synthetic juror to improve the students' argument skills and not as a way to learn how to engage with the judicial system. In these cases, a carefully created scaffold to draw boundaries is necessary.

Soft Boundaries Require Maintenance Of our three cases, Ana sits closest to the soft boundary because it surfaces existing community voices through quoted excerpts. However, the representational problem remains. The chatbot persona, by lending mediated authority to the indexed material, performs legitimization work. This instability is intrinsic to soft boundaries. As community members continue to engage with such a system, as organizational staff cite its outputs in communications, and as it becomes the default interface to the community dataset, the system accrues representational status that was never formally claimed for it.

Indeed, the legal recognition of this drift has been estab-

lished. In *Moffatt v. Air Canada*, the British Columbia Civil Resolution Tribunal held the airline responsible for its chatbot's statements, rejecting the argument that the chatbot was a separate legal entity from the company it served (Lifshitz and Hung 2024). The tribunal recognized that the deployment of a chatbot in a customer-facing role makes the chatbot a functional representative.

Maintaining the soft boundary, therefore, requires active design and governance work, and importantly, it is not a property the system possesses on its own. Interfaces should clarify and explicitly emphasize the system's role in indexing data rather than being a voice. Outputs that quote or summarize community data should route attribution back to the community sources rather than aggregate into a single chatbot persona. Importantly, evaluation processes should test for representational drift over time and, building on participatory design tradition, ask process-oriented questions. This includes asking questions like *Whose voices and processes are informing the system?* A system that is informed by the researcher's analytic process is differently situated than one that purports to involve a community's participatory process. *Does the system create or reduce opportunities for different stakeholders to participate in their own terms?* A system that points researchers toward community voices they had not heard expands participation. *What would the represented community say about the system, in conditions where their assessment carries weight?* This shifts the evaluative locus from the researcher's judgment of fidelity to the community's judgment of the system's relation to them.

Limitations

To conclude, given the experimental nature of the cases available for studying, we acknowledge that our case materials are drawn from the documentation published by the organizations. Since a lot of this material is targeted towards their potential clients, some of their claims may be more aspirational than material. The inability to interact with the three AI agents directly is another limitation which could have added more analytical depth to our analysis. Ana chatbot was available for interaction, but its terms of use limit us from discussing the generated responses.

Acknowledgments

An earlier version of this work was presented at the symposium 'Chat Token Vector: Questioning Models of Language and Neo-Structuralism in AI' organized by the ERC project AIMODELS at the Department of Philosophy and Cultural Heritage of Ca' Foscari University of Venice in partnership with Cambridge Language Sciences and Cambridge Digital Humanities. We are thankful to the symposium for their valuable feedback. We would like to express our gratitude to Takuya Maeda for the productive discussions around this paper. We would also like to thank the anonymous reviewers and the meta-reviewer for their helpful comments which helped in shaping this paper.

Declaration on Use of AI

We acknowledge the use of Claude (Anthropic) as a copy-editing tool to improve spelling, grammar, punctuation, clarity, and concision. Gemini (Google) was also used for web-search.

References

- Abbey, S. 2024a. Legal slack: Synthetic Juror.
- Abbey, S. 2024b. SyntheticJuror™ and Lairm™ Workbench: Revolutionizing Trial Consulting: Synthetic Juror.
- Agarwal, C.; Tanneru, S. H.; and Lakkaraju, H. 2024. Faithfulness vs. Plausibility: On the (Un)Reliability of Explanations from Large Language Models. ArXiv:2402.04614 [cs].
- Agnew, W.; Kapania, S.; Schroeder, H.; Aubin Le Quéré, M.; Fox, S. E.; and Heidari, H. 2026. Workshop on Developing Standards and Documentation For LLM Use as Simulated Research Participants. In *Proceedings of the Extended Abstracts of the 2026 CHI Conference on Human Factors in Computing Systems*, 1–4. Barcelona, Spain: ACM. ISBN 979-8-4007-2281-3.
- Albrecht, E. 2025. Does the United Nations need agents? Testing the role of AI agent generated personas in humanitarian action.
- Alvarado Garcia, A.; Shukla, N. A.; Azmat, M.; and Wong-Villacres, M. 2025. Disciplinary Practices in the Generation of Text Synthetic Data: A Critical Discourse Analysis. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 8(1): 152–165.
- Arawjo, I. 2026. That's Enough About AI Replacing Users in User Research. In *Proceedings of the Extended Abstracts of the 2026 CHI Conference on Human Factors in Computing Systems*. In Review.
- Argyle, L. P.; Busby, E. C.; Fulda, N.; Gubler, J. R.; Rytting, C.; and Wingate, D. 2023. Out of One, Many: Using Language Models to Simulate Human Samples. *Political Analysis*, 31(3): 337–351.
- Baeyaert, J. 2025. Beyond Personhood: The Evolution of Legal Personhood and Its Implications for AI Recognition. *Technology and Regulation*, 2025: 355–386.
- Ball, S.; Allmendinger, S.; Kreuter, F.; and Kühl, N. 2025. Human Preferences in Large Language Model Latent Space: A Technical Analysis on the Reliability of Synthetic Data in Voting Outcome Prediction. ArXiv:2502.16280 [cs].
- Bronner-Martin, G. 2025. Ethical AI as a Moral Agent is Impossible: Why Algorithmic Fairness is an Illusion and We Need Democratic AI Governance.
- Burrows, A.; Mitchell, V.; and Nicolle, C. 2015. Cultural probes and levels of creativity. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*, MobileHCI '15, 920–923. New York, NY, USA: Association for Computing Machinery. ISBN 978-1-4503-3653-6.
- Bødker, S.; Ehn, P.; Kammersgaard, J.; Kyng, M.; and Sundblad, Y. 1987. A UTOPIAN Experience: On Design of Powerful Computer-Based Tools for Skilled Graphic Workers. In Bjørknes, G.; Ehn, P.; and Kyng, M., eds., *Computers and Democracy*, 251–278. Avery. ISBN 0 566 05476 0.
- Chakraborty, M.; Wang, L.; and Jurgens, D. 2025. Structured Moral Reasoning in Language Models: A Value-Grounded Evaluation Framework. ArXiv:2506.14948 [cs.HC].
- Elavsky, F. 2026. On genAI: Was prototyping really a bottleneck?
- Ersoz, A. B. 2026. 12 Angry AI Agents: Evaluating Multi-Agent LLM Decision-Making Through Cinematic Jury Deliberation. ArXiv:2605.01986 [cs].
- Gault, M. 2025. The UN Made AI-Generated Refugees.
- Gaver, B.; Dunne, T.; and Pacenti, E. 1999. Design: Cultural probes. *Interactions*, 6(1): 21–29.
- Gunkel, D. J.; and Wales, J. J. 2021. Debate: what is personhood in the age of AI? *AI & SOCIETY*, 36(2): 473–486.
- Habermas, J. 1996. *Between facts and norms: contributions to a discourse theory of law and democracy*. Studies in contemporary German social thought. Cambridge, Mass: MIT Press, 1. mit press paperback ed., 3. print edition. ISBN 978-0-262-08243-3 978-0-262-58162-2.
- Hoogerhuis, E. 2024. Beyond Mock Trials: How SyntheticJurors deliver Data-Driven Insights for Winning Legal Cases: Synthetic Juror.
- Horn, A. 2025. New Sun Rising launches Virtual Support Center – New Sun Rising.
- Jiang, L.; Hwang, J. D.; Bhagavatula, C.; Bras, R. L.; Liang, J.; Dodge, J.; Sakaguchi, K.; Forbes, M.; Borchardt, J.; Gabriel, S.; Tsvetkov, Y.; Etzioni, O.; Sap, M.; Rini, R.; and Choi, Y. 2022. Can Machines Learn Morality? The Delphi Experiment. ArXiv:2110.07574 [cs.CL].
- Jowsey, T.; Braun, V.; Clarke, V.; Lupton, D.; and Fine, M. 2025. We Reject the Use of Generative Artificial Intelligence for Reflexive Qualitative Research. *Qualitative Inquiry*, 10778004251401851.
- Juror, S. 2025. About- Synthetic Juror.
- Kaplan, D. M.; Palitsky, R.; and Raison, C. L. 2025. The “machinal bypass” and how we’re using AI to avoid ourselves. *Proceedings of the National Academy of Sciences*, 122(51): e2518999122.
- Kim, D.; Yun, Y.; and Hyun, K. H. 2026. Layer-Specific LLM Participation in Spatial Usability Evaluation. In *Proceedings of the Extended Abstracts of the 2026 CHI Conference on Human Factors in Computing Systems*.

- Kuric, E.; Demcak, P.; and Krajcovic, M. 2026. Synthetic Participants Generated by Large Language Models: A Systematic Literature Review. In *Proceedings of the Extended Abstracts of the 2026 CHI Conference on Human Factors in Computing Systems*. In Review.
- Laflamme, A. S.; College, A.-L.; and Bruneault, F. 2023. Including a Social Perspective In AI Ethics: The Contribution of a Dialogue Between American Pragmatism and Critical Theory.
- Larson, J.; Mattu, S.; Kirchner, L.; and Angwin, J. 2016. How We Analyzed the COMPAS Recidivism Algorithm.
- Leibo, J. Z.; Vezhnevets, A. S.; Cunningham, W. A.; and Bileschi, S. M. 2025. A Pragmatic View of AI Personhood. ArXiv:2510.26396 [cs.AI].
- Lewis, S.; Dontcheva, M.; and Gerber, E. 2011. Affective computational priming and creativity. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 735–744. Vancouver BC Canada: ACM. ISBN 978-1-4503-0228-9.
- Lifshitz, L. R.; and Hung, R. 2024. BC Tribunal Confirms Companies Remain Liable for Information Provided by AI Chatbot.
- Loru, E.; Nudo, J.; Di Marco, N.; Santirocchi, A.; Atzeni, R.; Cinelli, M.; Cestari, V.; Rossi-Arnaud, C.; and Quattrocchi, W. 2025. The simulation of judgment in LLMs. *Proceedings of the National Academy of Sciences*, 122(42): e2518443122.
- Madden, D.; Cadet-James, Y.; Atkinson, I.; and Watkin Lui, F. 2014. Probes and prototypes: a participatory action research approach to codesign. *CoDesign*, 10(1): 31–45. eprint: <https://doi.org/10.1080/15710882.2014.881884>.
- Maden, W. v. d. 2026. Three Evaluation Problems for LLM-Simulated Research Participants. In *Proceedings of the Extended Abstracts of the 2026 CHI Conference on Human Factors in Computing Systems*.
- Maeda, T.; and Quan-Haase, A. 2024. When Human-AI Interactions Become Parasocial: Agency and Anthropomorphism in Affective Design. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency*, 1068–1077. Rio de Janeiro Brazil: ACM. ISBN 979-8-4007-0450-5.
- Maeda, T.; and Stark, L. 2025. Anthropomorphism as Social Affordance: Charting the Co-Animation of Chatbots into Social “Agents”. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 8(2): 1661–1673.
- Mayne, H.; Kearns, R. O.; Yang, Y.; Bean, A. M.; Delaney, E.; Russell, C.; and Mahdi, A. 2025. LLMs Don’t Know Their Own Decision Boundaries: The Unreliability of Self-Generated Counterfactual Explanations. *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*.
- Narasimhan, S. 2026. The Flight Simulator for Management: Generative-Agent Simulations for Faster, Safer Decisions.
- Nayak, A. 2021. Boredom, Time-Perception and Algorithmic Governmentality. *Journal of AI Humanities*, 7: 11–24.
- New Sun Rising Pittsburgh. 2025. AI-driven Assistance: Introducing the NSR Virtual Support Center.
- Nguyen, D. C.; and Welch, C. 2026. Generative Artificial Intelligence in Qualitative Data Analysis: Analyzing—Or Just Chatting? *Organizational Research Methods*, 29(1): 3–39.
- Orr, J. E. 1996. *Talking about machines: an ethnography of a modern job*. Collection on technology and work. Ithaca, N.Y: ILR Press. ISBN 978-1-5017-0739-1.
- Premananda, H. 2026. *Non-Venatus: On the Production of Wars and Universities*. Bloomsbury. ISBN 978-93-6952-647-5.
- Redelmeier, D. A.; Katz, J.; and Kahneman, D. 2003. Memories of colonoscopy: a randomized trial. *Pain*, 104(1): 187–194.
- Reshef Kera, D.; Navon, E.; Wellner, G.; and Kalvas, F. 2024. Governance in Silico: Experimental Sandbox for Policymaking over AI Agents.
- Riedl, M. O.; and Desai, D. R. 2025. AI Agents and the Law. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 8(3): 2189–2198.
- Rising, N. S. 2025. Community Voice – New Sun Rising.
- Rouvroy, A.; and Berns, T. 2013. Algorithmic governmentality and prospects of emancipation: Disparateness as a precondition for individuation through relationships? *Réseaux*, 177(1): 163–196.
- Sampson, P.; Lurie, E.; and Metaxa, D. 2026. Surveying the LLM Simulation of Marginalized Identities. In *Proceedings of the Extended Abstracts of the 2026 CHI Conference on Human Factors in Computing Systems*.
- Sanders, E. B.-N.; Brandt, E.; and Binder, T. 2010. A framework for organizing the tools and techniques of participatory design. In *Proceedings of the 11th Biennial Participatory Design Conference*, 195–198. Sydney Australia: ACM. ISBN 978-1-4503-0131-2.
- She, J.; Seepersad, C. C.; Holttä-Otto, K.; and MacDonald, E. F. 2018. Priming Designers Leads to Prime Designs. In Plattner, H.; Meinel, C.; and Leifer, L., eds., *Design Thinking Research*, 251–273. Cham: Springer International Publishing. ISBN 978-3-319-60966-9 978-3-319-60967-6. Series Title: Understanding Innovation.
- Simile.ai. 2026. Home | Simile.
- Simonsen, J.; Karasti, H.; and Hertzum, M. 2020. Infrastructuring and Participatory Design: Exploring Infrastructural Inversion as Analytic, Empirical and Generative. *Computer Supported Cooperative Work (CSCW)*, 29(1-2): 115–151.
- Smith, B. C. 2019. *The Promise of Artificial Intelligence: Reckoning and Judgment*. Cambridge: MIT Press. ISBN 978-0-262-04304-5 978-0-262-35520-9.
- Smith, K. L.; Akanbi, O.; Maqsood, S.; Obar, J. A.; Redden, J.; and Bannerman, S. 2025. Crafting an Intersectional Response to Bill C-27 for the Standing Committee on Industry and Technology. *Canadian Journal of Communication*, 50(1): 54–67.
- Sommers, S. R. 2006. On racial diversity and group decision making: Identifying multiple effects of racial composition on jury deliberations. *Journal of Personality and Social Psychology*, 90(4): 597–612.

- Stark, L. 2024. Animation and Artificial Intelligence. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency*, 1663–1671. Rio de Janeiro Brazil: ACM. ISBN 979-8-4007-0450-5.
- Strawson, P. 1962. Freedom and Resentment. *Proceedings of the British Academy*, 48: 187–211.
- Su, R. 2026. Toward Risk-Calibrated Standards for LLM-Simulated Participants in Human Subjects Research. In *Proceedings of the Extended Abstracts of the 2026 CHI Conference on Human Factors in Computing Systems*.
- Suchman, L. A. 1987. *Plans and situated actions: the problem of human-machine communication*. Cambridge [Cambridgeshire] ; New York: Cambridge University Press. ISBN 978-0-521-33137-1 978-0-521-33739-7.
- Tan, C. T.; Atmosukarto, I.; Tandianus, B.; Shen, S.; and Wong, S. 2025. Exploring the Impact of Avatar Representations in AI Chatbot Tutors on Learning Experiences. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, 1–12. Yokohama Japan: ACM. ISBN 979-8-4007-1394-1.
- Timaite, G.; and Castelle, M. 2025. Agents Without Agency: Anthropological and Sociological Lessons for Contemporary AI Research and Policy. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 8(3): 2493–2507.
- Wang, A.; Morgenstern, J.; and Dickerson, J. P. 2025. Large language models that replace human participants can harmfully misportray and flatten identity groups. ArXiv:2402.01908 [cs].
- Weizenbaum, J. 1976. *Computer power and human reason: from judgment to calculation*. San Francisco: Freeman. ISBN 978-0-7167-0464-5 978-0-7167-0463-8.
- Xiao, Y.; Cheng, Y.; Fu, J.; Wang, J.; Li, W.; and Liu, P. 2024. How Far Are LLMs from Believable AI? A Benchmark for Evaluating the Believability of Human Behavior Simulation. ArXiv:2312.17115 [cs].
- Yang, J. 2026. Carrier: Infrastructure for Validating LLMs as Simulated Research Participants. In *Proceedings of the Extended Abstracts of the 2026 CHI Conference on Human Factors in Computing Systems*.
- Young, I. M. 2000. *Inclusion and Democracy*. Oxford political theory. Oxford: Oxford university press. ISBN 978-0-19-829755-0.