Reflexive Data Curation: Opportunities and Challenges for Embracing Uncertainty in Human-AI Collaboration

ANNE ARZBERGER, Department of Software Technology, TU Delft, Netherlands
MARIA LUCE LUPETTI, Department of Architecture and Design, Politecnico di Torino, Italy
ELISA GIACCARDI, Department of Design, Politecnico di Milano, Italy

CCS Concepts: • **Human-centered computing** → **HCI design and evaluation methods**; **HCI theory, concepts and models**.

Additional Key Words and Phrases: human-AI collaboration, machine-learning uncertainty, Research through Design, reflexive data curation, human-AI reflexive practices

This article presents findings from a Research through Design investigation focusing on a reflexive approach to data curation and the use of generative AI in design and creative practices. Using binary gender categories manifested in children's toys as a context, we examine three design experiments aimed at probing how designers can cultivate a *reflexive human-AI practice* to confront and challenge their internalized biases. Our goal is to underscore the intricate interplay between the designer, AI technology, and publicly held imaginaries and to offer an initial set of tactics for how personal biases and societal norms can be illuminated through interactions with AI. We conclude by proposing that designers not only bear the responsibility of grappling critically with the complexities of AI but also possess the opportunity to creatively harness the limitations of technology to craft a *reflexive data curation* that encourages profound reflections and awareness within design processes.

1 INTRODUCTION

The increasing availability of user-friendly and accessible AI solutions like Midjourney, Dall-E, and ChatGPT is transforming creative practices. These off-the-shelf AI tools are being implemented to analyze user data and automate design workflows [65], generate stock pictures [136] and storyboards [20], develop design concepts [70], and much more. Human creativity is supplemented, at times replaced, by algorithmic counterparts whose speed and effectiveness create general excitement within the creative industries and communities. In a growing number of outlets, AI is portrayed as the ultimate force contributing to an ongoing "escalation of creativity" where everyone, especially non-professionals, is enabled to create media, whether it is music, images, or poetry.

However, the widespread adoption of AI in society, spanning areas such as employment, housing, education, healthcare, and law enforcement, has raised concerns about biases and inequalities on an unprecedented scale [8]. The incorporation of algorithmic decision-making processes in both public and commercial sectors has exposed new intersectional forms of protected identities and algorithmic discrimination [83, 121], often resulting from the social power structures and prejudices that are baked in the training data [86, 87]. These technologies

Authors' addresses: Anne Arzberger, a.arzberger@tudelft.nl, Department of Software Technology, TU Delft, Netherlands; Maria Luce Lupetti, Department of Architecture and Design, Politecnico di Torino, Italy, maria.lupetti@polito.it; Elisa Giaccardi, Department of Design, Politecnico di Milano, Italy, elisa.giaccardi@polimi.it.

ACM 1557-7325/2024/8-ART
https://doi.org/10.1145/3689042

frequently violate fundamental rights, promoting discrimination, privacy violations, stereotyped depictions, and economic inequality. In turn, this is causing an exacerbation of distributive inequities, with marginalized and underprivileged populations being disproportionately affected [8, 50]. A striking example is the recent scandal that affected the tax authorities in the Netherlands, which adopted an algorithmic decision-making system for assessing the risk of inaccuracy or fraud of individuals applying for childcare benefits. The system, which used data like individuals' nationality, resulted in racially biased and discriminatory claims, with severe consequences for the affected families, such as parents losing jobs or breaking up and children being put in foster care [64].

Even if more subtly and with longer-term effects, AI brings unintended and undesirable consequences for society even when used for creative purposes. For instance, it is largely debated whether creative practitioners should welcome AI technologies that are likely to replace the creative workforce and bypass copyright regulations [38], as stock photos, logos, music, and more can be generated automatically. Most critically, research has shown how generative AI tends to exacerbate existing issues related to biased representations of different social groups. In this regard, a recent experiment reported by Bloomberg [93] has revealed that images of professionals generated using Stable Diffusion (a popular text-to-image model) would depict a rather prejudiced reality regarding the race and gender of workers in various occupations. For instance, when prompting the word "judge," only 3% of times the image generated by Stable Diffusion would show a woman, while in many countries, the percentage of women judges is significantly higher, such as in the US where it is 34%.

To mitigate these issues, many initiatives are now being explored with the ambition of "debiasing" data and models [14]. The strategies implemented, however, reflect technocentric approaches. These focus on addressing biases through data processing and algorithm refinement, and thus often fail to address the power structures and prejudices embedded in the training data, while the societal, cultural, institutional, and economic factors that are required to understand biases and their effect on the world remain overlooked [16, 72]. Taking the use of gendered language in job descriptions as an example, a technocentric approach to debiasing LLMs might focus solely on removing explicitly gendered terms from the training data, aiming to achieve a gender-neutral model. However, in many societies, certain professions have traditionally been associated with specific genders. For example, words like "nurse" or "teacher" might be more commonly associated with women, while "engineer" or "CEO" might be associated with men. Even if these gendered terms are not explicitly present in job descriptions, the historical and societal associations can still influence the model's understanding. The difficulty in navigating responsibly open datasets and generative AI tools is reflected also by the commercial work of creative practitioners [56], which often seem to lack the criticality and skills for addressing the limits and possible effects of AI in society, including the biases and uncertainties that are inherent in computational data and models [15].

In search for a reflexive approach to data curation in the day-to-day use of AI tools, this paper reports on a Research through Design (RtD) project, in which we embrace the biases and uncertainties inherent to prototyping with data, algorithms, and models as a fundamental shift in how we understand the role of the prototype in design-led inquiries [44]. In this context, our intention is not to argue that AI is indispensable for fostering reflection or that reflection cannot be attained by other means. Rather, we seek to reconsider our approach to data, contemplating how a reflexive stance enables us to perceive them as subjective and subject to interpretation. This orientation resonates with the critical discourse within HCI surrounding AI data collection [24, 68, 100] and reflexive design approaches [12, 27, 29, 95], as they intersect an expanding realm of RtD practices that actively engage with data and artificial intelligence [77, 91, 124]. In this project, the first author has crafted a process of *reflexive data curation* to challenge the binary gender categories conventionally baked into children's toys. Selecting the task of creating *queer toys* is deliberate in its ability to surface how social norms and stereotypes are integrated into products. It also provided a context for the practitioner (i.e., the first author) to actively and reflexively address her preconceived ideas and biases through the conscious use of data curation involving classification algorithms and generative adversarial networks. Within this context, the paper unpacks three design experiments to examine how a *human-AI reflexive practice* emerged from the first author's data curation

process and illustrates the fundamentally situated, reflexive, and recursive interactions between the first author and the AI. This process of unpacking is carried out in line with practices of situated data analysis in both HCI [33, 71] and data science [67, 72].

We describe how the first author attempted to displace her perception, surface internalized values and norms, and better understand her biases and fixations. She aimed to engage in an ongoing dialogue with the personal, cultural, and historical contexts in which the self, the AI, and collective imaginaries are intricately positioned and entwined, considering them as *"design material"* [104]. These craft-oriented explorations are not foreclosing broader reflections; rather, they present a starting point for surfacing and dismantling the manifestation of broader collective imaginaries in personal biases. Experiencing confrontation through the AI exacerbation of one's biases provides insights into the intricate workings of classification and generation tools. Three design tactics for reflexive data curation— *auto-confrontation*, *shift in perspective*, and *clash of expectations*—are outlined. The discussion concludes by highlighting designers' capacity to transform data curation into a meaningful space for reflection and awareness within daily practice.

In addition to contributing concepts and tactics to the repertoire of algorithmic RtD practices (this call), critical HCI approaches to ML systems design [72], [15], and reflexive data science [9, 11, 59, 99], the work described in this paper more broadly resonates with the call for a radical rethinking of what is to do design under the conditions brought about by digital technologies and the most recent technological advancements in the field of AI [47, 116]. Aligning with recent work that considers "AI as a mirror", we investigate how the training process of AI models could contribute to the designer's self-reflection [82, 124]. Instead of aiming to eliminate the inherent subjectivity of AI, we view it as a starting point for exploration. By encountering the uncertainty of ML algorithms and computational models through the specificity of this project, the work illuminates what may be interpreted as a shift toward relational design practices, that is, practices that move away from ideas of linear progression from problem framing to implementation ([118]) and towards sustained processes of negotiation and reconfiguration [39, 46, 63, 126].

2 RELATED WORK

2.1 Debiasing AI

As the potential risks associated with the use of artificial intelligence in all spheres of public life become more apparent, there has been an increase in discussion and awareness about the expanded role of AI in our society, particularly its effects on marginalized people. Hence, reducing bias in AI systems to improve overall fairness has become a significant challenge. As a result, commercial businesses and governmental organizations, respectively, offer ethical principles and guidelines for the use of AI. Transparency, justice and fairness, non-maleficence, responsibility, and privacy, are the values that are most commonly mentioned. Yet, empirical studies reveal that reports on ethical standards have little discernible impact on software developers' decision-making since they are typically abstract and difficult to operationalize [88]. Ethics can be thought of as a type of "soft law" that has no system of enforcement. Hence, there are no repercussions for breaking any of the various ethical guidelines [49, 66]. Moreover, when examining the methods utilized to preserve and advance such ethical principles, we see that most of them rely on technical approaches like standards or explicit normative encoding [66].

Academics and organizations, however, have cautioned that the adoption of AI could result in the creation or amplification of societal imbalances. They have also advised against putting technology at the center of solutions to this issue, understanding that these systems are interconnected with wider systems of institutionalized oppression. Policymakers have, however, narrowly concentrated on the potential discriminatory implications of AI systems because of biased data and algorithms. Most of the approaches addressing the societal impact of AI primarily stem from the field of computer science and intervene either in the data or in the algorithms to mitigate or prevent biases and harms. As Zhou and colleagues [133] explain, these interventions can be grouped into three main

approaches: pre-processing, which focuses on modifying data before they are utilized for training; in-processing aims to eliminate prejudice at the algorithmic level, by modifying the algorithm itself; post-processing adjusts decision output to reflect a rationale that is fairness-weighted. These types of interventions constitute different ways of "debiasing" either data or algorithms. Balayn and Gürses [8] define debiasing as "the application of select methods to address bias by achieving certain forms of statistical parity (e.g., making sure that the accuracy of a recidivism prediction system is similar for Black and White people by rebalancing a training dataset and re-training a machine learning model)." The prevailing opinion is that the inherent biases in data, which benefit the privileged group when judgments are made, are the root source of algorithmic prejudice. However, relying exclusively on technocentric-oriented debiasing strategies prevents us from truly engaging with the structural prejudice brought on by AI-based systems [8].

The "technocentric" approach is furthermore predicated on the premise that "bias" or technical flaws in the system architecture are the primary causes of the negative effects resulting from the adoption of AI. However, it is not the algorithm that is inherently bad or dangerous [87]. Failures of implementations have suggested that the algorithms pick up on all patterns in the data that they are trained on, necessarily including any of all human biases that guided and influenced the process of data collection and structuring. Algorithms are doomed to reinforce past biases of society [55]. In this sense, these technologies exhibit "function creep" [26]: they promise one thing, but as a result of the design, something else – often less ethical – sneaks in. These things, whether deliberately so or not, are political entities that raise certain ethical and philosophical issues that we must also explore via design [111].

Debiasing techniques run the risk of rendering social, structural, and political issues technical. As such, those challenges are left to the private technology sector to solve, which is far more driven by commercial interests that are not based on value- or principle-based ethics. [8, 49]. Algorithmic bias, then, should be reframed from a mere technical problem to a social challenge. Yet, the complexity of the technical, institutional, and economic arrangements required to introduce AI into the world remains largely overlooked [8].

2.2 Reflexive approaches in art, design, and data science

According to Agre, every technology has a place in a vast network of social practices. Some are used in specific settings, allowing us to trace their impact easily. Others, such as AI, are so widespread and influential across various practices that they pose a challenge in understanding their overall effects on society [3]. Yet, despite this uncertainty, data are increasingly mobilized across society and our everyday worlds, lending themselves to breaking down and producing unintended effects but also (because of this) to creative and critical appropriations [35]. As awareness of the situated and entangled nature of AI workings and inherent biases grows, design is becoming more engaged in conversations and practices concerning the mutually shaping relationship between humans and technology.

Unlike the majority of Human-Computer Interaction (HCI) researchers who view ambiguity in the human-AI relationship as a problem requiring resolution, Research through Design (RtD) practitioners are beginning to see uncertainty as an inherent quality of machine learning materiality [15]. Caramiaux and Fdili Alaoui [23], for instance, illustrate how artists embrace errors and surprises that emerge when engaging with AI and look at unpredictability as an expressive quality. By embracing this quality, artists position themselves as opposed to traditional HCI and industrial research, driven by values of accuracy, productivity, and performance [23]. By getting rid of the rigid norms of engineering AI practices, creative investigations explore unconventional configurations of AI and ML that unveil the underlying assumptions a designer makes about AI and the world [106]. Especially when prototyping with small data, and shallow models, designers produce novel configurations and behaviors that can facilitate the experience and discussion of AI and ML, enabling pluralistic perspectives to emerge [106]. Akten et al. [4], for instance, illustrate how an instrument using live camera input and embedding

an artificial neural network to generate "visual predictions" can be designed with the explicit intent to confront the audience with the idea that a trained model that looks at the world through a camera "can only see what it already knows" [4].

Van der Burg and colleagues [123] further engage with this space on "AI ways of seeing" as a composite process where human interpretation is coupled with algorithmic recognition to generate meaning. What usually is considered a mistake, a misclassification of the algorithm, in the authors' work is rather seen as a site for surprise and generativity. Taking the metaphor of "AI as a mirror", [124] further explores how the process of training AI models might play a part in the designer's self-reflection, taking the inherent subjectivity of AI not as something to eliminate but as a starting point. The reflection witnessed in this AI mirror would be shaped by the choices we made in building it. Relatedly, Hoggenmueller and colleagues [60] reflect upon their exploration of text-to-image algorithms for robot ideation and highlight how, on the one hand, the generated images perpetuate stereotypical ideas of robots embedded in the datasets, while on the other hand also help surface the designers' expectations and pre-conceived ideas especially when encountering what the authors describe as "pleasant surprises", instances that break free from traditional robot imaginaries by yielding "some form of robot hybridity, i.e., the blending of robot characteristics or functionalities with unrelated concepts in novel ways" [60]. The notion of hybridity as a result of creative engagement with AI is also central to the work of Turtle [122], who engages with disciplinary critiques of AI norms where queering is explored as a strategy to "subvert existing sociotechnical systems and codings of hegemonic worldviews that reinscribe unsustainable, unethical and apolitical design practices". Here the author leverages their first-person experience of experimenting with AI to argue how notions of hybridity and mutation can help us contest dominant and stagnant framings of AI as something separate from the social, cultural, and biological systems.

These few examples show how creative and designerly perspectives in HCI research are opening the discourse surrounding human-AI collaboration to alternative models of inquiry and viewpoints. These, often grounded on concepts and approaches from the social sciences and humanities, such as intersectional feminism, take a distinctively critical stand and advocate for more contextualized, inclusive, and accountable design of AI and ML [72, 92]. The conversation on AI limitations and biases is enriched with a strong focus on the personal and subjective experience of the researcher designing and using the algorithms, following the overarching belief that "the reality we perceive in our minds is not a mirror image of the outside world, but a reconstruction based on and limited by our physiology, expectations, and prior beliefs" [4].

These approaches possess a distinctive character that contributes to the extensive body of critical and reflective Human-Computer Interaction (HCI) research developed in recent decades in response to an increased understanding of the interconnected relationship between humans and artificial entities [39]. There has been a focus on the potential role of design in fostering reflection [37, 101], enabling HCI to encompass various facets of human experiences, including emotion, family life, sexuality, culture, and religion [11]. There has also been a call for reflexive practices in HCI that involve rethinking dominant metaphors and values. In these, researchers adopt a critical stance, delving into the political and ethical implications of the design process [32, 107, 108]. These reflexive design approaches illuminate the unconscious practices, values, and identities embedded in our technological designs, resonating with the more recent critical discourse surrounding AI data collection [24, 68, 100].

In creative explorations of AI and ML [4, 23, 106] and the expanding realm of RtD practices that actively engage with data and artificial intelligence [77, 91, 122–124], the specific emphasis on reflexivity not only acknowledges but also encourages the recognition and mobilization of humans as interpretative entities shaped by distinct social, cultural, and individual contexts. This mobilization aims to influence our understanding of these technologies as well as their application to the creative process. While creative approaches show promise in addressing issues related to biases in AI and ML, potentially diverging from technocentric ideas of debiasing, further investigation

is needed to understand how biases and stereotypes are integrated into the collective imaginaries that AI relies upon and what scope designers have to work with them, or better *through* them.

3 METHODOLOGICAL APPROACH

In the pursuit of a reflective method for curating data ethically and responsibly in the everyday application of AI tools, this study aims to shed light on how the convergence of design practices and reflective approaches in art and HCI can offer insights into addressing the issue of AI biases. Methodologically, it adopts a uniquely exploratory and creative standpoint, diverging from conventional techno-centric approaches like debiasing. Instead, it embraces the inherent biases and uncertainties associated with prototyping using data, algorithms, and models as design materials.

The article provides a comprehensive account of a data-enabled Research through Design (RtD) investigation in which the process of making has been approached as a site for introspection. In particular, we build on related positions from the HCI field [4, 71, 122] to approach the crafting and interpretation of data as an intimate and reflexive process that is "not only critically grounded in a designer's own situation but also offers modes of imagining the world otherwise." [71] (page 1). Building on this epistemological foundation, we welcomed the inherent uncertainty involved in prototyping with data, algorithms, and models. This allowed us to transform the issue of biases in AI datasets from a problem to be eliminated into a quality to actively engage with.

3.1 RtD process and reflective model

We approach this inquiry with the methodological perspective of RtD, wherein the practice of design serves as a means to pose broader questions that extend beyond the confines of a specific design problem [134]. Additionally, RtD serves as a means to complement traditional scientific research, especially when traditional methods struggle to capture and address the potential impact of emerging technology [135] and navigate the more imaginative and uncertain boundaries of design [96]. As a practice, it can be applied to contexts that are selected for their theoretical and topical potential [42]. In this instance, we specifically selected the context of binary gender categories in children's toys. This deliberate choice was made to bring the integration of social norms and stereotypes into products to the forefront. It enabled the practitioner (i.e., the first author) to actively and reflexively examine how collective imaginaries are ingrained in her preconceived ideas and biases. This examination involved the conscious use of data curation techniques, including classification algorithms and generative adversarial networks that depend on these collective imaginaries.

This RtD process unfolds as a *reflexive inquiry* concerned with matters of biases in the context of human-AI collaboration. Within the paper, we examine the design artifacts produced through this process, considering them as embodiments of the first author's assessments regarding the implicit possibilities and tensions within the given context. We engage in reflection on these assessments to articulate insights across topical, procedural, pragmatic, and conceptual dimensions. By employing Schön's reflective model [103], we examine how the practitioner's reflective process unfolded *in-action*, seamlessly integrated into the design experiments and reflecting the first author's perspective, and *on-action*, occurring retrospectively to the design process and its outcomes. Initially conducted by the first author, this reflection later illuminates the shared interpretation of the inquiry space by all three authors.

Drawing upon the ideas of Dewey, and particularly the concept of critical inquiry and experiential learning [31], Schön's work builds upon and extends Dewey's ideas, emphasizing the importance of reflection and conversation in professional practice, including the design situation [103]. By extending Dewey's concepts into the specific context of professional design, Schön emphasizes the ongoing, adaptive, and reflective nature of the design process [104]—attributes that form the methodological point of interest in this work. He describes the internal processes of *reflection-in-action* as a conversation with the materials of a situation. This conversation allows

the designer to shape and reshape the situation or activity on which they are working while it is unfolding. It is generally associated with the experience of surprise. By conducting experiments and being responsive to unexpected and unintended changes, designers generate both new knowledge and an intentional change in the situation at hand. *Reflection-on-action* is instead the exploration of what happened in that particular situation after a situation has occurred. Reflection-on-action is often associated with reflective writing.

We leverage this epistemological grounding, and the reflexive principles specifically, to bring forth the argument that data-enabled Research through Design (RtD) holds great potential for encouraging and supporting designers' reflexivity. To illustrate and unpack this potential, the first author set up design experiments focused on developing *queer toys* as sites for reflection-in-action that were further reflected upon by the three authors to articulate pragmatic and conceptual insights for the HCI community to use. These insights are offered and discussed in section 5 and section 6, respectively. But why *queer toys?*

3.2 Children's toys and gender as a design context

Queer toys were chosen as a design context within which the practitioner would conduct her experiments. Toys are cultural signifiers [5] of contemporary society norms, values, and stereotypes. They represent the ultimate gendered products: the vast majority of toys are designed, advertised, and commercialized, targeting either boys or girls, leveraging historically established social norms and attitudes prescribing separate spheres for women and men [36]. Bold and dark colors are typically attributed to toys marketed for boys, whereas pastel, delicate colors are used for the ones intended for girls [7]. Typologies of toys also perpetuate stereotypical attribution of social roles, such as girls' toys being associated with physical attractiveness, nurturing, and domestic skills, while boys' toys support more competitive, exciting, and adventurous activities [17]. Only very recently, toy companies started searching for new narratives that somehow reflect new social realities, from new strong female Disney heroes to Barbie advertisements presenting new female futures [5]

In choosing to design queer toys, the first author set out to dismantle the binary categories of femininity and masculinity, but also, and foremost, to confront her own gender biases and the collective imaginaries embedded within them. This viewpoint should be understood within the socio-cultural context of the first author, specifically situated within a Western framework. We conceptually draw on a growing body of research engaged in similar endeavors, especially from the queer HCI community, that advocates for "queerness" as a meaningful framing to understand all things that are "indeterminate, ambiguous and always in relation, denoting flexible spaces for the expression of all aspects of non, anti, contra, straight, cultural production and reception" [74, 112]. Particularly relevant to the scope of this article is the work by Turtle [122], who illuminates how adopting queerness as a conceptual framework for designers engaged with AI can create opportunities to celebrate diverse and unconventional possibilities. This approach provides more responsible, inclusive, and liberating models for engaging with AI. Queerness, in this context, establishes a space for liberation, experimentation, and confrontation—offering a generative and instrumental foundation for the objectives outlined in the presented work.

A queering approach brings a radically different perspective into matters of gender and AI. Traditionally, AI has used the identification of an individual's gender based on facial features and behavior, distinguishing between masculinity and femininity, for various applications, such as video surveillance, law enforcement, demographic research, online advertising, human-computer interaction [76], as well as in security and forensic evidence collection [28, 129]. These practices were found not only ethically questionable but also inadvertently contributing to perpetuating biased decisions, even when the underlying computational processes were well-intended. Recent studies, for instance, highlight significant inaccuracies in gender classification systems, particularly when it comes to dark-skinned female faces [21]. In this pursuit, our focus is directed toward scrutinizing the practitioner's implicit assumptions, conscious judgments, and creative attempts to challenge traditional norms in the design of

children's toys. These actions occur as the practitioner enters into a reflective conversation with the collective imaginaries arising from interactions with the AI as design materials [104]. We are not delving into the inherent framing of gender by AI systems.

3.3 Situated data analysis

When we create and experiment with situated artifacts–things intended for a specific use scenario, the research process can open new opportunities and leave room for unexpected events to be considered in the design [103]. This is fundamental to understanding the role of artifacts in RtD [114]. However, the empirical realities of the world in which queer toys are situated are not observable phenomena in a conventional sense [115]. They are intimate, complicated experiences—thoughts and feelings of the first author that are entrenched deeply in data, computational models, and the collective imaginings that feed them, and that unfold and become observable in interaction with the AI.

Every step in the unfolding of the experiments has been captured and annotated by the first author on a Miro board. This choice was made in an attempt to situate the experience and support reflection both *in* and *upon* action. Visual mapping and interpretation were employed to make sense of text, image, and video data. This approach was undertaken to counter the prevalent issue of insufficient documentation, which frequently impedes the rigor in Research through Design (RtD) practices [18]. As argued in [2], RtD can be "hindered by the lack of any fundamental documentation of the design process which produced them. Too often, at best, the only evidence is the object itself, and even that evidence is surprisingly ephemeral" (p. 121). Nonetheless, it is important to be clear on the nature of the documentation that can be produced. This is not hard evidence but a discussion starter, next to the designed artifacts themselves. As mentioned, design experiments are a conversation with the materials of a situation, and hence intrinsically reflective [103]. They are based on what is referred to as "learning-by-doing" [102]. Because of this, large parts of the knowledge on which design is based remains "tacit" [128]. Tacit knowledge cannot be communicated via words, and though it can be made conscious, it usually remains hidden [84, 103, 128]. The experience, however, may be conveyed and represented in certain ways, such as through written accounts and other forms of representation of the subjective insights and understandings that were generated in relation to specific circumstances [61]. Such intermediate forms of knowledge can range from annotated portfolios [41] to conceptual constructs [117], experiential qualities [78], and strong concepts [62]. According to [61], however, the field still has difficulty communicating information and ideas related to aesthetics, politics, ethics, and other intangible elements in design practice.

The designed artifacts in this project are representations of hybrid forms of human-AI practice in uncertain and unpredictable design spaces [44, 110]. In and by themselves, the designed artifacts do not speak—they cannot express the knowledge gained through the process in its entirety. Together with the written, reflective accounts in the Miro board (by the first author) and in this article (by the three authors, collaboratively), they serve as conversation starters for a discussion on reflexive human-AI interaction and the data curation and tactics that can be crafted to mobilize it. These written accounts reflect in a structured way the introspective, iterative processes of tracking and reflecting on one's thoughts, mental images, sensations, and actions [130]. They build on the understanding that, as human beings, we are always able to relate and combine what we have previously learned with the new materials of a situation [34]. But they also acknowledge that our experiences are infused with collective expectations and cultural norms [1]. Through reflection and introspection, the first author attempted to displace perception, surface internalized values and norms, and better understand biases and fixations. In further reflecting together on the first author's human-AI creative practice and its outcomes, we entered into a conversation with the personal, cultural, and historical relations in which the self, the AI, and collective imaginaries are positioned and entangled.

In this sense, our RtD process, methods, and forms of documentation align more closely to situated data analysis in reflective and critical HCI (and the feminist epistemology on which it draws) than to annotated portfolios and other intermediate forms of knowledge. As nicely pointed out in [72] in HCI, drawing on the seminal work of Donna Haraway in the field of science and technology studies, knowledge production is a "partial, situated, embedded and embodied" process that is implicated in a broader nexus of power relations [51] and that involves many actors, both human and technological [54], none of which is either objective or neutral [52]. "Situating" data, algorithms, and computational models as a method [33, 67] "allows contextual factors to be taken into account in a more systemic way and also for technologies to be developed that are more accountable for their own specific partial ways of generating knowledge" [72] (p. 1529). In the context of RtD, it allows embracing uncertainty as a design opportunity rather than an obstacle [15]. Grappling with "design as a probabilistic outcome" in algorithmic forms of RtD means that the artifact assumes an agential role that "profoundly challenges the stabilizing character of the prototype inherited from previous practices of skillful material crafting and industrial design manufacturing" RtD [44] (p. 145). The queer toys in this project become more than a material embodiment [117] or a manifestation [75] of the ideas, skills, and knowledge of the designer, for others to experience—a sketch, a mock-up, or polished material outcome confronting the world of ideas and skills of the designer with the world-out-there before a final artifact exists [19]. They are the expression of fundamentally situated, reflexive, and recursive interactions between the first author and the AI. We refer to it as *reflexive human-AI practice*.

4 DESIGN EXPERIMENTS

In this section, we describe three design experiments oriented to the creation of queer toys: the *robo-doll*, the *dino-unicorn*, and the *hair-drill*. In these experiments, the first author used various off-the-shelf AI tools to support the ideation of the toys and explore and reflect on the unintentional biases and uncertainties of machine learning (ML).

To resist the idea of binary gender distinctions, *queerness* was used as a conceptual framing to design with because of its capacity to denote "flexible spaces for the expression of all aspects of non, anti, contra, straight, cultural production and reception" [122]. As a non-normative stand, queerness is particularly suited for imagining alternative futures and configurations –including toys– where the emphasis is "less about expanding a range of choices (liberal freedom) than it is about transforming the kinds of beings we desire to be while embracing the [..] multitudes and entangled lifeworlds that make up the pluriverse'[122]". In partnering up with AI for creative purposes, the ambition here is to intentionally get lost in ambiguity, as a way to redefine what the "masculine," the "feminine," and the in-between and beyond mean to the first author. The experiments, then, provided a platform for the first author to explore distinct introspective strategies to build reflexivity through human-AI collaboration. She deliberately manipulated ML data inputs and outputs to surface personal biases and used ML uncertainties, biases, and ambiguity to explore queerness in toy design. Because of their iterative and experimental nature, these introspective experiments provided space and time for understanding whether and how AI limitations would help the first author build a critical stand toward her own assumptions and worldviews.

4.1 The robo-doll

In the robo-doll experiment, the first author explored how to design a toy that would challenge the binary ideas of gender as represented and materialized by the association *Robot = Masculinity* and *Doll = Femininity*, using Teachable Machine [81]– a web-based ML model– as a mirror to her own assumptions.

The design process began with the sketching of initial ideas (Figure 1, left) as a way to understand what features and choices could contribute to perceiving the artifact as gendered. Through this exercise, the first author built awareness of the role that the proportions of the different body elements have on her personal perception

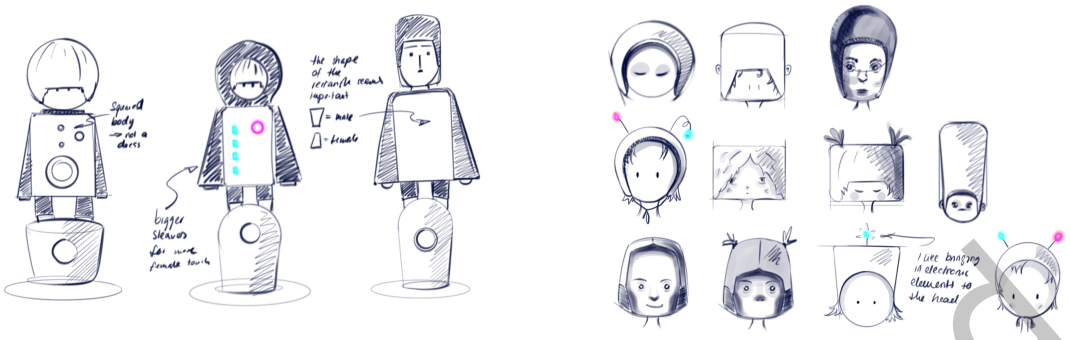

Fig. 1. These early design sketches show the initial explorations of the first author's gender associations with shapes. The sketches on the left depict early brainstorming ideas for the robo-doll with the first author's annotations of their perceived gender. On the right, more detailed studies of robo-doll heads, as they were identified to convey most gender cues.

of the toy's gender. For instance, a rectangle with a wider top than bottom was perceived as wide shoulders associated with masculinity, while the opposite proportions – wider bottom compared to the top– were perceived as dress and associated with femininity. In addition to body proportions, the head was also identified as a critical component in conveying gender cues. Therefore, the head was studied more in detail through further sketching (Figure 1, right). The first author's reflection and articulation of personal assumptions regarding gender further expanded from matters of shape to also consider the materiality of the object. On the one hand, the robot's electronic components were seen as a manifestation of masculinity, while the warmth and naturalness of the wood were seen as feminine. These considerations were used to mix and match morphological and material features with the elements to be designed for the character toy composition (elements shared by both the robot and the doll, such as the head, chest, legs, arms, and so on).

The very choice of using a classification algorithm may sound inappropriate for the scope of this work, that is, to challenge the narrow social categories of femininity and masculinity. As a matter of fact, a classification algorithm is a function that balances the properties of the input so that the output divides one class into positive values and the other into negative values. To find the weights (and functions) that best separate the two classes of data and offer the most accurate results, a classifier is trained on labeled data such as images [90].

Classification algorithms, thus, necessitate assigning labels to specific objects or images. In the case of teachable machine, the classification algorithm employed here learns through pre-trained MobileNet models. This can lead to the algorithm picking up on problematic patterns like gender stereotypes due to unevenly distributed training data or biases stemming from biased human selection and labeling [1]. This might seem like these classification algorithms are inherently unsuitable for designing "queer things"–toys in this case. Yet, in contrast to current classification practices, which were found to perpetuate and exacerbate gender bias, we emphasized friction as

[1]It's important to acknowledge that biases evident in interactions with the AI tools utilized here stem from both the training data of the first author and the data used to pre-train these models. Nevertheless, this doesn't prevent introspection, for two reasons: Firstly, personal gender biases often mirror broader societal norms, meaning biases from pre-training are likely to align with those of the first author. Secondly, the first author aimed to confront the subconscious gender associations inherent in her designs. The catalyst for this reflection is irrelevant; what matters is engaging in the process of labeling data herself, which serves as a crucial step in personally confronting the outcomes of these models. This principle applies to all experiments outlined here.

Fig. 2. This dataset created by the first author comprises images and objects ranging from concrete and context-specific objects like toys to more abstract depictions of landscapes and art that were found online. When making quick and intuitive decisions, the first author divided these images into two distinct categories: "masculinity" and "femininity."

opposed to efficiency and seamlessness. As a consequence, aspects like queerness and neutrality can evolve from an interpretive process of the human about the frictions and biases arising from the classification algorithm.

The very process of data labeling was engaged by the first author as an opportunity to critically examine her own thinking about these social categories, thus as an opportunity to question narrow and tacit assumptions. Furthermore, in the process of designing the robo-doll, labeling was experienced as a deliberate action first, when *teaching the algorithm*, and as a judgment after, when *running the algorithm*. Initially, in fact, the first author would deliberately teach her own gender biases to the algorithm by providing image data labeled in gender binaries, masculine vs feminine (Figure 2). She selected a collection of images depicting toys but also other objects and more abstract visuals see Figure 2. About 60-80 images were obtained from internet searches on non-gendered search terms such as "architecture," "art," or "household tools." Additional searches along the lines of male and female gender were done to obtain images of "boy toys" and "girl toys." Roughly 20-30 images per category were then generated through the camera record function in "teachable machine" itself, holding objects the first author collected from her personal environment in front of the camera. Machine bias (machine learning unwanted patterns from the images) was attempted to keep low by training the classification algorithm on images of hands in both categories of masculine and feminine (as these were shown on images that were self-recorded and should not be used to make a distinction of masculinity or femininity), and deliberately selecting images from the internet that had plane white backgrounds. These images were then labeled individually and intuitively based on whether the first author perceived them as having more feminine or masculine qualities. This process resulted in a binary data set of masculine and feminine images. The online platform "eachable machine" was then trained on these selected and categorized images. After that, the first author would confront herself and her design ideas by running the algorithm and showing sketches of the first three robo-doll concepts (the algorithm would "see" the sketches by having them uploaded as jpg in the online ML environment). This last step, in which the first author intentionally challenged herself, was repeated several times, each time presenting the algorithm with variations of the robo-doll design.

The first author iteratively edited design features by looking at how the algorithm would classify the sketches in terms of feminine and masculine categories. Her ambition was to get to a 50%-50% ratio between the two categories, as she assumed this would be the point where the highest ambiguity–and queerness– would be achieved. Following this assumption, the first author noticed how the combination of a more masculine body

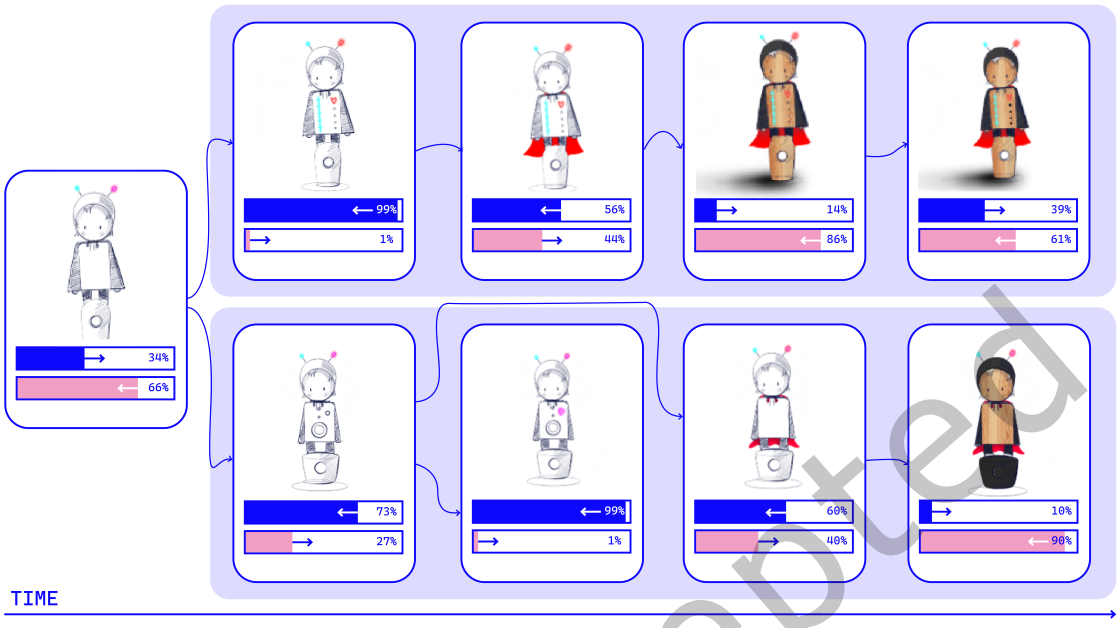

Fig. 3. Evolution of design sketches towards more gender ambiguity as a result of the first authors' experiment with the classification algorithm. Each design idea is presented to the algorithm, creating a measure of femininity and masculinity based on the author's personal perception of gender. Aiming to reach a high ambiguity (50%-50% ratio between categories), too masculine or too feminine designs are being discarded or countered with colors or shapes associated with the opposite gender. As the sketches evolve through each test, eventually, some texture and shading are added.

with a more feminine head would lead the algorithm to assign the more "balanced" classification results, that is, a ratio of 34% masculinity and 66% femininity.

Iterations on the sketches and confrontation with the algorithm also functioned as a way for the first author to scaffold the logic of the algorithm. Understanding the "what" in a picture gets picked up from training images and then becomes a metric for recognizing gender attributes, which is anything but straightforward. So the first author progressed in her sketching iteration, increasing the level of details little by little and focusing more and more on specific details to test whether these would play a significant role in the classification. She incrementally made changes to each design variation to stir a design evolution guided by this understanding. Some iterations showed lower ambiguity than the previous design, and in those cases, she reverted to the previous step and tried a different design variation (see Figure 3).

After selecting two sketches with promising ambiguity (close to the 50%-50% ratio), the process shifted from sketching to physical prototyping (see Figure 4). To capture both the three-dimensionality and texture of the designs, she created an easy-to-assemble prototype that allowed for quick changes in the scale and size of the body parts, such as arms, legs, and chests. Several design variations were laser-cut, sanded, and painted. While some pieces could simply be snapped in place, others were held together by magnets, allowing for the rotation of arms and head and a quick iteration of design variations.

The various alternative body parts and configurations were then systematically tested with the classifier (by showing them through the camera view connected to the online model), as illustrated in Figure 5.

Fig. 4. Prototyping different design variations. Design elements such as the pedestal, capes, corps, arms, and legs were laser cut, glued together, and composed in different variations to be tested with the classification algorithm.

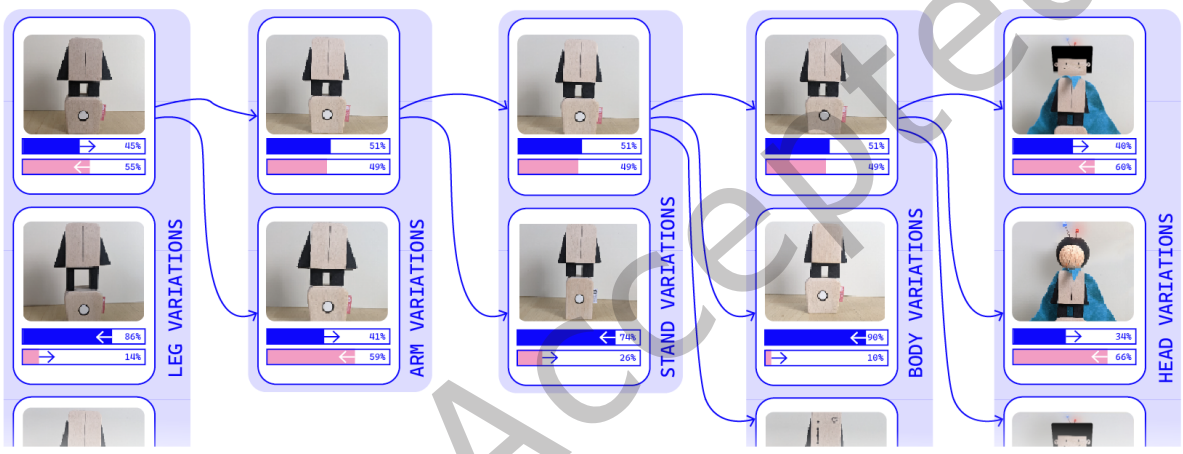

Fig. 5. Further challenging prototyped design variations in front of the classification algorithm. For each body component, all possible variations are tested. The most queer design of each of those body categories was then chosen and used as the baseline for the next iteration.

While the final results did not yet show the desired level of ambiguity, the experiment's setup allowed the first author to identify the elements in her design that needed improvement. As seen in Figure 5, for example, the cape showed to be a less ambiguous feature than intended. This invited the first author to critically question her choice of adding a cape to the design.

Ultimately, the reflections of the first author mediated by the classification algorithm led to the choice of a combination of body parts deemed the best in terms of ambiguity. This configuration, developed into a tangible prototype, was completed with the addition of electronic components, as depicted in Figure 6.

4.2 The Dino-Unicorn

In the second experiment, the first author focused on the stereotypical depictions of gender in plush toys, specifically the *Dinosaur = Masculinity* and the *Unicorn = Femininity*. Despite the same goal of using ML to challenge her binary ideas of gender, unlike the first experiment, the first author here used a StyleGAN instead of a classification algorithm. A StyleGAN is an extension of a progressive generative adversarial network (GAN),

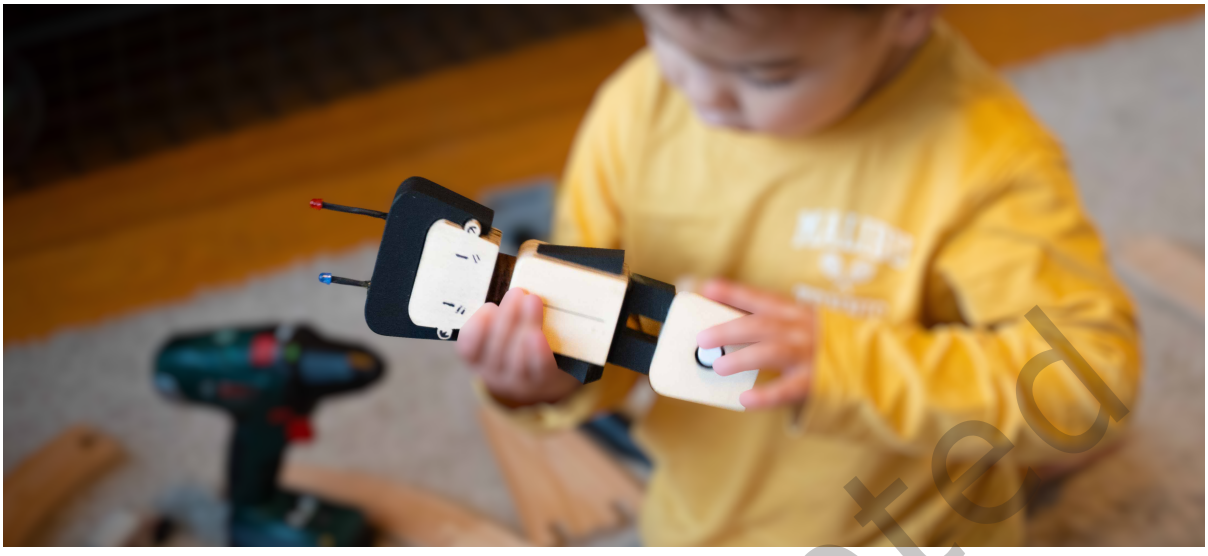

Fig. 6. The resulting queer robo-doll toy inspired by the interactions with the classification algorithm.

an architecture that allows the generation of high-quality and high-resolution images. The objective of the experiment was to understand whether using the StyleGAN's ability to generate infinite and interconnected images would help envision more intermediate configurations between the two initial references (the dinosaur and the unicorn).

Fig. 7. The first author's selection of data for the dino-unicorn experiment. The internet is scraped for images of dinosaurs and unicorn stuffed animals that fit the first authors' perceived associations with femininity=unicorns and masculinity=dinosaurs.

Similar to the first experiment, this exploration started with the first author creating a data set comprising images of dinosaur and unicorn plush toys (Figure 7) as material for teaching a StyleGAN on her personal bias. The images for training were again obtained from internet searches on search queries such as "unicorn plush" and "dinosaur plush." No additional notions of gender were used in the search terms. A scraping algorithm was

used to obtain an initial data set containing images of unicorns and dinosaur plushes. The first author then deliberately chose images for each category from these initially scraped images that matched her understanding of femininity in unicorns (e.g., color pink) and masculinity in dinosaurs (e.g., darker colors, sharp teeth, or spikes). Again, attention was paid to only selecting images with white and neutral backgrounds and no hands or other irrelevant objects in the images. Because of the small number of images that were obtained through this process (roughly 100 per category, in contrast to the classification algorithm, which can provide classifications on smaller amounts of data (even if then very much over-fitted), styleGANs require larger amounts of data to generate interpretable images), the first author used a python script for data augmentation (rotating, coloring or flipping images) in order to enlarge the amount of suitable data. By selecting pictures of existing plush toys that strongly reflected her ideas of masculinity (dinosaurs) and femininity (unicorns), the first author eventually collected approximately 300-400 images in total and used *RunwayML* [98] to fine-tune with these a Style-GAN pre-trained on bird illustrations (deemed the most suitable model among the ones available).

Fig. 8. The first iteration of training results (left) shows high ambiguity and little similarity with the initial categories of unicorns and dinosaurs. After additional training time, the second training iteration (right) shows more similarity with the training data and is thus deemed more appropriate for the scope of the exploration.

The fine-tuned StyleGAN was then used to generate images. These would show some similarities with the training data but were not yet identifiable as plush toys (see Figure 8, left). To explore the potential for generating more ambiguous and challenging creatures, the first author tested multiple times this first iteration of fine-tuning by exploring the generated images in the latent space [2]. While some of these explorations resulted in intriguing and unexpected outcomes, the generated creatures either lacked femininity or were too far removed from the characteristics of both dinosaurs and unicorns.

This led the first author to run a second iteration for fine-tuning the model, which consisted of additional hours of training on the same data, which would improve the similarity between the training data and the

[2]The latent space is a representation of compressed data, where alike data points are located near to one another and less alike data points further away from each other [120].

Fig. 9. This graphic captures the latent walks between dinosaurs and unicorns, visualizing the space between the binary categories of femininity (unicorn) and masculinity (dinosaur). The first author deliberately chose images from the latent space that strongly reflected her ideas of masculinity and femininity to interpolate between.

AI-generated images. The quality of the images resulting from this iteration was higher than the previous and deemed appropriate for the scope of the exploration (see Figure 8, right). Further proceeding with the testing, the first author looked again into the latent space in search of images that would represent the most in-between creatures. Although results were perceived as slightly less surprising than the one from the first iteration, the images were less "alien" and, thus, easier to make sense of. Furthermore, by tweaking the latent space's parameter, the first author could illustrate more or less similarity between neighboring images. The first author employed the latent walk –the process of sampling a location in the latent space and gradually altering the latent representation [119] to help herself visualize the vast number of images that lay between the two categories, as illustrated in Figure 9. The images collected from the latent space and the latent walk exploration were collected into an inspiration board and used for ideating the dino-unicorn toy. Based on these, several design ideas were sketched and self-assessed, with the final design chosen for its perceived hybridity between the features of a unicorn and a dinosaur, thus a high ambiguity (while remaining relatable to both categories). Based on these choices, a layout paper model was prepared for prototyping, together with a set of choices regarding the use of colors and fabrics to emphasize the interplay of categories. Dinosaur-like elements were rendered in feminine colors, while unicorn features were colored in darker, more blueish hues. The fabrics were then cut out and sewn together into the final dino-unicorn plush toy prototype (Figure 10).

4.3 The Hair-Drill

Inspired by the dino-unicorn design process, the third experiment set out to explore even more the AI's ability to queer between different categories. The idea was to diverge from the initial category of masculine toys by

Fig. 10. The resulting queer dino-unicorn toy inspired by the latent space and latent walk of the StyleGAN.

Fig. 11. The data set for the third experiment consists of a primary category comprised of images of toy drills that are associated with masculinity and two secondary categories associated with femininity with images of hairdryer toys and vases.

infusing the stereotypical concept of a drill toy with attributes from objects usually associated with femininity, such as hairdryers. The model used for this experiment was again the StyleGAN from RunwayML [98]. However, this time, the first author used a different approach to the selection and use of the data.

To begin, the first author created a primary dataset consisting of images of toy-drills. Next, she compiled a second batch of data featuring images of toy-hairdryers. Similar to the previous experiment, these images were obtained from internet searches on search queries such as "toy-drill" and "toy-hairdryer." Again no further gendered terminology was used in those queries. Eventually, these two categories of images were selected because of the similar morphology and some shared interaction affordances (e.g., the way the toy affords a certain hold)

to the drill that coexists with the stereotypical association of the toy with the opposite gender (see Figure 11). The secondary dataset was slightly smaller in size compared to the primary dataset. Nonetheless, both datasets comprised hundreds of images of plain objects. To facilitate the data collection process and enrich the datasets, the first author again used the data augmentation program developed in Python that enabled her to generate multiple versions of the same data by rotating, flipping, and coloring the images. This program allowed her to generate sufficient amounts of clean data in a short amount of time and provided her with greater control over the data than data scraping [3].

Fig. 12. StyleGAN generated images of drills based on the first training round with the masculine image data of toy drills.

Next, the first author trained a StyleGAN from RunwayML on the primary dataset and tested the training results by generating a series of images. These, shown in Figure 12, exhibit drills similar to those present in the primary dataset.

The first author then proceeded to re-train her algorithm, previously trained on drill images, this time adding the secondary dataset, which contained images of hairdryers (see Figure 11, "femininity 1" dataset). At this stage, she limited the training session to a brief period (only half an hour). While still possessing qualities that would recall the image of a drill and maintaining the affordances of holding and directing the object, the context of use started to fade away, creating ambiguity (Figure 13, left). In the first author's view, the generated images were inspiring and generative: these could be looked at as meta-objects (or meta-toys) intended to train specific motor skills, for instance, instead of implying a stereotypical use case. Compared to the previous experiments, this process stimulated reflections on how "queering" toys could be more than an intervention in appearance and aesthetics, it could also open opportunities for breaking free from constraining ways of playing while generating new configurations where much is left to the imagination.

To further engage with this idea of subverting not only the image of the drill but also its associated form of play, the first author started exploring other categories of objects, textures, and features as ways towards hybrid forms of play. She conducted a third training round in which she re-trained the model originally trained on images of drills, to also add images of glass vases as secondary data (see Figure 11, "femininity 2" dataset). The assumption behind the use of this dataset was that glass vases can be associated with feminine characteristics

[3]data scraping (or web scraping) is the act of extracting "information from one or many websites and to process it into simple structures such as spreadsheets, database or CSV file" [69]

Fig. 13. On the left, the StyleGAN-generated images, which are based on six hours of training with drill-toy images and half an hour of training with hairdryer-toy images. The images illustrate how the masculine features of the drills fade away by adopting feminine features from the hairdryer dataset. On the right are the StyleGAN-generated images, which are on the same training with drill-toy images but with an additional half-hour training of vase images. While the images generated exhibit a strong visual resemblance to drills, they no longer afford to grasp and press a button.

such as sinuosity, fragility, and passivity, which are visually encoded in their appearance. Moreover, vases possess defined contours that are distinct from simple fur or glitter textures, making the training process and the outcome more readable. The images generated after this training exhibited a strong visual resemblance to drills, yet the original function was made impossible, especially by the sinuosity of the shapes contrasting with the rigid geometry needed by the tool to function (Figure 13, right). Despite the interesting results, these images still failed to convey the fragility and femininity associated with glass vases that the first author aspired to achieve. Therefore, she decided to discard this approach as she felt it did not contribute enough to the overall aim of conveying ambiguity.

She returned to her initial design experimentation with the hairdryers' datasets and further explored the idea of creating hybridity not only in the typology of toys but also in the modality of play. To this end, she moved from generating images to prototyping physical components. Somewhat similar to the first experiment, the first author looked for what archetypical elements would be necessary to convey the desired affordances and which features would be helpful for blending the identity of the toy. This led to the definition and prototyping (mostly laser-cut plywood) of simple volumes and abstracted details, such as the hairdryer's grid and fan were simplified and abstracted in shapes that, when added to the solid body that would slightly remind the drill, would create a sense of estrangement. The final drill-hairdryer prototype (Figure 14) is an object that feels familiar (at least to the authors of this article) and invites a clear interaction modality (grasping and directing), yet it is hard to describe in terms of the context of use and known typological categories.

Fig. 14. The resulting hair-drill toy is inspired by the merging of drill and hairdryer images with a StyleGAN.

5 REFLECTIVE EXAMINATION OF THE DESIGN EXPERIMENTS

This section presents a first-person account of the three design experiments. It builds on Schön's idea of "*reflection-on-action*" [105], wherein the designer takes a moment to retrospectively examine their actions. Additionally, it embraces the feminist viewpoint that emphasizes the significance of elucidating personal and contextual perspectives as a means of comprehending experiences [30]. In the process of co-creating with AI, the first author engaged in intuitive and implicit "*reflection-in-action*" [105]. Here, in collaboration with her co-authors, she retraces her thought processes, establishes connections with the results of the explorations, and articulates her lived and felt encounter with AI as a creative partner.

5.1 The allure of simplified classifications

The emergence of powerful computational techniques for data labeling and classification has fostered the misleading notion of algorithms being inherently "objective" [48]. Adopting a divergent strategy, the first author intentionally participated in the process of selecting and labeling data, imparting her individual unconscious perspective on gender to the algorithm. In essence, she deliberately introduced bias into the data to mirror her gendered perception of the world, actively engaging with the algorithm's subjectivity rather than dismissing or neglecting it. Assigning gender to everyday objects, colors, and shapes became a potent method that compelled the first author to surface and confront her preconceived ideas and biases related to societal categories.

Initially oblivious to the depth of implicit biases influencing her swift and habitual decision-making, she was *confronted with how easily she assigned gender, even to abstract objects like art, within mere seconds.* This initial phase of data selection and labeling presented a chance to delve into the fundamental assumptions and potential biases embedded in the cognitive processes of classification. This realization prompts contemplation on the widespread manifestation of such a cognitive process. In her exploration, the first author was surprised by how easy it was to force categorizations also onto "unreasonable" features. For instance, assigning architectural examples to a particular gender revealed a systemic issue. Unknowingly, we all partake in this mechanism on a larger scale, automatically categorizing things or individuals based on visual features, even when there is

no real correlation between those features and, for instance, a certain identity or behavioral trait. In creative explorations, this may lead to pleasant surprises. Yet, in real-world applications, this too often opens the way to errors, including the possibility of infringing human dignity (see the example of Google [132] and Facebook [80] algorithms classifying black people as "gorillas" and "primates" respectively).

Engaging with one's own preconceived notions and biases provided the designer with an avenue to encounter these invisible cognitive mechanisms and cultivate a sensitivity to identify when they come into play.

5.2 Reclaiming control through data curation

This first step of the process (in all three explorations) proved to be both intense and time-consuming, as numerous images needed to be searched on the internet for most algorithms. Additionally, the selection process was not as straightforward as copying and pasting the first 200 entries for a particular category, such as unicorn plush toys. Through several rounds of trial and error, the first author discovered that the algorithms were highly sensitive to image content and struggled to differentiate between background and foreground elements. Following this understanding, the author started pre-processing her chosen images to remove unintended effects she had previously identified, such as hands in pictures or busy backgrounds.

Training times varied significantly, ranging from mere seconds for classification using Teachable Machine (first experiment) to 7-8 hours for StyleGANs in RunwayML (second and third experiment). The first author found this extended training time to be a hindrance to the otherwise swift iterations in the traditional design process. It became apparent that problems associated with unintended biases in the data surfaced only after several hours of training.

By documenting and reflecting on the complexities and challenges encountered during the data selection and labeling process in the Miro board, the first author gained a deeper understanding of the intricate interplay between (un)intended biases, classification algorithms, and design decisions. For instance, after scrutinizing a significant volume of data, all sharing common characteristics, the primary author started recognizing inherent patterns concerning shapes and colors linked to gender. An example of this is the association of the color red with femininity in the context of child-toy design. This insight subsequently guided the later phases of the design process, facilitating more informed and critically engaged iterations with the ability to navigate toward specific outcomes.

Delving deeply into data work, the designer extracted valuable insights into the prevalent gender-related associations within the dataset. This not only confronted her personal biases but also challenged broader societal norms and imaginings that influenced her gender-related associations. This invites us to acknowledge the importance of engaging with technology and its functioning as a means to critically comprehend and influence its dynamics and limitations. This perspective diverges significantly from traditional data labor, where workers are typically tasked with swiftly labeling data without understanding how their actions impact AI operations. In this context, data work emerges as a path toward literacy and an avenue for asserting control over algorithmic behavior.

5.3 Surfacing and questioning norms and beliefs through reflexivity

Examining the patterns within extensive image datasets served as an indirect method to make the designer more aware of biases embedded in her decision-making. Nevertheless, after the initial experiment involving a classification algorithm, quantifying abstract concepts like "masculinity" and "femininity" made personal gender biases more tangible and provided actionable insights. These AI-generated measures not only guided the decisions of the first author but also subjected her ideas to examination as the AI functioned as an external entity. When the metrics aligned with the first author's expectations, they accentuated otherwise implicit gender associations, such as in the instance of the color red. The first author utilized red fabric to alter the classification of a design,

shifting it from a very masculine categorization to a more queer measure. Conversely, when the metrics appeared illogical, the first author was prompted to reassess her sensibilities and implicit tendencies.

Challenging the constraints set by the classification algorithm, the designer gained the capacity to foster a critical mindset and generate insights that influenced the subsequent design iterations.

Through iterative engagement with the classification algorithm, the designer discovered insights not just into the presence of biases within the dataset but also into how her own biases manifested in her design concepts. For instance, deliberately incorporating the color red into a design concept initially labeled as overly masculine by the algorithm aimed to introduce a more feminine quality. However, upon reflection, she recognized that while the color red might be linked to femininity in the realm of toys, she might have employed the same color to signify masculinity in a car design, underscoring the importance of contextualizing such exercises.

In the first experiment, the designer actively confronted and addressed her own biases, resulting in a redefined understanding of social categories and more deliberate reflection on specific design elements. In the second and third experiments instead, the first author lacked a clear measure of her bias. In the latent space, specifically through the latent walk, she grappled with her own biases, diverging from the framework of a masculinity-femininity measure. Instead, she confronted a spectrum of images that resisted straightforward gender identification. Exploring an infinite space of interconnected yet varied images fostered a newfound awareness of the irrationality embedded in categorical thinking, prompting the first author to critically examine her own implicit beliefs about gender norms in toys.

The generation of more ambiguous AI outputs was perceived as both novel and intriguing. Yet, the abstract quality of these outputs occasionally presented difficulties in translating them into specific design ideas that matched the intended toy design. Nonetheless, when coupled with additional training and the production of more tangible AI outputs, the first author regarded the ambiguous images as valuable sources of displacement and inspiration. This, in turn, prompted the exploration of more inclusive design directions.

5.4 Leveraging algorithmic challenges

The design experiments marked the beginning of an introspective journey for the first author, resulting in toys that embody a sense of gender ambiguity for the designer who created them. While the first author personally experienced a confrontation with her own biases throughout the process, it is important to acknowledge that the perception of gender ambiguity may vary among individuals who did not participate in the design process. In retrospect, the first author also recognized that the toys might not be entirely novel in terms of their concept and aesthetics. It may be possible to achieve similar outcomes through a conventional design process focused on creating toys that occupy an intermediate space between traditional gender categories.

Nevertheless, it is crucial to note that confronting various algorithms posed significant and deliberate challenges for the designer. Interacting with these algorithms required the acquisition of new skills and knowledge, honed through multiple rounds of iteration and testing. It became apparent that proficiency in data curation, an aspect typically not emphasized in design education, played a pivotal role in the effective utilization of these algorithms. This experience underscored the importance of delving into the functionalities and constraints of the algorithms. Understanding the intricacies and limitations of the algorithms proved crucial, allowing the designer to devise innovative strategies to navigate through data constraints and leverage the unique features of AI.

In essence, this introspective journey served as a first exploration of the intersection between creativity, biases, and technology, ultimately leading to a deeper understanding of the complexities inherent in design practice with generative AI tools.

6 TOWARD REFLEXIVE HUMAN-AI PRACTICES

The design experiments discussed above not only invited the first author to engage with AI in a non-prescriptive, introspective manner but also created a platform for the three co-authors to examine and elaborate on what it takes to reflexively engage with AI in design. In the following, we describe this practice as *reflexive data curation* and discuss how specific tactics can be leveraged to promote an introspective engagement between creative practitioners and AI. We elaborate on each tactic and explore their implications for algorithmic RtD practices and HCI more broadly.

To reiterate, it is important to clarify that reflexive data curation is not intended as a novel method of self-reflection, one that is meant to supersede existing methods or practices. Instead, we see data curation as a practice that holds particular significance in endeavors involving AI and the application of these technologies in society. In line with artists and scholars concerned with open and creative interaction around data and AI models [23, 25, 57, 58, 109, 131], we acknowledge and celebrate the interpretive and subjective nature of data curation as a means to cultivate a richer, more intimate connection with data.

In this discussion, we first acknowledge the first author's positionality in the context of this investigation, provide a definition of reflexive data curation, and outline a first set of tactics for reflexive data curation. We close with a consideration of the broader significance of reflexive human-AI practice for HCI.

6.1 First author's positionality

To fully grasp the potential for reflexivity inherent in the process of data curation, it's crucial to delineate the personal standpoint of the primary author and her predisposition and sensitivity towards gender biases, which significantly influenced the way reflexivity unfolded within the process. The first author is a young researcher with a background in industrial and interaction design. She is interested in feminist discourse surrounding gender-related issues and the impact of gender norms and roles on individuals' perspectives and behaviors. This interest, manifested in previous projects already, prompted the selection of gender in toy design as the investigation's context.

Despite her sensitivity towards gender issues, she was raised in a socio-cultural milieu where a binary conceptualization of gender is the norm. For a considerable time, she hardly perceived how these predominant ideas of masculinity and femininity affected her daily perceptions and behaviors. Through this investigation, she was instead confronted with this personal stance. This manifested in a different predisposition to her design interventions, transitioning from contemplating "how and what technologies to design for girls" (project TIG from the first author [6]) to understanding the material manifestations of gender initially.

While we acknowledge the situatedness and personal nature of this investigation, we contend that the significance of the data curation tactics outlined in this section extends far beyond the individual experience of the primary author. They can inform an emerging body of reflexive practices situated at the convergence of design and AI.

6.2 Reflexive data curation

We define *reflexive data curation* as an inquisitive and introspective approach to human-AI collaboration where the design process constitutes a site for the designer to *surface, dismantle, and defamiliarize* one's own norms and expectations but also those of society at large. The approach centres on the appropriation of the "hidden craft" of data curation [22] as a site for self-confrontation and explication of a designer's own worldview, and potentially even the broader structures of power and prejudice reflected in one's worldview, aimed to critically bridge designer and collective imaginings. Whether it is images, textual labels, or other media, *data is approached as the materialization of one's views on which the designer can act*—experimenting with variations and features to imagine things differently.

In contrast to current data curation practices, where data selection and labeling are often siloed and performed on massive scales, often even automatically, our data curation forms a very personal interaction between the training data as a reflection of self and the broader social and cultural hierarchies one is implicitly influenced by. Data is viewed as the materialization of one's perspectives, offering the designer an opportunity for deliberate intervention. This approach builds on a growing body of experiences that, especially stemming from artistic practices, provide alternative conceptualizations of algorithmic mistakes and ML uncertainty. In these, computational properties such as adaptive learning and probabilistic uncertainty are engaged and embraced to intentionally generate unpredictable behaviors and aesthetics [23, 106]. The focus is on embracing friction rather than prioritizing efficiency and seamlessness. We argue that characteristics like queerness and neutrality often aspired to in datasets and models, are not inherent to AI but rather result from the interpretative process of humans about AI, making them rational properties subject to examination. As such, these practices are positioned as diametrically opposed to engineering approaches that look at algorithmic uncertainty and unexpected results as problems to be solved, believing that a neutral/queer AI could be made. To these experiences, the notion of *reflexive data curation* adds a distinctively introspective dimension. As the reflections of the first author surface, introspection and explicit reflexivity in the use of AI as a creative partner have the potential to sensitize the designer to matters of technological inconsistency and unpredictability but also, and foremost, to confront oneself with one's own worldviews. This not only exposes how personal perceptions and biases permeate data but also provides a platform for imagining alternative worldviews. Moreover, it facilitates a broader reflection on cultural norms and social hierarchies that often remain hidden from awareness.

In this practice, the final artifacts and the AI tools used are of relative importance. The "hidden craft" [22] that is at work in algorithmic practices of human-machine collaboration not only concerns the selection, categorization, and labeling of the data. It also includes how the designer engages with the model's outcomes and confronts herself with their ambiguity and norm-breaching aspects. Data curation, when performed by designers critically and reflexively, becomes re-conceptualized from a hidden and exploitative form of labor [127] to a meaningful site for reflection and awareness.

In what follows, we describe three tactics for *reflexive data curation* that emerged from the experimental experiences of the first author.

6.3 Tactics for reflexive data curation

As we navigate the world with a unique lens, we shape our perception and understanding of reality. This lens, our worldview, influences how we interpret and make sense of the world around us. However, within this framework lies a profound paradox: while we confidently perceive the world through our chosen lenses, we remain oblivious to the existence of our blind spots. In other words, our worldviews harbor implicit blind spots, rendering us unaware of what we may be missing.

Reflecting both *in-action* and *on-action*, the first author derived three tactics. These illustrate means to recognize and challenge personal blind spots, unlock new perspectives, and unveil a richer understanding of the world around us. In this attempt, we hope to illuminate an alternative approach to human-AI collaboration in design, one more concerned with reflexivity, diversity, and inclusion than fast creativity. Further, these tactics pose counterarguments to current data curation practices centered around values such as precision and accuracy. Here, the proposed strategies emphasize the interpretative power that humans hold when engaging with modern AI. Using the textual and visual materials in MIRO, she analyzed and grouped underlying strategies that evoked her reflection. These initial strategies were then further developed in conversation with both literature and the other authors to further generalize and abstract the knowledge gained in the experiments.

6.3.1 Autoconfrontation. In the robo-doll experiment, the first author intentionally sought a deeper understanding of her blind spots, specifically the gender cues and implicit biases that unintentionally materialized in her design

sketches. This process was crafted by systematically iterating each sketch through the classification algorithm, recognizing when a dataset was problematic or inappropriate [10], and tracing nuances in the decision-making and ideation process back to the sketches, annotated by classification's measures of femininity and masculinity.

We consider this process of surfacing and challenging one's perception and internalized norms as a deliberate tactic of *autoconfrontation*. Through autoconfrontation, a valuable avenue emerged in which the first author could articulate her personal preferences and cultural exemplars and how these were codified in her tacit knowledge as a designer in a mindful and nonjudgmental way [103]. This autoconfrontation triggered and supported the first author's introspection and reflective processes in relation to what was being designed that was captured in the data. But it also facilitated the recall and reflection of more nuanced aspects, such as why design elements like shapes or colors were added and how they were associated with gender, that surpassed what was explicitly recorded in the data by allowing the first author to slow down and observe herself, and thus offering her a feedback loop on her own process and practice [85]. This presented her with critical insights, such as context-specific gender associations with color and shapes. For example, the first author was able to successfully identify and articulate blind spots that perpetuated binary gender notions, such as the inadvertent use of colors or shapes or how particular combinations of elements in the head of the robo-doll appeared to be gender-associated.

In addition to pinpointing weaknesses and blind spots in her designs, autoconfrontation encouraged and prompted the first author to "take action." It provided a type of "self-information" [73] that helped the first author build awareness and foster change. Autoconfrontation acted as a catalyst for further interacting with the classification algorithm and exploring tangible measures that could help her thoughtfully detect gender biases, rectify design flaws, and surpass what was experienced as limiting her current practice and ability to imagine otherwise.

6.3.2 Change of Perspective. Our perception of the world is shaped by our worldview. However, due to inherent blind spots, we remain unaware of our assumptions and how reality may be "more uncertain, more nuanced or more complex than originally assumed or regarded" [45] (p. 104). As often as not, we require the perspective of others to point out the blind spots in our worldview [40] and "offer different ways of understanding what we know and what we do, humans and nonhumans alike" [45] (p. 100). Next to autoconfrontation, another tactic developed by the first author to become keenly aware of her blind spots and "problematize the design space" [45] (p. 126) entailed an engagement with the perspective of the "other"—in this case, the AI. We refer to this tactic as *change of perspective*.

One way in which the first author achieved a change in perspective involved shifting her attention from "first-order observations" to "second-order observations" [79]. When you look at an object in the outside world, you are a first-order observer. A second-order observer observes how a first-order observer observes the outside world. Here, the focus of attention is shifted to how one looks instead of what one sees. This is an important shift because it opens the possibility to think through questions like: why is someone doing or saying things the way s/he/they do or say? For example, in the first design experiment, the first author trained the classification algorithm on her perception of gender. In doing so, the algorithm acted as a second-order observer, which examined the first author's design ideas and reported on observed gender biases through distinct femininity and masculinity measures. This prompted the first author to shift her attention from making intuitive design decisions to the criteria according to which she was intuitively classifying certain images as masculine or feminine. Patterns in her decision-making processes, such as colors and shapes, could thus be surfaced and brought to attention.

Another way that the first author used to change her perspective entailed using the use of metaphors. Metaphors are powerful means to free one's mind from internalized norms, roles, and values [53], and can provide new lenses through which we can perceive the world differently [89]. For example, in the dino-unicorn experiment, the first author used the latent space of the StyleGAN and latent walk function to create a lens –a from the norm deviating idea of a toy– that would allow her to deliberately explore ideas beyond the boundaries of the gender

norms and roles that appeared ingrained in her worldview. This allowed the first author to challenge the binary ideas of femininity and masculinity conventionally attributed to dinosaurs and unicorns and instead explore "spaces and objects of design that are not constituted yet but emerge in response to non-human perspectives" [45] (p. 102).

6.3.3 Clash of Expectations. Often reflection is prompted by a critical incident involving an error, a difficult situation, or an unexpected result of one's actions [94]. "Errors" occur when one's understanding of reality clashes with the actual reality. Such a clash with reality reveals that one's assumptions about the world are incomplete or incorrect. The moment in which the expectations that we based on certain assumptions clash with reality and collapse produces a sense of discomfort that urges us to reflect and rethink. In her human-AI practice, we see that the first author deliberately used these moments of discomfort as opportunities to disrupt her practice and open up unexpected directions for idea generation. We refer to this tactic as *clash of expectations*.

In the hair-drill experiment, for example, the first author was training the model by adding images of delicate and fragile glass vases as secondary data on top of the drill in an attempt to explore measures of femininity and masculinity further and subvert norms around the idea of drill toys. When the transformations generated by the StyleGAN failed to convey the fragility and femininity associated with glass vases, the first author felt a sense of discomfort. The images generated by the StyleGAN appeared familiar at first, but upon closer inspection, the fading masculinity of the drill became apparent, leaving the first author with an object that no longer matched her expectation of a drill. This clash between the first author's idea of a drill and the hair-drill generated by the AI using glass vases as secondary data exposed the incorrectness of the first author's assumptions about the world while at the same time exposing the limitations of the algorithm.

In this moment of discomfort, the author felt confronted by the hair-drill object that combined the two categories of masculine and feminine that are so often thought of as being exclusionary. We structure the world around us in categories as a way of simplifying its complexity. These categories become a model of reality but never truly capture reality as such. It is, therefore, not surprising that we are, now and then, presented with situations that puzzle us, with moments when we are not able to assign a known category to things or people. It is such a moment when we wish to escape the unpleasant feeling of discomfort produced that we can actually break up with a too-narrow view of the world [13, 43]. In the case of the hair-drill, the images of hybrids generated by the StyleGAN inspired the final toy design by visualizing a combination of feminine and masculine design cues –such as feminine coloring of masculine shapes– that were previously unknown to the first author.

6.4 Beyond AI systems as deterministic tools

Reflexive data curation marks a significant departure in how designers and researchers engage with AI. It diverges from conventional HCI and industry-centric research paradigms by challenging the perception of AI systems as deterministic tools. Instead, it embraces the inherent uncertainties of machine learning, regarding errors and surprises as valuable insights rather than failures. This departure from rigid engineering norms fosters creativity and experimentation in configuring AI and machine learning systems, particularly in contexts involving limited datasets and simplistic models.

Reflexive data curation acknowledges and encourages active human engagement as interpretative entities shaped by diverse social, cultural, and individual contexts. This mobilization of reflexivity seeks to deepen comprehension of AI technologies and influence their integration within the creative process, promoting a nuanced and socially conscious approach to AI system design and deployment. This shift towards relational design practices signifies a departure from linear problem-solving frameworks towards dynamic processes of negotiation and reconfiguration.

In its essence, reflexive data curation embodies several key characteristics. It begins with a recognition of the inherent uncertainties present within AI systems. Rather than shying away from these uncertainties, designers are

encouraged to embrace them, viewing them as opportunities for exploration and experimentation with different machine-learning configurations.

This approach also encourages a departure from rigid engineering norms, instead fostering a culture of creative exploration and innovation in AI design. Particularly in situations where datasets are limited and models are relatively simple, there's a push for designers to think outside the box and explore unconventional solutions.

Furthermore, reflexive data curation challenges the traditional view of errors and surprises as setbacks. Instead, they are seen as valuable learning experiences, offering insights that can lead to improvement and growth. This perspective contributes to a broader discourse on AI, machine learning, and associated ethical concerns by valuing diverse viewpoints and embracing the complexity inherent in these systems.

7 CONCLUSION

In this paper, we explored *reflexive data curation* as a deliberate craft for *human-AI reflexive practice*, which re-positions off-the-shelf generative AI tools from platforms of hidden and potentially exploitative labor to a meaningful site for reflection and awareness.

As evident in the public debate around what to consider as responsible use of AI, the data used to train machine learning has the propensity to incorporate and replicate the power structures and prejudices through which the training data was produced. And while we are witnessing growing excitement for AI tools such as Midjourney, Dall-E, and ChatGPT, which use available data to turn the probabilistic inferencing of machine learning models into a novel source of fast creativity, the pressing question of how designers might deal responsibly with the biases and uncertainties intrinsic to computational data and models remains unanswered for the most part.

This Research through Design (RtD) project has explored how bias exacerbation, inconsistency, and unpredictability might play a generative role in exposing a designer's implicit biases and inherited prejudices and help the designer become more fluent and intentional in how social categories are materialized in the design outcome. The paper has described the introspective process of the first author and how she curated data to surface her own biases and to design a series of queer children's toys that challenge the binary categories of masculinity and femininity conventionally baked into toy design.

To illustrate her introspective process of reflexive data curation, we have unpacked three design experiments: the robo-doll, the dino-unicorn, and the hair-drill. Based on these, we have derived three design tactics for reflexive data curation: *autoconfrontation*, *shift in perspective*, and *clash of expectations*. Then, we have concluded by discussing how these experiments illuminate an alternative approach to the deployment of generative AI tools in design.

Methodologically, the work described and discussed in this paper has combined a data-enabled RtD process with introspective design research methods, embracing the uncertainty of prototyping with data, algorithms, and models as a fundamental shift in how we understand the role of the prototype in algorithmic RtD processes [44]. The unpacking of this work has been carried out in line with HCI practices of situated data analysis [33, 67, 71], as the expression of fundamentally situated, reflexive, and recursive interactions between the first author and the AI.

The contribution of the paper is articulated in terms of both conceptual resources (*reflexive data curation* and *human-AI reflexive practice*) and practical resources (design tactics for reflexive data curation, i.e., *autoconfrontation*, *shift in perspective*, and *clash of expectations*). These concepts and tactics add to the repertoire of algorithmic RtD practices (this call), critical HCI approaches to ML systems design [15, 72], and reflexive data science research moving beyond technocentric solutions of debiasing algorithms and datasets in the development and deployment of AI [8, 14].

More broadly, the work described in this paper resonates with the call for a radical rethinking of what is to do design under the conditions brought about by digital technologies and the most recent technological

advancements in the field of AI [47, 116]. By encountering the distinct materiality of machine learning algorithms and computational models through the specificity of this project, the work illuminates what may be interpreted as a shift toward relational design practices, moving away from ideas of linear progression from problem framing to implementation [63, 118] and towards sustained processes of negotiation and reconfiguration [46, 125].

Of course, the paper has limitations concerning how individual practices of *reflexive data curation* on simple design objects can provide a scalable solution to AI-driven systemic inequality and address the full extent of the social, economic, and political problems being encountered in the deployment of AI systems in all areas of public life [97]. Moreover, the algorithms used in this project focus on visually encoded information. Though images are most impactful in a design process, we know that biases are multi-dimensional [113], and not all of them (and maybe the most relevant ones) cannot be captured through images only.

Yet, the contributions of this paper offer conceptual resources and concrete tactics to the design and HCI communities to engage critically and reflexively with data and generative AI tools in the context of various forms of human-AI collaboration, hopefully, more concerned with equity, diversity, and inclusion than producing on-demand images in five minutes.

8 ACKNOWLEDGEMENTS

This project was carried out within the DCODE Labs of the DCODE Network (https://www.dcode-network.eu/). The DCODE Network has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 955990. Special thanks go to Simone Rebaudengo from oio.studio for his expertise and assistance in various aspects of our study.

REFERENCES

[1] Tony E Adams, Carolyn Ellis, and Stacy Holman Jones. 2017. Autoethnography. *The international encyclopedia of communication research methods* (2017), 1–11.

[2] Kenneth Agnew. 1993. The spitfire: Legend or history? An argument for a new research culture in design. *Journal of Design History* 6, 2 (1993), 121–130.

[3] Philip E Agre. 2014. Toward a critical technical practice: Lessons learned in trying to reform AI. In *Social science, technical systems, and cooperative work*. Psychology Press, 131–157.

[4] Memo Akten, Rebecca Fiebrink, and Mick Grierson. 2019. Learning to see: you are what you see. In *ACM SIGGRAPH 2019 Art Gallery*. 1–6.

[5] Danielle Barbosa Lins de Almeida. 2017. On diversity, representation and inclusion: New perspectives on the discourse of toy campaigns. *Linguagem em (Dis) curso* 17 (2017), 257–270.

[6] Anne Arzberger. 2021. TIG. Tech-toolkit for girls. https://ifdesign.com/en/winner-ranking/project/tig/323482

[7] Carol J Auster and Claire S Mansbach. 2012. The gender marketing of toys: An analysis of color and type of toy on the Disney store website. *Sex roles* 67 (2012), 375–388.

[8] Agathe Balayn and Seda Gürses. 2021. Beyond Debiasing: Regulating AI and its inequalities. *EDRi report* (2021).

[9] Jo Bates, David Cameron, Alessandro Checco, Paul Clough, Frank Hopfgartner, Suvodeep Mazumdar, Laura Sbaffi, Peter Stordy, and Antonio de la Vega de León. 2020. Integrating FATE/critical data studies into data science curricula: where are we going and how do we get there?. In *Proceedings of the 2020 conference on fairness, accountability, and transparency*. 425–435.

[10] Eric PS Baumer, Vera Khovanskaya, Mark Matthews, Lindsay Reynolds, Victoria Schwanda Sosik, and Geri Gay. 2014. Reviewing reflection: on the use of reflection in interactive system design. In *Proceedings of the 2014 conference on Designing interactive systems*. 93–102.

[11] Eric PS Baumer and M Six Silberman. 2011. When the implication is not to design (technology). In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2271–2274.

[12] Anaëlle Beignon, Emeline Brulé, Jean-Baptiste Joatton, and Aurélien Tabard. 2020. Tricky Design Probes: Triggering Reflection on Design Research Methods in Service Design. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference*. 1647–1660.

[13] Genevieve Bell, Mark Blythe, and Phoebe Sengers. 2005. Making by making strange: Defamiliarization and the design of domestic technologies. *ACM Transactions on Computer-Human Interaction (TOCHI)* 12, 2 (2005), 149–173.

[14] Rachel KE Bellamy, Kuntal Dey, Michael Hind, Samuel C Hoffman, Stephanie Houde, Kalapriya Kannan, Pranay Lohia, Jacquelyn Martino, Sameep Mehta, Aleksandra Mojsilović, et al. 2019. AI Fairness 360: An extensible toolkit for detecting and mitigating

algorithmic bias. *IBM Journal of Research and Development* 63, 4/5 (2019), 4–1.
[15] Jesse Josua Benjamin, Arne Berger, Nick Merrill, and James Pierce. 2021. Machine learning uncertainty as a design material: a post-phenomenological inquiry. In *Proceedings of the 2021 CHI conference on human factors in computing systems*. 1–14.
[16] Abeba Birhane. 2021. Algorithmic injustice: a relational ethics approach. *Patterns* 2, 2 (2021), 100205.
[17] Judith E Owen Blakemore and Renee E Centers. 2005. Characteristics of boys' and girls' toys. *Sex roles* 53 (2005), 619–633.
[18] Boudewijn Boon, Ehsan Baha, Abhigyan Singh, Frithjof E Wegener, Marco C Rozendaal, and Pieter Jan Stappers. 2020. Grappling with diversity in research through design. (2020).
[19] Marion Buchenau and Jane Fulton Suri. 2000. Experience prototyping. In *Proceedings of the 3rd conference on Designing interactive systems: processes, practices, methods, and techniques*. 424–433.
[20] Michał Bucholc. 2023. Storyboarding with AI. https://bettermarketing.pub/storyboarding-with-ai-d1534c4d91d5
[21] Joy Buolamwini and Timnit Gebru. 2018. Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability and transparency*. PMLR, 77–91.
[22] Baptiste Caramiaux. 2023. *Machine Learning in Interaction: Tool, Material, Culture.* Ph. D. Dissertation. Université Paris-Saclay.
[23] Baptiste Caramiaux and Sarah Fdili Alaoui. 2022. " Explorers of Unknown Planets" Practices and Politics of Artificial Intelligence in Visual Arts. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–24.
[24] Toby Chong, Nolwenn Maudet, Katsuki Harima, and Takeo Igarashi. 2021. Exploring a makeup support system for transgender passing based on automatic gender recognition. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–13.
[25] Kate Crawford and Trevor Paglen. 2019. Excavating ai. *The AI Now Institute, NYU. Accessed January* 24 (2019), 2021.
[26] Johanne Yttri Dahl and Ann Rudinow Sætnan. 2009. "It all happened so slowly"–On controlling function creep in forensic DNA databases. *International Journal of Law, Crime and Justice* 37, 3 (2009), 83–103.
[27] Peter Dalsgaard. 2017. Instruments of inquiry: Understanding the nature and role of tools in design. *International journal of design* 11, 1 (2017).
[28] Meltem Demirkus, Kshitiz Garg, and Sadiye Guler. 2010. Automated person categorization for video surveillance using soft biometrics. In *Biometric Technology for Human Identification VII*, Vol. 7667. SPIE, 236–247.
[29] Audrey Desjardins and Cayla Key. 2020. Parallels, Tangents, and Loops: Reflections on the'Through'Part of RtD. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference*. 2133–2147.
[30] Audrey Desjardins, Oscar Tomico, Andrés Lucero, Marta E Cecchinato, and Carman Neustaedter. 2021. Introduction to the special issue on first-person methods in HCI. , 12 pages.
[31] John Dewey. 2023. *Logic the theory of inquiry.* Balaji Publications.
[32] Paul Dourish, Janet Finlay, Phoebe Sengers, and Peter Wright. 2004. Reflective HCI: Towards a critical technical practice. In *CHI'04 extended abstracts on Human factors in computing systems*. 1727–1728.
[33] Claude Draude, Goda Klumbyte, Phillip Lücking, and Pat Treusch. 2020. Situated algorithms: a sociotechnical systemic approach to bias. *Online Information Review* 44, 2 (2020), 325–342.
[34] Fred Dretske. 1994. Introspection. In *Proceedings of the Aristotelian Society*, Vol. 94. JSTOR, 263–278.
[35] Melisa Duque, Robert Willim, Minna Ruckenstein, and Sarah Pink. 2018. Broken data: Conceptualising data in an emerging world. *Big Data and Society* 5, 1 (2018).
[36] Cordelia Fine and Emma Rush. 2018. "Why does all the girls have to buy pink stuff?" The ethics and science of the gendered toy marketing debate. *Journal of Business Ethics* 149 (2018), 769–784.
[37] Rowanne Fleck and Geraldine Fitzpatrick. 2010. Reflecting on reflection: framing a design landscape. In *Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction*. 216–223.
[38] Catherine Flick and Kyle Worrall. 2022. The Ethics of Creative AI. In *The Language of Creative AI: Practices, Aesthetics and Structures*. Springer, 73–91.
[39] Christopher Frauenberger. 2019. Entanglement HCI the next wave? *ACM Transactions on Computer-Human Interaction (TOCHI)* 27, 1 (2019), 1–27.
[40] Hans-Georg Gadamer and Günter Figal. 2007. *Truth and Method.* Akademie Verlag Berlin.
[41] Bill Gaver and John Bowers. 2012. Annotated portfolios. *interactions* 19, 4 (2012), 40–49.
[42] William Gaver. 2012. What should we expect from research through design?. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 937–946.
[43] William W Gaver, Jacob Beaver, and Steve Benford. 2003. Ambiguity as a resource for design. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 233–240.
[44] Elisa Giaccardi. 2019. Histories and futures of research through design: From prototypes to connected things. *International Journal of Design* 13, 3 (2019), 139–155.
[45] Elisa Giaccardi. 2020. Casting things as partners in design: towards a more-than-human design practice. *Relating to things: design, technology and the artificial. Bloomsbury, London* (2020), 99–132.
[46] Elisa Giaccardi and Johan Redström. 2020. Technology and more-than-human design. *Design Issues* 36, 4 (2020), 33–44.

[47] Elisa Giaccardi, Chris Speed, Johan Redström, Somaya Ben Allouch, Irina Shklovski, and Rachel Charlotte Smith. 2022. AI and the conditions of design: towards a new set of design ideals. *DRS2022: Bilbao* 25 (2022).

[48] Tarleton Gillespie. 2014. The relevance of algorithms. *Media technologies: Essays on communication, materiality, and society* 167, 2014 (2014), 167.

[49] Thilo Hagendorff. 2020. The ethics of AI ethics: An evaluation of guidelines. *Minds and machines* 30, 1 (2020), 99–120.

[50] Alexa Hagerty and Igor Rubinov. 2019. Global AI ethics: a review of the social impacts and ethical implications of artificial intelligence. *arXiv preprint arXiv:1907.07892* (2019).

[51] Donna Haraway. 1988. Situated knowledges: The science question in feminism and the privilege of partial perspective. *Feminist studies* 14, 3 (1988), 575–599.

[52] Donna Haraway. 1996. Modest witness: Feminist diffractions in science studies. (1996).

[53] Donna Haraway and A Cyborg Manifesto. 1991. SCIENCE, TECHNOLOGY, AND SOCIALIST-FEMINISM IN THE LATE TWENTIETH CENTURY," IN SIMIANS, CYBORGS AND WOMEN: THE REINVENTION OF NATURE (NEW YORK; ROUTLEDGE, 1991), PP. 149-181. (1991).

[54] Donna J Haraway. 2000. A cyborg manifesto: Science, technology, and socialist-feminism in the late twentieth century. In *Posthumanism*. Springer, 69–84.

[55] Enninga Heidi. 2023. Are new technologies keeping US stuck in old biases? - newsroom: University of St. Thomas. https://news.stthomas.edu/are-new-technologies-keeping-us-stuck-in-old-biases/

[56] Melissa Heikkilä. 2023. Machine Learning in Interaction: Tool, Material, Culture. *MIT Technology Review* (2023).

[57] Drew Hemment. 2020. Reordering the assemblages of the digital through art and open prototyping. *Leonardo* 53, 5 (2020), 529–536.

[58] Drew Hemment, Ruth Aylett, Vaishak Belle, Dave Murray-Rust, Ewa Luger, Jane Hillston, Michael Rovatsos, and Frank Broz. 2019. Experiential AI. *AI Matters* 5, 1 (2019), 25–31.

[59] Simon David Hirsbrunner, Michael Tebbe, and Claudia Müller-Birn. 2022. From critical technical practice to reflexive data science. *Convergence* (2022), 13548565221132243.

[60] Marius Hoggenmueller, Maria Luce Lupetti, Willem Van Der Maden, and Kazjon Grace. 2023. Creative AI for HRI Design Explorations. In *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*. 40–50.

[61] Kristina Höök, Jeffrey Bardzell, Simon Bowen, Peter Dalsgaard, Stuart Reeves, and Annika Waern. 2015. Framing IxD knowledge. *Interactions* 22, 6 (2015), 32–36.

[62] Kristina Höök and Jonas Löwgren. 2012. Strong concepts: Intermediate-level knowledge in interaction design research. *ACM Transactions on Computer-Human Interaction (TOCHI)* 19, 3 (2012), 1–18.

[63] Kristina Höök and Jonas Löwgren. 2021. Characterizing interaction design by its ideals: A discipline in transition. *She Ji: The Journal of Design, Economics, and Innovation* 7, 1 (2021), 24–40.

[64] Amnesty International. 2021. Xenophobic machines: Discrimination through unregulated use of algorithms in the Dutch childcare benefits scandal. (2021). https://www.amnesty.org/en/documents/eur35/4686/2021/en/

[65] Hannah Jaye. 2023. 7 Ways UX Designers Can Use AI to Their Advantage. https://designlab.com/blog/how-to-use-ai-as-a-ux-designer/

[66] Anna Jobin, Marcello Ienca, and Effy Vayena. 2019. The global landscape of AI ethics guidelines. *Nature Machine Intelligence* 1, 9 (2019), 389–399.

[67] Christine Kaeser-Chen, Elizabeth Dubois, Friederike Schüür, and Emanuel Moss. 2020. Positionality-aware machine learning: translation tutorial. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. 704–704.

[68] Shivani Kapania, Alex S Taylor, and Ding Wang. 2023. A hunt for the Snark: Annotator Diversity in Data Practices. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–15.

[69] Moaiad Ahmad Khder. 2021. Web Scraping or Web Crawling: State of Art, Techniques, Approaches and Application. *International Journal of Advances in Soft Computing & Its Applications* 13, 3 (2021).

[70] Jingoog Kim and Mary Lou Maher. 2023. The effect of AI-based inspiration on human design ideation. *International Journal of Design Creativity and Innovation* 11, 2 (2023), 81–98.

[71] Brian Kinnee, Audrey Desjardins, and Daniela Rosner. 2023. Autospeculation: Reflecting on the Intimate and Imaginative Capacities of Data Analysis. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–10.

[72] Goda Klumbytė, Claude Draude, and Alex S Taylor. 2022. Critical tools for machine learning: Working with intersectional critical concepts in machine learning systems design. In *2022 ACM Conference on Fairness, Accountability, and Transparency*. 1528–1541.

[73] Ian Li, Jodi Forlizzi, and Anind Dey. 2010. Know thyself: monitoring and reflecting on facets of one's life. In *CHI'10 Extended Abstracts on Human Factors in Computing Systems*. 4489–4492.

[74] Ann Light. 2011. HCI as heterodoxy: Technologies of identity and the queering of interaction with computers. *Interacting with computers* 23, 5 (2011), 430–438.

[75] Youn-Kyung Lim, Erik Stolterman, and Josh Tenenberg. 2008. The anatomy of prototypes: Prototypes as filters, prototypes as manifestations of design ideas. *ACM Transactions on Computer-Human Interaction (TOCHI)* 15, 2 (2008), 1–27.

[76] Feng Lin, Yingxiao Wu, Yan Zhuang, Xi Long, and Wenyao Xu. 2016. Human gender classification: a review. *International Journal of Biometrics* 8, 3-4 (2016), 275–300.
[77] Joseph Lindley, Haider Ali Akmal, Franziska Pilling, and Paul Coulton. 2020. Researching AI legibility through design. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.
[78] Jonas Löwgren. 2009. Toward an articulation of interaction esthetics. *New Review of Hypermedia and Multimedia* 15, 2 (2009), 129–146.
[79] Niklas Luhmann. 1993. Deconstruction as second-order observing. *New literary history* 24, 4 (1993), 763–782.
[80] Ryan Mac. 2015. Facebook Apologizes After A.I. Puts 'Primates' Label on Video of Black Men. https://www.nytimes.com/2021/09/03/technology/facebook-ai-race-primates.html
[81] Teachable Machine. 2022. https://teachablemachine.withgoogle.com/
[82] Nirav Malsattar, Tomo Kihara, and Elisa Giaccardi. 2019. Designing and Prototyping from the Perspective of AI in the Wild. In *Proceedings of the 2019 on Designing Interactive Systems Conference*. 1083–1088.
[83] Monique Mann and Tobias Matzner. 2019. Challenging algorithmic profiling: The limits of data protection and anti-discrimination in responding to emergent discrimination. *Big Data & Society* 6, 2 (2019), 2053951719895805.
[84] Claudia Mareis. 2012. The epistemology of the unspoken: On the concept of tacit knowledge in contemporary design research. *Design Issues* 28, 2 (2012), 61–71.
[85] Nikolas Martelaro and Wendy Ju. 2018. Cybernetics and the design of the user experience of AI systems. *interactions* 25, 6 (2018), 38–41.
[86] Dan McQuillan. 2018. People's councils for ethical machine learning. *Social Media+ Society* 4, 2 (2018), 2056305118768303.
[87] Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. 2021. A survey on bias and fairness in machine learning. *Comput. Surveys* 54, 6 (2021), 1–35.
[88] Jessica Morley, Libby Kinsey, Anat Elhalal, Francesca Garcia, Marta Ziosi, and Luciano Floridi. 2021. Operationalising AI ethics: barriers, enablers and next steps. *AI & SOCIETY* (2021), 1–13.
[89] DS Murray-Rust, Iohanna Nicenboim, and Dan Lockton. 2022. Metaphors for designers working with AI. In *DRS Conference Proceedings 2022*.
[90] Theoden I. Netoff. 2019. Chapter 14 - The Ability to Predict Seizure Onset. In *Engineering in Medicine*, Paul A. Iaizzo (Ed.). Academic Press, 365–378. https://doi.org/10.1016/B978-0-12-813068-1.00014-2
[91] Iohanna Nicenboim, Giaccardi Elisa, and Johan Redström. 2023. Designing more-than-human AI: Experiments on situated conversations and silences. *diid disegno industriale industrial design* 80 (2023), 32–43.
[92] Iohanna Nicenboim, Elisa Giaccardi, and Johan Redström. 2022. From explanations to shared understandings of AI. (2022).
[93] Leonardo Nicoletti and Dina Bass. 2023. Humans are biased. Generative AI is even worse. https://www.bloomberg.com/graphics/2023-generative-ai-bias/
[94] Ikujirō Nonaka and Hirotaka Takeuchi. 2007. The knowledge-creating company. *Harvard business review* 85, 7/8 (2007), 162.
[95] Suvi Pihkala and Helena Karasti. 2016. Reflexive engagement: enacting reflexivity in design and for'participation in plural'. In *Proceedings of the 14th Participatory Design Conference: Full Papers-Volume 1*. 21–30.
[96] Isabel Prochner and Danny Godin. 2022. Quality in research through design projects: Recommendations for evaluation and enhancement. *Design Studies* 78 (2022), 101061.
[97] Bogdana Rakova and Roel Dobbe. 2023. Algorithms as Social-Ecological-Technological Systems: An Environmental Justice Lens on Algorithmic Audits. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (Chicago, IL, USA) *(FAccT '23)*. Association for Computing Machinery, New York, NY, USA, 491. https://doi.org/10.1145/3593013.3594014
[98] Runway. 2023. Everything you need to make anything you want. https://runwayml.com/
[99] Jeffrey Saltz, Michael Skirpan, Casey Fiesler, Micha Gorelick, Tom Yeh, Robert Heckman, Neil Dewar, and Nathan Beard. 2019. Integrating ethics within machine learning courses. *ACM Transactions on Computing Education (TOCE)* 19, 4 (2019), 1–26.
[100] Nithya Sambasivan, Shivani Kapania, Hannah Highfill, Diana Akrong, Praveen Paritosh, and Lora M Aroyo. 2021. "Everyone wants to do the model work, not the data work": Data Cascades in High-Stakes AI. In *proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–15.
[101] Corina Sas and Alan Dix. 2009. Designing for reflection on experience. In *CHI'09 Extended Abstracts on Human Factors in Computing Systems*. 4741–4744.
[102] Roger C Schank, Tamara R Berman, and Kimberli A Macpherson. 1999. Learning by doing. *Instructional-design theories and models: A new paradigm of instructional theory* 2, 2 (1999), 161–181.
[103] Donald Schön. 1983. The reflective practitioner basic books. *New York* (1983).
[104] Donald Schön and John Bennett. 1996. Reflective conversation with materials. In *Bringing design to software*. 171–189.
[105] Donald A Schön. 2017. *The reflective practitioner: How professionals think in action*. Routledge.
[106] Hugo Scurto, Baptiste Caramiaux, and Frédéric Bevilacqua. 2021. Prototyping machine learning through diffractive art practice. In *Designing Interactive Systems Conference 2021*. 2013–2025.

[107] Phoebe Sengers, Kirsten Boehner, Shay David, and Joseph'Jofish' Kaye. 2005. Reflective design. In *Proceedings of the 4th decennial conference on Critical computing: between sense and sensibility*. 49–58.
[108] Phoebe Sengers, John McCarthy, and Paul Dourish. 2006. Reflective HCI: articulating an agenda for critical practice. In *CHI'06 extended abstracts on Human factors in computing systems*. 1683–1686.
[109] Caroline Sinders. 2020. Feminist data set. *Clinic for Open Source Arts. https://carolinesinders. com/wp-con tent/uploads/2020/05/Feminist-Data-Set-Final-Draft-2020-0517. pd f* (2020).
[110] Robert Soden, Laura Devendorf, Richmond Wong, Yoko Akama, Ann Light, et al. 2022. Modes of Uncertainty in HCI. *Foundations and Trends in Human–Computer Interaction* 15, 4 (2022), 317–426.
[111] Marie Louise Juul Søndergaard and Lone Koefoed Hansen. 2018. Intimate futures: Staying with the trouble of digital personal assistants through design fiction. In *Proceedings of the 2018 designing interactive systems conference*. 869–880.
[112] Katta Spiel, Os Keyes, Ashley Marie Walker, Michael A DeVito, Jeremy Birnholtz, Emeline Brulé, Ann Light, Pınar Barlas, Jean Hardy, Alex Ahmed, et al. 2019. Queer (ing) HCI: Moving forward in theory and practice. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–4.
[113] Ramya Srinivasan and Ajay Chander. 2021. Biases in AI systems. *Commun. ACM* 64, 8 (2021), 44–49.
[114] Pieter Jan Stappers and Elisa Giaccardi. 2017. Research through design. In *The encyclopedia of human-computer interaction*. The Interaction Design Foundation, 1–94.
[115] Pieter Jan Stappers, Froukje Sleeswijk Visser, and Ianus Keller. 2014. The role of prototypes and frameworks for structuring explorations by research through design. In *The Routledge companion to design research*. Routledge, 163–174.
[116] Niya Stoimenova and Rebecca Price. 2020. Exploring the nuances of designing (with/for) artificial intelligence. *Design Issues* 36, 4 (2020), 45–55.
[117] Erik Stolterman and Mikael Wiberg. 2010. Concept-driven interaction design research. *Human–Computer Interaction* 25, 2 (2010), 95–118.
[118] Erik Stolterman and Mikael Wiberg. 2020. Compositional interaction design—changes in design practice and its implications for teaching and research. *Digital Creativity* 31, 1 (2020), 44–63.
[119] Keras Team. 2022. Keras Documentation: A walk through latent space with stable diffusion. https://keras.io/examples/generative/random_walks_with_stable_diffusion/#:~:text=Latent%20space%20walking%2C%20or%20latent,frame%20in%20the%20final%20animation.
[120] Ekin Tiu. 2020. Understanding latent space in machine learning. https://towardsdatascience.com/understanding-latent-space-in-machine-learning-de5a7c687d8d#:~:text=The%20latent%20space%20is%20simply,representations%20of%20data%20for%20analysis.
[121] A. Tsepko. 2021. Emerging Trends: How Artificial Intelligence Will Affect Creative Decisions. https://www.forbes.com/sites/forbestechcouncil/2021/03/10/2021-emerging-trends-how-artificial-intelligence-will-affect-creative-decisions/?sh=74c988ad5327
[122] Grace Leonora Turtle. 2022. Mutant in the mirror: Queer becomings with AI. (2022).
[123] Vera van der Burg, AA Akdag Salah, and RSK Chandrasegaran. 2022. Ceci n'est pas une Chaise:: Emerging Practices in Designer-AI Collaboration. In *Proceedings of Design Research Society International Conference (DRS2022), Bilbao*. Design Research Society.
[124] Vera van der Burg, Gijs de Boer, Alkim Almila Akdag Salah, Senthil Chandrasegaran, and Peter Lloyd. 2023. Objective Portrait: A practice-based inquiry to explore Al as a reflective design partner. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*. 387–400.
[125] Ron Wakkary. 2020. Nomadic practices: A posthuman theory for knowing design. *International Journal of Design* 14, 3 (2020), 117.
[126] Ron Wakkary. 2021. *Things we could design: For more than human-centered worlds*. MIT press.
[127] Meredith Whittaker, Kate Crawford, Roel Dobbe, Genevieve Fried, Elizabeth Kaziunas, Varoon Mathur, Sarah Mysers West, Rashida Richardson, Jason Schultz, Oscar Schwartz, et al. 2018. *AI now report 2018*. AI Now Institute at New York University New York.
[128] WLP Wong and David F Radcliffe. 2000. The tacit nature of design knowledge. *Technology analysis & strategic management* 12, 4 (2000), 493–512.
[129] Wenying Wu, Pavlos Protopapas, Zheng Yang, and Panagiotis Michalatos. 2020. Gender classification and bias mitigation in facial images. In *Proceedings of the 12th ACM Conference on Web Science*. 106–114.
[130] Haian Xue and Pieter MA Desmet. 2019. Researcher introspection for experience-driven design research. *Design Studies* 63 (2019), 37–64.
[131] Martin Zeilinger. 2021. *Tactical entanglements: AI art, creative agency, and the limits of intellectual property*. meson press.
[132] Maggie Zhang. 2015. Google Photos Tags Two African-Americans As Gorillas Through Facial Recognition Software. https://www.forbes.com/sites/mzhang/2015/07/01/google-photos-tags-two-african-americans-as-gorillas-through-facial-recognition-software/
[133] Yan Zhou, Murat Kantarcioglu, and Chris Clifton. 2021. Improving fairness of ai systems with lossless de-biasing. *arXiv preprint arXiv:2105.04534* (2021).
[134] E Zimmerman. 2003. Play as design: The iterative design process. In *Design research*. MIT Press.
[135] John Zimmerman and Jodi Forlizzi. 2008. The role of design artifacts in design theory construction. *Artifact: Journal of Design Practice* 2, 1 (2008), 41–45.

[136] Martina Šimkovičová. 2023. On the Parallel Struggles of Photography and GAN-generated Imagery. *European Journal of Media, Art and Photography* 11, 1 (2023), 66–71.

Received 3 July 2023; revised 13 May 2024; accepted 9 July 2024