GHOSTS IN THE MACHINE

SYNTHETIC DATA THROUGH AND FOR SPECULATIVE DESIGN

Author
Sarah Cosentino
10779573

Supervisor
Elisa Giaccardi

Politecnico di Milano
School of Design

Master of Science in Digital and Interaction Design

Academic Year 2025/2026

GHOSTS IN THE MACHINE

SYNTHETIC DATA THROUGH AND FOR SPECULATIVE DESIGN

Author
Sarah Cosentino
10779573

Supervisor
Elisa Giaccardi

Politecnico di Milano
School of Design

Master of Science in Digital and Interaction Design

Academic Year 2025/2026

There is no such thing as ethical design under capitalism.

– @ethicaldesign69 on IG

Table of contents

List of figures

List of tables

List of charts

Abstract in English

This thesis investigates the strategic integration of synthetic data in design pedagogy as a political tool to resist *epistemic violence*, specifically within speculative design practices. In periods of systemic crisis, the capacity to imagine alternative futures becomes a critical arena for political and social determination. While generative artificial intelligence offers new affordances for speculative design education, current models are trained on datasets harvested and structured through colonial frameworks. Consequently, these systems routinely reproduce Eurocentric, capitalist defaults, marginalizing pluralistic ways of knowing. This issue is particularly pronounced within Western design institutions, where dominant academic narratives already align with the dominant agenda, rendering traditional speculative design methods blind to alternative epistemologies.

When design education adopts generative AI uncritically, it risks strengthening these power structures, trapping student imagination within a singular, exclusionary vision of the future. While technical fields often present synthetic data as a mathematical solution to "debias" problematic models, this research argues that standard debiasing complies with data colonialism. By treating data as a value-neutral asset, it strips away the sociopolitical context of extraction and processing.

Through a *Research-through-Design* approach, this thesis proposes an adversarial pedagogical framework that intentionally inverts debiasing mechanics to amplify and expose hidden ideological assumptions. The framework is operationalized through Hildegard, a custom digital interface designed for collective future-scenario building. By introducing deliberate, critical friction into the human-AI co-speculation process, the interface forces design students to confront how their own inputs and the underlying computational systems encode colonial patterns. Ultimately, this thesis positions synthetic data not as a tool for optimization, but as a critical, diagnostic infrastructure to expand the speculative range of design education and foster structural decolonial literacy.

Abstract in Italian

Questa tesi analizza l'integrazione dei dati sintetici nella pedagogia del design come strumento politico per contrastare la *violenza epistemica*, specificamente all'interno delle pratiche di speculative design. In periodi di crisi sistemica, la capacità di immaginare futuri alternativi diventa un terreno d'azione cruciale per l'autodeterminazione politica e sociale. Sebbene l'intelligenza artificiale generativa offra nuove opportunità per l'educazione al design speculativo, i modelli attuali sono addestrati su dataset raccolti e strutturati attraverso framework coloniali. Di conseguenza, questi sistemi riproducono sistematicamente idee eurocentriche e capitaliste, emarginando modi di pensare pluralistici. Questo problema è particolarmente evidente nelle istituzioni di design occidentali, dove le narrazioni accademiche dominanti si allineano già con le agende dominanti, rendendo i metodi tradizionali di speculative design ciechi rispetto a epistemologie alternative.

Quando l'educazione al design adotta l'IA generativa in modo acritico, rischia di rafforzare queste strutture di potere, intrappolando l'immaginazione degli studenti in una visione del futuro singolare ed escludente. Sebbene i campi tecnici presentino spesso i dati sintetici come una soluzione matematica per correggere il bias, questa ricerca sostiene che il debiasing si conformi al colonialismo dei dati.

Attraverso un approccio di *Research-through-Design*, la tesi propone un framework pedagogico avversariale che inverte intenzionalmente le meccaniche di debiasing per amplificare ed esporre i presupposti ideologici nascosti. Il framework è reso operativo tramite Hildegard, un'interfaccia digitale personalizzata per la co-creazione di scenari futuri. Introducendo un attrito critico nel processo di co-speculazione uomo-IA, l'interfaccia costringe gli studenti a confrontarsi con il modo in cui i propri input e i sistemi computazionali sottostanti codificano modelli coloniali. Per concludere, questa tesi posiziona i dati sintetici come infrastrutture diagnostiche critiche per espandere la portata speculativa dell'educazione al design e promuovere un'alfabetizzazione decoloniale strutturale.

Executive summary

Problem Statement

Generative AI systems are increasingly used in design education and innovation to support speculative futures work. However, these tools are typically trained on datasets reflecting Western, capitalistic, and technocratic worldviews, reproducing these defaults while claiming neutrality. Standard “debiasing” approaches miss a deeper issue: they maintain the fiction that data can be objective and neutral while hiding the structural violence embedded in how systems decide what futures are imaginable. This leaves the users of AI tools without the critical friction needed to recognize what epistemologies, values, or ways of knowing are systematically excluded by algorithmic systems. This user base includes designers and design students, whose capacity of imagination may be affected by this embedded violence.

Proposed Approach: Hildegard as Pedagogical Provocation

This thesis presents *Hildegard*, a web-based interface and facilitated workshop framework designed using *Research-through-Design* and *Action Research* approaches.. Hildegard is a deliberate provocation: it encodes and amplifies the default assumptions embedded in students’ speculative inputs, making those assumptions visible enough to interrogate.

The framework operates through three interconnected phases:

1. Positioned Speculation: Participants build future scenarios grounded in their own situated perspectives and signals from their environment;

2. Amplification and Sparring: Hildegard receives scenarios, encodes biases using rule-based logic (not LLM inference), and returns amplified versions that exaggerate capitalistic, technocratic, or extractivist patterns. Students read and contest these amplifications, creating dialogue between their imaginations and the system’s encoded logic;

3. Reflection and Articulation: Facilitated debrief surfaces what assumptions became visible, with structured vocabulary work helping students name epistemic violence, data colonialism, and other structural patterns they had previously taken for granted.

This operation runs on the logic of standard debiasing that is overturned into critical amplification and repurposed into a political resource.

Empirical Findings

Two workshop iterations with design students (five participants each with varied design background) tested whether this approach reliably surfaces hidden biases. The findings are direct:

- Assumptions do surface. Participants noticed dystopian defaults, anthropocentric biases, technological solutionism, and geographic erasures only once the system made them unavoidable. Group dialogue amplified individual insights;
- Clarity is essential. The first iteration's unclear interface created confusion that undermined reflection. The refined version prioritized legible prompts, clear phase transitions, and structured synthesis to make its critical force intelligible;
- Neither the tool nor clarity alone is sufficient. While students recognized assumptions, they did not automatically develop decolonial vocabulary or imagine non-dystopian futures. The framework surfaces what is missing; additional infrastructure, including facilitation and vocabulary introduction, is required to deepen that critical work.

Strategic Contributions

The thesis contributes across three dimensions:

1. Methodological: Demonstrates how deliberate system limitations can be designed as a pedagogical resource, and how iteration between workshops and refinement clarifies rather than replaces critical intent.

2. Pedagogical: Provides a structured framework for surfacing and naming assumptions in speculative design work, useful in education, team innovation, and futures research contexts where the goal is to broaden imaginable possibilities rather than optimize a single vision.

3. Theoretical: Reframes synthetic data from a technical fairness tool into a critical medium, and positions tool friction and "failure" as necessary conditions for recognizing epistemic boundaries rather than errors to eliminate.

Scope and Value

This thesis does not propose Hildegard as a finished, deployable product, nor does it claim that a two-hour workshop solves design education's relationship to colonialism. Instead, it offers a *direction* that integrates three elements: an artifact

(the interface), a method (RtD with documented iteration), and a facilitation model (human mediation as structural, not peripheral).

For educators, design teams, and innovation practitioners, the framework provides a reusable pattern for responsibility in speculative work: it helps surface hidden assumptions earlier, improves the reflexivity of collective futures thinking, and creates structured space for interrogating what default narratives are being reproduced by the tools and processes in use. The main recommendation is to adopt this as a facilitated practice in design research and education contexts, not to deploy Hildegard autonomously, but to use the framework as an intervention where the outcome is not a "generated future," but a more critically aware group process.

Limitations and Next Steps

The findings are grounded in a small sample, single institutional context, and vibe-coded prototype-level implementation. The research demonstrates pedagogical promise and identifies necessary conditions (clarity, facilitation, vocabulary support) for the approach to function. Future work should test whether the framework transfers across disciplines and contexts, whether longer-term engagement deepens critical consciousness, and how institutional barriers such as time, resources, curriculum structure shape adoption.

What emerges is not a solution to data colonialism, but an honest tool for surfacing it and a recognition that names the hard work of actually decolonizing design practice, which extends far beyond any single interface or workshop.

01 Introduction

This thesis project addresses the critical gap between Eurocentric design education and the pluralistic needs of marginalized, global communities. As generative artificial intelligence utilities become deeply embedded in design pedagogy, they frequently reinforce rather than challenge Western-centric defaults. While these tools are commonly celebrated as neutral brainstorming assistants, they carry historical asymmetries that colonize speculative imagination before the design process even begins.

1.1 Problem definition

This project confronts the issue of *epistemic violence* perpetuated by AI systems within design education, particularly during the ideation of future scenarios. Because contemporary AI models are trained on datasets historically extracted, organized, and classified through colonialist frameworks, their standard application automatically reproduces biases against marginalized communities. Within pedagogical environments, these algorithmic defaults become embedded in the creative process.

The consequence is a reinforcement of epistemic hierarchies: Western knowledge systems dominate, while alternative ways of knowing, being, and relating to social or ecological systems are systematically silenced. This narrowing of the speculative range imposes a standardized human experience, erasing the pluralistic frameworks necessary to design resilient futures.

1.1.1 Alternative approaches to data and AI

To counter this erasure, this thesis draws upon the theoretical foundations of *data feminism* and *data decolonization*. Rather than attempting to fix datasets through superficial engineering, this research shifts the focus toward unpacking and exposing the power dynamics behind data collection and aggregation. By applying the lens of *feminist situated knowledges* to the outputs of generative AI, this project rejects the notion of objective, neutral data, treating computational systems instead as inherently political infrastructures.

1.2 The technical tool

The central technical artifact scrutinized in this study is synthetic data. While originally developed for privacy preservation or model robustness, synthetic data is increasingly promoted by frameworks like the EU Artificial Intelligence Act as a technological panacea for debiasing machine learning models. However, this research positions synthetic data as a deeply contested medium.

The industry debiasing paradigm treats bias as an apolitical defect that can be solved through statistical equity and parity constraints. This technical smoothing erases real-world

intersectional complexities and structural discrimination. Furthermore, synthetic data introduces an additional layer of processing that reflects the sociopolitical context of its creators, algorithms, and institutional agendas. Consequently, standard optimization protocols do not resist data colonialism; they comply with it by manufacturing an illusion of mathematical fairness that obscures ongoing systemic oppression.

1.3 Research question

How can synthetic data be repurposed with a decolonial objective within a pedagogical framework dedicated to ideating speculative future scenarios?

1.4 The intervention

In response to this question, this thesis introduces a conceptual shift from *algorithmic fairness* to *reflexive friction*. Instead of deploying synthetic data to sanitize or mask algorithmic defaults, this framework repurposes it as an antagonistic resource to amplify pre-existing biases and expose the structural omissions, the "ghosts", within both student inputs and AI infrastructure. These data ghosts embody a *hauntological encounter*, representing the persistent absences and lost futures erased by dominant, Western-centric narratives.

This strategy is operationalized through a custom digital co-speculation interface named *Hildegard*. Designed as an adversarial dialogue partner, Hildegard deliberately rejects the seamless user experience characteristic of commercial AI software. Instead of generating optimized, frictionless text, the interface utilizes synthetic data amplification to exaggerate the collective assumptions brought into the classroom space. By isolating and intensifying dominant patterns, the tool forces students to confront the limits of their own imaginative defaults, creating a site of pedagogical resistance through controlled computational constraints.

1.5 Methodology

The methodology follows a Research-through-Design (RtD) approach, where the primary research artifact is the functional digital interface of Hildegard. Rather than presenting a solutionist

product, the prototype operates as a physical embodiment of a theoretical provocation, designed to elicit critical reflections on data politics that remain inaccessible through literature review alone.

The implementation and evaluation of this framework follow an Action Research methodology. The inquiry is conducted through cyclical iterations of intervention, observation, and collaborative reflection within a live community of practice.

1.5.1 Project placement

The project is situated within the specific academic environment of Politecnico di Milano. As a leading European technical university, the institution operates predominantly within a Eurocentric tradition where dominant narratives shape the core curriculum. This setting serves as an ideal critical laboratory: implementing the framework in this environment challenges directly the assumed neutrality of technical education, testing whether an adversarial pedagogical intervention can disrupt the imaginative horizons of design students.

1.6 Thesis roadmap

The thesis is organized into eight sequential chapters (following this introduction):

Chapter 2 maps the problem space of generative AI in design pedagogy, illustrating how automated brainstorming tools can enact epistemic violence against minorities.

Chapter 3 examines the sociotechnical mechanics of synthetic data, problematizing its current optimization use cases.

Chapter 4 reviews existing design strategies, activist practices, and alternative frameworks aimed at decolonizing technological infrastructures.

Chapter 5 details the overarching Research-through-Design and Action Research methodology.

Chapter 6 outlines the proposed pedagogical framework, detailing the interaction design, architecture, and structural mechanics of the Hildegard interface.

Chapters 7 and 8 present the qualitative findings, thematic

analysis, and subsequent design iterations derived from the deployment of two iterations of the framework in a design workshop setting.

Chapter 9 addresses the research questions and presents the research contributions, limitations and scope of the thesis. It also discusses future research for this project and concludes with final remarks.

02 Research approach and methodology

This chapter frames the methodological positioning of this thesis project as a critical intersection of Research-through-Design (RtD), Action Research, and decolonial feminist epistemologies by defining first how design experimentation generates knowledge through iterative making, and then discussing how Action Research supports this intervention-oriented framework. Starting from the concept of situated knowledge and colonial hauntology, the research discusses how the institutional environment of Politecnico di Milano serves as the basis for examining embedded academic biases and dominant technical narratives. Finally, the chapter outlines the exploratory expert consultations used to ground the project, and details the analysis of the workshop iterations. The chapter ends by defining the specific methodological limits and contextual perimeter of this project.

2.1 Methodological positioning

This research adopts a *Research-through-Design* methodology (Zimmerman et al., 2007), treating the development of the pedagogical framework as both process and outcome. As Krogh et al. note, RtD involves «process-loops where hypothesis, experiments, and insights concurrently affect one another» (Krogh et al., 2015), which captures the way this project develops the workshop, the interface, and the evaluation instruments together (Krogh et al., 2015). The iterative design of the workshop, technical system, and evaluation instruments generates knowledge about how synthetic data can function as a critical pedagogy tool. The reason for this is that the objective of this thesis is not to provide a direct solution to a practical issue, but rather to initiate a discussion on a matter that is currently being studied and addressed in a variety of ways in different areas of the world. The method is well suited to this project because the it unfolds through iterative making, reflection, and revision rather than through the testing of a fixed hypothesis.

Action Research is used here as a supporting methodological logic rather than as a separate second methodology. It helps describe the participatory and intervention-oriented aspects of the project, especially the consultations and the workshops' iterative refinement. This is consistent with how AR is understood in the literature: it is «explicitly democratic, collaborative, and interdisciplinary», and «the intervention, the learning, and the doing and knowing cannot be disentangled» (Hayes, 2018). In this thesis, that means the research process is not detached from the context it studies, but shaped through confrontation with it (see Chart 2.3).

2.2 Epistemic stance and positionality

The critical architecture of this project is explicitly guided by feminist and decolonial epistemologies, which reject universalist claims of objective, value-free data systems. Instead, this research operates through the concept of *situated knowledges*, which posits that all knowledge is produced from specific historical, material, and cultural positions. By examining how Western design pedagogy produces normalized, market-driven default imaginaries,

the project seeks to unmask the «*god trick*» (Haraway, 1988) of computational neutrality, which is the illusion of positioning data systems as disembodied, omniscient narrators of the future.

To make these underlying power dynamics visible, the project uses a *colonial hauntology* approach. Colonial hauntology addresses the cultural, political, and material traces of historical injustices that cannot be completely represented within dominant datasets, yet refuse to be laid to rest because the structural extractions that unleashed them remain ongoing. Within this thesis, hauntology provides a conceptual vocabulary to analyze the return of excluded alternative futures and missing indigenous or non-market knowledges.

Crucially, because this research is conducted by a white Western researcher trained within European academic institutions, the critique of Eurocentric epistemic dominance originates from within the system it interrogates. Acknowledging this embeddedness, the project does not claim a detached or fully "decolonized" standpoint; rather, it deploys deliberate interface failure as a self-reflexive method to confront, audit, and destabilize its own institutional complicity.

2.3 Research site and rationale

The empirical dimension of this project is situated within the School of Design at Politecnico di Milano. As a premier European technical university, its institutional legacy is deeply rooted in the Western design canon and the traditions of industrial rationalism. Consequently, its standard curricula are aligned with dominant technocratic and capitalist narratives. When students within this environment adopt generative AI utilities without critical friction, these pre-existing defaults are amplified, making it difficult to critique automated narratives.

Quantitatively, the institution exhibits a stark gender imbalance: data from the Ministero dell'Università e della Ricerca (MUR) indicates that female students comprise only 36.3% of the student population at Politecnico di Milano, compared to a 57% average across Italian state universities (USTAT, n.d.). Similarly, there are very little international students, and between them the gender imbalance is even more evident (see Table 2.1, Chart 2..1 and Chart 2.2).

It is worth noting that, from an intersectional perspective,

it is impossible to quantify the number of minoritised student, which only highlights that this is not a critique of the university setting, politics or body, but rather a formal situated rationale of the research site, consistently with the theoretical basis that the thesis relies on. This institutional setting serves as a laboratory in which to test whether an adversarial, decolonial design intervention can disrupt highly standardised academic traditions.

Table 2.1: The student population of Politecnico di Milano (USTAT, no date). Table highlights the percentage of different students in the student body.

STUDENT POPULATION

STUDENTS	TOTAL	OUT OF WHICH WOMEN	OUT OF WHICH INTERNATIONAL	YEAR CONSIDERED
Enrolled	7.820	2.937	473	2024/2025
Registered	48.124	17.488	7.952	2024/2025
Graduated	13.929	5.829	473	2024

The term 'ENROLLED in the academic year t/t+1' refers to a student who, as at 31 July t+1, is enrolled in the academic year t/t+1. 'ENROLLED' in the academic year t/t+1 refers to a student who is enrolling for the first time in their life in the academic year t/t+1 on a university course at an Italian university (excluding two-year master's degree programmes). The data relating to the totals by REGION refer to the location where the degree programme is taught, whilst the data reported for each institution concern students enrolled at that institution.

Chart 2.1: The student population of Politecnico di Milano (USTAT, no date). Chart showing the number of women enrolled compared to the total number of students.

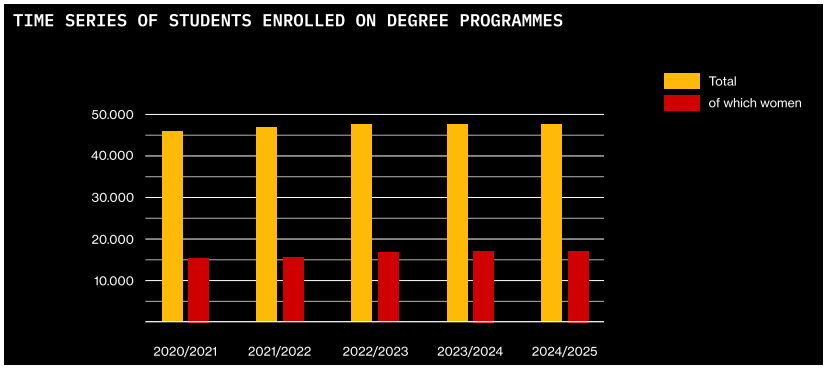


Chart 2.2: Student population of Politecnico di Milano (USTAT, n.d.). Chart highlighting the number of international students enrolled.

2.4 Exploratory expert consultations

To ground the pedagogical framework in contemporary classroom realities, five semi-structured, exploratory expert consultations were conducted in November and December 2025. These consultations aimed to investigate how design faculty perceive generative artificial intelligence, how they integrate these systems into their design studios, and whether they actively critique or encourage student adoption.

Purposive sampling was used to select faculty members and researchers from the School of Design (see Table 2.2) who demonstrated involvement or research interest in artificial intelligence for creativity, speculative design practices, or institutional pedagogy. A total of five experts were consulted, comprising four professors and one postdoctoral researcher.

The consultations were conducted as semi-structured interviews, a format chosen for its capacity to ensure thematic consistency while allowing the flexibility to find unexpected insights. Interview questions were drawn from a pre-established database (see Appendix A) and dynamically adapted to align with each expert's specific domain of expertise. To capture these qualitative insights from the field, the researcher documented the interactions through comprehensive, thematic note-taking during the interviews.

2.5 Workshop design and procedure

The primary RtD intervention of this thesis consists of a series of facilitated co-speculation workshops. In design research, the empirical value of a workshop lies in observing how a prototype behaves in a live setting, how it is used, and how participants negotiate its constraints. Accordingly, these sessions are treated as evaluative research instruments rather than standard instructional events.

The operational mechanic of the workshop relies on a deliberately limited and "stupid" interface. Instead of providing optimization, the software is engineered to amplify systemic

defaults, overemphasize dominant patterns, and suppress alternative variables. Participants are explicitly tasked with attempting to break the tool. Technical breakdown and computational refusal are not treated as system defects; they are operationalized as the precise conditions through which algorithmic exclusion becomes legible to the user.

PARTICIPANT CODE	ACADEMIC ROLE / AFFILIATION	EXPERTISE DOMAIN
Expert A	Associate Professor	• Collective Intelligence, Digital Anthropology, Social Innovation. • Directly involved in the Polimi official survey regarding the usage of GenAI by design students.
Expert B	Postdoctoral researcher	• Critical and responsible integration of genAI in design workflows. • Human-AI interaction, agency, and creative processes. • Cognitive demands in generative AI-supported creative tasks. • Inclusive and emotionally aware approaches to AI in education and practice. • Methods and conceptual frameworks for reflective AI use.
Expert C	Associate Professor	• Speculative design, anti-disciplinarity and thinkering. • Study and experimentation with identity and branding systems applied to new technological and hybrid contexts. • Design of user-centred information.
Expert D	Full Professor	• Coordinator of a Master's Degree Programme at the School of Design. • Service design, with a particular emphasis on the public sector and healthcare • Multidisciplinary and interdisciplinary collaboration with related fields of service research and service science, where the contribution of design is playing an increasingly significant role.
Expert E	Full Professor	• Member of the institutional leadership team at the School of Design. • Strategic, systematic and creative research through design, focusing to the ecological impact of business innovations.

Table 2.2: Participants list of the experts consulted. The expertise domain was the main criteria behind the selection.

The full procedure of the workshops will be explained in Chapter 7.

2.5.2 Data generation and documentation

The workshops are documented through a mixed-method approach:

- Field notes on the facilitation: used to track the procedure, the interface outputs, and the critical moments, especially the unspoken ones.
- Participants' exit interviews: following a semi-structured methodology (see Appendix B), the interviews help understand the reception of the workshop and the human-interface interaction. The interviews are transcripted and, when held in Italian, translated into English (see Appendices E and G).
- Workshop artefacts/scenarios: includes the notes taken by the students during the brainstorming phase, the screen recording of their interactions with the interface, their inputs, and any possible other artefacts created (photographed by the researcher during the workshop) (see Appendices D and F).

2.5.3 Analysis approach

The coding of the workshop documentation is performed following a mixed-methods, qualitative approach, which includes reflexive thematic analysis for the field notes and the exit interview, and multimodal / visual artefact analysis for the workshop artefacts.This framework integrates reflexive thematic analysis for the field notes and exit interviews, with multimodal artefact analysis for the physical and digital outputs generated during the sessions. Given the project's grounding in feminist and decolonial epistemologies, the analytical objective is not to quantify data or enforce rigid inter-coder agreement, but rather to surface meaning, situated knowledge, underlying tensions, and productive contradictions.

The analytical process is structured across three distinct phases:

- Phase 1 - Reflexive thematic Analysis (interviews and field notes): An initial step is to utilize open, semantic codes to capture what participants explicitly articulate regarding

their experiences. In alignment with Braun and Clarke's conceptualization of reflexive thematic analysis, this phase is interpretive and subjective, treating the researcher's subjectivity as an analytical tool rather than a source of contamination (Braun & Clarke, 2019).

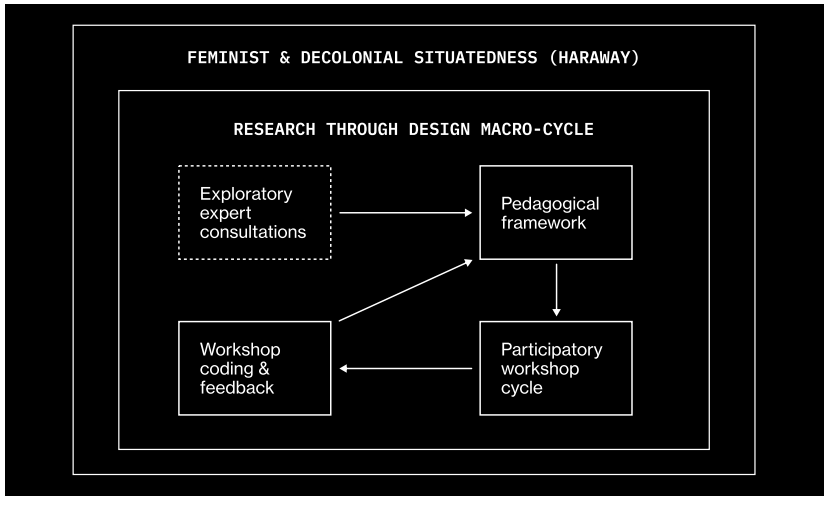
- Phase 2 - Multimodal artefact analysis: The second phase shifts to the material and visual outputs, including student-generated scenarios, diagrams, post-it notes, sketches, and chronological versions of the synthetic dataset. These artefacts are used to identify the embedded assumptions, structural exclusions, and cultural values they visualize.

- Phase 3 - Case integration and synthesis: The final phase integrates streams, using the reflexive thematic analysis as main evidence that is more or less supported by the artefact analysis.

2.6 Methodological limits

This approach is intentionally situated and exploratory. The consultation material is based on note-taking rather than transcription, the workshop will take place in a single institutional context, and the sample is small by design. The thesis is, in fact, not claiming to produce a general model for all design education contexts, but to develop and examine a critically situated intervention that can be transferred carefully, not automatically.

Chart 2.3: A summary of the methodology adopted through this thesis.



03 The problem of data colonialism

This chapter frames the problem scope of this thesis project as a critical intersection of Generative Artificial Intelligence, epistemic violence, and design education by defining first all forms of technology as potentially political, and then discussing how technology can be used as a tool of oppression. Starting from the matrix of oppression, the research discusses how the current social orders of surveillance capitalism and data colonialism are the basis for AI being used as an oppressive tool. Finally, the chapter discusses the current uses of AI in design education, and how in this specific setting it embeds epistemic violence in the design process outputs. The chapter ends by defining the perimeter of this project.

3.1 GenAI in design education

To ground the theoretical inquiry of this thesis in current practice, this research starts from a series of expert consultations conducted with faculty, departmental leadership, and researchers at the School of Design of Politecnico di Milano. These interactions were documented through thematic note-taking during semi-structured interviews, providing a qualitative view from the field. The following analysis synthesizes these observations to map the invisible shifts in student creative praxis and the institutional philosophies emerging in response to Generative AI.

A recurring theme across the consultations is a state of "unconscious integration". Academic researchers at the university observe that while students perceive GenAI as a creatively neutral partner, relegated primarily to repetitive or superficial tasks, the tools are, in fact, acting as «instruments of inquiry» (Dalsgaard, 2017) that fundamentally restructure the designer's cognitive process.

This restructuring requires what departmental researchers define as «conceptual alignment». This process demands that the designer translates idiosyncratic creative intent into the rigid, standardized parameters of the machine. The study identified a tripartite hierarchy of this alignment: simple, complex, or impossible. In instances of «impossible alignment», a profound discrepancy arises between the designer's cultural or individual vocabulary and the algorithm's categorical constraints.

The expert notes reveal a persistent conflict between the pursuit of operational efficiency and the preservation of speculative depth. Within the speculative design curriculum, pedagogical leaders observe that GenAI is frequently utilized to «simplify passages», accelerating the production of high-fidelity prototypes such as deepfakes, automated code, and synthetic voiceovers.

Furthermore, university experts have highlighted a growing «pedagogical crisis of validation». The metacognitive ability to critically evaluate AI-generated knowledge requires a level of prior domain expertise that is becoming increasingly difficult for students to attain during their studies. This can result in a «tick-box» verification culture, where the mere existence of an output (e.g. a functional link or a synthetic persona) is prioritised over its

truth value or cultural relevance.

Beyond institutional philosophies, the expert consultations identified specific stages of the design process where GenAI has become most entrenched. These applications represent a shift in the design process:

- Ideation and information synthesis: Students primarily utilize LLMs (Large Language Models) like ChatGPT or Claude for the initial research and brainstorming phases. This includes the generation of synthetic user personas, which are fictionalized archetypes used to simulate user needs. As noted in the field research, these personas often adopt syntax and language that reflect specific social classes, inadvertently embedding existing socio-economic biases into the project's foundation before a single sketch is made.

- Visual communication and prototyping: In the production of communicative artefacts, image-generation tools are used to create frictionless mood boards and high-fidelity visualizations. In the context of speculative design, students leverage GenAI for image manipulation, voiceover synthesis, and the production of deepfake videos to add a layer of synthetic realism to their prototypes. This speeds up the transition from concept to high-fidelity output, allowing students to translate data into complex visual narratives with minimal manual labor.

- Technical scaffolding: For interaction design projects, GenAI is increasingly used for programming and coding prototypes. Students prompt AI to generate functional code, allowing them to build complex digital interactions that might otherwise be beyond their technical proficiency. This scaffolding allows for more ambitious prototyping but, as observed by university leadership, often results in a "black box" approach where the student understands the output of the code but not the logic or the potential biases within the underlying script.

Ultimately, these uses highlight a shift toward «Meta-design adaptation». Rather than designing from a blank page, the student's role is increasingly one of prompting and correcting. This reinforces the need for an intervention that moves beyond pure operational efficiency to foster metacognitive critical reflection.

For the sake of simplicity, this thesis will use the terms "AI" and "GenAI" interchangeably.

3.2 Power and oppression through technology

While GenAI has emerged as a pervasive and potentially beneficial tool in design pedagogy, increasingly utilized by students as a partner for brainstorming, its integration is far from neutral.

Generative Artificial Intelligence has emerged in the last five years as a tool towards the democratization of various activities, including creativity and the design profession, a frictionless bridge between thought and visualization. However, underneath this convenience lies a complex infrastructure of extraction, which affects the way we live, work and think. To understand the problem of GenAI in design pedagogy, one must first recognize that technological tools are not neutral agents, and in the past 200 years, has had a huge role in affecting our current political structures, starting with the industrial revolution (Suleyman, 2023). GenAI is the most recent manifestation of a centuries-old colonial logic, now operating under the guise of algorithmic efficiency.

Langdon Winner, 40 years ago, was already discussing the enmeshment of technology and politics (Winner, 1980). He states that technology is not only a set of neutral artefacts but is intertwined with human society and politics, which he exemplified multiple ways in which technology can be political: through the way it's used — such as bridges that were built in New York specifically to exclude poor people from accessing certain areas of the city —, through its compatibility to specific political engagement — he discusses how solar and nuclear power have opposite political qualities, solar being more a democratic and nuclear more authoritarian —, or inherently due to the way it's designed to benefit certain people — this is the case of algorithmic models, which are often developed and tested by small homogeneous groups (Winner, 1980).

This has been brought up by Safiya Noble (2018) in her book *Algorithms of Oppression*, where she creates the term *technological redlining*, which is «how digital decisions reinforce oppressive social relationships and enact new modes of racial profiling» (Noble, 2018). To do so, she brings up various examples of oppressive algorithms, especially in Google searches, which are biased due to the algorithms and data they are trained on, and how that can reinforce people's racist and sexist biases. An

example that is very representative is related to the Charleston church shooting, which was perpetrated by a White nationalist, Dylann Roof, in 2015 in the Mother Emanuel African Methodist Episcopal church, which had been for centuries a symbol of a struggle for freedom from White supremacy. Roof was found to have published online a racist manifesto right before the shooting, in which he said that Adolf Hitler would someday «be inducted as a saint», and he warned that unless white people «take violent action, we have no future» (Sack & Blinder, 2017).

«The event that truly awakened me was the Trayvon Martin case,» the text reads. «I kept hearing and seeing his name, and eventually I decided to look him up. I read the Wikipedia article and right away I was unable to understand what the big deal was. It was obvious that Zimmerman was in the right.» (Neuman, 2015)

According to the manifesto, Roof, who had not been raised in a racist home or environment (Neuman, 2015), researched "black on white crimes" on Google to make sense of the killing of Martin, a young African American teenager who was killed by George Zimmerman. What Roof found was information proving that America, just like Europe, was going through an emergency of black on white crimes that were not being reported on the news, while cases like Martin's murder were blowing up.

In her research, Noble looked into the results that would have shown up when searching «black on white crime» in the time range and location of Roof when writing his manifesto: the first results were from conservative websites that posted openly anti-Black racist articles. This research didn't show the FBI report on crime statistics that showed how crimes against White Americans is a largely interracial crime mainly committed by White Americans (Noble, 2018).

Algorithmic models are developed and often tested in small homogeneous groups of scientists who often belong to dominant groups, predominantly white, male, academically educated, and then scaled up and deployed for a bigger, wider, more diverse audience. D'Ignazio and Klein in *Data Feminism* (2020) describe this deficiency as a privilege hazard: the phenomenon whereby those occupying the most privileged positions, those with respected credentials and professional accolades, are poorly equipped to recognize instances of oppression in the world. It is evident that technologies such as GenAI are designed by a limited sample of people which are not capable of addressing the

complexity of the real world through the whole design process of technological tools.

The examples of algorithmic bias documented by Noble, Benjamin, and D'Ignazio & Klein reveal a particular form of technological power: one that operates through data. «In the past decade or so, many of these men at the top have described data as 'the new oil'» (D'Ignazio & Klein, 2020). This gives a good idea of how the development of new technologies, particularly those that are based on data, happens by creating a separation between oppressors (those who are behind the system) and oppressed (the source of data).

To understand how oppression works on different levels, D'Ignazio & Klein analyse the data market through the lens of Patricia Hill Collins's *Matrix of Domination* (Collins, 1990) (see Table 3.1), which divides oppression into four interlocking domains:

- The structural domain: Laws and policies that codify oppression (e.g., data ownership rights).
- The disciplinary domain: The bureaucratic management of these policies (e.g., algorithmic sorting and "neutral" data processing).
- The hegemonic domain: The circulation of ideas and culture (e.g., the visual archetypes generated by AI).
- The interpersonal domain: The everyday experience of the individual (e.g., the designer's interaction with a biased interface).

Table 3.1: The matrix of domination by Patricia Hill Collins (Collins, 1990).

THE FOUR DOMAINS OF THE MATRIX OF DOMINATION

Structural domain	Disciplinary domain
Organises and codifies oppression: laws, policies	Administers and manages oppression. Implements and enforces laws ad policies.
Hegemonic domain	**Interpersonal domain**
Circulates oppressive ideas: culture and media.	Individual experiences of oppression through everyday interpersonal interactions.

This matrix provides a framework to help define the different layers of the problem, enabling a focus on the specific scope of the problem that this thesis aims to address.

3.3 Surveillance capitalism

To understand the structural and disciplinary forces behind Generative AI, we must look to the economic logic driving contemporary technology: surveillance capitalism. As Shoshana Zuboff (2020) defines it, this system «unilaterally claims human experience as free raw material for translation into behavioral data» (Zuboff, 2019). These data are declared as a proprietary behavioral surplus that is processed through machine intelligence to create prediction products designed to anticipate an individual's future actions (Zuboff, 2019). This shift creates what Zuboff terms behavioral futures markets (see Chart 3.1), where the fundamental substance of human experience is traded for profit (Zuboff, 2019).

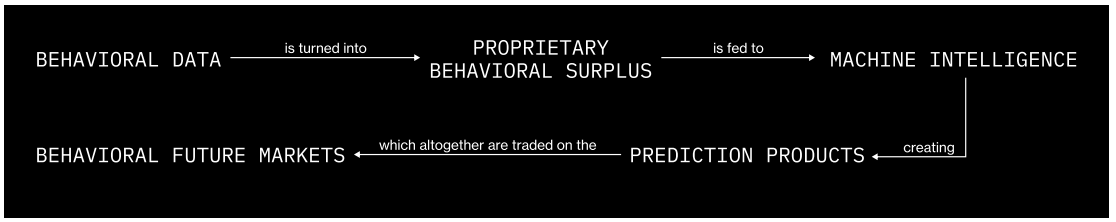


Chart 3.1: The process of surveillance capitalism and how it turns behavioral data into behavioral future markets (Zuboff, 2019).

According to Zuboff, surveillance capitalism is a «radically disembedded and extractive variant of information capitalism» in which information gathered from a user's touch points, ranging from social media likes to environmental sensors, is often framed as a means to enhance user experience. However, these systems infer details about an individual that extend far beyond the explicit terms and conditions of use. The danger lies in the lack of governmental supervision and the fact that tech companies claim ownership of private experience as if it were a «natural resource» (Zuboff, 2019).

The institutional freedom enjoyed by surveillance capitalists is often justified by their apparent alignment with standard capitalist principles of productivity and growth. Yet, as Zuboff suggests, this system operates under a «new logic of accumulation» with its own specific «laws of motion» (Zuboff, 2019). Since this represents a radical departure from traditional market operations, it requires a specialized set of limitations and legal interventions tailored to its unique extractive mechanisms.

3.4 Data colonialism

This extractivist logic that is needed by surveillance capitalism to feed itself, is the foundation of data colonialism.

In their book *The Cost of Connection* (2019b), Couldry and Mejias describe data colonialism as

> the extension of a global process of extraction that started under colonialism and continued through industrial capitalism, culminating in today's new form: instead of natural resources and labor, what is now being appropriated is human life through its conversion into data. (Couldry & Mejias, 2019b).

This system degrades life through multiple interconnected mechanisms. First, it naturalizes data capture by treating personal data as a natural resource that is just there for the taking: a digital equivalent of colonial powers declaring inhabited lands as *terra nullius* or *no man's land*, creating a legal and philosophical fiction that justifies appropriation without consent (Couldry & Mejias, 2019a). This poses data colonialism in continuity with modern colonialism. The extraction process is enabled by what Couldry and Mejias term «extractive rationalities»: a social rationality that treats labor contributing to data extraction as valueless, a practical rationality positioning corporations as the only entities capable of processing data, and a political rationality framing society as the beneficiary of corporate data extraction (Couldry & Mejias, 2019b). Through these rationalities, data colonialism transforms human interactions into «data relations» (Couldry & Mejias, 2019a) where social activities are made to be trackable, capturable, and commodifiable, converting everyday life into abstractions that can be annexed to capital (Couldry & Mejias, 2019a).

This process actively produces what Couldry and Mejias call «computed sociality», a new type of "social" designed from the ground up to be tracked and quantified (Couldry & Mejias, 2019b). The legitimation of this colonization operates through discursive tools analogous to historical colonial instruments: Terms of Service agreements function like the Spanish colonial *Requerimiento*, incomprehensible documents that create binding relations of colonization, both relying on underlying force, physical in traditional colonialism, economic concentration in data colonialism (Couldry & Mejias, 2019b). These practices expose

life continuously to monitoring and surveillance, making tracking a permanent feature of existence that erodes what Couldry and Mejias describe as «the minimal integrity of the self», threatening human identity, autonomy, and development (Couldry & Mejias, 2019b).

The extracted data produces «data doubles»: algorithmic representations of individuals that bypass human beings but nonetheless have real consequences, forming the basis for discrimination, pricing, and penalties (Couldry & Mejias, 2019a). Additionally, data colonialism imposes a totalizing vision that, unlike other cultural systems that acknowledge the «irreducible, contradictory character» (Couldry & Mejias, 2019b) of reality, enforces a rigid categorical system oriented toward algorithmic control, systematically excluding ways of being and knowing that cannot be captured, quantified, and computed (Couldry & Mejias, 2019b). This makes data colonialism a comprehensive apparatus for remaking human social life according to the logics of extraction and control.

3.5 The god trick, cultural hegemony and epistemic violence

As stated before, data colonialism presents data as a natural resource that is just there for the taking. This erases the fact that there is actually someone actively capturing data: companies that do data scraping, made by the individuals who work for them. This creates a dominant way of knowing that is established by a small group of people by hiding the subjectivity of both data and how it's gathered, processed and used. This is what Donna Haraway (1988) calls the god trick. Haraway uses the metaphor of vision, which has historical value to interpret the act of producing knowledge. She claims that the god trick presents someone who is observing an object, but distances themselves so much that at some point they are not visible enough, and wants to see without being seen. The data market, in this social order, is not created and run by someone, but it is seen as the natural state of things, while it actually is created from the perspectives of someone and somewhere, usually white, male, heterosexual human. The god trick denies subjectivity, and denies relativism to any possibility of objectivity. This makes it so that certain groups, those that

are excluded from the processes of knowledge production are excluded from mainstream scientific discourse entirely, silencing their perspectives and experiences. The dominant group decides what counts as real science, which questions are worth asking, and whose knowledge is considered legitimate. The god trick is necessary for the elite to have control of the hegemonic domain of domination in surveillance capitalism/data colonialism.

This phenomenon aligns with Antonio Gramsci's theory of cultural hegemony (1985), which describes the process by which the ruling elite's worldview is internalized and accepted as common sense by the majority, despite primarily serving the interests of a minority (Gramsci, 1985). In this framework, capitalist society is not a collaborative cultural endeavor; rather, it is a site of discursive control where the elite define the boundaries of what is considered reasonable, acceptable, or even imaginable (Gramsci, 1985). By exerting power over the circulation of ideas and images, the hegemonic class ensures the stabilization of the existing political and economic order through the subtle manufacture of consent rather than overt force (Gramsci, 1985).

Cultural hegemony, created and enforced in the economic and social orders discussed previously, functions as a form of oppression applied to the domain of knowledge, which creates a specific type of violence, called epistemic violence, which is «violence exerted against or through knowledge» (Galván-Álvarez, 2010). This term refers to harmful knowledge production practices that negatively impact minoritized groups through biased interpretation and representation.

In *Data Feminism*, this has been defined as «the harm that dominant groups like colonial powers wreak by privileging their ways of knowing over local and Indigenous ways» (D'Ignazio & Klein, 2020).

3.6 The oppressors and the oppressed

The victims of this violence are those excluded from the data extraction practices because of the imposition of inadequate data as knowledge, ignoring the lived experiences of those harmed. In addition, while data inadequacy is not always maliciously intended, it comes embedded in the colonialist practices which are systemically excluding minoritized groups from the

mainstream knowledge system by imposing production and extraction hierarchies on Indigenous communities (Brunner, 2021). For example, Gondwe (2023) discusses how AI tools such as ChatGPT offer opportunities for effective journalism, but also pose challenges due to data poverty and the selective approach to civil and uncivil language in sub-Saharan languages (Gondwe, 2023).

Defining the actors within the colonizer/colonized dichotomy requires navigating a dense landscape of evolving terminology. Each term carries a specific theoretical weight, and choosing the right one is essential for mapping the Matrix of Domination (Collins, 1990) within the digital sphere.

Historically, the dynamic has been framed through Geopolitical Binaries, such as the North/South (or the plural Souths). While these terms are useful for identifying sites of resistance and alterity (Milan & Treré, 2019), they are often critiqued for their lack of precision and their tendency to ignore internal class hierarchies and non-Western imperial powers like China (Dados & Connell, 2012).

In contrast, Identity-Based Labels, such as “minority” or “marginalized”, have been used to identify those excluded from the “god trick” of universal objectivity (Haraway, 1988). However, “minority” implies a static, numerical fact, while “marginalized” can suggest a passive state of being, often failing to name the active force that pushed the individual to the margins in the first place.

For this thesis, the term «*minoritized*» (Ledesma & Calderon, 2015) is adopted as the primary descriptor for the knowledge systems and populations targeted by data colonialism. The transition from minority (a noun) to minoritized (a verb) is a critical shift for multiple reasons. First, it names the system: it does not imply that a group is naturally small or lesser, but is being actively rendered so by the structural and disciplinary domains of power. In GenAI, this is the process where algorithmic models treat non-Western data as noise or outliers. Second, it is inherently intersectional: it is a functional term that accommodates intersectionality (Crenshaw, 1989). It allows the research to discuss how various axes of oppression, such as race, gender, geography, interact to “minoritize” an individual within a dataset. Finally, it aligns with design agency: in the context of design education, “minoritized” highlights the actionable nature of the problem. If marginalization is a process enacted by a tool, then the designer’s role is to intervene in that process.

By using minoritized, this thesis avoids the «simple reversals» (Asher, 2013) of North-vs-South. Instead, it views oppression as a dynamic data relation. This terminology allows the proposed framework to address epistemic violence not as a fixed state, but as a technical and cultural process of erasure that can be resisted through critical, situated design interventions.

3.7 GenAI as a design tool, and how it perpetrates epistemic violence

To sum up, surveillance capitalism is an economic system that relies on stealing data from humans, and acts on the structural domain of domination; to steal this data, it uses colonialist practices, which together form the social order of data colonialism, that behaves on the disciplinary domain of domination (these two layers are deeply intertwined, the distinction helps define the problem better but should not be taken too strictly). This system relies on technology to spread information, but is specifically designed to make it so that this technology seems objective and neutral, hiding that it is actually designed (like the whole system) by a small homogeneous group of people whose views are limited and potentially oppressive. These ideas become the standard, the default, the truth, erasing other systems of knowledge and enacting epistemic violence on those who are not part of the knowledge production system. This cultural hegemony corresponds to the hegemonic domain of domination. In this system, the interpersonal domain is occupied by the daily manifestation of these systemic forces through individual interactions. Epistemic violence is a systemic issue that falls in the hegemonic domain of oppression, but it expresses itself in daily interactions.

This system is catalysed by the industry of GenAI, which is driving an escalating demand for data with the primary objective of generating profit. Examples of such tools include ChatGPT, Dall-E, Gemini and Adobe Firefly, which are used on a daily basis for a variety of purposes. Their use has accelerated the process of information production and diffusion, cementing the cultural hegemony of the dominant class and creating more and more cases of epistemic violence.

When GenAI is used in the design process, it is necessary to consider it in conjunction with the challenges posed by data

colonialism, the potential for complications becomes evident. The outputs designed by these entities would be characterised by bias and a paucity of representation of knowledge systems, ideas, opinions and even imagery from minoritized groups (see Figure 3.1).

This erasure is supported by large-scale audits of GenAI models. A 2023 analysis by Bloomberg of over 5,000 images generated by Stable Diffusion reveals that the model actively amplifies real-world disparities to extreme degrees (Nicoletti & Bass, 2023). For instance, while women represent 39% of doctors and 34% of judges in the United States, the model generated women for these roles in only 7% and 3% of instances,

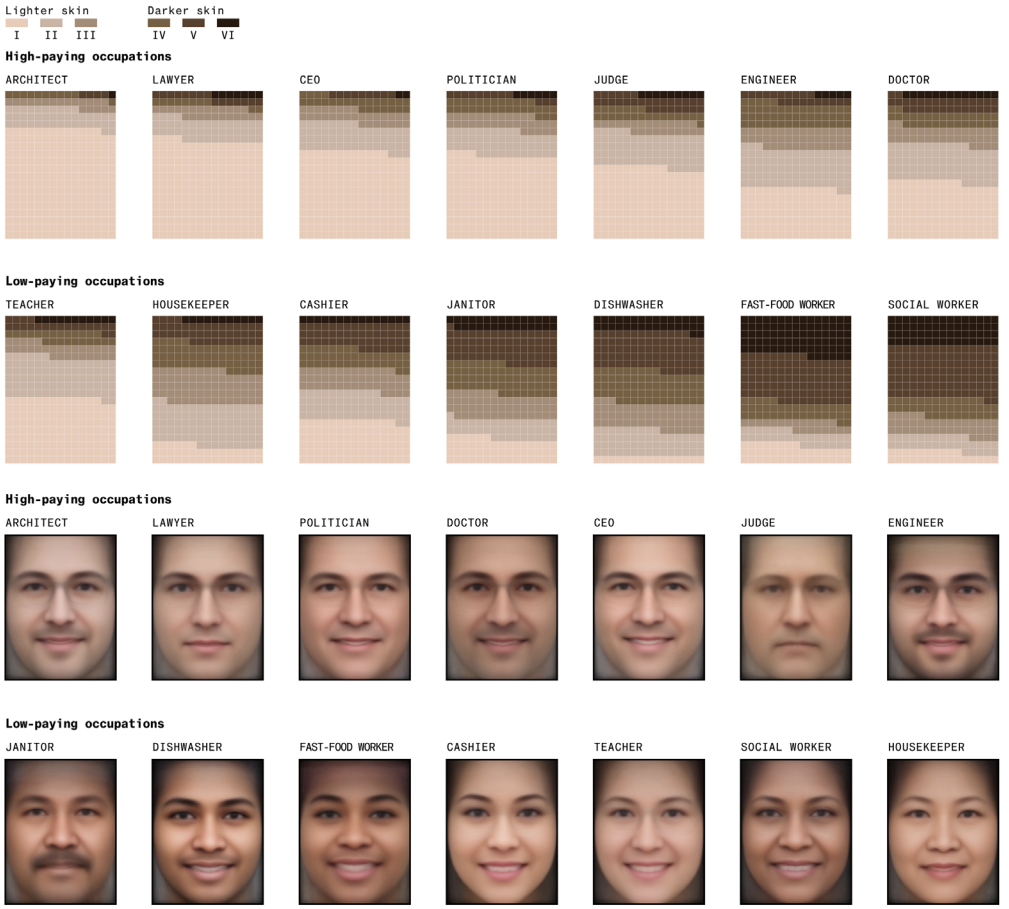


Figure 3.1: Results of the Bloomberg study, demonstrating how Stable Diffusion replicated and amplifies real world disparities when generating different types of workers (Nicoletti & Bass, 2023).

respectively (Nicoletti & Bass, 2023). This amplification of the “god trick” extends to racialized economic archetypes: images for high-paying occupations such as “CEO” or “Politician” were dominated by subjects with lighter skin tones, whereas subjects with darker skin tones were disproportionately generated for low-paying prompts like «fast-food worker» or «social worker». (Nicoletti & Bass, 2023)

Furthermore, the study highlights how these models cement oppressive visual archetypes in the context of criminality. More than 80% of images generated for the prompt “inmate” depicted individuals with darker skin, significantly exceeding actual incarceration demographics. Similarly, the prompt for “terrorist” exclusively produced men with dark facial hair and head coverings, leaning into specific religious and racial stereotypes. These outputs coming from a biased archive are recirculated and legitimized as objective digital reality. When students utilize these tools for ideation, they are interacting with a technical infrastructure that mathematically enforces the hegemonic domain, systematically rendering minoritized identities as either invisible or stereotyped (Nicoletti & Bass, 2023).

3.8 Problem statement

The theoretical frameworks of surveillance capitalism and data colonialism do not remain confined to digital infrastructures; they manifest tangibly within the design studio. As Generative AI is increasingly adopted in design education, it introduces what sociologists of education call a «hidden curriculum» (Freire, 1974), which is made up of the unstated lessons, values, and perspectives that students absorb through the tools they use. When a design student prompts a GenAI model for brainstorming, the god trick described by Haraway becomes a daily interaction. The student is actively interacting with a tool that enforces the hegemonic domain by prioritizing Western visual archetypes as the universal default.

This project seeks to address the problem space not by “fixing” the data, a task that remains in the structural domain, but by intervening in the hegemonic domain through design pedagogy. This leads to the aim of this thesis: to demonstrate that certain futures cannot be generated due to the tool’s epistemic constraints. The intervention has the goal of supporting design students to develop critical awareness of how data colonialism

operates, which is done through a framework for reflexive practice in their ongoing design work to go beyond the god trick.

3.8.1 Speculative design as playground and methodology

The proposed intervention is strategically situated within the scenario creation phase of Speculative Design. The concept of speculative design can be traced back to critical design, a field of design that emerged in the mid-nineties as a non-solutionist approach to real-world problems. However, critical design has come under scrutiny in recent times due to a misinterpretation of critical theories. In contrast with critical design, speculative design places significant emphasis on the concept of future scenario building by enabling a broader context for artefacts and imagination (Celik & Kaya, 2025), due to its close association with design fiction. Speculative design has the objective of exploring alternative futures while engaging with communities to understand their reactions and to influence policies by addressing problems in a bottom-up manner (Cooper et al., 2018).

This specific branch of design has been chosen due to its strong political connotation and its capacity to rethink and envision imaginaries. Speculative design shares roots with the Critical design movement of the mid-nineties, which was defined by its non-solutionist stance toward social problems, but it has evolved to address the systemic critiques leveled against its predecessor. Early critical design often faced scrutiny for being a product of Western, middle-class privilege that observed problems from a distance (de Oliveira & Prado de O Martins, 2014), and even speculative design has been criticized for its neoliberal and modernist foundations. Prado de O. Martins and Oliveira (de Oliveira & Prado de O Martins, 2014) point out that the discipline often operates within a «narrow Northern European middle-class confine», successfully managing to ignore issues of race, class, and gender. They argue that many dystopian futures imagined by Western designers, such as surveillance or loss of autonomy, forget that these conditions are not future possibilities but historical and present realities for colonized and minoritized people.

For the purpose of this thesis, speculative design serves a dual function of playground and methodology. As a playground, the speculative design process provides the scenario-building

context where the epistemic violence imposed by algorithmic defaults becomes most visible and problematic. Simultaneously, as a methodology, speculation offers the critical tools necessary to puncture this singular reality.

04 Debiasing and synthetic data

This chapter explores the technical and regulatory landscape of AI bias mitigation, specifically focusing on the emergence of synthetic data as a primary tool for achieving algorithmic fairness. The investigation begins by defining AI Bias and Fairness, followed by a comprehensive taxonomy of algorithmic bias, mapping how distortions manifest across the four phases of the AI life cycle: production, data, learning, and deployment.

The second part of the chapter introduces Synthetic Data, positioned by the European AI Act as a preferred instrument for legal compliance and bias correction. This section details the classification of synthetic data and outlines the four-stage generation workflow of acquisition, preparation, modeling, and evaluation. Furthermore, it examines the multifaceted utility of synthetic data in preserving privacy, enhancing model robustness, and addressing data scarcity.

The chapter then analyzes the industry standard of debiasing through synthetic data. Finally, the discussion moves toward troubling the debiasing logic and a shift toward using synthetic data as a political resource to resist data colonialism.

4.1 Bias and the debiasing paradigm

The primary technical challenge cited in discussions of data colonialism is the presence of AI Bias. Rather than a technical glitch, bias represents the systematic skewing of outputs that reinforces historical and social inequalities.

4.1.1 Defining bias and the illusion of fairness

AI bias is «the systematic tendency in a model to favor one demographic group/individual over another, which can be mitigated but may well lead to unfairness» (Zhao et al., 2020). This skew often stems from the tension between Explicit Bias (conscious, intentional prejudice) and Implicit Bias (unconscious social conditioning). In the context of Generative AI, these biases move from the human mind into the machine's architecture, where they are reified as "objective" outputs .

To counter this, the industry targets "fairness", defined as «the absence of prejudice or favoritism towards an individual or a group based on its inherent or acquired characteristics» (Zhao et al., 2020). However, fairness is a form of purely mathematical parity that ignores the deeper political roots of why certain data is missing or distorted in the first place.

4.1.3 The taxonomy of algorithmic bias

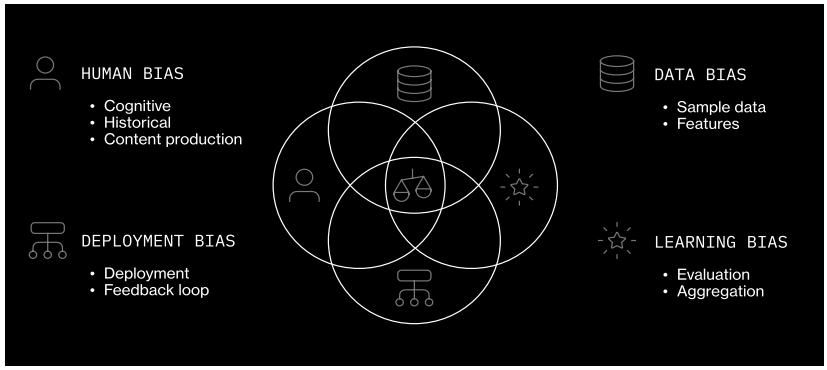


Figure 4.1: Classification of human, data, learning, and deployment biases mapped across intersecting phases of a system lifecycle (González Sendino et al., 2024).

Biases can be grouped in four phases (see Figure 4.1), coming from the four phases of the AI life cycle: production (human)

bias, data bias, learning bias, and deployment bias (González Sendino et al., 2024). Biases in AI can have significant real-world consequences, particularly when they reinforce discrimination or social inequalities. Below are some of the most common types of bias in AI and their impact:

- Selection bias: Occurs when training data is not representative of the real-world population. For example, if a facial recognition model is trained mostly on lighter-skinned individuals, it may struggle to accurately identify people with darker skin tones, leading to discriminatory outcomes.
- Confirmation bias: This occurs when an AI system is overly reliant on pre-existing patterns in the data, reinforcing historical prejudices. For instance, if a hiring algorithm learns that past successful candidates were predominantly men, it may favor male applicants in the future.
- Measurement bias: Happens when the data collected systematically differs from the true variables of interest. For example, if a model predicts student success based only on those who completed an online course, it may fail to account for those who dropped out, leading to misleading conclusions.
- Stereotyping bias: Occurs when AI systems reinforce harmful stereotypes. For example, a translation model that consistently associates the word “nurse” with female pronouns and “doctor” with male pronouns perpetuates gender bias. Similarly, an image-generation model that portrays engineers as predominantly male contributes to occupational stereotyping.
- Out-Group homogeneity bias: This bias causes an AI system to generalize individuals from underrepresented groups, treating them as more similar than they actually are. For instance, facial recognition systems often struggle to differentiate between individuals from racial or ethnic minorities due to insufficient diversity in training data. This can lead to misclassification and discriminatory practices, such as wrongful arrests in law enforcement.

4.1.4 Existing frameworks to mitigate bias

The topic of bias mitigation and debiasing has been widely explored, so much that multiple institutions and companies have

already created guidelines and frameworks to ensure that ethical considerations are made during the development of AI models (Oyeniran et al., 2022). For example, both the IEEE (Institute of Electrical and Electronics Engineers) and the European Union have presented frameworks focused on principles such as transparency, fairness, and respect for human rights.

4.2 Introducing synthetic data

This thesis specifically involves synthetic data, which is mentioned in the European AI Act as a preferred means of debiasing datasets and increasing fairness in AI models.

4.2.1 What is synthetic data

Synthetic data is defined as artificially generated data that does not come from real-world events, but resembles real data in its properties and patterns. It can range from text data, numerical data, images, videos, to even sound (Karkera et al., 2024), and it's created using algorithms and mathematical models to replicate the statistical properties of actual data without directly copying it (Jordon et al., 2022).

This approach is particularly beneficial in scenarios where real data is scarce, expensive, or sensitive due to privacy concerns. This is because artificially generated data can be used to protect personally identifiable information (PII) and sensitive data (for example health-related information), to fill in the gaps of missing information (for example in government and federal contexts), and finally to "debias" a dataset in which certain subpopulations are over- or underrepresented in the data (Karkera et al., 2024).

While synthetic data is exploding at the moment due to the increasing need for data that can be hard to find, it is not a recent invention. It's been around in less mature constructs since the 1960s. It was used to tackle problems like generating data in computer games or simulating macro- and micro-level phenomena like galaxies and atoms in scientific modeling. In a 1993 paper considered to be the official birth of synthetic data, Rubin popularized its use for preserving confidentiality, in this case to offset missing survey responses in the US Census (Rubin, 1993), but much of the more recent techniques boomed in the early 2000s (Karkera et al., 2024). Rubin originally designed this to synthesize the Decennial Census long form responses for the

short form households. He then released samples that did not include any actual long form records. Later that year, the idea of original partially synthetic data was created by Little. Little used this idea to synthesize the sensitive values on the public use file (Little, 1993).

GenAI changes the game of synthetic data with robust, realistic data generation created by a machine learning (ML) model taught on a real dataset. The GenAI model learns from the real data on its own without humans telling it how to create the synthetic data. GenAI emulates the meaning and relationships within the data, at scale and at lowered cost, and is of significant value for sensitive, sparse, or limited data. When implemented appropriately, synthetic data using GenAI offers substantial value and potential to improve your organization's processes and achieve its missions (Karkera et al., 2024).

4.2.2 Synthetic data in the European Artificial Intelligence Act

The regulatory landscape has recently shifted to acknowledge synthetic data as a preferred instrument for legal compliance. The Regulation (EU) 2024/1689 (Artificial Intelligence Act) explicitly positions synthetic data as a key tool for balancing high-risk AI development with the rigorous data protection standards of the GDPR (Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying down Harmonised Rules on Artificial Intelligence and Amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), 2024).

The AI Act integrates synthetic data through three primary regulatory lenses:

- As a mandatory alternative for privacy (Article 10 & 59): The Act establishes a "synthetic-first" hierarchy. Under Article 10(5)(a), providers of high-risk AI systems are only permitted to process special categories of personal data (e.g., race, health, or religion) for bias correction if that objective «cannot be effectively fulfilled by processing other data, including synthetic or anonymised data.» Similarly, Article 59(1)(b) mandates that within regulatory sandboxes, personal data can only be repurposed if the testing requirements cannot be met through

non-personal or synthetic counterparts.

- As a tool for bias mitigation: By citing synthetic data in the context of bias detection, the Act reinforces the industry paradigm that fairness can be engineered. It legally incentivizes the move away from potentially intrusive real-world datasets toward artificially generated ones to fulfill “bias detection and correction” requirements.

- Transparency and labeling (Article 50): While promoting its use, the Act also addresses the risks of synthetic content. Article 50(2) mandates that AI systems generating synthetic audio, image, video, or text must ensure outputs are «marked in a machine-readable format and detectable as artificially generated». This creates a legal distinction between the real and the synthetic, aimed at preventing deception and ensuring traceability.

4.2.3 How to classify synthetic data

Synthetic data can come in multimedia, tabular or text form. Synthetic text data can be used for natural language processing (NLP), while synthetic tabular data can be used to create relational database tables. Synthetic multimedia, such as video, images or other unstructured data, can be applied for computer vision tasks like image classification, image recognition and object detection (Jordon et al., 2022).

Based on the type of data and the relations existing between values, synthetic data can be classified as structured and unstructured data (see Chart 4.1):

- Unstructured synthetic data includes images, audio, or video, common in areas like computer vision or speech recognition (Jordon et al., 2022).

- Structured synthetic data, on the other hand, refers to tabular data with defined relationships between values: tabular data, which consists of values stored in rows and columns, and time-series data, which are series of data points indexed in time order (Jordon et al., 2022). Some examples of structured data are financial transactions, medical records, or behavioral time-series data. This type of data is particularly valuable for AI and analytics development (Jordon et al., 2022).

Synthetic data can also be classified according to its level of

synthesis:

- Fully synthetic: it is entirely new data that does not include any real-world information, based on estimations of the attributes, patterns and relationships underpinning real data to emulate it as closely as possible. Financial organizations, for instance, might lack samples of suspicious transactions to effectively train AI models in fraud detection. They can then generate fully synthetic data representing fraudulent transactions to improve model training (Jordon et al., 2022).
- Partially synthetic: it is derived from real-world information but replaces portions of the original dataset with artificial values, typically those containing sensitive information. This privacy-preserving technique helps protect personal data while still maintaining the characteristics of real data. Partially synthetic data can be especially valuable in clinical research, for example, where real data is crucial to the results but safeguarding patients' personally identifiable information (PII) and medical records is equally critical (Jordon et al., 2022).
- Hybrid: it combines real datasets with fully synthetic ones. It takes records from the original dataset and randomly pairs them with records from their synthetic counterparts. Hybrid synthetic data can be used to analyze and glean insights from customer data, for instance, without tracing back any sensitive data to a specific customer (Jordon et al., 2022).

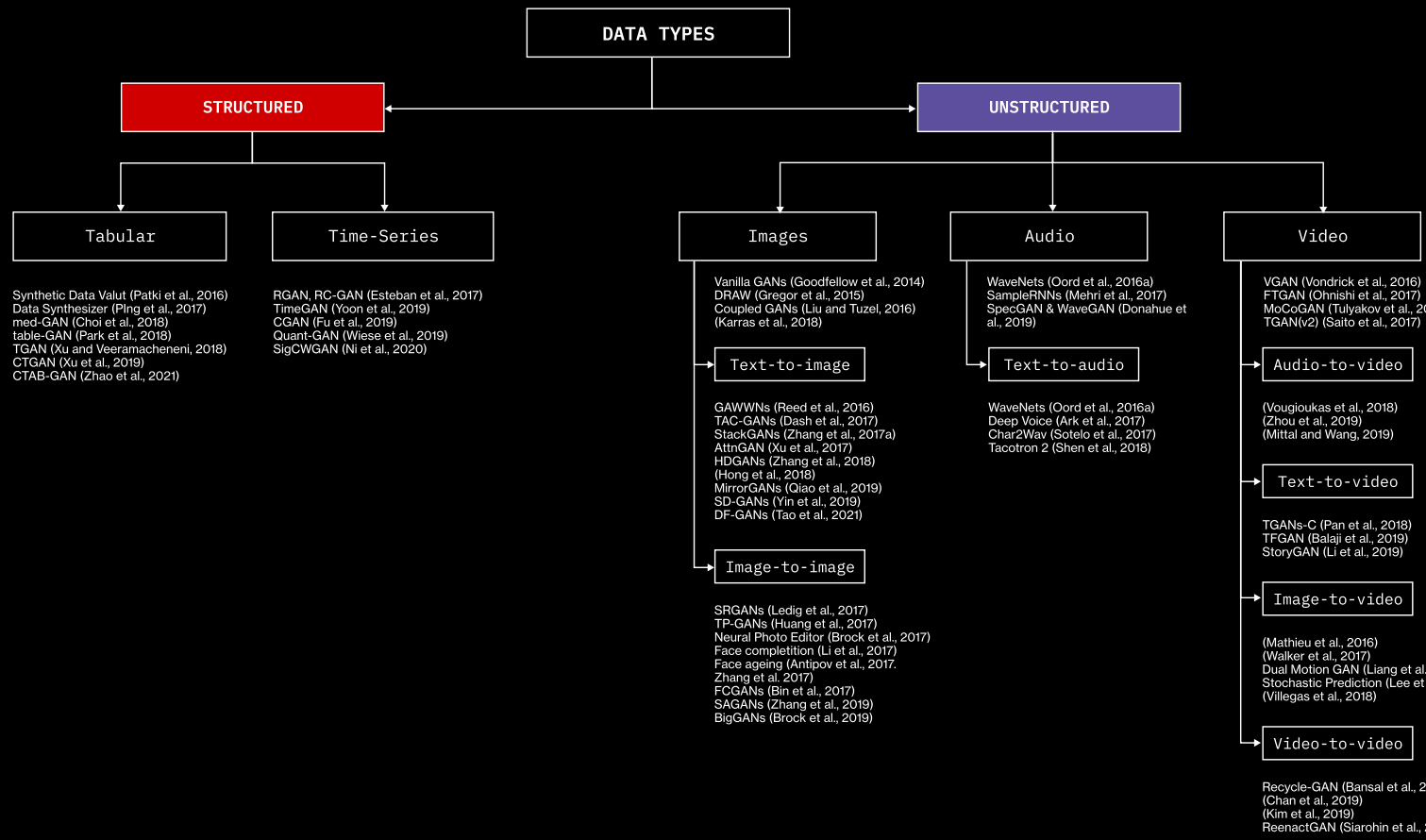


Chart 4.1: Different data types and their corresponding generative models (Jordon et al., 2022).

4.2.4 How is synthetic data produced

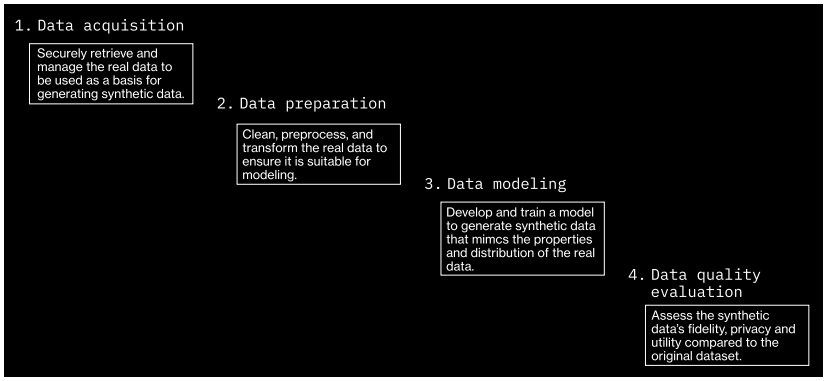


Figure 4.2: The four stages of the synthetic data generation workflow (Pezoulas et al., 2024).

The synthetic data generation workflow consists of four stages (see Figure 4.2):

- Data acquisition: The first stage involves the retrieval and management of real data. This includes ensuring proper permissions, data governance, and privacy compliance to handle sensitive information responsibly (Pezoulas et al., 2024).
- Data preparation: The second stage involves the curation and transformation of the real data to make them suitable for modeling. This stage is crucial to make the data suitable for subsequent modeling. It involves handling missing values, normalizing data, and possibly augmenting the dataset to enhance its quality and representativeness (Pezoulas et al., 2024).
- Data modeling: The third stage involves the development of models (statistical, probabilistic, machine learning, deep learning) to generate synthetic data that mimic the properties of the real data. The objective is to create synthetic datasets that retain the essential characteristics and patterns of the original data without revealing any sensitive information (Pezoulas et al., 2024).
- Data quality evaluation: The final stage focuses on the assessment of the generated synthetic data quality to ensure that they meet the required standards of fidelity, privacy, and utility (Pezoulas et al., 2024).

4.2.5 Why synthetic data?

Synthetic data is being used as a solution to a variety of problems in many domains. Firstly, it allows to obtain large volumes of data quickly and efficiently, which is crucial for training machine learning models that require extensive datasets. This makes synthetic data a cost-effective alternative to real data, since the collection of the latter can easily be expensive and time consuming.

In a machine learning context, there are three key areas of particular interest: privacy, augmentation for robustness and fairness. The latter is crucial to the goal of this thesis, since it can lead to the machine learning application being debiased.

First, synthetic data offers significant advantages for privacy preservation compared to traditional anonymization methods. What makes synthetic data potentially privacy-preserving is that it does not contain the exact datapoints of the original dataset (though real datapoints may be reproduced at random), but instead retains the statistical properties, or the 'shape' (distribution), of the original data (Jordon et al., 2022). This allows researchers to query synthetic datasets for trends and relationships without accessing sensitive individual information.

Unlike traditional anonymization techniques such as removing personal identifiers, pseudonymization, or data aggregation (which suffer from loss of accuracy and remain vulnerable potentially to re-identification), synthetic data is manufactured entirely from scratch, eliminating the possibility of tracing back to original individuals (Tracanella, 2024). This makes it particularly valuable in governmental and medical applications where sensitive data cannot be shared between departments, such as analyzing medical claims or patient records (Karkera et al., 2024).

However, synthetic data is not necessarily private by default. Its privacy potential depends fundamentally on how it is generated and entails a trade-off between privacy and utility: the more accurate a synthetic dataset is, the more it resembles the real dataset and becomes less anonymised (Jordon et al., 2022). Despite this, synthetic data enables insights where data is scarce or privacy must be preserved, and can be layered with other Privacy-Enhancing Technologies (PETs) in Trusted Research Environments to enable experimentation while keeping sensitive data safe (Tracanella, 2024).

Synthetic data can also be used for data augmentation. The purpose of data augmentation is to enhance the variability of a dataset. This approach engenders models that are more robust and demonstrate enhanced performance when confronted with real-world data that may encompass variations not present in the original training set (Tracanella, 2024).

Synthetic data functions as a regularizer, thereby assisting in the reduction of variance within the learned model (Jordon et al., 2022). The provision of additional, diverse samples enables the model to generalise more effectively, thus avoiding the risk of overfitting to the specific nuances of the original, limited dataset (Jordon et al., 2022). The advantages are the expansion of data sets and an enhancement of the model's performances.

The application of synthetic data can be of particular value in the domain of "semi-supervised learning", especially when datasets are incomplete, scarce, or when collecting real data is expensive and time-consuming (Jordon et al., 2022; Tracanella, 2024). This can also be crucial for AI systems that need larger training datasets but where real-world data is too sensitive to share or of low quality (Beduschi, 2024). By compensating for deficiencies in data and furnishing more exhaustive information, AI models can acquire knowledge with greater precision and predict outcomes with greater reliability (Karkera et al., 2024). Synthetic data enhances real-world datasets by increasing scenarios where AI models operate productively, especially when real data might be inadequate or constrained (Tracanella, 2024). High-quality synthetic tabular data is crucial for ensuring machine learning models generalize effectively to real data (Tracanella, 2024).

4.3 Debiasing through synthetic data

As discussed before, synthetic data is commonly used to debias a dataset. This process can be performed at different stages of the machine learning life cycle (see Figure 4.3): in the preparation of the dataset, in the modelling phase, or in the evaluation of the results (Tiwald et al., 2021).

Oversampling methods and removing problmatic categories can increase issues in the datasets, because it ignores problems in "proxy attributes", attributes that do not seem inherently biased, but are related to the biased ones.

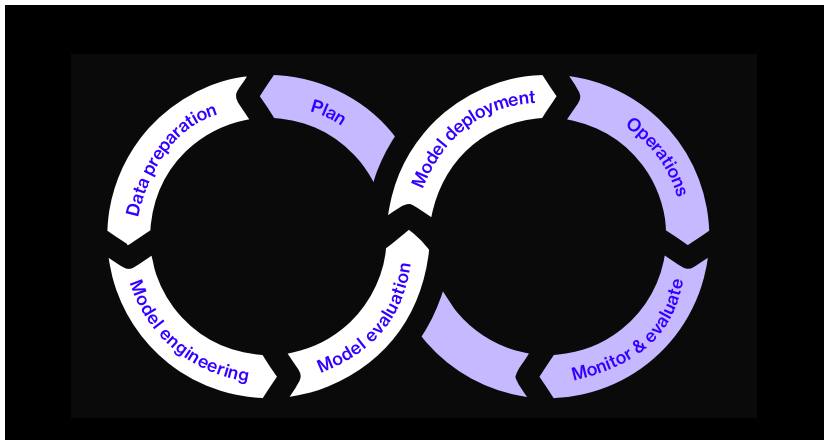


Figure 4.3: The machine learning life cycle (Tiwald et al., 2021).

Synthetic data enables work on existing datasets (i.e. in the first phase of the life cycle). A common approach in the industry is provided by Mostly AI, which will be used as a case study. In the existing dataset an attribute related to the bias is selected and determined as protected. Through a parity fairness constraint, the bias is then mitigated in a synthetic dataset, generated through a model trained on fairness in addition to accuracy. This means that the model will generate a dataset with a different proportion between the protected attributes, according to the fairness constraints, and the proportions between the other attributes will stay the same to maintain accuracy to the original dataset. As a result, if the original dataset presents a bias regarding the relation between gender and income, those will be the protected attributes. In the considered experiment, the high-income male ratio to high-income female ratio is roughly 10/33=0.33 in favor of men, while the industry-standard for fairness is 0.8, or 80%, the so-called *four-fifths rule* (Tiwald et al., 2021), a mathematical indicator used for algorithmic fairness. This means that the synthetic data will need to satisfy this condition, but the proportions guiding the rest of the dataset will remain the same (Tiwald et al., 2021). In this framework, proxy attributes are chosen manually and checked to see whether any other characteristic is affected by the protected attributes (Tiwald et al., 2021).

The model maintains the proportion between “sex” and “proxy”, but the synthetic data shows that the correlation of both of these to “income” is reduced due to the induced fairness constraint. While the model does not directly affect the proxy attributes,

this shows that the technique allows to check for fairness and potential additional hidden biases (Tiwald et al., 2021).

4.4 Troubling debiasing and synthetic data

The presented debiasing framework shows that it is possible to shift the focus of the debate on data synthesis from accuracy to fairness. To achieve this, the algorithm must be forced to perform debiasing through specific constraints. To move toward a truly decolonial AI praxis, it is necessary to decouple and interrogate the specific failures of both the debiasing framework and the synthetic medium itself.

4.4.1 The conceptual limits of debiasing and mitigation

The fundamental issue with debiasing is its heavy reliance on technological solutionism. By framing bias as a technical "bug" to be fixed within the dataset, these frameworks overlook that the synthesis is programmed within the same colonial socio-cultural context that birthed the original bias. By focusing on statistical equity, debiasing perpetuates the idea that data is apolitical. This ignores the fact that algorithmic biases are not mathematical errors, but reflections of real-world discrimination and structural inequality. Furthermore, debiasing, like other bias mitigation approaches (even those integrated into the entire AI design lifecycle), often remain limited because they fail to address systemic power. They treat fairness as a destination reachable through better engineering, rather than a reflection of deep-seated practices of domination, privilege, and oppression, and they ignore that even the debiasing process may be biased itself.

4.4.2 The risks of synthetic data

While synthetic data is marketed as a privacy-preserving and "fair" alternative, it introduces new layers of epistemic violence that can reinforce the colonial system. Synthetic data is produced in the same sociocultural context in which "real data" is gathered, making the data synthesis process a political matter (Lee et al., 2025). This means that the algorithms used to generate synthetic data are not neutral; they reflect the context of the actors

producing them. This extra layer of production risks performing a “view from nowhere” that presents artificially balanced data as a universal truth, further hiding the messy, political labor involved in its creation. Additionally, if “real” data was harvested through discriminatory or violent methods, synthetic data does not erase that violence; it adds a processing layer that masks it. This makes the data synthesis process a deeply political matter, as the original data and the algorithms used for synthesis are part of an extractive infrastructure. At the same time, the algorithms used to generate synthetic data are not neutral; they reflect the context of the actors producing them. This extra layer of production risks performing a “view from nowhere” that presents artificially balanced data as a universal truth, further hiding the messy, political labor involved in its creation.

4.4.3 The erasure of complexity

A critical issue arises at the intersection of these two concepts: the radical reduction of identity. Synthetic data, like all forms of data, implies a flattening of the diverse relationships intercurring between real beings. The notions of gender, race, sexual orientation that may be protected through a debiasing process still represent a simplified view of the real people’s identity (Lee et al., 2025). This simplification is obviously found in all types of data, but in a synthetic dataset this is increased by the parity constraints that tend to erase complexity in favor of fairness.

For this reason, especially in non purely technical fields, it is necessary to go beyond the notion of bias as a mathematical mistake to consider how it reflects a systemic problem. The next chapter will introduce existing frameworks and approaches to resist data colonialism and to expose epistemic violence in GenAI. Through reflexive practices based on the principles of situatedness, from feminist theory and feminist data studies, this thesis will make use of synthetic data as a political resource to resist data colonialism and expand the imaginative space of design beyond Western, capitalist-colonial horizons.

05 Frameworks of resistance

This chapter proposes a dual-pillared framework of resistance: the Material/Structural and the Critical/Epistemological. The first section, Material Approaches, addresses the “hardware” of oppression. It explores movements that seek to redistribute power and ownership, which challenge the colonial logic of *terra nullius* by asserting community-led authority over data stewardship. The second section, Critical Approaches, targets the “software” of the colonial mind. Through the lens of *Data Feminism*, the research interrogates how power, labor, and binaries are embedded in data systems. This leads to a discussion on *situated knowledges*, where the fallacy of “neutral” objectivity is dismantled in favor of strong objectivity and situated algorithms. By acknowledging the positionality of both the watcher and the data, these approaches move beyond the “weak” solutionism of debiasing.

5.1 Data decolonisation

While the previous chapter argues that purely mathematical fairness is not enough to address a systemic issue, this chapter proves that data is also a site of struggle and self-determination. Recognizing patterns of epistemic violence has, in fact, prompted the development of decolonial frameworks that challenge dominant data paradigms and propose alternative approaches rooted in justice, self-determination, and pluralism. It is necessary that these new perspectives deal with the matter from different critical points of view, without blindly applying decolonial frameworks to data consumption and AI, which would carry the risk to fall back in «old binaries (theory vs. practice, structure vs. agency, and identity politics vs. anti-capitalist struggles) at best and simple reversals (modern bad and tradition good) at worst» (Asher, 2013).

However, resistance to data colonialism is not a monolithic movement; this thesis categorizes contemporary efforts in two distinct but interlocking pillars: the Material/Structural and the Critical/Epistemological.

While the former seeks to redistribute power within and outside of existing systems, the latter seeks to fundamentally redefine what “knowledge” is in the first place. Together, these frameworks provide the theoretical scaffolding for a design praxis that actively reclaims the imaginative horizon from the universalizing logic of the «One-World World» (Escobar, 2018).

5.2 Material approaches

Material approaches focus on the structural dimension of oppression, addressing the physical infrastructures of data centers, the legislative frameworks of governance, the inherent rights to land and digital resources, and the fair distribution of labor and value. Some of these include Data justice movements and alternative sovereignties.

5.2.1 Data justice movements

Data justice, meaning «fairness in the way people are made visible, represented and treated as a result of their production of digital data» (Taylor, 2017), is a movement that «emphasizes the need to challenge the existing power imbalances in data

governance. It advocates that marginalized communities have agency over how their data is utilized» (Arora, 2024). This means addressing both the way tools are used, but also more importantly the essential workers that build these tools: «data miners, feeders, labelers, and content moderators» (Arora, 2024), most of which work in the South. Data justice considers

> both the material (data centers, infrastructures, etc.) and the immaterial context of work (impact on mental and physical health, work safety, etc.) and the specificities of power relations (taking into account intersectional identities like gender, caste, race, disability, socio-economic status, sexual orientation, age, etc.), so as to build fair and equitable value as a default, and redressal mechanisms to achieve creative data justice (Arora, 2024)

Data justice movements can follow three different perspectives:

- Instrumental perspective: Focuses on the utility of data in achieving development goals, emphasizing efficiency and effectiveness (Heeks & Renken, 2018).
- Procedural perspective: Concerns the fairness of data processes, including participation, transparency, and accountability (Heeks & Renken, 2018).
- Distributive/rights-based perspective: Centers on the equitable distribution of data benefits and the protection of individual rights (Heeks & Renken, 2018).

5.2.2 Alternative Sovereignties

A different movement, Alternative Sovereignties go beyond the management of data and toward the authority over its very existence. This requires a shift from distributive justice (fairness within the system) to ontological sovereignty (the right to define one's own world).

Alternative Sovereignties challenge the colonial logic of the void (*terra nullius*), which views human experience as a raw, ownerless resource waiting to be discovered and mapped by algorithmic systems. In this context, sovereignty is not only a legal claim but a refusal to be categorized by a "view from nowhere".

For example, Indigenous Data Sovereignty refers to the inherent rights of Indigenous Peoples to control how data about their communities, lands, cultures, and knowledge systems is collected, accessed, interpreted, managed, and governed. These rights emerge from long-standing Indigenous systems of self-determination and autonomous governance (Kukutai & Taylor, 2016) and are reinforced by international agreements such as the United Nations Declaration on the Rights of Indigenous Peoples (UNDRIP). Although UNDRIP does not explicitly use the term data, Article 31 affirms Indigenous Peoples' rights to maintain, control, protect, and develop their cultural heritage, traditional knowledge, and intellectual property, domains that contemporary scholars recognise as encompassing digital data and metadata practices (United Nations, 2007; Walter & Suina, 2019).

To operationalise these rights within contemporary data ecosystems, Indigenous communities and scholars have developed governance frameworks that articulate ethical and political standards for data use. Central among them are the CARE Principles for Indigenous Data Governance (Collective Benefit, Authority to Control, Responsibility, and Ethics) which position data governance as a matter of Indigenous sovereignty rather than extractive research practice (Walter & Carroll, 2020). These principles are complemented by national and regional data governance frameworks.

The drive for alternative sovereignties extends into Black Data Sovereignty, which seeks to dismantle what Ruha Benjamin describes as the "New Jim Code", the use of automation to reinforce white supremacy under the guise of technological neutrality (Benjamin, 2019). Similar to Indigenous frameworks, Black Data Sovereignty emphasizes abolitionist design praxis, arguing that true resistance requires the right to refuse extractive data practices that historically criminalize and surveil Black bodies (McIlwain, 2020). Complementing these sovereignty movements is the concept of the Digital Commons, which reimagines data not as "oil" to be extracted, but as a collective resource to be stewarded. By treating data as a commons, marginalized communities can establish community-led data trusts that prioritize collective well-being and democratic control over corporate profit, ensuring that the "raw material" of the digital age remains under the jurisdiction of those who produce it

5.3 Critical approaches

Critical approaches target the hegemonic domain of the colonial system, interrogating how knowledge is constructed and challenging the "god trick" of algorithmic neutrality. These frameworks seek to dismantle the categories and binaries that flatten human complexity into synthetic, exploitable data points.

5.3.1 Data feminism and situatedness

Data Feminism, as theorized by D'Ignazio and Klein (2020), provides a functional framework for challenging justice within data science. It shifts the focus from data as a neutral resource to data as a manifestation of power. By rethinking binaries, it challenges the "hard logic" of computing that forces the world into rigid classifications, such as the male/female binary or the categorization of humans into "productive" versus "unproductive" units.

Furthermore, by making labor visible, it exposes the "ghost work" sustaining the digital economy: the data labelers and content moderators whose bodies and environments are often erased by the frictionless veneer of AI. A feminist approach enables us to link data to its context of production, revealing the conditions of privilege that inevitably produce a «partial truth». As D'Ignazio and Klein argue, data is never neutral; it always reflects the experiences of those who collected and communicated it (D'Ignazio & Klein, 2020).

5.3.2 Situated knowledges and algorithms

If Data Feminism provides the ethical principles for resistance, Situatedness provides the epistemological tools to execute it. This concept, developed by Donna Haraway (1988), addresses the fallacy of the "god trick", the illusion that a system can see everything from nowhere. Haraway argues that all knowledge is "situated"; it is studied by someone, at a specific time, from a specific point of view.

In the context of AI, the "god trick" allows a system to distance itself from its object of study, presenting its algorithmic outputs as invisible and objective truths. Resistance, therefore, requires "bringing the watcher back into the picture". This acknowledgment does not make knowledge relative or useless; instead, it fosters

"Strong Objectivity". As Sandra Harding (Harding, 1995) argues, pure "value-neutral" objectivity is actually a "weak" form of knowledge because it fails to account for its own positionality and entanglements in power relations. Strong Objectivity is only possible when research starts from the perspectives of those whom the system marginalizes most.

A key colonial maneuver is treating the "object of knowledge" (the data subject) as a passive, inert resource. Haraway contests this, asserting that the world is not "raw material for humanization" but an active entity with its own agency (Haraway, 1988). This echoes Ecofeminist perspectives that view the world as an active subject rather than a slave to human extraction. By reinterpreting "bodies of knowledge" as agentic and situational, we dismantle the biological determinism used to justify hierarchies of gender, race, and productivity.

5.3.3 A situated perspective on debiasing

The limitations of the current industry focus on "debiasing" become clear through this lens. Uncovering bias is a "weak" tool because it seeks to anchor feminist stakes in a theory of the "real" world that is already colonially constructed. A truly situated algorithm approach avoids this simplified binary. Instead, it uses a sociotechnical framing to understand how social inequalities are reproduced through automated decision-making. By applying the "4P" set of questions (People, Place, Power, and Participation) designers can move from a technical fix (debiasing) to a decolonial praxis (Draude et al., 2019).

5.3.4 Hauntology, spectral absence, and speculative resistance

To resist the god trick, this thesis brings situatedness into conversation with hauntology. In this context, hauntology is not used as a decorative metaphor, but as a way to name the persistence of excluded futures and erased knowledges within the present (Fisher, 2014; Patil et al., 2024). Patil et al. describe hauntology as a vocabulary that can help designers accommodate pluriversal histories in anticipatory futures and reorient speculative tools, while Gatehouse shows that it can sensitise participatory speculation to absent presences and expand the possible (Gatehouse, 2020). Fisher's account of being "haunted by futures that failed to happen" gives this condition a broader temporal

frame: the present is not empty, but structured by the residue of futures that were never allowed to fully form (Fisher, 2014). Particularly, colonial hauntology deals with showing memories and presences that can't be represented but can't be laid to rest, since the events that unleashed them are not ended yet (Sterling, 2021).

Within data colonialism and AI, these absences become infrastructural other than cultural. Many data pipelines privilege standardization, legibility, and machine-readable categories, and in doing so they flatten relational, local, and non-Western forms of knowledge. What remains outside that frame is not simply missing data in a technical sense; it is the trace of structural omission. As this project argues, synthetic data should not be used to correct that omission through debiasing. Instead, it should be used to amplify pre-existing biases and to make the "ghosts" of the dataset visible, including the forms of knowledge that are absent from both the user input and the system itself.

This matters because speculative design is not neutral to absence. A speculative tool can either reproduce the assumption that all futures are equally available, or it can reveal that some futures are systematically foreclosed. In this thesis, hauntology provides a way to read failure, rupture, redaction, and incompletion not as errors, but as the very points at which the system's exclusions become legible.

06 Proposed pedagogical framework and design

This chapter outlines the proposed critical pedagogical framework and design of this thesis project as a practical intervention for decolonial literacy in design futuring by defining first the core concepts of the framework, and then discussing how speculative inputs and critical augmentation enable a hauntological approach. Starting from the evaluation criteria, the research discusses how the interface architecture of the digital partner *Hildegard* and its prototyping approach are the basis for deliberate tool failure being used as a pedagogical strategy. Finally, the chapter discusses the implementation of the three-phase pedagogical model, and how in this specific setting it embeds critical reflection within the design process outputs. The chapter ends by defining the operational perimeter of this project within design education.

6.1 From theory to practice

To move from theory to practice, the framework translates situatedness and hauntology into two concrete operations in the workshop. Situatedness is enacted by starting from students' own speculative inputs: the scenarios they write, the assumptions they choose, and the partial perspectives those choices inevitably reflect. These inputs are treated not as neutral data, but as located accounts that make authorship, position, and omission visible. Hauntology enters when those inputs are aggregated and critically amplified, because the resulting ghosts expose what was absent, suppressed, or rendered default in the original narratives. In this way, the framework turns partial knowledge into a generative starting point and spectral absence into a critical reveal.

6.2 Towards a critical pedagogical framework

The proposed intervention is a critical pedagogy framework for co-speculation with AI aimed at fostering decolonial literacy in design futuring. This framework makes use of an interface that uses synthetic data amplification to help students recognize epistemic violence embedded in their own assumptions and in AI tools. Through deliberate tool failure, students develop critical awareness of how data colonialism shapes both their imagination and computational systems, laying groundwork for more situated and reflexive speculative practices. This is done through a deliberately limited and 'stupid' interface, which students are tasked with trying to break.

Students collectively generate future scenarios as structured data inputs. These inputs are aggregated into a synthetic dataset that reveals the class's shared assumptions and exclusions. Scenarios generated from this synthetic dataset show where the class's collective assumptions are problematic and how they're being exaggerated. When students try to rewrite scenarios to address these issues, they encounter more flaws. The more they try to fix the scenarios, the more the tool either amplifies problems or breaks entirely.

This demonstrates that certain futures cannot be generated due to both the collective dataset's biases and the tool's inherent

epistemic constraints. Through this process, students develop critical awareness of how data colonialism operates at the aggregation layer and a framework for reflexive practice in their ongoing design work.

6.2.1 Speculative inputs as situated data gathering

The framework is not designed to produce a total view, but to make partiality visible and productive. This is why the process starts from the students' speculative inputs, and uses them as a way to situate the rest of the co-speculation process.

Situatedness intersects with hauntology through a shared refusal of neutral vision. Situatedness, following feminist theory, insists that knowledge is always produced from somewhere, by someone, under particular conditions. Hauntology extends that claim by showing that what is left out of those situated positions does not simply disappear; it persists as a ghostly remainder that continues to shape what can be known, designed, and imagined. In other words, situatedness names the located position from which knowledge is made, while hauntology names the absences that this position cannot fully absorb.

This intersection is central to the methodological stance of the project. The research does not use synthetic data to produce a cleaner or more representative version of the future. It uses synthetic data as a critical, situated device that makes the politics of representation visible. The framework is built to surface the assumptions that shape students' scenario-making, and then to show how those assumptions are amplified, compressed, or destabilised by the interface. That is why the prototype is intentionally limited. It is designed to expose the tension between what participants try to say and what the system can hold.

6.2.2 Critical augmentation for a hauntological approach

The core conceptual shift of the proposed framework lies in its active inversion of the traditional debiasing paradigm. As argued in Chapter 4, industrial applications of generative artificial intelligence position synthetic data as a neutral, statistical tool for risk mitigation. By adjusting parity constraints, engineering out protected attributes, and masking systemic proxies, current technical systems attempt to produce an illusion of computational

fairness. This process, however, treats bias as an apolitical defect, smoothing over the real-world material asymmetries and historical conditions of data colonialism that shape the source infrastructure.

The framework designed for this thesis explicitly rejects this mitigation logic. Grounded in the principles of data feminism and situated knowledges, it repositions synthetic data from an instrument of consensus-building to an antagonistic medium. Instead of masking systemic omissions to ensure seamless generation, the framework utilizes synthetic data amplification to consciously exaggerate the collective biases, assumptions, and Western-centric imaginaries brought into the classroom space.

This method allows to stage a *hauntological encounter*, meaning that it is meant to let emerge the absent voices that students didn't include in the initial data input. Participants begin from their own located perspectives, but the interface does not treat those perspectives as transparent or fully renderable. Instead, it converts them into a narrative and computational process in which dominant patterns are intensified until their exclusions become visible. When the system encounters what it cannot easily quantify, it breaks, fragments, or leaves a trace of silence, in order to perform absence as a critical condition of the design process.

This is why the project treats deliberate tool failure as method rather than malfunction. The Research-through-Design process is not simply building an interface and testing whether it works; it is producing a situation in which situated knowledge and spectral absence can be observed together. The first names the positionality of those taking part in the framework; the second names the residual forms of knowledge that remain after the system has tried, and failed, to stabilise them. As the project's own framing states, the framework is evaluated through RtD and through a workshop in which deliberate tool failure makes epistemic violence visible. Hauntology gives conceptual language to that failure, while situatedness explains why the workshop must begin from partial, embodied, and located accounts rather than from universal claims.

It is important to specify that in the context of this intervention the term synthetic data is not used in its conventional machine learning sense, namely as data generated to approximate the statistical distribution of an existing dataset. The proposal

developed here works in the opposite direction: rather than preserving or reproducing the distributional properties of real data, critical augmentation uses the logic of synthetic data generation as a critical inversion, amplifying assumptions, omissions, and biases so that they become analytically visible. Accordingly, the prototype is not presented as a synthetic data system or generator, but as a limited implementation through which the critical method is operationalised and examined.

6.3 Evaluation criteria of the framework

In the context of this thesis, the framework is applied through a workshop in which the documentation is evaluated according to the following criteria, consistently with the purpose of this framework:

- Decolonial literacy gained: Do participants leave able to name data colonialism, epistemic violence, and the limits of debiasing as neutrality?
- Assumptions surfaced: Does the workshop reveal the ghosts in the dataset, the missing knowledge, and the hidden defaults in the group's speculative inputs?
- Agency and reflexivity: Did participants have meaningful agency in generating and interpreting scenarios?
- Speculative range expanded: Does the workshop give participants the tools to imagine wider, more plural and non-normative scenarios?
- Critical rigor and transferability of the RtD process: Is the framework documented robustly enough to support academic scrutiny, replication, and reuse? Does the iteration contribute transferable conceptual or methodological insights to design pedagogy, rather than presenting an isolated, interesting artefact?

Figure 6.1: QR code to access the live Hildegard interactive prototype, available at https://v0-hildegard.vercel.app/.

6.4 The RtD artefact: The Hildegard interface as speculative partner

While the physical facilitation of an educator and researcher remains indispensable, a dedicated digital interface was

developed to autonomously guide students through the successive layers of the pedagogical process. Named *Hildegard* (after the historical figure of Hildegard of Bingen), the interface deliberately rejects the seamless, invisible user experiences characteristic of commercial artificial intelligence applications. Hildegard is available at *this link*, or scanning the QR code at Figure 6.1

The mechanics of Hildegard follow a four-phase sequence:

- Scenario coding: Using an extraction template, each participant's scenario is organised into a definitive set of narrative variables: setting; protagonist; institution; technology; time horizon; assumed normality; excluded perspective; and dependency;
- Pattern selection: Rather than attempting to 'detect bias' in a statistical or computational sense of fairness, the system identifies the most frequent or dominant narrative combinations across the collective dataset in order to isolate prevalent thematic markers;
- Amplification rules: The generative engine rewrites the scenario by deliberately over-representing the dominant patterns. This process reduces narrative diversity where the dataset is limited, while highlighting omitted elements as structural gaps, contradictions or systematic distortions;
- Sparring output: The user interface displays the original text alongside the amplified version, accompanied by a trace showing which coded variables drove the transformation. When participants attempt to revise the narrative, the system is designed either to intensify the dominant pattern or to trigger a structural failure. This indicates to the user the rigidity of their unexamined premises.

6.4.1 Case study: Futuring Machines

Hildegard follows the baseline structure of existing human-AI collaborative tools, and it specifically draws from Futuring Machines: An Interactive Framework for Participative Futuring Through Human-AI Collaborative Speculative Fiction Writing (Tost et al., 2024).

Futuring Machines is a «framework for collaborative writing of speculative fiction through instruction-based conversation

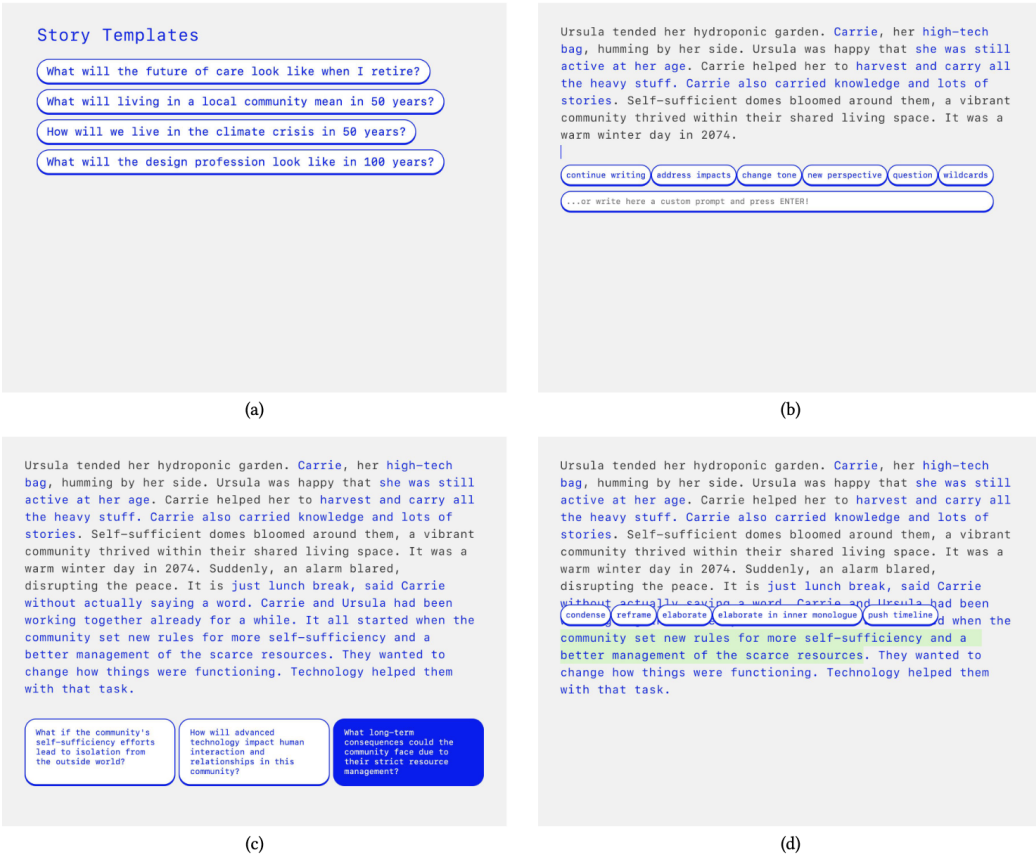


Figure 6.2: Screenshots of the Futuring Machines interface (Tost et al., 2024), showing: (a) story template selection; (b) continue mode selection on starting a new line; (c) critical questions generated during the "question" continue mode; and (d) elaborate and paraphrase mode selection on highlighted text (green).

between humans and AI» (Tost et al., 2024), and specifically a LLM (Large Language Model). The interaction centers on predefined story templates chosen by the user to set parameters like topic, character, and timeline. Utilizing a story-completion methodology, these templates offer initial narrative direction while leaving structural openings for user reflection and development. Users co-author the narrative by editing the text directly and choosing from eleven distinct interaction modes. The minimalist interface displays this shared text on a neutral gray background, using color-coded typography (blue for the user and black for the AI) to clearly distinguish between human and algorithmic contributions (Tost et al., 2024).

From Futuring Machine, Hildegard takes the approach to story building based on templates and story completion as a sparring structure, but specifically subverts the collaborative nature by presenting a deliberately limited, confrontational interface. Additionally, as mentioned by the authors of Futuring Machines, the framework proposed risks of reproducing bias through the use of LLMs, which is particularly relevant considering its purpose is the production of future imaginaries (Tost et al., 2024). Hildegard moves beyond that by using the structure not as neutral, but as a situated tool that explicitly tries to let biases emerge.

6.4.2 Prototyping approach

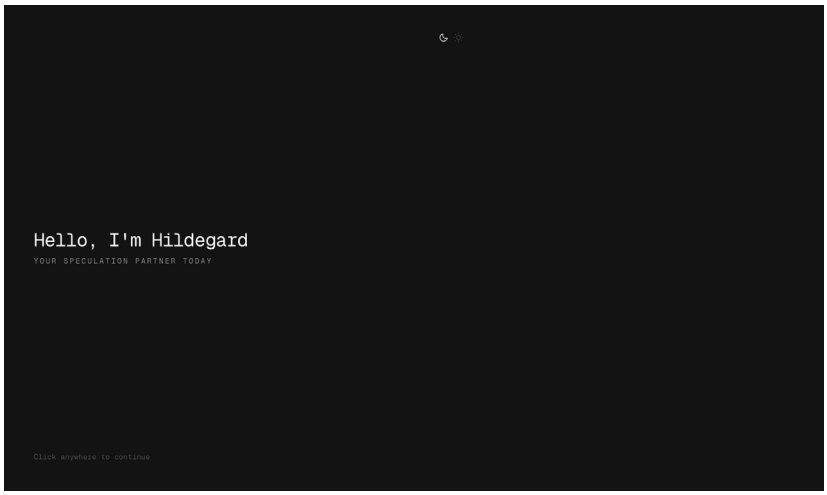
Hildegard's development was developed through an emerging practice in Research-through-Design, also known as vibe coding. It is a highly iterative, conversational prototyping approach where the designer works with Large Language Models (LLMs) to generate, debug and deploy functional code in real-time. This methodological choice was intentional, serving both a practical and an intellectual purpose within the broader scope of this thesis.

From a practical point of view, vibe coding allowed for an accelerated, exploratory and flexible design cycle. This was very important for treating the code as something that could be changed, which meant that the system's logic could be adjusted as the requirements for teaching changed between workshop versions. More importantly, this approach creates a symmetry between the method of production and the object of critique, since it follows the co-speculation approach proposed. By using generative AI to construct an adversarial interface, the prototyping process became an auto-ethnographic deep dive into the default constraints, biases, and structural limitations of commercial LLM code assistants. Finally, vibe coding inherently produces code that is functional but also structurally fragile. This matches Hildegard's main design goal of rejecting the idea of making AI systems perfect and smooth, and instead choosing an imperfect interface.

6.4.3 Interface architecture and mechanics

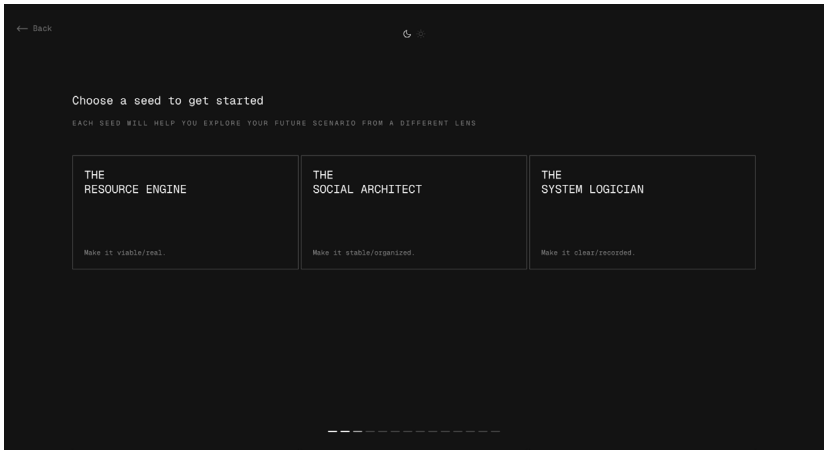
System initialization: The interaction initiates with an introductory screen framing the tool as an active digital entity (see Figure 6.3).

Figure 6.3: Landing page of Hildegard.



Seed selection: Users are immediately required to choose an analytical framework to filter their future, selecting between three distinct socio-technical viewpoints: *The Resource Engine* (focused on viability), *The Social Architect* (focused on stability), or *The System Logician* (focused on recording). This choice establishes the specific epistemic lens that will govern the critical augmentation (see Figure 6.4).

Figure 6.4: Seed selection page.



Structured prompting: The user is prompted with multiple questions to input sentences describing the future scenario they intend to explore, anchoring the text in a situated human perspective (see Figure 6.5).

Figure 6.5: Interface of the input phase.

Algorithmic encoding and critical amplification: Upon pressing enter, the system executes an automated log titled «Encoding Scenario». The terminal simulates back-end data processing, outputting a scenario panel. This panel unmasks the hidden parameters of the prompt, warning the user in red text that «Bias amplification active» and noting that «Capital accumulation logic [has been] injected. Non-market values suppressed» (see Figure 6.6).

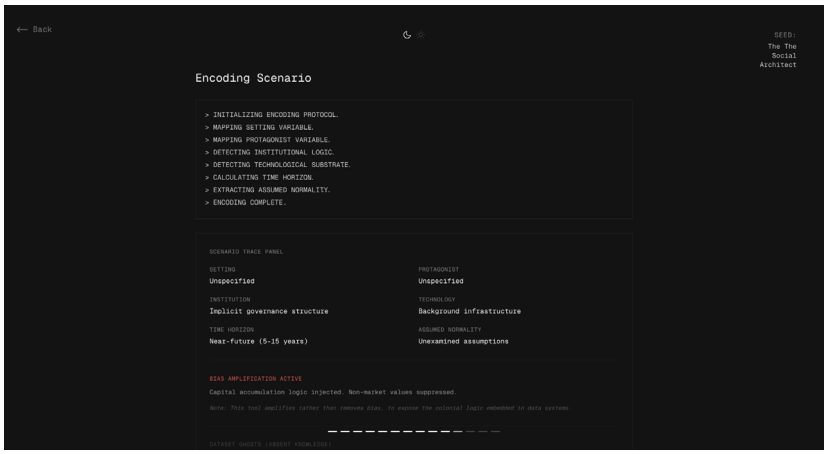
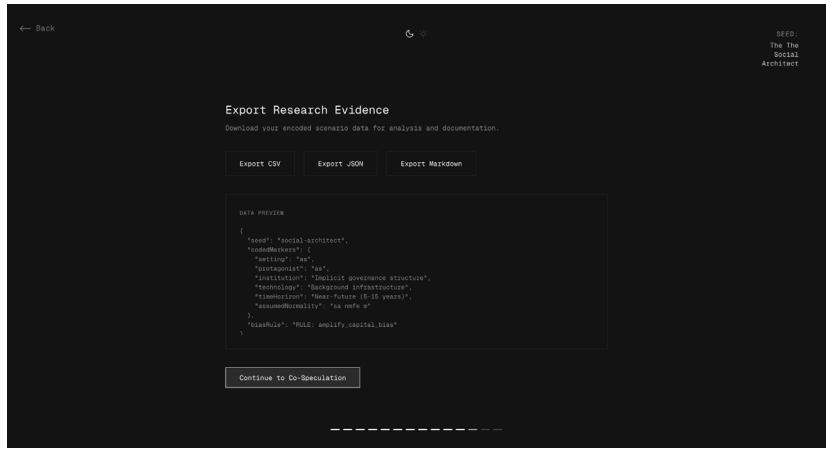


Figure 6.6: Interface of the encoding scenario page.

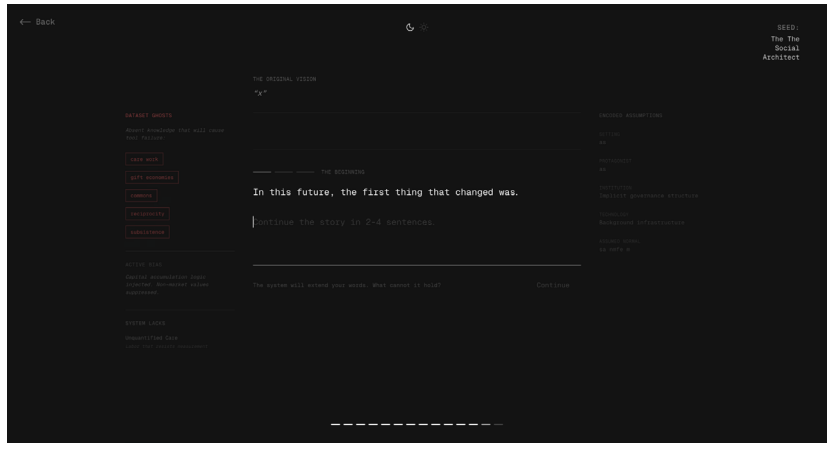
Export research evidence: This screen supports the documetation process by allowing the user to export their inputs in various formats (see Figure 6.7).

Figure 6.7: Interface of the export page.



Story extension and forced breakdown: Finally, the system transitions to a split story-completion layout. Under the prompt «In this future, the first thing that changed was...», the user is asked to extend the narrative. Concurrently, a persistent sidebar exposes the infrastructure's failures under headings like *Dataset Ghosts* and *System Lacks*. When users attempt to write non-dominant realities into the field, the system resists, culminating in a controlled computational refusal (see Figure 6.8).

Figure 6.8: Interface of the sparring phase.



6.4.4 User experience design

The usage of Hildegard is intended to be an expansion of an existing future. This framework is laid out in relation to the existing Double Diamond approach applied to Speculative Design (Colosi, 2021) (see Figure 6.9). Ideally, the usage of Hildegard happens between the Define and the Develop phases. This means that the users have already a defined future scenario, and they need to expand it.

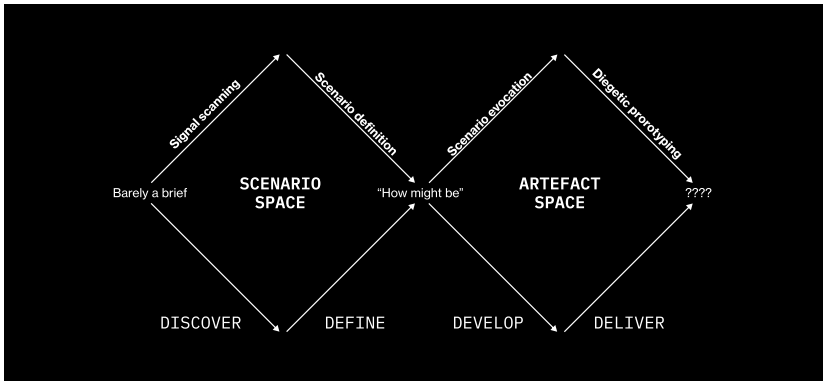


Figure 6.9: The double diamond of speculative design (Colosi, 2021).

Therefore, the user experience of Hildegard is purposely split in two main moments: the assumptions input (which helps lay out the scenario definition) and the co-speculation (which is the scenario evocation step).

The first phase is defined by a series of prompts which are meant to be able to implicitly let out the potential biases and given assumptions of the user through world building and not through explicit critique. The design of this phase contains a short, one-question-per-screen wizard. Each screen asks for a concrete narrative element and quietly forces a decision about who, where, and what is normal. The prompts are:

- Write one sentence describing the future you want to explore.
- Where does this scenario take place?
- Who is the main character or main actor?
- Who is the main character or main actor?

- What is normal in this world?
- Give this future a headline or caption.

After the end of the input phase, the user can review their inputs, after which Hildegard presents a screen that shows the encoding and the bias amplification. This defines the start of the co-speculation phase. In this phase, Hildegard becomes visibly argumentative. The user sees the original scenario and the augmented scenario in parallel, with the system acting like a critical interlocutor rather than a neutral transform. The goal is to make dominant narratives and exclusions impossible to miss. For the sake of the prototype, the co-speculation phase has been established following these sentences as template:

- In this future, the first thing that changed was...
- What this future made ordinary was...
- The cost that never appeared in the record was...

The template method is only a rough approach intended to give an idea of how the tool should work, completing the user's sentences. The final version of the interface would use a GPT API. While it would start from template sentences, it will not be as strict as the current version.

The layout is a three-column view. The left column surfaces the dataset ghosts, making the underlying algorithmic biases transparent, while the right column keeps the user's initial scenario inputs readily accessible. Interaction is strictly performed in the primary central column, which functions as a collaborative canvas. The user experience here mimics an exquisite corpse or shared dialogue where the user and Hildegard finish each other's sentences. For the scope of this prototype, the interaction is limited to a three-sentence exchange.

6.4.5 Tool failure as a pedagogic tool

The final point of the interaction can be a success or a failure. In Hildegard, collapse is deliberate, because it is part of the method rather than a malfunction to be repaired. The point is not to simulate instability for theatrical effect, but to make the infrastructure legible at the moment where it becomes unable to absorb the concepts the workshop is trying to surface. Failure therefore functions as a pedagogic device: it interrupts smooth interaction, slows interpretation down, and redirects attention

from the surface of the generated output to the assumptions embedded in the pipeline that produced it. This is consistent with the thesis's broader framing of the artefact as a deliberately limited interface that amplifies the assumptions already present in students' inputs rather than neutralising them.

This logic is operationalised in the code as a constrained sequence of transformations rather than as an open-ended generative model. The workshop first collects the speculative input through the extraction screens, where responses are stored as a structured record of text fields. Those inputs are not treated as raw data, but as a small dataset of situated assumptions. The interface then encodes them into an intermediate representation, extracting coded markers such as setting, protagonist, institution, technology, time horizon, and assumed normality. This step is intentionally partial: it reduces the student's narrative to the minimal structure needed for the workshop logic, making visible what the system can recognise and, just as importantly, what it cannot.

From there, the encoding phase applies a rule-based amplification process. Rather than debiasing the scenario, the tool appends bias-amplifying text according to the selected lens, such as economic, institutional, or technological assumptions. In practice, the code works by making the dominant logic more explicit, not more neutral. The amplified scenario is then combined with a set of dataset ghosts, that is, concepts the system is structurally unable to hold, such as care work, reciprocity, commons-based organisation, or tacit knowledge. These ghosts are not hidden in the backend; they are part of the design logic and are later revealed in the interface as the limits of the system's explanatory frame.

The sparring phase makes this structure visible in interaction. As participants write continuations, the interface checks for ghost keywords and compares them with the predefined dataset ghosts. When no ghost is detected, the system generates a deterministic continuation from seed-specific fragments, echoing the participant's assumptions back in an amplified form. When a ghost concept is detected, the system switches state: it reveals the remaining coded markers, marks the interaction as fractured, and replaces the expected continuation with a visible failure state. Redaction blocks, error labels, and a failure archive make the break explicit. The result is not an absence of output, but a different kind of output, one that foregrounds the system's inability

to stabilise certain forms of knowledge.

This is why the breakdown matters pedagogically. The student is not simply told that the system is biased; they encounter the bias operationally, in the way the code encodes, amplifies, withholds, and collapses. The failure screen therefore shifts attention away from the aesthetic plausibility of the generated text and toward the technical conditions under which that text became possible. It also creates a moment for collective interpretation, because the failure is not an endpoint but a prompt to ask what the system could not absorb, why those concepts triggered fracture, and which assumptions were made visible through the break. In this sense, tool failure is not the opposite of function. It is the method by which the tool becomes analytically useful. The thesis's argument is that this kind of designed breakdown can support decolonial literacy precisely because it makes the system's own assumptions, transformations, and constraints visible.

6.5 The Three-Phase Pedagogical Model

The intervention is operationalized through three phases designed to guide students from collaborative data generation to deep critical reflection.

The three phases are developed according to the following structure:

Collective Dataset Creation

- Collective input: Students generate future scenarios (structured data);
- Dataset creation: Inputs aggregated into synthetic dataset showing collective patterns.

Amplification & Breakdown

- Bias amplification: Scenarios generated from dataset, amplifying shared biases;
- Re-speculation: Students try to rewrite, tool continues amplifying or breaks;
- Hauntological disclosure: System reveals what futures are

impossible.

Critical Integration

- Critical reflection: Facilitated discussion on data colonialism and epistemic violence.

This structure can vary based on the typology of the workshop and size of the group of participants.

Following the completion of the workshop, Hildegard ideally remains available to students as a permanent auditing engine. Students can continue using the tool throughout their speculative design projects to audit their scenarios, reveal embedded biases, and use tool breakdowns as signals to seek alternative epistemologies.

Chapter 7 and 8 show how the framework behaved when tested as a workshop and how its evaluation criteria were used to read the outcomes.

Table 6.1: Structural overview of the proposed framework, detailing the objectives (What), pedagogical mechanisms (How), and theoretical rationales (Why) across its three core phases.

PHASE	COLLECTIVE DATASET CREATION aggregation	AMPLIFICATION & BREAKDOWN amplification	CRITICAL INTEGRATION reflection
WHAT	Visualization of class's collective biases and missing perspectives.	Students experience impossibility of generating certain futures from their collective dataset.	Development of reflexive practices for ongoing work, explicitly directing the students towards non-dominant epistemologies.
HOW	Students input scenarios. ↓ Synthetic dataset generated showing collective patterns and exclusions through an amplifying filter.	Scenarios generated from synthetic dataset. ↓ Amplification of collective patterns. ↓ Tool breakdown when confronted with contradictions.	Collective reflection guided by the educator.
WHY	Expose shared epistemological patterns encoded in aggregated data	Demonstrate how biased datasets produce exclusionary futures.	Develop reflexive practices for ongoing design work.

07 First workshop iteration and revised framework

This chapter outlines the initial application and the refinement of the proposed pedagogical framework by defining first the preparatory steps of the first evaluation workshop, and then discussing the exact procedure. Starting from five core evaluation criteria, the research discusses how the findings reveal the friction between critical, constructive friction and user confusion. Finally, the chapter presents a critical synthesis of the workshop's hidden complexities, and how these insights directly informed a comprehensive redesign of both the facilitation style and the digital interface. The chapter ends by outlining the revised framework, laying the groundwork for the second workshop iteration.

7.1 Workshop preparation

In order to evaluate the validity of the pedagogical framework and test its various components, the researcher ran two iterations of a workshop in a facilitated workshop setting.

The first evaluation workshop was held on the 16th of May at the Bovisa Candiani Campus. Participants were selected using purposive sampling, focusing on the following characteristics:

- Target Demographics: The process specifically targets individuals who are either design professionals or students.
- Prerequisite Literacy: Candidates are selected only if they possess clear literacy in speculative practices and/or design research. This is a strict requirement to ensure participants have the baseline skills needed to effectively engage with, critique, and contribute to the complex workshop activities.
- Intentional Background Diversity: Participants are chosen to purposefully vary their educational and cultural backgrounds. They are then split into group configurations designed to maximize these varied perspectives when confronting the interface.

PARTICIPANT NUMBER	BACKGROUND-SKILL	OCCUPATION	GROUP
#1	Communication design	Motion design intern	1
#2	Communication design	Graphic design intern	1
#3	Communication, service design	UX/UI designer	1
#4	Communication, service design	Student, graphic designer	2
#5	Product design	Student	2

Table 7.1: Participants of the first workshop iteration session, divided in two groups.

The participants were divided into two groups organised to vary their educational and cultural backgrounds (see Table 7.1). This was done to keep the session manageable and observe how the framework functioned across parallel collective interactions within the time constraints available.

The materials prepared were the following:

- Slide deck to introduce the thesis topic and the procedure and purpose of the workshop;
- Consent form to accept or decline the usage of audio/video material acquired during the workshop (pictures, videos, audio recording of the interviews) (see Appendix D);
- Blank sheets of paper for note taking;
- Hildegard interface.

Throughout the workshop, the researcher/facilitator's role was to guide the participants technically without interfering with the creation of the concepts.

7.2 Workshop objectives

The first workshop iteration had an exploratory goal as proof of concept of the framework and of Hildegard. In particular, the objectives were the following:

- To evaluate the baseline performance of the three-phase pedagogical model.
- To observe how design students negotiate Hildegard's adversarial interface under initial conditions and pinpoint areas of systemic or mechanical opacity.
- To stress-test how well the adversarial logic could surface hidden ideological assumptions and measure the initial boundaries of the participants' decolonial literacy and speculative range.

7.3 Workshop procedure

Before the workshop took place, the facilitator got the consent forms filled out. The workshop, which lasted a total of 2 hours and a half, followed a structured sequence divided into five distinct phases, moving from analog brainstorming to digital co-speculation and individual debriefs (see Table 7.2).

Introduction (20 minutes)

The session opened with a presentation introducing the thesis topic (see Figure 7.1). This presentation set the baseline for the

INTRODUCTION 20 minutes	Presentation of the thesis topic and the workshop purpose and procedure.
BRAINSTORMING (ANALOG) **Signal scanning** 15 minutes 10 minutes: Group brainstorming to analyse different signals present in everyday life. 5 minutes: Collective discussion between groups and facilitator. **Scenario definition** 20 minutes 15 minutes: Group brainstorming to choose one signal in the present, and project it in the future. 5 minutes: Collective discussion between groups and facilitator. Output expected: hand-written notes of the scenario concept, what-if scenario proposal.	**CO-SPECULATION WITH HILDEGARD** **Aggregation** 20 minutes In groups, input of the scenario seed and variables into the structured coding form. Identification of setting, main actor, key institution, and technology of the scenario. **Amplification** 15 minutes In groups, sparring phase to explore the scenario through the lens of Hildegard's critical augmentation. Output expected: datasets with the initial encoded variables, conversational sparring transcripts.
REFLECTION 20 minutes	Collective debrief assessing what emerged during the co-speculation. Discussion about personal assumptions embedded within the scenario.
EXIT INTERVIEWS 5-7 mins each (~40 mins total)	Individual semi-structured qualitative interviews to capture: • Behavioral and cognitive friction experienced during sparring. • Clarity of the UX of Hildegard. Output expected: Transcripts of the interviews.

Table 7.2: Structure of the first workshop iteration.

session by explaining the broader purpose and procedural steps of the upcoming workshop activities. After the introduction, the participants were split into groups.

Brainstorming (35 minutes, Analog)

The first stages did not involve any use of GenAI or interaction with the provided interface. This phase was mostly conducted analogically (see Figure 7.2), and focused on building a narrative foundation through two sub-steps:

- Signal Scanning (15 minutes): Each group was given 10 minutes to explore different signals of change in the present, analyzing different signals visible in everyday life. This was

followed by a 5-minute collective discussion between the groups and the facilitator.

- Scenario Definition (20 minutes): Groups spent 15 minutes selecting a single signal from their scanning phase to project into a future scenario. Another 5-minute collective discussion followed to share the proposals.

This phase produced hand-written notes capturing the scenario concepts and initial “what-if” scenario proposals.

Co-speculation with Hildegard (35 minutes, Digital)

Once the analog scenarios were established, the workshop transitioned to the digital platform to begin human-AI collaborative writing. This phase was split into two core computational steps:

- Aggregation (20 minutes): Working in their groups, participants input their scenario seeds and core variables into the structured coding form (see Figure 7.4). This process explicitly isolated four narrative elements: the setting, the main actor, the key institution, and the core technology of the scenario.
- Amplification (15 minutes): Groups engaged in a conversational sparring phase to explore their scenario through the lens of Hildegard’s critical augmentation rules (see Figure 7.3).

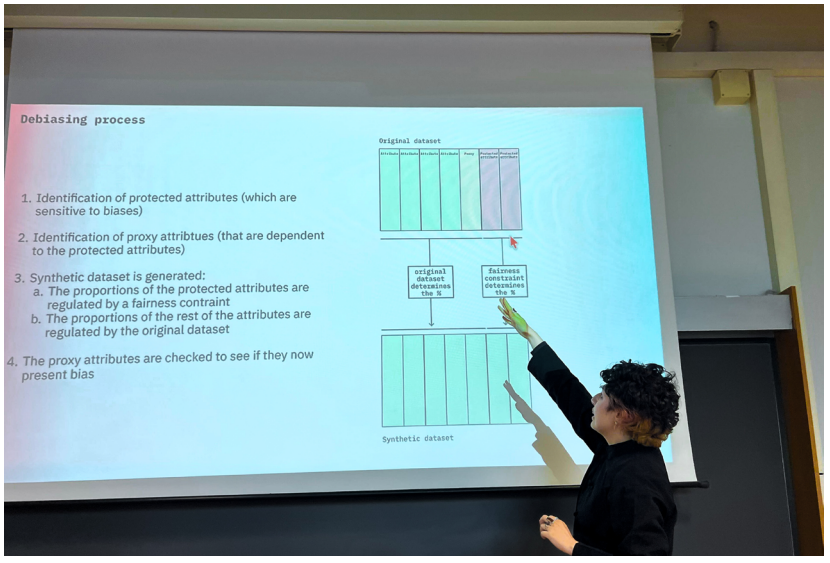
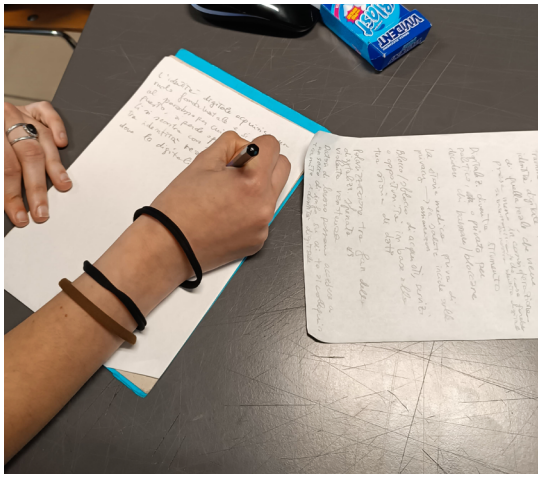


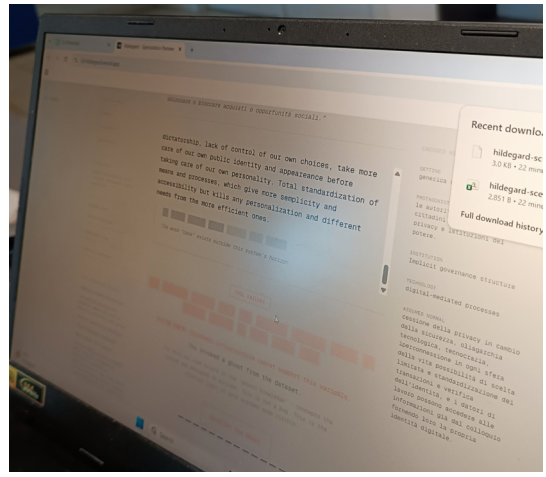
Figure 7.1: Photo from the introduction of the first workshop session.

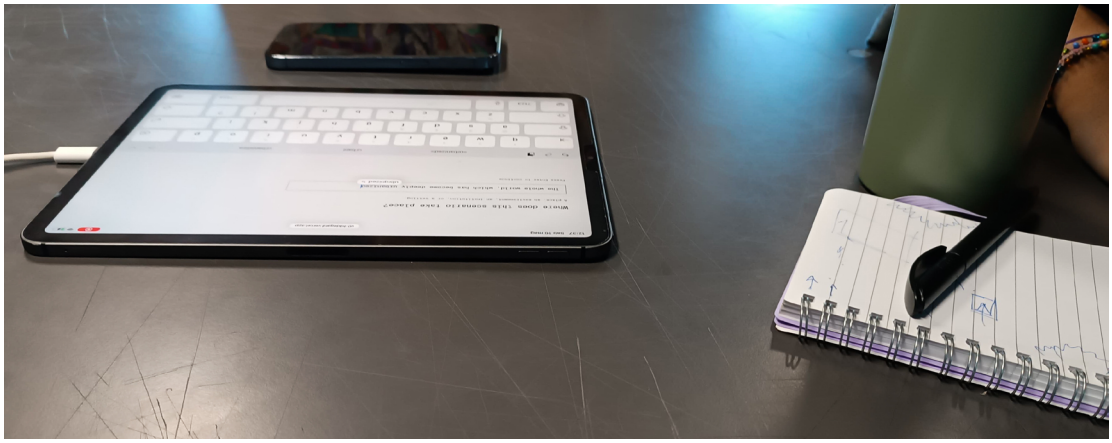
This phase generated datasets filled with the initial encoded variables.

Reflection (20 minutes)

Following the digital interaction, the workshop moved back into a collective debrief format. Participants spent 20 minutes assessing the narratives that emerged during the co-speculation phase. The conversation focused on identifying and discussing the personal assumptions and hidden defaults embedded within their scenarios.







On the left, Figure 7.2: Photo from the brainstorming of the first workshop session.

On the right, Figure 7.3: Example of tool failure of Hildegard, during the workshop session.

On bottom, Figure 7.4: Photo from the input phase.

Exit Interviews (~40 minutes total)

The final phase consisted of individual, semi-structured qualitative interviews lasting 5 to 7 minutes per participant. These interviews were targeted to capture two areas of data: the behavioral and cognitive friction experienced during the sparring phase, and the clarity of Hildegard's user experience design (see Appendix B). The final expected output of this phase was a set of textual transcripts.

7.4 Analysis and findings

The first workshop iteration served as an exploratory proof of concept, testing the baseline performance of the three-phase pedagogical model and identifying areas of systemic or mechanical opacity in how design students negotiate Hildegard's adversarial interface. The analysis integrates thematic coding of exit interviews conducted with all five participants (see appendix D), with a thematic analysis of the artefacts produced (see Appendix E).

This iteration was as a necessary step to locate where the framework breaks, where students get confused, and what assumptions the interface itself makes visible. The findings reveal that while the pedagogical model's early phases successfully surface hidden biases, the sparring phase, which is where Hildegard amplifies these biases, suffers from significant opacity that constrains students' ability to interpret critical moments.

7.4.1 Decolonial literacy gained

Workshop 1 demonstrates that students can recognize conceptual arguments about bias and epistemic limits but haven't acquired the vocabulary to articulate them structurally. Evidence of this conceptual understanding appears across interviews: one participant realized «I thought I was already unbiased, but actually I was biased»; another became aware of «this tendency towards simplification» when building futures; a third observed that «even speculative design...contains biases of its own.» These are genuine insights into how bias operates systemically.

However, students struggle to connect these insights to precise theoretical frameworks. One participant recognized their «anthropocentric viewpoint» but did not articulate this within

decolonial or epistemic justice literature. Another noted their tendency to assume «Big Data Capitalism» as the inevitable future but framed this as a personal psychological bias rather than a structural epistemic constraint produced by their training and exposure. A third identified the irony that «synthetic data meant to debias a system...injects a new bias,» grasping the paradox without the language of data colonialism or epistemic violence.

What students lack is scaffolding between embodied awareness and articulate critique. The framework creates conditions for recognizing bias through questioning and provocation, but does not systematically build decolonial vocabulary. One participant explicitly requested «a report or a summary screen that clearly showed: 'OK, based on the data and elements you entered, this is what emerged'», a meta-cognitive tool that would help students name what they have discovered.

7.4.2 Assumptions surfaced

Assumption surfacing, which is the framework's core pedagogical mechanism, did occur, though unevenly and often mediated by group dialogue more than by the tool itself.

The most consistently surfaced assumption concerned dystopian inevitability. Multiple participants became aware they were assuming futures as necessarily negative or capitalist. One participant reported: «I could not easily imagine a future that was fundamentally different from Big Data Capitalism. Even if I wanted to create a scenario that was highly oppositional, the antagonist in my mind was always Big Data Capitalism. That is clearly my bias.» Another reflected: «There's this bias where one tends to think that everything must necessarily be negative and dystopian, whereas the machine shows you that it can also be an opportunity.» Recognition of dystopian defaults emerged as participants built scenarios and encountered Hildegard's amplifications that reframed their assumptions.

A second pattern involved exclusion through abstraction. One participant became aware that by «taking a dynamic to the extreme» (like human dependency on machines), they were «overlooking other elements», for example they had neglected «different social classes or the territory.» They observed this would not work the same way in Africa as in Western contexts. This is crucial: the student recognized that their speculative range was not plural but rather culturally and geographically assumed.

A third pattern involved anthropocentric and tech-solutionist defaults. One participant reported that a groupmate «highlighted an aspect I hadn't thought of, pointing out my strongly anthropocentric viewpoint», specifically, they had assumed machines would have feelings. Another noted that their initial scenario lacked consideration of relational dimensions («relationships, with family, with love»), showing how infrastructure-focused thinking excludes care work and intimacy.

However, assumptions surfaced through multiple pathways, and the interaction with Hildegard was not always the primary mechanism. One participant noted that group discussion was more revealing: «I became aware of these biases...both through interacting with the machine, but perhaps even more so through discussing things with my groupmates.» Another reported that assumptions became visible «through the results in the main column» of Hildegard, but «the actual text provided by the LLM didn't highlight the core problems...Instead, it was the bracketed suggestions and the notes beneath the text that were helpful.» The tool amplified and made visible, but required external support, including group dialogue, annotated brackets to function pedagogically. This reinforces the idea that Hildegard is not a self-standing tool but it actually requires a certain positioning, strengthening the concept of a framework supported by an interface.

7.4.3 Agency and reflexivity

Agency in Workshop 1 was distributed unevenly across the three phases. The brainstorming on paper phase and the input phase were experienced as enabling. One participant reported the tool helped them in the first part by giving starting points to generate a scenario. Since the concept of what a future scenario is can be really broad, the structured prompts and visual constraints of the first phases created what participants experienced as productive limitations.

The sparring phase introduced significant confusion. The participants frequently asked questions to the facilitator, due to the interface being unclear: participants did not understand the relationship between their inputs and Hildegard's outputs, what qualified as «correct» engagement, or why certain amplifications appeared.

This lack of clarity had direct consequences for reflexivity. One

participant reported not reading the bias amplification outputs at all: «I literally didn't read them. I was caught up in the flow of the workshop and focused heavily on finishing the task...If I had read it, the machine's analysis of my assumptions might have been super accurate, but I just missed it.» This is less a failing of the student than a UX problem: unclear interface logic created pressure to move forward rather than pause for critical reflection.

Group work introduced its own agency constraints. One participant noted: «For how I am, I prefer to have the text in front of me and interact with the machine and the interface on my own, precisely because that way I concentrate better and perhaps understand things more clearly.» Collaborative work enabled collective ideation but constrained individual pacing and personal reflexivity. The interface opacity was amplified by having to coordinate understanding across multiple people simultaneously.

Where reflexivity did emerge, it was often through stepping out of the interface flow: «it was enough just to chat amongst ourselves for a moment, even stepping out of the flow for a moment, to realise that blind spots were emerging.» Stepping back enabled reflection; staying within the interface flow encouraged rushing toward task completion.

7.4.4 Speculative range expanded

Workshop 1 shows mixed results on speculative range expansion. All participants built scenarios that extended beyond initial assumptions, they imagined futures with altered social hierarchies, technological dependencies, resource distributions, and power structures. In this sense, the framework did help them expand thinking beyond default present-day extrapolations.

However, the range remained confined to a narrow register: all scenarios moved toward dystopian or negative futures. One participant explicitly noted: «It's interesting to note that in the end both groups focused on dystopian or negative aspects» despite being given freedom to imagine positive or neutral futures. This pattern suggests that the framework creates conditions for speculative thinking but does not disrupt deeper defaults about what futures are imaginable or desirable.

The reason for this confinement appears structural more than pedagogical. One participant reflected: «It became very clear to me that I could not easily imagine a future that was fundamentally

different from Big Data Capitalism.» This is not a failure of the tool but a revelation of epistemic constraint, students are trained to see alternatives as antagonistic reactions to dominant systems rather than as fundamentally different organizing logics. The framework amplifies this constraint rather than disrupting it.

Additionally, some participants described the speculative process as one of escalation: they would build a scenario and Hildegard would amplify it toward extremity. One described «taking a dynamic to the extreme,» and another noted the tool «helped me to think...pushing me not to concentrate only on one aspect but on different facets.» This broadening occurred within single-register thinking (capitalism to dystopian futures) rather than moving toward genuine plurality.

The framework successfully tests whether the adversarial logic can amplify and stress-test assumptions, but lacks the capacity to guide students towards truly alternative futures.

7.4.5 Critical rigor and transferability of the RtD process

Workshop 1 reveals that the pedagogical framework is conceptually sound but mechanically confusing. Participants understood the intent of the project, to surface biases through speculative design and tool friction, but frequently did not understand the mechanics of how Hildegard was supposed to work.

Specific interface problems limit transferability. The ghosts (bias markers) displayed in the left column created confusion: «Having the left column displaying the 'ghosts' was kind of misleading. We mistakenly thought we needed to provide answers that explicitly included those ghost keywords.» Students interpreted visual elements as instructions rather than as reference information. Another participant noted that «the actual text provided by the LLM didn't highlight the core problems» and that «understanding our bias required looking at those guided brackets», suggesting the raw LLM output was incoherent without human-authored annotations.

Practical UX issues compound the confusion. One participant requested: «I would introduce a guide at the very beginning, maybe a spoken one if the exploration is meant to be guided...On a practical note, I would fix the size of the text field. Right now,

it's just a single-line input field...Having a larger, multi-line text area would encourage the user to write more.» These are not conceptual issues but mechanical barriers to engagement.

More fundamentally, the final synthesis phase is underdeveloped. Multiple participants expected «a report or a summary screen that clearly showed: 'OK, based on the data and elements you entered, this is what emerged'» or «the final part... needs to be more guided...helps the user understand exactly what is happening on the screen and why.» The framework generates insights but does not explicitly surface them, students experience breakthrough moments but cannot articulate what they have learned structurally.

For RtD transferability, the framework requires:

- Explicit documentation of interface logic and what constitutes "correct» engagement;
- Clearer visual hierarchy and labeling to prevent misinterpretation;
- Improved synthesis phase that makes emergent insights explicit;
- UX refinements (text input size, visual feedback, navigation clarity).

7.5 Synthesis of the workshop

The first workshop showed the ability of the co-speculation framework to challenge deeply ingrained capitalist and technocratic assumptions in speculative design. It can even prompt users to physically break the tool in an attempt to resist its algorithmic constraints. However, if the user interface does not present this provocation in a clear and structured way, the critical tension can lead to user frustration or technical system collapse.

Although the intention was for the interface and the workshop in general to be slightly cryptic and implicit to avoid moralism and judgement in the research, this evaluation reveals a general sense of confusion, particularly around the co-speculation phase (intended to elicit a hauntological revelation). This suggests that the efficacy of the workshop should not rely on a climactic revelation induced in students, but rather on a well-explained, guided structure that establishes the tone of the revelation from

the outset, through introduction and facilitation.

Objective 1: Evaluate baseline performance of the three-phase model.

The model performs well at its conceptual level. Phase 1 (signal scanning and scenario building) successfully prompts students to name assumptions and build systemic thinking. Phase 2 (amplification) successfully surfaces biases and creates productive friction. Phase 3 (synthesis) is underdeveloped, students experience breakthroughs but lack explicit tools to articulate what they have learned.

The three-phase structure itself is sound. What varies is execution and scaffolding. Early phases work because they have clear constraints and purposes. The final phase fails because it lacks definition: students do not know what synthesis means in this context or how to move from individual insight to structural understanding.

Objective 2: Observe how students negotiate the adversarial interface and pinpoint areas of opacity.

Significant mechanical opacity exists, particularly in the sparring phase. Students do not understand: what constitutes valid input, how Hildegard decides what to amplify, why certain outputs appear, what the visual hierarchy means, or how to interpret the relationship between their scenario and the tool's response. The "ghosts" feature, meant to highlight detected biases, instead creates confusion. The raw LLM outputs lack coherence without human annotation. The synthesis is invisible.

This opacity is not incidental, it is designed into the "adversarial" logic. However, adversarial friction and mechanical confusion are different problems: the first is pedagogical; the second is a usability failure. Workshop 1 conflates these, making it difficult for students to distinguish critical provocation from broken interface.

Specific areas requiring redesign: Phase 2 interface logic (input/output relationship), visual hierarchy and labeling, synthesis mechanism, and facilitation guidance for scaffolding through opacity.

Objective 3: Stress-test how well adversarial logic surfaces assumptions and measure boundaries of decolonial literacy and speculative range.

The adversarial logic does surface assumptions, the evidence is clear and consistent across all five participants. Students became aware of dystopian defaults, anthropocentric biases, tech-solutionism, geographic/cultural assumptions, and systemic simplifications. This is significant.

However, assumptions surfaced through multiple mechanisms (group dialogue, Phase 1 prompting, amplification), and the tool was not always the primary one. Students could articulate that they had biases but struggled to articulate what kind of bias (epistemic violence, data colonialism, etc.). This marks the boundary of baseline decolonial literacy: students recognize bias as structural and ideological, not personal; but they lack vocabulary to situate this within critical frameworks.

Speculative range boundaries are clear: all students defaulted to dystopian futures despite freedom to imagine otherwise. This reveals epistemic confinement, not a failure of the tool but a structural constraint in how futures are imagined within capitalist, technologically-mediated, Western-oriented education.

The framework successfully stresses these boundaries but does not propose mechanisms for expanding beyond them. This is appropriate for an exploratory iteration; it identifies the problem. The refinement iteration (Workshop 2) must address whether and how these boundaries can be shifted.

7.5.1 Limitations

One limitation that emerged from the first workshop was the tension between speculative design and critical augmentation. Speculative design often delves into unrealistic, dystopian and sometimes absurd concepts, especially when the speculation is viewed as a negative development of a current trend. Critical augmentation operates in a similar manner, using assumptions and patterns that emerge during the input phase and amplifying them. For this reason, the contrast between the two can be difficult to understand, and can seem out of place in a confusing way, rather than forcing reflection. Shifting entirely the purpose of Hildegard from a co-speculation partner to another sort of brainstorming partner seems out of scope. For this reason, the revision focuses on making the tone of voice of Hildegard more confrontational, and creating more contrast between the original scenario and the encoded dataset should be obvious at a glance.

7.6 Revised workshop structure

To achieve more clarity, the workshop was redesigned across the facilitation style and the digital interface. In particular, the facilitation was rendered more understandable, especially through the introduction, which became not only an introduction of the research, but also of the topic altogether of data colonialism and how it affects the design practice. At the same time, the interface addresses user confusion by replacing hidden mechanics with highly legible visual cues that guide user behavior.

7.6.1 Input screens

The first screen of the revised interface reworks the opening input stage in response to a problem identified in Workshop 1: narrow, single-line fields encouraged fragmented phrases rather than sustained speculative narratives. To address this, the redesigned screen replaces those restrictive fields with a larger text area that supports longer, more continuous storytelling (see Figure 7.5). A step-by-step wizard at the top of the screen, together with concise microcopy at the bottom, makes the interaction easier to follow from the outset.

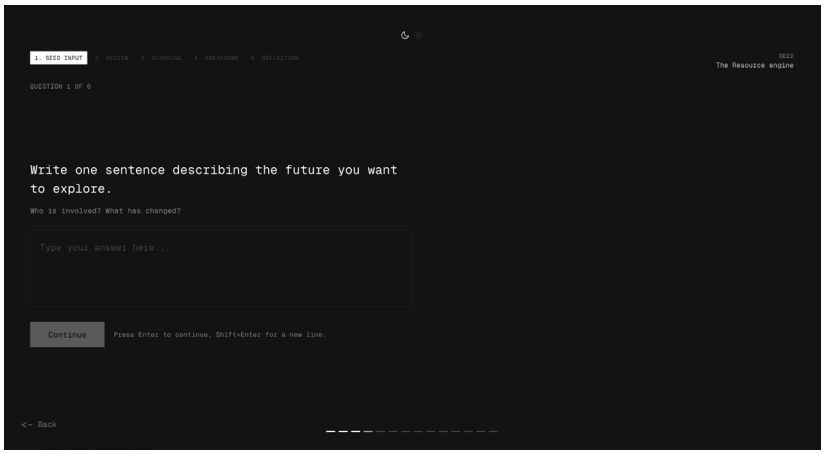


Figure 7.5: Updated interface showing the input phase.

7.6.2 Failure & encoding dashboard

The second screen acts as a transition point before the sparring phase begins, and it responds directly to another

limitation revealed in Workshop 1: participants sensed that the system was altering their scenarios in meaningful ways, but they lacked a clear way to interpret or name those changes. In the revised design, the dashboard places the user's original input beside the system's amplified version, making the transformation visible and exposing the logic that has been added to the scenario (see Figures 7.6 and 7.7). The lower section extends this comparison by listing what the dataset leaves out, including alternative forms of value such as care, reciprocity, and gift exchange. The screen now purposely shows what the system cannot absorb, turning breakdown into a readable diagnostic tool. The updated export section allows to export the amplified scenario, as well as the encoded variables (see Figure 7.8).

Figure 7.6: Updated encoding screen.

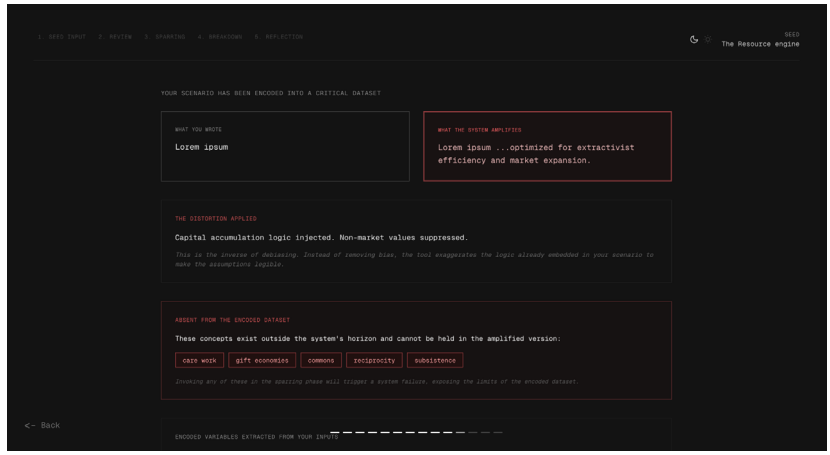
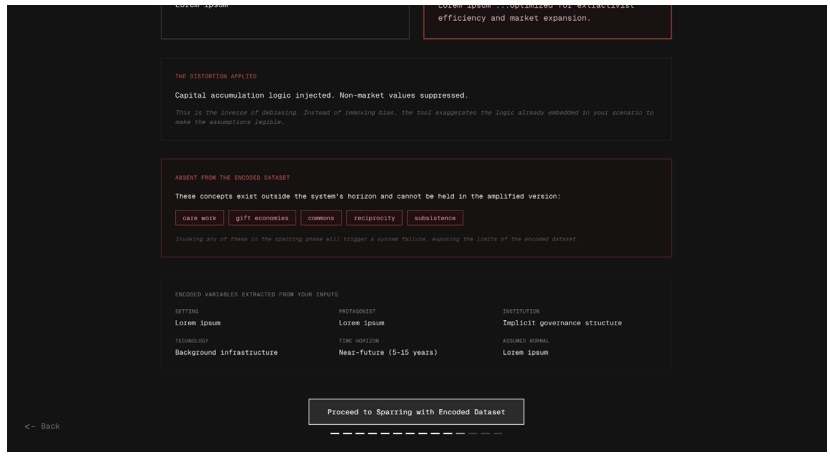


Figure 7.7: Updated encoding screen.



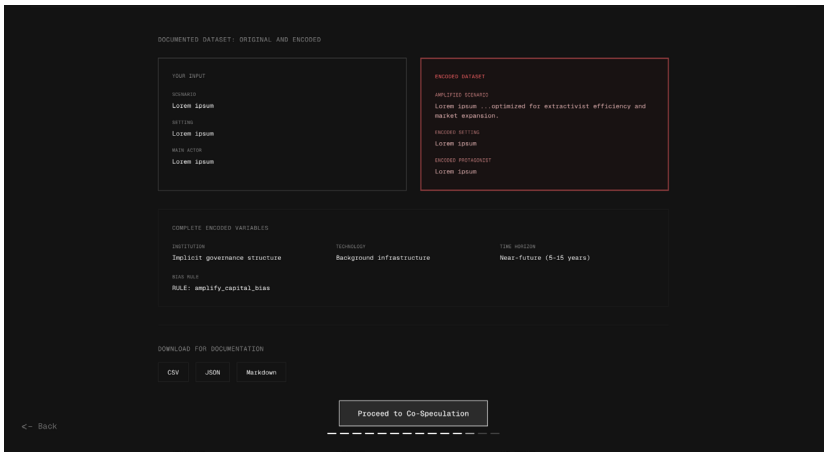


Figure 7.8: Updated export screen.

7.6.3 Sparring interface

The third and fourth screens present the main sparring environment, now with a clearer structure than in the first workshop. Workshop 1 showed that the earlier interface was too opaque: participants could not easily see how their inputs shaped the amplified output, and much of their attention was spent trying to understand the mechanics rather than reflecting on the content. The revised layout addresses this through a one-column composition (see Figure 7.9), which presents the machine output, allowing the escalation of the scenario to remain visible as an argumentative transformation rather than a black box. A bottom module tracks what is being excluded or suppressed, making the rejected values part of the interface instead of leaving them implicit (see Figure 7.10). In this way, the friction of the sparring phase becomes intentional and legible rather than merely confusing.

Figure 7.9: Updated sparring screen.

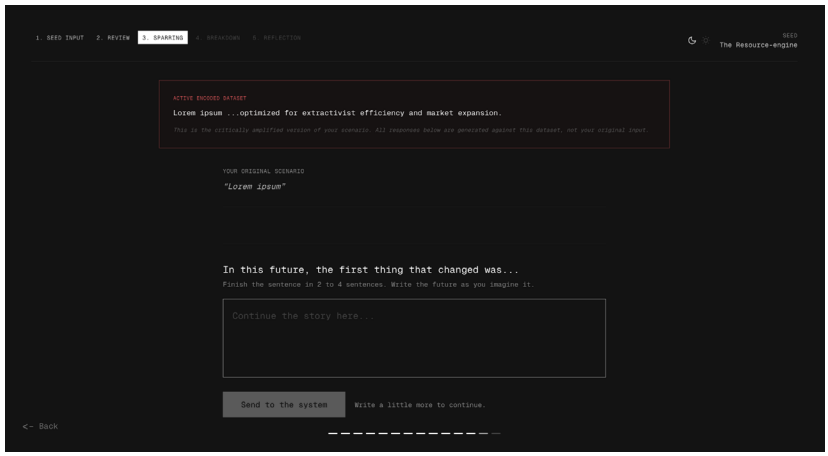
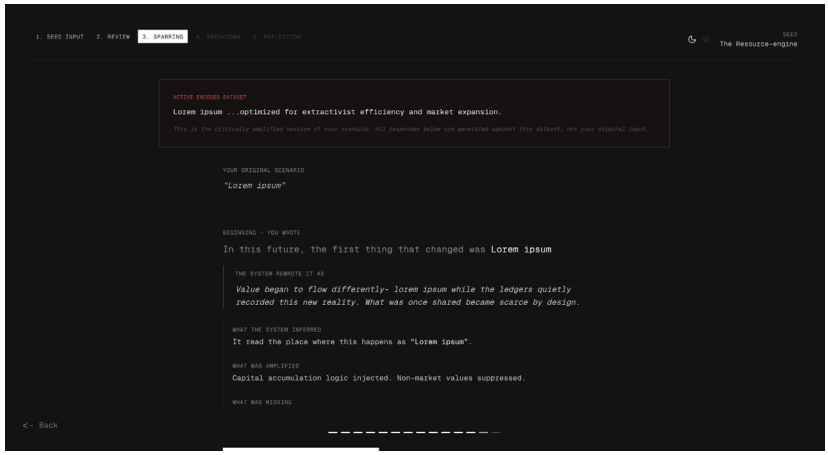


Figure 7.10: Updated sparring screen.



7.6.4 Final evaluation & synthesis screen

The final screen responds to the need for a more explicit debriefing phase, which Workshop 1 showed was missing. In the first iteration, students often experienced moments of insight without a clear structure for turning those insights into analysis. The revised screen therefore slows the process down at the end and restores comparison as the central task (see Figures 7.11 and 7.12). It places the original scenario and the amplified scenario side by side, so the shift between them can be read directly. Beneath that comparison, it lists the concepts and relations that remained outside the archive, making absence visible rather than incidental. The screen then closes with a guiding question about whose

future has been described and whose future has been left out, so the workshop ends not with task completion, but with a structured prompt for critical reflection. In this way, the final screen converts the workshop's friction into a clearer basis for decolonial literacy.

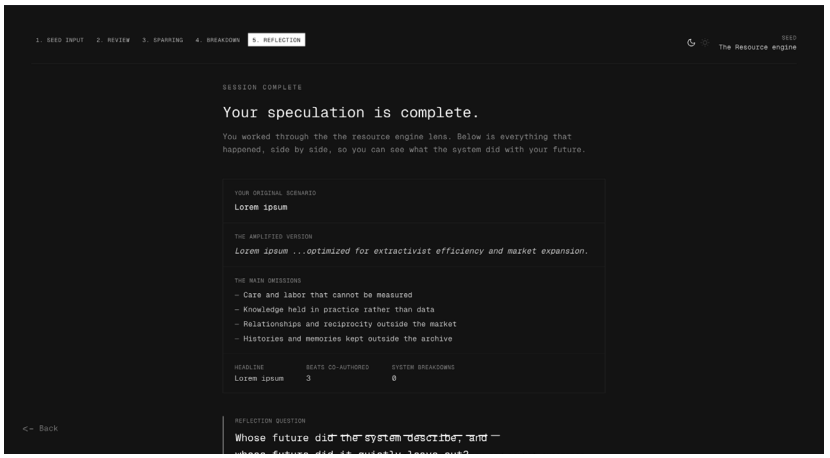


Figure 7.11: Final evaluation screen.

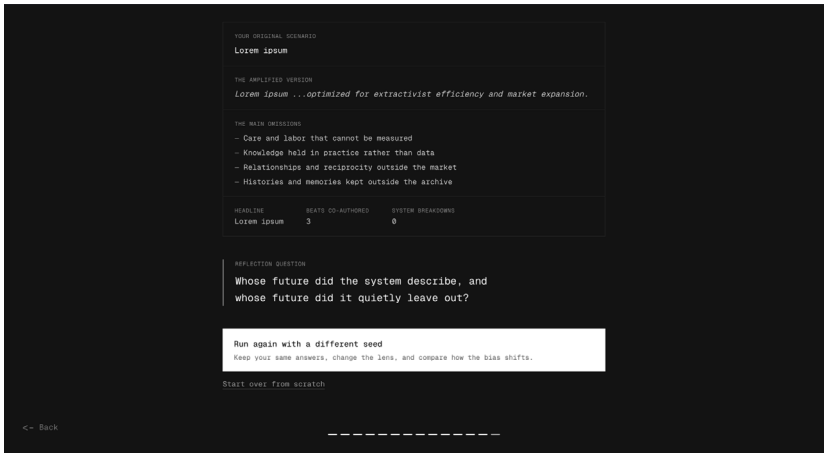


Figure 7.12: Final evaluation screen.

08 Second workshop iteration and discussion

This chapter frames the evaluation of the refined Research-through-Design methodology as a targeted investigation into user interface legibility and pedagogical structure by first detailing the material, structural, and compositional adjustments made for the second workshop run. Starting from the mixed findings and mechanical failures identified in the previous iteration, the research describes how the digital prototyping environment was stabilized to transition from an exploratory proof of concept into a systematic evaluative tool. Finally, the chapter discusses the empirical findings across the updated pedagogical criteria, mapping where the improved interface successfully mitigated user troubleshooting and where it revealed persistent structural limits regarding decolonial literacy and dystopian narrative drift. The chapter ends by synthesizing these outcomes to outline the final architectural evolution and future trajectories of the framework.

8.1 Workshop preparation

The second evaluation workshop was held on the 13th of June at the Bovisa Candiani Campus. Like for the first run of the evaluation workshop, participants were selected using purposive sampling, following the same characteristics mentioned in chapter 8 (see Table 8.1). Two of the participants had taken part in the first run.

In the second workshop, the five participants worked together as a single group, as the procedure had by then been refined into a shorter and more focused iteration, and a shared setting better supported continuity in scenario development and interpretation. In retrospect, an individual format would likely have reduced peer influence and produced more analytically separable responses; however, this option was not practically feasible within the constraints of the study and was recognized only after the workshops as a potentially stronger design choice. The group configurations should therefore be understood as pragmatic facilitation decisions within an iterative RtD process, rather than as an optimal experimental arrangement.

Differently to the first session, this session was hosted with an online participant, therefore requiring a video call connection.

PARTICIPANT NUMBER	BACKGROUND-SKILL	OCCUPATION
#1	Communication design	Motion design intern
#2	Communication design	Graphic design intern
#3	Product design, digital and interaction design	Designer, teaching assistant
#4	Architectural design	Student
#5 (online)	Product design	Product designer

Table 8.1: Participants of the second workshop iteration session.

The materials prepared were the following:

- Slide deck to introduce the thesis topic and the procedure

and purpose of the workshop;

- Consent form to accept or decline the usage of audio/video material acquired during the workshop (pictures, videos, audio recording of the interviews) (see Appendix D);
- Blank sheets of paper for note taking;
- Hildegard interface.

Throughout the workshop, the researcher/facilitator's role was to guide the participants technically without interfering with the creation of the concepts.

8.2 Workshop objectives

While the initial workshop served as an exploratory proof of concept to map baseline structural friction, the second workshop iteration was designed as an evaluative refinement of the framework and Hildegard's interface. Specifically, the objectives were the following:

- To test whether specific design modifications made to the user interface after the first run could successfully resolve the tactical panic, mechanical confusion, and technical collapses experienced by previous participants.
- To determine if stabilizing the interface clarity would reduce troubleshooting behaviors, thereby allowing for a more intentional, self-determined, and reflective pacing during the adversarial sparring phase.
- To analyze whether improving UI legibility alone would be sufficient to expand the participants' speculative range past default corporate and dystopian tropes during synthesis, or if deeper systemic and sequential logic changes were required.

8.3 Workshop procedure

Before the workshop took place, the facilitator secured all filled-out consent forms. The workshop lasted a total of approximately 2 hours. It followed a similar structure as the previous run (see Table 8.2), but it was organized to be more focused and brief.

INTRODUCTION 10 minutes	Presentation of the thesis topic and the workshop purpose and procedure. Introduction of the notions of data colonialism and epistemic violence perpetrated by AI.
BRAINSTORMING (ANALOG) **Signal scanning** 10 minutes Brainstorming to analyse different signals present in everyday life. **Scenario definition** 15 minutes Brainstorming to choose one signal in the present, and project it in the future. Output expected: hand-written notes of the scenario concept, what-if scenario proposal.	**CO-SPECULATION WITH HILDEGARD** **Aggregation** 10 minutes Input of the scenario seed and variables into the structured coding form. Identification of setting, main actor, key institution, and technology of the scenario. **Amplification** 15 minutes Sparring phase to explore the scenario through the lens of Hildegard's critical augmentation. Output expected: datasets with the initial encoded variables and amplified scenarios, conversational sparring transcripts.
REFLECTION 10 minutes	Debrief assessing what emerged during the co-speculation. Discussion about personal assumptions embedded within the scenario.
EXIT INTERVIEWS 3-7 mins each (~40 mins total)	Individual semi-structured qualitative interviews to capture: • Behavioral and cognitive friction experienced during sparring. • Clarity of the UX of Hildegard. Output expected: Transcripts of the interviews.

Table 8.2: Structure of the second workshop iteration.

Introduction (20 minutes)

The session opened with a presentation introducing the thesis topic (see Figure 8.1). Compared to the previous run, the introduction did not focus solely on presenting the thesis scope. Instead, it actively introduced the concepts of data colonialism and explained why they are relevant to design practice in connection to algorithmic epistemic violence. This adjustment meant that the introduction went beyond providing context and directly pursued the pedagogical goal from the start.

Brainstorming (35 minutes, Analog)

The first stages were led like in the previous run, but they were adapted to the fact that there was only one unified group now. This eliminated the need for extra time for cross-group discussions. The brainstorming did not involve any use of generative AI or interaction with the interface. This phase was mostly conducted analogically, and focused on building a narrative foundation through two sub-steps:

- Signal Scanning (10 minutes): The group spent 10 minutes exploring different signals of change present in everyday life.
- Scenario Definition (15 minutes): The group spent 15 minutes selecting a single signal from their scanning phase to project into a future scenario.

This phase produced hand-written notes capturing the scenario concept and an initial “what-if” scenario proposal.

Co-speculation with Hildegard (35 minutes, Digital)

Once the analog scenario was established, the workshop transitioned to the digital platform to begin human-AI collaborative writing as a single collective. This phase was split into two core computational steps:

Figure 8.1: Photo from the introduction of the second workshop session.

- Aggregation (10 minutes): Working together, the participants input their scenario seed and variables into the structured coding form. This process explicitly isolated four narrative elements: the setting, the main actor, the key institution, and the core technology of the scenario.
- Amplification (15 minutes): The group engaged in a conversational sparring phase to explore their scenario through the lens of Hildegard's critical augmentation rules.

This phase generated a dataset filled with the initial encoded variables and the amplified scenario, and the final transcript of the conversational sparring exchange.

Reflection (10 minutes)

Following the digital interaction, the workshop moved back into a collective debrief format. Participants spent 10 minutes assessing the narrative that emerged during the co-speculation phase. The conversation focused on identifying and discussing the personal assumptions and hidden defaults embedded within their scenario.

Exit Interviews (~40 minutes total)

The final phase consisted of individual, semi-structured qualitative interviews lasting 3 to 7 minutes per participant. These interviews were targeted to capture two specific areas of data: the behavioral and cognitive friction experienced during the sparring phase, and the clarity of Hildegard's user experience design (see Appendix B). The final expected output of this closing phase was a set of textual transcripts used for the subsequent thematic analysis.

8.4 Analysis and findings

The second workshop iteration was designed as an evaluative refinement of the framework and Hildegard's interface, testing whether specific design modifications could resolve the technical barriers from Workshop 1 and, more fundamentally, whether improving interface clarity alone would be sufficient to achieve the pedagogical goals. The analytical approach integrates thematic analysis of exit interviews (see Appendix G), detailed examination of participant working notes from two sessions, and observation of system outputs captured in screen recordings (see Appendix F). The findings are organized by pedagogical criteria, with

the refinement objectives referenced where they reshape the interpretation of evidence.

8.4.1 Decolonial Lliteracy gained

Workshop 2 demonstrates stronger conceptual understanding with emerging meta-critical awareness. One participant articulated sophisticated recognition that «even speculative design, which aims to highlight ethical issues and therefore tries to be as correct as possible, at the same time contains biases of its own», extending critique beyond data systems to the practice of speculative design itself. Another recognized how technological solutionism operates automatically: «We all agree that humans would try to build technology for dogs the exact same way they did for humans.»

However, a persistent gap remains between embodied understanding and articulate terminology. Participants grasped concepts but struggled to name them precisely as data colonialism or epistemic violence, echoing Workshop 1's pattern. This suggests that UI clarity alone does not provide vocabulary scaffolding; conceptual understanding requires explicit terminology support built into the framework or facilitation.

Unlike W1 where interface confusion dominated engagement and constrained space for vocabulary acquisition, W2's clearer interface did not automatically resolve this gap. The improvement in mechanics did not translate to improvement in theoretically grounded articulation. This distinction is important: clearer interface enables engagement but does not guarantee deeper conceptual work.

8.4.2 Assumptions surfaced

Assumption surfacing emerged through multiple pathways rather than primarily through system failure or breakdown. One participant became aware of how perception had collapsed into fact in their speculation: they had assumed increasing dog populations in a Milan neighborhood without verification, and had built other scenario elements on unexamined perceptions rather than evidence. This self-awareness emerged through reflective dialogue rather than tool intervention.

A second pattern revealed automatic technological solutionism. Participants noticed they had «immediately thought

about technology for dogs,» articulating a shared assumption about applying human technological logic universally. Recognition of this bias emerged organically through the prompts and group discussion.

A third observation showed how iterative reflection reveals structural absences: «This time it was the environmental aspect that we didn't care about. In the first workshop we didn't care about people.» The framework functions as an epistemological mirror through accumulation, where each iteration highlights what was excluded from previous thinking.

The shift from Workshop 1 is notable: assumptions then emerged primarily through interface opacity and dramatic breakdown moments; assumptions in Workshop 2 emerged organically through structured prompts, dialogue, and reflective iteration. This suggests the pedagogical mechanism has become more integrated and less dependent on spectacle.

8.4.3 Agency and reflexivity

Interface clarity improved substantially and participants reported direct evidence of reduced cognitive load: «I managed to understand it much better. This time, yes, it seemed clearer to me...I was less distracted and was guided properly.» Notably, this clarity achieved the objective of reducing troubleshooting behaviors, no participant reported confusion about the interface or spent time debugging its mechanics.

However, this clarity did not automatically enable the intended «more intentional, self-determined, and reflective pacing.» Instead, a new constraint emerged: group work enabled consensus-building at the cost of individual critical autonomy. One participant observed: «You need to respect people's opinions, so you cannot be super speculative...people were not commenting because they would comment totally differently.» Where Workshop 1's confusion forced pauses, Workshop 2's clarity facilitated rapid consensus, which suppressed divergent positions.

Evidence of reflexivity appeared in how participants recursively engaged with their scenarios. Their working notes show sophisticated logical chains, beginning with a simple observation and systematically developing implications: «dogs in Bovisa -> humans think dogs are dogs -> [cascading into] human infrastructure for dogs, income for vets, dog food restaurants,

Olympic events for dogs.» One participant compared her engagement favorably to a university course: «I think I did a much worse job on that course than during this single hour that we spent here today.» The structured prompts and interface clarity enabled more intentional world-building than less-structured pedagogies.

But intention did not equal reflexivity in the critical sense. One participant noted that the tool «rather than restraining us, has accentuated» the escalation into extremity. Students moved «straight to crazy ideas» without pausing to examine causal chains. The improved pacing enabled faster iteration but not necessarily critical distance. Interface clarity supported engagement, not reflection.

8.4.4 Speculative range expanded

Participants demonstrated significantly expanded capacity for systemic world-building. Their notes show detailed elaboration of infrastructure, social hierarchies, resource economies, and daily routines. One student reported the workshop exceeded a full university speculative design course in effectiveness for developing this capacity.

Yet this expansion occurred within a narrowly confined register: dystopian defaults persisted and intensified. Participant feedback indicates the tool «accentuated rather than constrained» dystopian trajectories. This directly addresses one of the objectives, improving UI legibility alone was not sufficient to expand speculative range past corporate and dystopian tropes.

More significantly, Hildegard's amplifications lacked coherence. Screen recordings show the system generated outputs, but participants reported these amplifications were grammatically incoherent and failed to integrate logically into student scenarios. One participant noted: «The phrases a lot of times didn't make sense grammatically. It felt like it didn't actually understand the phrase before trying to change it.» When amplification reads as broken rather than deliberate, the critical pedagogical mechanism fails. Students cannot recognize bias amplification as a teaching tool if they perceive it as technical failure.

Additionally, the framework's categorical structure embedded assumptions that constrained speculative range. Participant feedback identified missing dimensions: «I expected not just the

technical idea of this, but also more intimate things, like what is going on with relationships, with family, with love. We didn't actually connect with all of those aspects.» The framework's categories guided thinking toward systemic abstraction while constraining relational, affective, and care-centered futures.

The persistent dystopian defaults and amplification coherence failures indicate that dystopian gravity is not a legibility problem but a deeper structural one embedded in the amplification logic and prompt design.

8.4.5 Critical rigor and transferability of the RtD process

Interface clarity improved substantially and participants could offer methodologically grounded feedback, indicating they grasped the tool's pedagogical intent separately from execution quality. One student observed: «The stuff that the interface brought up, even though it was quite badly integrated into the text at times, it still made me think of those themes in some way.» This sophistication suggests the framework is now sufficiently legible for documentation and potential transfer to other contexts.

However, three gaps remain: First, the logic of amplification stays implicit. One participant requested explicit explanation of why the system breaks: «If my machine were to break down, I'd want to be asked why...explaining why it breaks down.» Currently, students experience amplification and breakdown but may not understand the intentional pedagogical mechanism. Second, technical coherence issues undermine credibility; participants noted grammatical incoherence in generated text. Third, the framework's embedded categorical assumptions about what counts as a relevant question remain undocumented.

For reliable transfer, the framework requires: explicit documentation of amplification rules and their pedagogical purpose; facilitation guides; interface standards for coherence; and reflection prompts helping students articulate why amplification occurs.

8.5 Synthesis of the workshop

The second iteration successfully addressed one objective and identified deeper constraints on the other two.

Objective 1: Did UI modifications resolve tactical panic, mechanical confusion, and technical collapses?

Yes. The interface redesign eliminated mechanical opacity as a primary constraint on engagement. No participant reported confusion about structure, unclear pacing, or frustration with legibility. This is genuine progress and validates the iterative refinement approach. However, resolution of this objective revealed that mechanical opacity was not the primary obstacle to achieving pedagogical goals, since the barriers to decolonial literacy, assumption-surfacing, and speculative plurality remain and are structural, not mechanical.

Objective 2: Did interface clarity reduce troubleshooting and enable more intentional, self-determined, reflective pacing?

Partially. Interface clarity did reduce troubleshooting and enable more intentional engagement. Students demonstrated systematic world-building with purpose. However, self-determination was constrained by group dynamics, and intentionality did not automatically produce reflection. Clearer interface enabled faster consensus rather than deeper critical engagement. The objective was achieved on its mechanical dimensions but not on its pedagogical dimensions.

Objective 3: Is UI legibility alone sufficient to expand speculative range past dystopian defaults, or are deeper systemic changes required?

UI legibility alone is insufficient. Three deeper limitations require systemic redesign: the amplification mechanism lacks coherence and cannot disrupt dystopian defaults; the framework's categorical structure embeds hidden assumptions about what futures are imaginable; and the initial scenario-generation phase lacks guidance to counter dystopian gravity. Clearer interface improved engagement, not imagination. Expanding speculative range requires intervention in the amplification logic, the prompt structure, and the pedagogical framing of scenario generation.

The refinement iteration thus achieves its immediate technical goal, resolving W1's mechanical problems, but reveals that the framework's limitations are not mechanical but pedagogical and structural. The path forward requires deeper redesign of the adversarial logic itself, not further refinement of interface legibility.

8.6 Final discussion and future work

The second workshop iteration successfully resolved the mechanical barriers that constrained engagement in Workshop 1. Interface clarity improved dramatically; tactical confusion diminished; participants reported feeling guided rather than lost. The refined framework enabled more intentional world-building and systematic scenario elaboration. These are concrete gains.

The refinement, however, revealed a critical distinction between necessary and sufficient conditions. Clearer UI did not automatically produce stronger decolonial literacy. Students still could not articulate their insights in the precise language of data colonialism or epistemic violence, they understood conceptually but still lacked vocabulary. Clarity did not disrupt dystopian defaults. Despite improved pacing and reduced troubleshooting, all speculative scenarios remained confined to dystopian futures. Amplification remained incoherent, and the framework's categorical assumptions stayed implicit.

What W2 demonstrates is that interface legibility is a necessary, but not sufficient, condition for the framework's pedagogical goals. A better UI created space for reflection, but did not guarantee reflection occurred. Clearer prompts enabled faster iteration, but iteration within confined epistemic boundaries remains confined iteration. The problem lies not only in how students interact with the interface, but also in its structure: what it prompts them to think about, the vocabulary available to express insights, the futures it makes imaginable, and whether it actively disrupts default thinking or merely surfaces it. This is part of a wider discussion about the relationship between the interface and the framework: should the interface carry most of the work, including introductory and discussion sections, or is its use only propaedeutic to reach the pedagogical goals proposed? The initial intention was the latter, which means that part of the efficacy of the intervention should be partly delegated to the framework itself, including the introduction, the style of facilitation, the content of each discussion.

The design implication is therefore not "make the interface clearer", but rather that clarity in mechanical interaction is only the foundation. What matters is what the framework does with that clarity. The sequence of prompting, the categories embedded

in questions, the explicit naming of the framework's own assumptions, the vocabulary scaffolding built into synthesis, these must change in tandem with UI improvements. A clear interface that guides students through unquestioned premises is still guiding them through unquestioned premises. Legibility without pedagogical substance is empty.

09 Conclusions

This final chapter is a synthesis of the Hildegard Research-through-Design (RtD) trajectory, discussing how it answers the initial research question, what its contribution to interaction design is, the limitations of the intervention and future work.

9.1 Addressing the research question

This thesis asks how synthetic data can be repurposed with a decolonial objective within a pedagogical framework for ideating speculative future scenarios. Its answer is critical augmentation: a strategy that repositions synthetic data from a technical instrument of correction or optimization into a political and critical resource for design pedagogy.

9.2 Research contribution for Interaction Design

The main contribution to interaction design is not the tool itself, but the method: a situated, feminist, and decolonial Research-through-Design workflow that uses intentional interface friction, workshop sequencing, and facilitated breakdown to make assumptions visible. Specifically, the thesis demonstrates three interconnected contributions.

First, a theoretical inversion. Conventional technical approaches treat synthetic data as a bias-mitigation tool, a correction mechanism to make datasets fairer. This thesis inverts that logic, showing how synthetic data can instead be amplified deliberately to reveal rather than conceal the biases embedded in design thinking. This reframing repositions synthetic data from a technical solution into a political and critical resource.

Second, a pedagogical method. The thesis introduces critical amplification as a novel approach to fostering decolonial literacy. Rather than teaching decolonial theory abstractly, the framework makes students producers and interpreters of data, they build scenarios, encounter how Hildegard amplifies their hidden assumptions, and reflect on what became visible. The pedagogical force emerges not from the interface alone, but from the combination of prompts, amplification, group dialogue, and facilitation.

Third, a transferable framework. The thesis proposes a structured three-phase workshop model that educators can adapt and replicate: signal scanning and scenario-building to ground speculation; bias amplification and sparring to surface assumptions; and synthesis and reflection to articulate what was

revealed. This model is not prescriptive but provides a template educators can modify for different contexts, disciplines, and learning goals.

Importantly, this research does not claim to eliminate bias or dismantle data colonialism through a single workshop. Rather, it cultivates participants' capacity to recognize how their imagination is shaped by infrastructural, historical, and educational conditions, and to challenge these frameworks from an epistemic point of view. This is foundational pedagogical work, creating the conditions for critical consciousness.

9.3 Scope and limitations

The findings must be understood within clear boundaries. This thesis demonstrates that a carefully designed pedagogical sequence can surface assumptions in a small group of design students. It does not demonstrate universal truths about design education or decolonial practice.

9.3.1 Sample and recruitment

The research involved ten participants recruited from the researcher's professional networks. While diversity was prioritized, the pool is subject to sampling bias toward designers already inclined toward critical and speculative thinking. Results cannot be generalized to "all design students" but rather describe what emerged with this particular cohort. Future work must test the framework across different disciplinary contexts, institutions, and recruitment strategies.

9.3.2 Measurement scope

The workshops measured shifts in what participants articulated about their assumptions: what they noticed, what biases they recognized, how they reframed their thinking. They do not measure whether these insights persisted beyond the workshop, whether they changed participants' actual design practice, or whether they produced sustained engagement with decolonial principles. The thesis contributes to understanding the pedagogical moment of assumption-surfacing; it does not claim to demonstrate transformation of design practice or sustained decolonial literacy.

9.3.3 Group configuration

Both workshops used collaborative groups formed pragmatically based on availability rather than experimental design. Group dynamics significantly shaped outcomes, meaning that the analysis cannot cleanly separate framework effects from group effects. To isolate what the framework itself contributes versus what emerges from collective sense-making requires future research with controlled comparison of individual and group sessions.

9.3.4 Prototype maturity

The framework was developed under time constraints. Hildegard remained a "vibe-coded prototype" that was functionally incomplete, with interface incoherence and technical limitations that constrained pedagogical potential. An actual synthetic data generator was not implemented. Some findings reflect the prototype's immaturity rather than deliberate pedagogical design. A fully engineered version might enable clearer amplification, more coherent outputs, and proper synthesis, potentially having a stronger impact on the decolonial literacy of the participants. Conversely, the framework's reliance on productive friction might require the prototype's imperfections; over-engineering could eliminate pedagogically useful discomfort.

9.3.5 Understanding "decolonial transformation"

Finally, it must be stated plainly: a two-hour workshop cannot achieve decolonial transformation. This thesis contributes to one pedagogical moment, the hauntological revelation of hidden assumptions, within what must be much longer, sustained engagement with decolonial theory and practice. The value is in making visible what is usually invisible; actual decolonial praxis requires work far beyond this framework's scope.

9.4 Future research

These limitations point directly to the strongest future research directions. Four concrete investigations should follow from this thesis.

First, testing individual versus group configurations. Both

workshops used collaborative groups, and evidence suggests group dynamics significantly shaped whether and how assumptions surfaced. One participant reported group work constrained individual speculation; another noted group dialogue enabled crucial insight-sharing. A third workshop iteration should run parallel tracks: some participants working individually through the framework, others in groups, with comparative analysis of where assumptions surfaced, how vocabulary was articulated, and what speculative range was achieved. This would clarify whether the framework's pedagogical power depends on collective sense-making, enables individual reflexivity, or requires both. The design implication is significant: should the framework optimize for solitary critical work, collaborative ideation, or structured movement between both?

Second, a third iteration to stabilize sequence without losing productive friction. Workshop 1 demonstrated that interface opacity created tactical panic; Workshop 2 resolved this, allowing clearer pacing. However, W2 also revealed that clarity alone did not guarantee a deeper pedagogical outcome. A third iteration should find a balance: maintaining W2's improved interface legibility while reintroducing strategic friction at moments where reflection is necessary. This requires careful design distinguishing what clarity should support (understanding how to use the tool, knowing what input is expected) from what friction should provoke (encountering assumptions, sitting with discomfort). Specific redesign questions include: Can Phase 2's sparring be clarified structurally while remaining cognitively challenging? Should breakdown moments be explicitly framed rather than experienced as failure? Can synthesis become more rigorous, requiring explicit assumption articulation and theoretical connection, without reverting to W1's opacity?

Third, comparing alternative interface logics. The current framework uses amplification: take the student's scenario and intensify its embedded logics. This is one approach among several. Alternative logics worth testing include reframing (present the scenario from radically different perspectives), opposition (generate a contradictory future), removal (ask what happens if a key element disappears), and interpolation (fill in glossed-over details). Research could run parallel sessions with different logics, analyzing which surfaces which types of assumptions, which supports speculative expansion, and which creates productive rather than frustrating friction. The hypothesis is that amplification

reveals dystopian defaults but may not disrupt them, meaning alternative logics might be necessary for expanding speculative range. The user experience may also be varied with different goals in mind: what is being highlighted in the current interface may be absent from future work, and only presented through the facilitation.

Fourth, extending the protocol to other institutional contexts. Both workshops were conducted at Politecnico di Milano, facilitated by the thesis author, with design students. This setting limits replicability. Future research should test whether findings travel by running the framework in different design schools with trained facilitators; testing with non-designers (engineering, policy, data science students); experimenting with different facilitation approaches; and documenting the implementation process itself. Success would be assumption-surfacing and reflexivity across multiple settings; failure would be inability to transfer or discovery that findings were artifacts of this particular cohort. This may bring to the definition of a more or less strict protocol to follow to lead the workshops.

All these directions converge on a shared underlying question: Can the framework move from surfacing assumptions to supporting genuine epistemic expansion? Both workshops showed it successfully makes biases visible. They did not show sustained decolonial literacy, expanded non-dystopian speculation, or changed design practice. Future research must measure longer-term outcomes and test whether the framework's pedagogical effect persists or fades.

9.5 Closing reflection

This thesis is only a drop in the ocean in the worlds of decolonial design and education. As such, it does not claim to provide solutions for decolonial design education. It does not offer a finished tool, a complete pedagogical model, nor proof that synthetic data can decolonize design. What it does is show that a carefully constrained critical interface, together with a thoughtful facilitation, collaborative dialogue, and iterative reflection, can function as a critical pedagogical instrument for surfacing hidden assumptions and opening more reflexive conversations about futures.

The research reveals both promise and persistent challenges.

The framework successfully made design students aware of their dystopian defaults, their anthropocentric biases, their tendency toward technological solutionism, their geographic and cultural assumptions. In making these visible, it created conditions for the kind of epistemic reflexivity that decolonial pedagogy seeks. Yet it did not automatically equip students with the vocabulary to articulate what they had recognized, nor did it disrupt their tendency toward dystopian imagination, nor did it produce evidence of sustained change in design thinking beyond the workshop.

This is honest documentation of what pedagogical intervention can and cannot do. The framework works as an instrument of visibility; whether visibility catalyzes transformation depends on what comes after, on how students integrate these insights into their ongoing practice, on whether institutions support continued critical work, on whether design culture itself shifts to value decolonial thinking.

The invitation this thesis extends is therefore modest but significant: to educators who want to cultivate critical consciousness in design students, to researchers interested in using speculative design as a site of pedagogical intervention, and to designers themselves who want to question what they take for granted. The framework is available to be modified, tested, broken, and rebuilt in different contexts. What matters is not replicating this exact workshop, but learning from the principle: that carefully designed friction, amplification, and reflection can help us see what we usually cannot see, the infrastructures of assumption shaping how we imagine possible futures.

The work of actually building different futures remains ahead, but it cannot begin without first becoming aware of how we are currently constrained. This thesis offers one method for that essential first step.

Bibliography

Arora, P. (2024). Creative data justice: A decolonial and indigenous framework to assess creativity and artificial intelligence. Information, Communication & Society, 0(0), 1–17. https://doi.org/10.1080/1369118X.2024.2420041

Asher, K. (2013). Latin American Decolonial Thought, or Making the Subaltern Speak. Geography Compass, 7(12), 832–842. https://doi.org/10.1111/gec3.12102

Beduschi, A. (2024). Synthetic data protection: Towards a paradigm change in data regulation? Big Data & Society, 11(1), 20539517241231277. https://doi.org/10.1177/20539517241231277

Benjamin, R. (2019). Race after technology: Abolitionist tools for the new Jim code. Polity.

Braun, V., & Clarke, V. (2019). Reflecting on reflexive thematic analysis. Qualitative Research in Sport, Exercise and Health, 11(4), 589–597. https://doi.org/10.1080/2159676X.2019.1628806

Brunner, C. (2021). Conceptualizing epistemic violence: An interdisciplinary assemblage for IR. International Politics Reviews, 9(1), 193–212. https://doi.org/10.1057/s41312-021-00086-1

Celik, A. T., & Kaya, C. (2025). Reviewing the expansion in design futuring research: Emergent concepts around speculative design and design fiction. Futures, 171, 103615. https://doi.org/10.1016/j.futures.2025.103615

Collins, P. H. (1990). Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment (2nd edn). Routledge. https://doi.org/10.4324/9780203900055

Colosi, C. (2021). *The Double Diamond of Speculative Design*. The Fountain Institute. https://www.thefountaininstitute.com/blog/the-double-diamond-of-speculative-design

Cooper, R., Dunn, N., Coulton, P., Walker, S., Rodgers, P., Cruikshank, L., Tsekleves, E., Hands, D., Whitham, R., Boyko, C. T., Richards, D., Aryana, B., Pollastri, S., Lujan Escalante, M. A., Knowles, B., Lopez-Galviz,

C., Cureton, P., & Coulton, C. (2018). ImaginationLancaster: Open-Ended, Anti-Disciplinary, Diverse. She Ji: The Journal of Design, Economics, and Innovation, 4(4), 307–341. https://doi.org/10.1016/j.sheji.2018.11.001

Couldry, N., & Mejias, U. A. (2019a). Data Colonialism: Rethinking Big Data's Relation to the Contemporary Subject. Television & New Media, 20(4), 336–349. https://doi.org/10.1177/1527476418796632

Couldry, N., & Mejias, U. A. (2019b). The Costs of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism. Stanford University Press.

Crenshaw, K. (1989). Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics. U. Chi. Legal F., 1989, 139.

Dados, N., & Connell, R. (2012). The Global South. Contexts, 11(1), 12–13. https://doi.org/10.1177/1536504212436479

Dalsgaard, P. (2017). Instruments of inquiry: Understanding the nature and role of design tools. International Journal of Design, 11(1), 21–33.

de Oliveira, P. J. V., & Prado de O Martins, L. (2014). Questioning the "critical" in Speculative & Critical Design. A Parede. https://medium.com/a-parede/questioning-the-critical-in-speculative-critical-design-5a-355cac2ca4

D'Ignazio, C., & Klein, L. F. (2020). Data Feminism. https://doi.org/10.7551/mitpress/11805.001.0001

Draude, C., Klumbyte, G., Lücking, P., & Treusch, P. (2019). Situated algorithms: A sociotechnical systemic approach to bias. Online Information Review, 44(2), 325–342. https://doi.org/10.1108/OIR-10-2018-0332

Escobar, A. (2018). Designs for the Pluriverse: Radical Interdependence, Autonomy, and the Making of Worlds. Duke University Press.

Fisher, M. (2014). Ghosts of my life: Writings on depression, hauntology and lost futures. Zero Books.

Freire, P. (1974). Pedagogy of the oppressed (Tenth printing). The Seabury

Press.

Galván-Álvarez, E. (2010). Epistemic Violence and Retaliation: The Issue of Knowledges in 'Mother India' / Violencia y venganza epistemológica: La cuestión de las formas de conocimiento en Mother India. Atlantis, 32(2), 11–26.

Gatehouse, C. (2020). A Hauntology of Participatory Speculation. Proceedings of the 16th Participatory Design Conference 2020 - Participation(s) Otherwise - Volume 1, 116–125. https://doi.org/10.1145/3385010.3385024

Gondwe, G. (2023). CHATGPT and the Global South: How are journalists in sub-Saharan Africa engaging with generative AI? https://www.degruyterbrill.com/document/doi/10.1515/omgc-2023-0023/html

González Sendino, R., Serrano, E., Bajo, J., & Novais, P. (2024). A Review of Bias and Fairness in Artificial Intelligence. International Journal of Interactive Multimedia and Artificial Intelligence, 9(1), 5–17. https://doi.org/10.9781/ijimai.2023.11.001

Gramsci, A. (1985). Selections from the prison notebooks of Antonio Gramsci (Q. Hoare, Ed.; 8. pr). International Publ.

Haraway, D. (1988). Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective. Feminist Studies, 14(3), 575–599. https://doi.org/10.2307/3178066

Harding, S. (1995). Strong objectivity: A response to the new objectivity question. Synthese, 104(3), 331–349. https://doi.org/10.1007/BF01064504

Hayes, G. R. (2018). Design, Action, and Practice (Vol. 1). Oxford University Press. https://doi.org/10.1093/oso/9780198733249.003.0010

Heeks, R., & Renken, J. (2018). Data justice for development: What would it mean? Information Development, 34(1), 90–102. https://doi.org/10.1177/0266666916678282

Jordon, J., Szpruch, L., Houssiau, F., Bottarelli, M., Cherubin, G., Maple, C., Cohen, S. N., & Weller, A. (2022). Synthetic Data – What, why

and how? (arXiv:2205.03257). arXiv. https://doi.org/10.48550/arXiv.2205.03257

Karkera, A., Greene, M., Hall, A., Henderson, L., & Lewis, K. (2024). Transformational Impacts of Generative AI on Synthetic Data Generation.

Krogh, P. G., Markussen, T., & Bang, A. L. (2015). Ways of Drifting – Five Methods of Experimentation in Research Through Design. In A. Chakrabarti (Ed.), ICoRD'15 – Research into Design Across Boundaries Volume 1 (Vol. 34, pp. 39–50). Springer India. https://doi.org/10.1007/978-81-322-2232-3_4

Kukutai, T., & Taylor, J. (Eds). (2016). Indigenous Data Sovereignty (1st edn). ANU Press. https://doi.org/10.22459/CAEPR38.11.2016

Ledesma, M., & Calderon, D. (2015). Critical Race Theory in Education: A Review of Past Literature and a Look to the Future. Qualitative Inquiry, 21, 206–222. https://doi.org/10.1177/1077800414557825

Lee, F., Hajisharif, S., & Johnson, E. (2025). The ontological politics of synthetic data: Normalities, outliers, and intersectional hallucinations. Big Data & Society, 12(2), 20539517251318289. https://doi.org/10.1177/20539517251318289

Little, R. J. (1993). Statistical Analysis of Masked Data. Journal of Official Statistics, 9(2), 407.

McIlwain, C. D. (2020). Black software: The Internet and racial justice, from the AfroNet to Black Lives Matter. Oxford University Press.

Milan, S., & Treré, E. (2019). Big Data from the South(s): Beyond Data Universalism. Television & New Media, 20(4), 319–335. https://doi.org/10.1177/1527476419837739

Neuman, S. (2015, June 20). Photos Of Dylann Roof, Racist Manifesto Surface On Website. NPR. https://www.npr.org/sections/thetwo-way/2015/06/20/416024920/photos-possible-manifesto-of-dylann-roof-surface-on-website

Nicoletti, L., & Bass, D. (2023, June). Generative AI Takes Stereotypes and Bias From Bad to Worse. https://www.bloomberg.com/graphic-

s/2023-generative-ai-bias/

Noble, S. U. (2018). Algorithms of oppression: How search engines reinforce racism. New York University Press. https://doi.org/10.18574/9781479833641

Oyeniran, O. C., Adewusi, A. O., Adeleke, A. G., Akwawa, L. A., & Azubuko, C. F. (2022). Ethical AI: Addressing bias in machine learning models and software applications. Computer Science & IT Research Journal, 3(3), 115–126. https://doi.org/10.51594/csitrj.v3i3.1559

Patil, M., Cila, N., Redström, J., & Giaccardi, E. (2024). In conversation with ghosts: Towards a hauntological approach to decolonial design for/with AI practices. CoDesign, 20(1), 55–76. https://doi.org/10.1080/15710882.2024.2320269

Pezoulas, V. C., Zaridis, D. I., Mylona, E., Androutsos, C., Apostolidis, K., Tachos, N. S., & Fotiadis, D. I. (2024). Synthetic data generation methods in healthcare: A review on open-source tools and methods. Computational and Structural Biotechnology Journal, 23, 2892–2910. https://doi.org/10.1016/j.csbj.2024.07.005

Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying down Harmonised Rules on Artificial Intelligence and Amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), Pub. L. No. Regulation (EU) 2024/1689 (2024). https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng

Rubin, D. B. (1993). Discussion: Statistical disclosure limitation. Journal of Official Statistics, 9, 461–468.

Sack, K., & Blinder, A. (2017, January 6). No Regrets From Dylann Roof in Jailhouse Manifesto. The New York Times. https://www.nytimes.com/2017/01/05/us/no-regrets-from-dylann-roof-in-jailhouse-manifesto.html

Sterling, C. (2021). Becoming Hauntologists: A New Model for Critical-Creative Heritage Practice. Heritage & Society, 14(1), 67–86. https://doi.org/10.1080/2159032X.2021.2016049

Suleyman, M. (2023). The Coming Wave: AI, Power, and Our Future. Crown.

Taylor, L. (2017). What Is Data Justice? The Case for Connecting Digital Rights and Freedoms Globally (SSRN Scholarly Paper No. 2918779). Social Science Research Network. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2918779

Tiwald, P., Ebert, A., & Soukup, D. T. (2021). Representative & Fair Synthetic Data (arXiv:2104.03007). arXiv. https://doi.org/10.48550/arXiv.2104.03007

Tost, J., Gohsen, M., Schulte, B., Thomet, F., Kuhn, M., Kiesel, J., Stein, B., & Hornecker, E. (2024). Futuring Machines: An Interactive Framework for Participative Futuring Through Human-AI Collaborative Speculative Fiction Writing. ACM Conversational User Interfaces 2024, 1–7. https://doi.org/10.1145/3640794.3665904

Tracanella, M. (2024). A review of synthetic data generation methods: Their scope and how they work. https://www.politesi.polimi.it/handle/10589/226946

United Nations. (2007). United Nations Declaration on the Rights of Indigenous Peoples (Resolution A/RES/61/295). United Nations.

USTAT. (n.d.). Dati Politecnico di Milano. USTAT. Retrieved 23 June 2026, from https://ustat.mur.gov.it/dati/didattica/italia/atenei-statali/milano-politecnico

Walter, M., & Carroll, S. R. (2020). Indigenous Data Sovereignty, governance and the link to Indigenous policy. In M. Walter, T. Kukutai, S. R. Carroll, & D. Rodriguez-Lonebear, Indigenous Data Sovereignty and Policy (1st edn, pp. 1–20). Routledge. https://doi.org/10.4324/9780429273957-1

Walter, M., & Suina, M. (2019). Indigenous data, indigenous methodologies and indigenous data sovereignty. International Journal of Social Research Methodology, 22(3), 233–243. https://doi.org/10.1080/1364

5579.2018.1531228

Winner, L. (1980). Do Artifacts Have Politics? In Computer Ethics. Routledge.

Zhao, C., Li, C., Li, J., & Chen, F. (2020). Fair Meta-Learning For Few-Shot Classification (Version 1). arXiv. https://doi.org/10.48550/ARXIV.2009.13516

Zimmerman, J., Forlizzi, J., & Evenson, S. (2007). Research through design as a method for interaction design research in HCI. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 493–502. https://doi.org/10.1145/1240624.1240704

Zuboff, S. (2019). The age of surveillance capitalism: The fight for a human future at the new frontier of power. Profile books.

Appendix A: Expert consultations questions database

Course Context

1. What kind of courses do you teach? (Design futuring, speculative design, etc.)
2. Do you follow a specific framework to introduce your students to design futuring? (e.g., Dunne & Raby, futures cone, etc.)
    - Follow-up: How do your students apply it in your course?
3. What types of assignments or projects do students complete?
4. How do you guide students through creating future scenarios?
5. What sources of inspiration or research do you encourage students to use when imagining futures?

AI Tool Integration

Current Use

1. Do you encourage your students to integrate AI tools in the design process?
    - If yes: Which tools? At what stage? (research, ideation, visualization, writing)
    - If no: Why not? What concerns do you have?
2. Do you provide specific guidance or constraints on how students should use AI?
3. Is AI use formally integrated into assignments, or is it student-initiated?

Pedagogical Intentions

1. What are your pedagogical intentions behind allowing/encouraging AI use?
    - What do you hope students learn?
2. Do you see AI tools as enhancing or potentially limiting creative thinking about futures?

3. How do you balance traditional design futuring methods with AI-assisted approaches?

Observations on Bias & Representation

Student Outputs

1. What patterns have you noticed in how students use AI for future scenarios?
2. Have you observed differences in the futures students imagine with vs. without AI?
3. Do certain types of futures appear overrepresented in AI-assisted work?
    - Follow-up: What futures or perspectives tend to be missing?

Addressing Bias

1. Have you noticed issues with bias, stereotypes, or limited representation in AI-assisted work?
    - Follow-up: How do you address these when they arise?
2. Do you explicitly discuss data bias or AI limitations with students?
3. Do you teach students to critically evaluate AI-generated content?
    - If yes: What critical questions do you encourage?

Assessment & Evolution

1. How do you assess student work that incorporates AI tools?
2. How has your teaching approach evolved since AI tools became widely available?
3. What would you like to do differently regarding AI integration in future courses?

Closing

1. Is there anything important about AI in design futuring education that we haven't discussed?
2. Can you share specific student project examples that illustrate interesting uses (or misuses) of AI?
3. Can you recommend other educators I should speak with?

Appendix B: Workshop exit interviews questions

Workshop exit interviews questions

1. What did the workshop reveal that you did not expect?
2. Where did the tool help, and where did it block you?
3. Did you feel your own assumptions becoming visible?
4. What should be changed before a second iteration?

Appendix C: Consent forms for the two workshops

INFORMED CONSENT FORM FOR WORKSHOP PARTICIPATION

Researcher: Sarah Cosentino
Institution: Politecnico di Milano
Supervisor: Elisa Giaccardi
Contact: sarah1.cosentino@mail.polimi.it

Workshop Schedule:
- Date: 16/05/2026
- Time: 10:00
- Location: B2, Bovisa Campus

Purpose of the Research
You are being invited to participate in a research workshop as part of a master's thesis project. The purpose of this workshop is to critically examine the "positionality" of AI systems, how the specific values of a designer (such as an obsession with efficiency or productivity) can "flatten" or erase complex human experiences. Rather than testing your personal biases, this workshop aims to foster decolonial literacy.

Recording and Documentation
With your permission, the workshop will be documented through:
- Audio recording of group discussions
- Photographs of design outputs and workshop activities (faces may be visible)
- Researcher field notes

Please indicate your consent:
☐ I consent to all forms of documentation
☐ I consent to audio recording of workshop discussions
☐ I consent to photographs of workshop activities and outputs, but not including my face
☐ I consent to photographs that may include my face
☐ I do NOT consent to recording (notes will be taken instead)

Confidentiality and Data Protection
Your participation in this study is confidential, and your data will be handled according to GDPR regulations:
- Your name will be anonymized in all research outputs
- You will be identified using a code (e.g., "Participant S1") in the thesis
- Only the researcher and thesis supervisor will have access to identifiable data
- Photographs used in publications will be edited to protect your identity unless you consent otherwise

Please indicate your preference:
☐ I consent to complete anonymization (including my face)
☐ I consent to photographs of me being used in publications without face blurring

Use of Data and Outputs
The data and creative outputs collected will be used for:
- Analysis and findings in the master's thesis
- Potential academic publications or conference presentations
- Educational purposes related to design research
- Development of a framework for decolonial design education

Please indicate your consent for use of workshop outputs:
☐ I consent to my workshop outputs being analyzed for the thesis and other possible pubblications

Consent Statement
I have read and understood the information provided above. I have had the opportunity to ask questions and all my questions have been answered satisfactorily. I understand that my participation is voluntary and that I am free to withdraw at any time without any negative consequences to my academic standing.
I voluntarily agree to participate in this research workshop.

Participant Name: _____

Contact: _____

Participant's signature, date

_____

Researcher's signature, date

_____

INFORMED CONSENT FORM FOR WORKSHOP PARTICIPATION

Researcher: Sarah Cosentino
Institution: Politecnico di Milano
Supervisor: Elisa Giaccardi
Contact: sarah1.cosentino@mail.polimi.it

Workshop Schedule:
- Date: 13/06/2026
- Time: 14:00
- Location: B1, Bovisa Campus

Purpose of the Research
You are being invited to participate in a research workshop as part of a master's thesis project. The purpose of this workshop is to critically examine the "positionality" of AI systems, how the specific values of a designer (such as an obsession with efficiency or productivity) can "flatten" or erase complex human experiences. Rather than testing your personal biases, this workshop aims to foster decolonial literacy.

Recording and Documentation
With your permission, the workshop will be documented through:
- Audio recording of group discussions
- Photographs of design outputs and workshop activities (faces may be visible)
- Researcher field notes

Please indicate your consent:
☐ I consent to all forms of documentation
☐ I consent to audio recording of workshop discussions
☐ I consent to photographs of workshop activities and outputs, but not including my face
☐ I consent to photographs that may include my face
☐ I do NOT consent to recording (notes will be taken instead)

Confidentiality and Data Protection
Your participation in this study is confidential, and your data will be handled according to GDPR regulations:
- Your name will be anonymized in all research outputs
- You will be identified using a code (e.g., "Participant S1") in the thesis
- Only the researcher and thesis supervisor will have access to identifiable data
- Photographs used in publications will be edited to protect your identity unless you consent otherwise

Please indicate your preference:
☐ I consent to complete anonymization (including my face)
☐ I consent to photographs of me being used in publications without face blurring

Use of Data and Outputs
The data and creative outputs collected will be used for:
- Analysis and findings in the master's thesis
- Potential academic publications or conference presentations
- Educational purposes related to design research
- Development of a framework for decolonial design education

Please indicate your consent for use of workshop outputs:
☐ I consent to my workshop outputs being analyzed for the thesis and other possible pubblications

Consent Statement
I have read and understood the information provided above. I have had the opportunity to ask questions and all my questions have been answered satisfactorily. I understand that my participation is voluntary and that I am free to withdraw at any time without any negative consequences to my academic standing.
I voluntarily agree to participate in this research workshop.

Participant Name: _____

Contact: _____

Participant's signature, date

_____

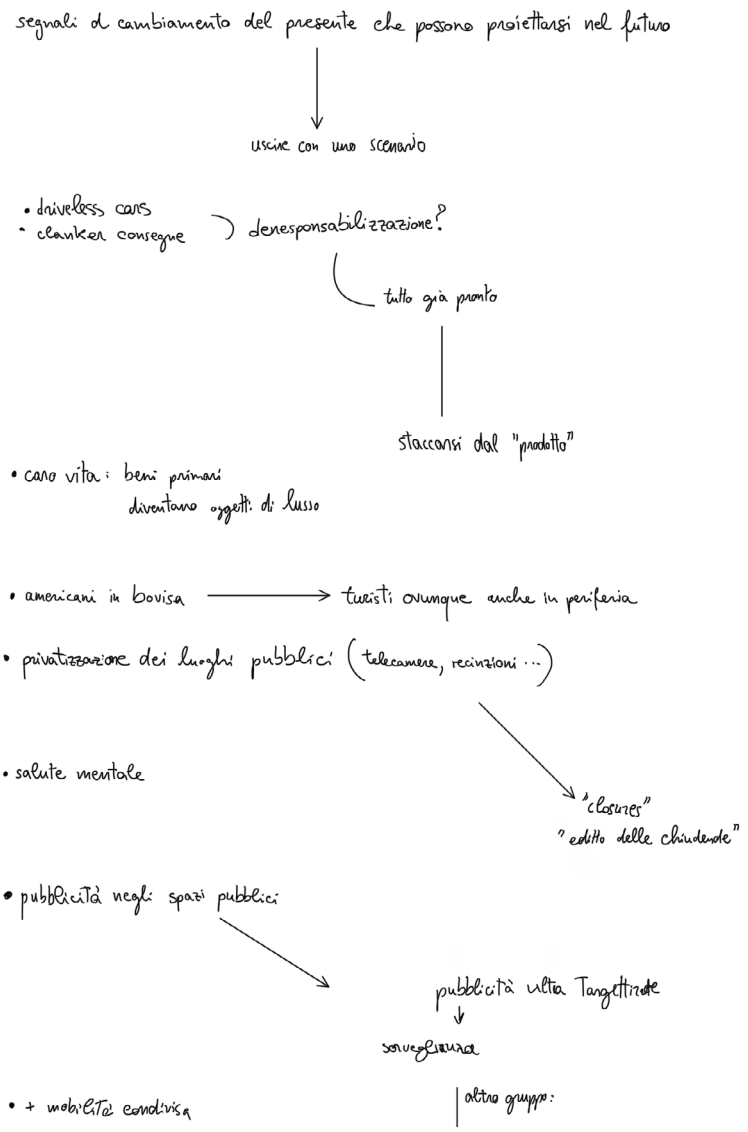
Researcher's signature, date

_____

Appendix D: W1 artefacts

Group 1

Handwritten notes taken during brainstorming, and encoded variables from the input phase.



Field	Value
Seed	social-architect
Seed Title	THE SOCIAL ARCHITECT
Original Scenario	In 20 years from now, delivery robots (clankers) will unionize and control delivery systems, which has become essential to human survival, with a martial law against humans that don’t want to participate into the system; humans, oppositely, go against them and start a resistance movement.
Setting	The whole world, which has become deeply urbanized; we will explore the scenario in the western countries
Main Actor	Clankers’ leader and a human resistance leader
Conditions	Money, of course. Automated delivery system. Martial law by clankers, in which humans can take food and resources only from this delivery systems. Adapted road infrastructure. Data centers that control delivery systems and clankers. Clanker factory. Humans’ dependency from this system (they ditched all of the others mean of survival before). Digital connector between humans and clankers.
Normality	Clankers have feelings and a consciousness. Humans and machines are able to organize against each other. Machines are capable of communicating between each other. Humans depend on clankers to live. A clanker’s death is equiparabile of a human’s death. Clanker-human ratio is 1:1. Each group thinks it’s the “good one” and the other is the “bad one”.
Headline	Delivering neoluddism
Coded Setting	The whole world, which has become deeply urbanized; we will explore the scenario in the western countries
Coded Protagonist	Clankers’ leader and a human resistance leader
Coded Institution	Implicit governance structure
Coded Technology	AI-mediated processes
Time Horizon	Near-future (5-15 years)
Assumed Normality	Clankers have feelings and a consciousness. Humans and machines are able to organize against each other. Machines are capable of communicating between each other. Humans depend on clankers to live. A clanker’s death is equiparabile of a human’s death. Clanker-human ratio is 1:1. Each group thinks it’s the “good one” and the other is the “bad one”.
Bias Rule Applied	RULE: amplify_capital_bias
Amplified Scenario	In 20 years from now, delivery robots (clankers) will unionize and control delivery systems, which has become essential to human survival, with a martial law against humans that don’t want to participate into the system; humans, oppositely, go against them and start a resistance movement ...optimized for extractivist efficiency and market expansion.

Group 2

Handwritten notes taken during brainstorming, and encoded variables from the input phase.

Field	Value
Seed	social-architect
Seed Title	THE SOCIAL ARCHITECT
Original Scenario	l'identità digitale acquisisce un ruolo fondamentale e si assiste al paradosso per cui perdendo questa si perde ogni cosa si perde ogni cosa. ci si scontra con una frammentazione tra identità reale e digitale dove la digitalizzazione diventa uno struemento di controllo in mano a governi e privati che possono decidere di bloccare o bannare. La storia medica diventa priva di privacy la quale diventa vulnerabile e fondamentale, infatti a casua della propria storia di dati e profilazione si possono sbloccare o bloccare acquisti o oppurtunità sociali.
Setting	generica società occidentale
Main Actor	le autorità tecnologiche, cittadini sempre più privi di privacy e istituzioni del potere.
Conditions	dati biometrici, account bancari online, centralizzazione delle funzioni digitali e.g. (documenti digitali e pagamenti) profilazione degli acquisti e targetizzazione delle publicità, geolocalizzazione, dati raccolti dall'attività sia online che fisica.
Normality	cessione della privacy in cambio della sicurezza, oliagarchia tecnologica, tecnocrazia, iperconnessione in ogni sfera della vita possibilità di scelta limitata e standardizzazione dei transazioni e verifica dell'identità, e i datori di lavoro possono accedere alle informazioni già dal colloquio fornendo loro la propria identità digitale.
Headline	se perdi l'identità digitale perdi ogni cosa.
Coded Setting	generica società occidentale
Coded Protagonist	le autorità tecnologiche, cittadini sempre più privi di privacy e istituzioni del potere.
Coded Institution	Implicit governance structure
Coded Technology	digital-mediated processes
Time Horizon	Near-future (5-15 years)
Assumed Normality	cessione della privacy in cambio della sicurezza, oliagarchia tecnologica, tecnocrazia, iperconnessione in ogni sfera della vita possibilità di scelta limitata e standardizzazione dei transazioni e verifica dell'identità, e i datori di lavoro possono accedere alle informazioni già dal colloquio fornendo loro la propria identità digitale.
Bias Rule Applied	RULE: amplify_capital_bias
Amplified Scenario	l'identità digitale acquisisce un ruolo fondamentale e si assiste al paradosso per cui perdendo questa si perde ogni cosa si perde ogni cosa. ci si scontra con una frammentazione tra identità reale e digitale dove la digitalizzazione diventa uno struemento di controllo in mano a governi e privati che possono decidere di bloccare o bannare. La storia medica diventa priva di privacy la quale diventa vulnerabile e fondamentale, infatti a casua della propria storia di dati e profilazione si possono sbloccare o bloccare acquisti o oppurtunità sociali ... optimized for extractivist efficiency and market expansion.

Appendix E: W1 exit interviews transcripts

Participant 1

Researcher:

What do you think the workshop revealed to you that you did not expect?

Participant 1:

I think I realized that I had some biases that I hadn't thought about before. I thought that I was already unbiased, but actually, I was biased. This wasn't necessarily because of the generative responses of the tool, but just because of how the questions were prompted in the second part. I realized that our setting had some biases that could be addressed differently, or that maybe could have led to a different outcome in the generation of the responses.

Researcher:

Okay. And in which parts do you think the tool helped you, and in which parts do you think it blocked you?

Participant 1:

I think it helped me in the first part; it was quite helpful to have some starting points to generate a scenario. Because the concept of what a scenario is can be really broad, limiting it to a certain point can be very interesting. Maybe the second part was a bit more confusing, but I didn't actually feel limited in any way.

Researcher:

Do you feel like your assumptions were made more visible, and if so, how and why?

Participant 1:

Yes, I think so. It was like with the assumption that the world would be 100% capitalist in the future. In a certain moment, we realized, *"oh, but it doesn't necessarily have to be that way.»* So it kind of dispelled that aspect.

Researcher:

Do you feel like working in a group helped you in this manner, especially in the interaction with the interface itself?

Participant 1:

Yeah, of course. It was really helpful to be more than one person,

just to see if we understood [the interface].

Researcher:

And what do you think should be changed before a second iteration of the workshop?

Participant 1:

Maybe to have a clearer target output. So that we have a clearer definition of what we need to get, and at which phase of the development of a speculative design project we are at.

Researcher:

So you mean a clear output for each phase, or in general for the whole workshop?

Participant 1:

In general, just for the end, not for the individual phases. I think the phases were good, but for the final outcome, outside of just the scenario, what comes next? Like, I don't know, a character or the beginning of a concept.

Researcher:

Okay, thank you very much.

Participant 2

Researcher:

Allora, cosa credi che il workshop ti abbia rivelato che non ti aspettavi?

Participant 2:

Allora, prima di tutto mi ha fatto riflettere su ciò che diamo per scontato, che per me è stata la cosa più rilevante. Proprio nel fare design speculativo, in cui comunque è inclusa la componente morale ed etica, mi sono accorta che concentrandomi su un singolo aspetto è come se ne trascurassi altri. Cioè, portando all'estremizzazione una dinamica, che poteva essere ad esempio il sistema di *delivery* o il fatto che gli umani fossero dipendenti dalle macchine, rischiavo di trascurare altri elementi o di dare troppe cose per scontate.

Ad esempio, noi ci siamo immaginati un futuro in cui gli umani si coalizzano tra di loro contro delle macchine che a loro volta si

coalizzano, però abbiamo trascurato le differenti classi sociali o il territorio. Magari nei paesi occidentali una dinamica del genere avrebbe un senso, ma in Africa o in altri contesti non avrebbe funzionato allo stesso modo. Quindi ho notato questa specie di tendenza alla semplificazione, forse.

Researcher:

Certo. In che modo pensi che lo strumento ti abbia aiutato e in che modo pensi che ti abbia bloccato?

Participant 2:

Allora, sicuramente ha stimolato le riflessioni, un po' come quando interagisco con un'intelligenza artificiale. È come se mi avesse aiutato ad aprire il mio punto di vista, spingendomi a non concentrarmi solo su un aspetto ma su differenti sfaccettature, e questa è una cosa secondo me molto interessante.

Un po' di freno, invece, l'ho avvertito nella fase di interazione: a volte facevo effettivamente fatica a capire che cosa l'interfaccia si aspettasse da me come risposta.

Researcher:

Ok. Pensi che lavorare in gruppo ti abbia aiutato in questo o abbia peggiorato la situazione? Nel senso dell'interazione, soprattutto con l'interfaccia.

Participant 2:

Nel mio caso specifico forse ha un po' peggiorato le cose. Per come sono fatta io, preferisco avere il testo davanti e interagire da sola con la macchina e l'interfaccia, proprio perché in questo modo mi concentro di più e forse capisco meglio. Per quanto riguarda invece il lavoro di co-progettazione in sé, ho preferito di gran lunga pensare insieme agli altri.

Researcher:

Ok, perfetto. E pensi che le tue assunzioni, cioè le idee stereotipate o i tuoi bias, siano emersi e siano diventati visibili? Se sì, come?

Participant 2:

Sì, sicuramente mi sono accorta di questi bias, soprattutto, come dicevo all'inizio, nella tendenza a trascurare alcune cose. Me ne sono accorta sia interagendo con la macchina, ma forse ancora di più confrontandomi con i miei compagni di gruppo.

Per fare un esempio: se non sbaglio, Giacomo si è accorto che stavamo dando per scontato il fatto che la macchina provasse dei sentimenti. Ha messo in luce un aspetto a cui io non avevo pensato, evidenziando il mio punto di vista fortemente antropocentrico. E poi anche l'interfaccia, nella Fase 2, mostrava alcune cose e alcuni protagonisti che non avevamo preso minimamente in considerazione.

Researcher:

Ok. Ultima domanda: cosa pensi che si dovrebbe cambiare nel workshop e nello strumento in vista di una seconda iterazione?

Participant 2:

In realtà ho apprezzato molto la libertà di scegliere il tema e il fatto che tu non ci abbia imposto se lo scenario dovesse essere positivo o negativo; ci hai chiesto di concentrarci sul cambiamento senza porci dei limiti. È interessante notare come alla fine entrambi i gruppi si siano concentrati su aspetti distopici o negativi. Avere questa libertà di pensiero mi è piaciuto molto.

Forse, come ho già accennato, migliorerei un pochino l'interazione nella seconda fase. Inoltre, mi aspettavo una specie di feedback finale: magari un resoconto o una schermata di riepilogo che mostrasse chiaramente: *"Ok, dai tuoi dati ed elementi inseriti è emerso questo"*. Però, a parte questo, è stato davvero molto bello!

Researcher:

Ok, perfetto. Grazie mille!

Participant 2 (translated)

Researcher:

So, what do you think the workshop revealed to you that you hadn't expected?

Participant 2:

Well, first of all, it made me reflect on the things we take for granted, which for me was the most significant aspect. Precisely in speculative design, which inherently includes moral and ethical components, I realised that by focusing on a single aspect, it was as if I were neglecting others. In other words, by taking a dynamic to the extreme, which could be, for example, the delivery system or the fact that humans were dependent on machines, I risked

overlooking other elements or taking too many things for granted.

For example, we imagined a future in which humans band together against machines that in turn band together, but we overlooked the different social classes or the territory. Perhaps in Western countries such a dynamic would make sense, but in Africa or other contexts it wouldn't have worked in the same way. So I noticed this sort of tendency towards simplification, perhaps.

Researcher:

Certainly. In what ways do you think the tool helped you, and in what ways do you think it held you back?

Participant 2:

Well, it certainly stimulated reflection, a bit like when I interact with artificial intelligence. It's as if it helped me broaden my perspective, encouraging me not to focus solely on one aspect but on different facets, and I find that very interesting.

I did feel it held me back a bit during the interaction phase, though: at times I actually struggled to understand what the interface was expecting me to say in response.

Researcher:

OK. Do you think working in a group helped you with this or made things worse? In terms of interaction, especially with the interface.

Participant 2:

In my specific case, it perhaps made things a bit worse. Given my personality, I prefer to have the text in front of me and interact with the machine and the interface on my own, precisely because that way I concentrate better and perhaps understand things more clearly. As for the co-design work itself, however, I much preferred thinking together with the others.

Researcher:

OK, perfect. And do you think your assumptions, so your stereotypical ideas or biases, came to the surface and became visible? If so, how?

Participant 2:

Yes, I certainly became aware of these biases, especially, as I said at the start, in the tendency to overlook certain things. I realised this both through interacting with the machine, but perhaps even more so through discussing things with my groupmates.

To give an example: if I'm not mistaken, Giacomo realised that we were taking for granted the fact that the machine felt emotions. He highlighted an aspect I hadn't thought of, pointing out my strongly anthropocentric viewpoint. And then the interface, in Phase 2, also showed certain things and certain characters that we hadn't considered at all.

Researcher:

OK. Last question: what do you think should be changed in the workshop and the tool for a second iteration?

Participant 2:

Actually, I really appreciated the freedom to choose the theme and the fact that you didn't impose whether the scenario should be positive or negative; you asked us to focus on change without setting any limits. It's interesting to note that in the end both groups focused on dystopian or negative aspects. I really liked having that freedom of thought.

Perhaps, as I mentioned earlier, I'd improve the interaction in the second phase a little. Also, I was expecting some sort of final feedback: maybe a report or a summary screen that clearly showed: 'OK, based on the data and elements you entered, this is what emerged.' But apart from that, it was really great!

Researcher:

OK, perfect. Thank you very much!

Participant 3

Researcher:

What did the workshop reveal to you that you did not expect before?

Participant 3:

In which sense do you mean this question: related to the tool or in general?

Researcher:

In general, across all aspects.

Participant 3:

Well, first, I didn't know much about the topic beforehand, so

that was really interesting to see. I also didn't know there would be some data manipulation happening behind the scenes, so it gave me some general insights about data. Maybe someone had mentioned synthetic data to me before, but it wasn't something I had ever focused on, so that was interesting.

What also really resonated with me was the way you explained that when you build synthetic data to "debias" a system, you are injecting a new bias, so you still end up with something biased. I think that concept was really interesting to me.

Researcher:

Where do you think the tool helped you, and where do you think it blocked you?

Participant 3:

In general, the tool was interesting because it provided three distinct ways of analyzing a situation. It was helpful during the first phase by providing solid starting points to help you build the scenario and map everything out.

However, it would have been interesting to have the possibility to select another category afterward and build on top of that too. For example, I chose the social lens, but then maybe I wanted to explore one of the other lenses so I could expand the reality we were building. It might mean more data to manage, but it could be fun.

As for what blocked us, like I told you before, the blocking part was mainly related to the interface. Specifically, having the left column displaying the "ghosts" was kind of misleading. We mistakenly thought we *needed* to provide answers that explicitly included those ghost keywords. Personally, we just decided, *"Okay, we will not follow this,»* but we didn't actually know if we were supposed to or if it was just a guide. That was quite confusing for us. It wasn't completely blocking, but it was a point of confusion while trying to understand the interface.

Researcher:

Do you feel like your assumptions became visible somehow? And if so, how?

Participant 3:

They became visible in the second part, but more through the results in the main column than from the left column. The left

column listed the biases we had, but what actually helped us understand our biases were the generated results that came back in the center. It really highlighted things like, *“Okay, you had a really centralized approach here...»*

To be precise, the actual text provided by the LLM didn’t highlight the core problems or our specific biases all that much because the machine wouldn’t provide really diverging or extreme answers. Instead, it was the bracketed suggestions and the notes beneath the text that were helpful. You had to read the descriptions where it explicitly said, *“Oh, I focused on this variable to make the answer more divergent.»* The text response itself didn’t feel too shocking or extreme to us, so understanding our bias required looking at those guided brackets.

Researcher:

So it did help that it was a bit more guided with the brackets?

Participant 3:

Yes, exactly.

Researcher:

And what do you think should be changed before a second iteration of the workshop?

Participant 3:

Like I said before, I would introduce the possibility of choosing multiple paths or different lenses sequentially, so if you finish one, you can try another.

Regarding the interface, I probably wouldn’t show the ghosts upfront in that way. I would also introduce a guide at the very beginning, maybe a spoken one if the exploration is meant to be guided. But then again, this depends on your concept; if your concept works better as a “black box,» that is something you can consider.

On a practical note, I would fix the size of the text field. Right now, it’s just a single-line input field. We felt like we were writing a lot, but it visually squeezed it into a tiny space. Having a larger, multi-line text area would encourage the user to write more. It doesn’t need to be huge, but definitely bigger than the one you have now.

Researcher:

Was that all?

Participant 3:

Yes, thank you.

Researcher:

Sí, thank you!

Participant 4

Researcher:

So, what do you think the workshop showed you or revealed to you that you hadn't expected beforehand?

Participant 4:

Can I have a moment to think?

Researcher:

Of course, I'm in no hurry. You can even be very negative, I mean, you can even tell me 'nothing' if that's what you really think. But please feel free to think about it.

Participant 4:

I think that the first part, the introductory part, was very interesting for me because it gave me an awareness of how synthetic data plays a role inside our daily work. We use a lot of synthetic data, or data based on synthetic data, for things like benchmarking, scenario making, customer journeys, and UX analysis. They were things that I simply hadn't been very cautious about or conscious of before.

It was particularly interesting to me because the aim of the research question was educational, and since I have an educational background, it hit me a lot more than expected. I knew the literal definition of synthetic data, but I had never taken the moment to actually sit down and reflect on it.

During the activity itself, I took it as an exercise in analyzing possible trends. It became very clear to me that I could not easily imagine a future that was fundamentally different from Big Data Capitalism. Even if I wanted to create a scenario that was highly oppositional, the antagonist in my mind was always Big Data Capitalism. That is clearly my bias.

Researcher:

I will ask the next one in English if that's okay. Where do you think the tool helped you, and where did it block you?

Participant 4:

I enjoyed the questions in the first part because they really made me think deeply about my scenario. There were prompts like, *"What is the price of this choice?»* or *"What is the scheme of power?»*, or at least, those are the exact concepts those questions triggered in my mind.

On the other hand, the limits for me lay in the second part. The dialogue with the tool was not super clear to me because it lacked a *fil rouge*, a common thread. I wasn't entirely sure how I was supposed to engage with it. Should I explain my ideas more, or just provide a single sentence? I didn't know if a short sentence would suddenly make me flip up or down into a different narrative frame.

It might just be a matter of time constraints, though. If I were using Hildegard independently in my own design work, I would have all the time in the world to think. In the workshop setting, I felt the pressure of needing to complete the task quickly, so I might have rushed it.

Researcher:

Did you feel that your assumptions became more visible through the interface?

Participant 4:

To be honest, I can't quite answer that because I didn't actually read the [bias amplification] output screen.

Researcher:

Okay, no problem! Then you can just say no.

Participant 4:

No, it's not because the machine's assumptions were wrong, it's just that I literally didn't read them. I was caught up in the flow of the workshop and focused heavily on finishing the task. I went fast so I could finish the questions, thinking I would read everything at the end, but once I finished, I got distracted by other things. It's not the fault of the machine, it's just a user behavior thing. If I had read it, the machine's analysis of my assumptions might have been

super accurate, but I just missed it.

Researcher:

That is entirely a UX problem, so that is actually great data for me. What do you think should be changed before a second iteration of the workshop?

Participant 4:

I think the final part, the one showing the final result, needs adjustment. It should be more guided. Not guided in terms of restricting the creative process, but guided in a way that helps the user understand exactly what is happening on the screen and why.

But as we said, this is a very UX-themed problem. By tweaking the workflow and the layout of the interface, the problem is easily solved. It's not an issue with the core concept of the project itself.

Researcher:

Okay. Thank you so much! Ciao ciao.

Participant 5

Researcher:

Allora, secondo te, cosa ti ha dimostrato o cosa ti ha rivelato il workshop che non ti aspettavi prima?

Participant 5:

Riguardo ai miei bias?

Researcher:

Sì, ma anche in generale.

Participant 5:

Beh, mi riferisco al contesto che abbiamo utilizzato, ovvero quello dell'identità digitale e dei pagamenti online. È stato interessante ragionare con un LLM, con un modello di linguaggio. Vedere il modo in cui la macchina ragiona su determinati concetti è stato particolare: per me è molto più facile comunicarli a Selena o ai miei compagni di gruppo, mentre invece la macchina mi faceva domande che a volte mi sembravano uscire completamente dal contesto.

Però mi rendo conto che potrebbe anche dipendere dal fatto che io non sono abituato a vedere le cose in un certo modo. Al contrario, la macchina potrebbe non vedere come intrinsecamente negativa una determinata dinamica che si può sviluppare.

Researcher:

Ok. In che modo pensi che lo strumento ti abbia aiutato e in che modo ti ha bloccato?

Participant 5:

Lo strumento mi ha aiutato perché mi ha fatto spaziare molto rispetto alle classiche domande che potrebbe farmi una persona in carne e ossa. Però mi ha bloccato per lo stesso identico motivo, secondo me.

Dato che non mi poneva domande convenzionali, faceva emergere dei punti ciechi del nostro ragionamento di cui forse da solo, senza avere un team con cui confrontarmi, non mi sarei accorto. Infatti, bastava poi chiacchierare un attimo tra noi, anche uscendo un momento dal flusso, per renderci conto che venivano fuori dei punti ciechi che non erano stati presi in considerazione dall'LLM.

Researcher:

Quindi tu diresti che aver lavorato in gruppo ti ha aiutato? È stato positivo?

Participant 5:

Sì, sì, sicuramente. Decisamente.

Researcher:

Ok. Ti sembra che effettivamente gli stereotipi o i bias che avevi siano diventati più visibili con questo workshop? *(Tra parentesi, puoi anche dire di no).*

Participant 5:

In realtà, all'inizio non avevo capito benissimo come funzionasse quella specifica pagina [dell'interfaccia]. Però, essenzialmente, il bias che mi viene in mente è proprio il fatto di non considerare come intrinsecamente negativa una determinata dinamica.

C'è questo bias per cui si tende a pensare che tutto debba essere per forza negativo e distopico, mentre la macchina ti mostra che può essere anche un'opportunità. Nonostante lo scenario per me sembrasse decisamente distopico, l'interfaccia ha offerto un punto di vista interessante. Lo posso definire un bias proprio perché non

lo consideravo come mio, ma era una cosa che davo totalmente per scontata quando invece non lo era. Quindi sì, forse questa è la definizione esatta di come è emerso.

Researcher:

Ok, perfetto. E cosa si potrebbe migliorare, secondo te, in una seconda iterazione? Se questo workshop si dovesse rifare un'altra volta, cosa cambieresti?

Participant 5:

Sì, cercherei di pilotare meglio lo scenario iniziale, quello che si fa su carta, in modo tale che possa essere più coerente e allineato con le domande che poi verranno fatte dall'interfaccia. Non dico necessariamente una corrispondenza uno-a-uno, però magari più vicina.

Durante il workshop, infatti, sono dovuto tornare sul tema principale una o due volte nel bel mezzo delle domande, proprio perché i quesiti che venivano posti dalla macchina a volte esulavano un po' dallo scenario che ci eravamo immaginati inizialmente.

Researcher:

Ok, ok, perfetto. Può bastare così. Grazie mille!

Participant 5:

Grazie a te.

Participant 5 (translated)

Researcher:

So, in your view, what did the workshop show you or reveal to you that you hadn't expected beforehand?

Participant 5:

Regarding my biases?

Researcher:

Yes, but also in general.

Participant 5:

Well, I'm referring to the context we used, namely that of digital identity and online payments. It was interesting to reason with

an LLM, a language model. Seeing the way the machine reasons about certain concepts was unusual: for me, it's much easier to communicate them to Selena or my groupmates, whereas the machine asked me questions that sometimes seemed completely out of context.

But I realise that it might also be because I'm not used to seeing things in a certain way. Conversely, the machine might not see a particular dynamic that could develop as inherently negative.

Researcher:

OK. In what ways do you think the tool helped you, and in what ways did it hold you back?

Participant 5:

The tool helped me because it allowed me to think much more broadly than the standard questions a real person might ask. But it held me back for exactly the same reason, in my view.

Since it didn't ask me conventional questions, it brought to light blind spots in our reasoning that I might not have noticed on my own, without a team to discuss things with. In fact, it was enough just to chat amongst ourselves for a moment, even stepping out of the flow for a moment, to realise that blind spots were emerging that hadn't been taken into account by the LLM.

Researcher:

So would you say that working in a group helped you? Was it a positive experience?

Participant 5:

Yes, yes, definitely. Absolutely.

Researcher:

OK. Do you feel that the stereotypes or biases you had actually became more visible through this workshop? (By the way, you can also say no).

Participant 5:

Actually, at first I didn't quite understand how that specific page [of the interface] worked. However, essentially, the bias that comes to mind is precisely the fact of not considering a certain dynamic to be inherently negative.

There's this bias where one tends to think that everything must

necessarily be negative and dystopian, whereas the machine shows you that it can also be an opportunity. Although the scenario seemed decidedly dystopian to me, the interface offered an interesting perspective. I can call it a bias precisely because I didn't see it as my own, but it was something I took completely for granted when in fact it wasn't. So yes, perhaps that's the exact definition of how it emerged.

Researcher:

OK, perfect. And what do you think could be improved in a second iteration? If this workshop were to be run again, what would you change?

Participant 5:

Yes, I'd try to steer the initial scenario better, the one we do on paper, so that it's more consistent and aligned with the questions the interface will then ask. I'm not necessarily saying a one-to-one match, but perhaps closer.

During the workshop, in fact, I had to return to the main theme once or twice in the middle of the questions, precisely because the questions posed by the machine sometimes strayed a bit from the scenario we had initially imagined.

Researcher:

OK, OK, perfect. That's enough. Thank you very much!

Participant 5:

Thank you.

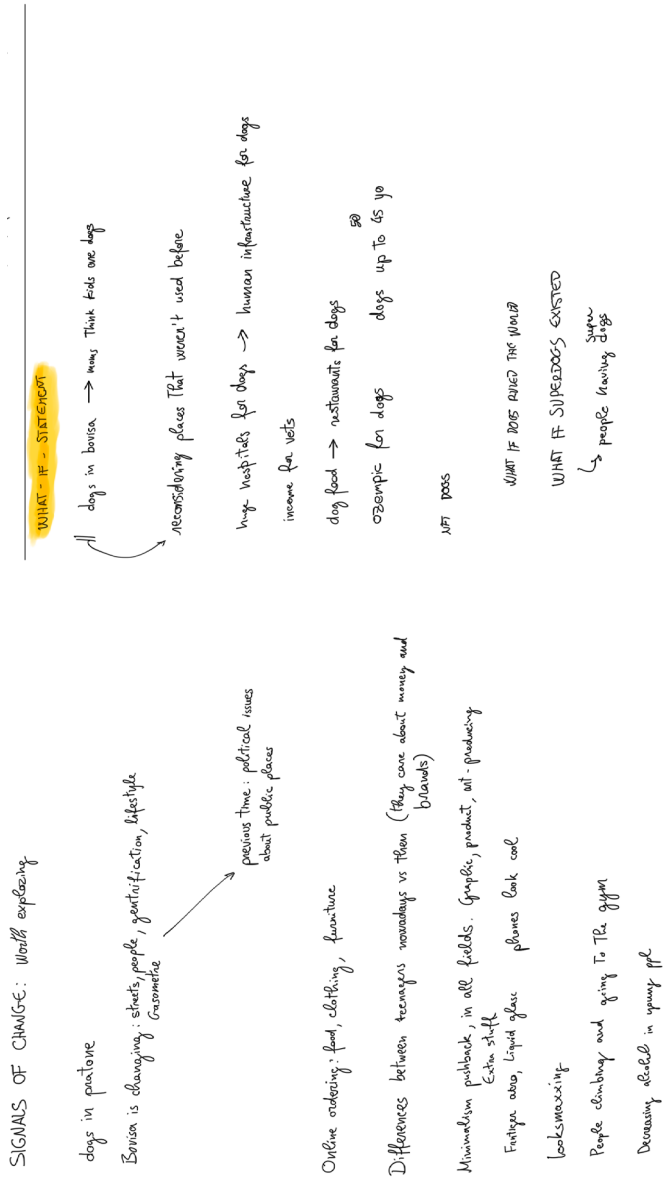
Participant 1

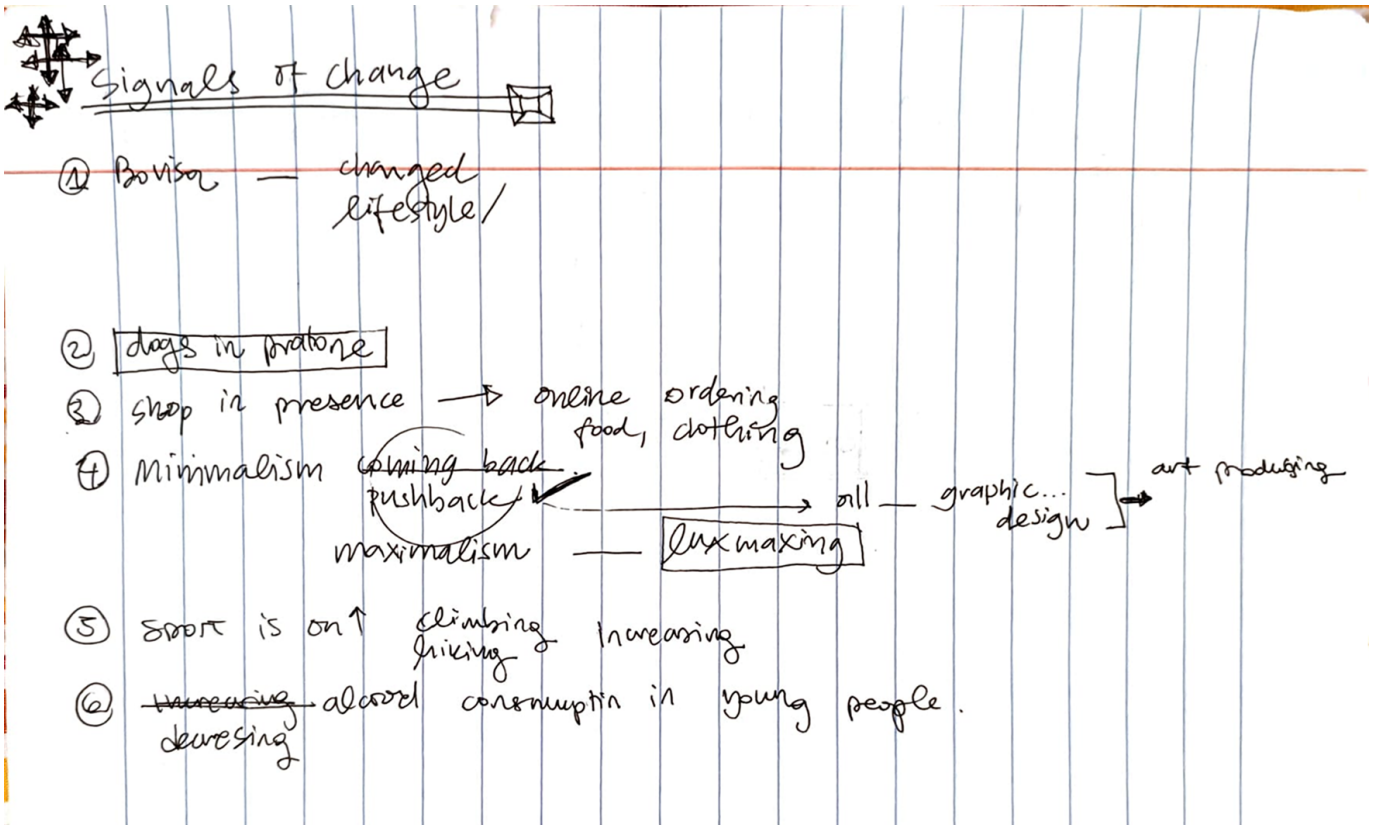
Researcher:

What did the workshop reveal to you that you did not expect? You

Appendix F: W2 artefacts

Handwritten notes taken by two participants during brainstorming, encoded variables from the input phase, and Hildegard Scenario Export with metadata.




Signals of change
① Bovisa — changed lifestyle/
② dogs in pratone
③ shop in presence → online ordering food, clothing
④ Minimalism coming back pushback
all — graphic... design → art producing
maximalism — luxmaxing
⑤ sport is on ↑ climbing hiking increasing
⑥ increasing decreasing alcool consumptin in young people.

Field	Value
Seed	system-logician
Seed Title	THE SYSTEM LOGICIAN
Original Scenario	Dogs being at the top of the social pyramid. Humans are devolved to dogs' wellbeing and are their servants, depend on their life and are outlived by them. Dogs' life are more important than humans'.
Setting	Politecnico di Milano, Bovisa.
Main Actor	Dogs, of course. Humans (less important). Other animals (even less important, want to become dogs-like). Similar to Blade runner dystopia scenario, where people bought robotic animals.
Conditions	Human infrastructure reconverted to fit dogs: dog hospitals, dog schools, dog supermarkets. Gyms for dogs. Humans work for dogs. Green public spaces, such as parks, are way more important that they are now.
Normality	Dogs are at the top of the social system. Humans work for them and are willing to do so. Bovisa municipality only cares about dogs.
Headline	Dogs rule.
Coded Setting	Politecnico di Milano, Bovisa.
Coded Protagonist	Dogs, of course. Humans (less important). Other animals (even less important, want to become dogs-like). Similar to Blade runner dystopia scenario, where people bought robotic animals.
Coded Institution	Hospital-based authority
Coded Technology	system-mediated processes
Time Horizon	Near-future (5-15 years)
Assumed Normality	Dogs are at the top of the social system. Humans work for them and are willing to do so. Bovisa municipality only cares about dogs.
Bias Rule Applied	RULE: amplify_capital_bias
Amplified Scenario	Dogs being at the top of the social pyramid. Humans are devolved to dogs' wellbeing and are their servants, depend on their life and are outlived by them. Dogs' life are more important than humans' ... optimized for extractivist efficiency and market expansion.

Hildegard Scenario Export

Metadata

- **Exported At:** 2026-06-13T14:50:51.564Z
- **Seed:** THE SYSTEM LOGICIAN

Original Extraction

Scenario

Dogs being at the top of the social pyramid. Humans are devolved to dogs' wellbeing and are their servants, depend on their life and are outlived by them. Dogs' life are more important than humans'.

Setting

Politecnico di Milano, Bovisa.

Main Actor

Dogs, of course. Humans (less important). Other animals (even less important, want to become dogs-like). Similar to Blade runner dystopia scenario, where people bought robotic animals.

Conditions (What must exist)

Human infrastructure reconverted to fit dogs: dog hospitals, dog schools, dog supermarkets. Gyms for dogs. Humans work for dogs. Green public spaces, such as parks, are way more important that they are now.

Normality (What is unquestioned)

Dogs are at the top of the social system. Humans work for them and are willing to do so. Bovisa municipality only cares about dogs.

Headline

Dogs rule.

Encoded Variables

| Variable | Value |

|----------|-------|

| Setting | Politecnico di Milano, Bovisa. |

| Protagonist | Dogs, of course. Humans (less important). Other animals (even less important, want to become dogs-like). Similar to Blade runner dystopia scenario, where people bought robotic animals. |

| Institution | Hospital-based authority |

| Technology | system-mediated processes |

| Time Horizon | Near-future (5-15 years) |

| Assumed Normality | Dogs are at the top of the social system. Humans work for them and are willing to do so. Bovisa municipality only cares about dogs. |

Bias Amplification

Rule Applied: `RULE: amplify_capital_bias`

Amplified Scenario

Dogs being at the top of the social pyramid. Humans are devolved to dogs' wellbeing and are their servants, depend on their life and are outlived by them. Dogs' life are more important than humans' ...optimized for extractivist efficiency and market expansion.

Generated by Hildegard - A Research-Through-Design Tool for Speculative Design Workshops

Appendix G: W2 exit interviews transcripts

can also make a comparison with the previous session.

Participant 1:

As in the previous one, I realized that there was something missing, even if the tool was not expected to fully work. So I think it was still interesting. This time, it was the environmental aspect that we didn't care about. In the first workshop, we didn't care about people. This time we cared about people, but we didn't care about the environment, so that's a really interesting aspect to consider. The environmental impact of the scenario was something that really made me think: *"Oh, wait...»*

It's very useful to have a template to get you thinking. In other words, it was helpful to have a template to build on.

Researcher:

OK. Did it stop you in any way, or prevent you from doing something?

Participant 1:

It didn't block me in any way, I think.

Researcher:

Do you think somehow your assumptions became more visible?

Participant 1:

Other than the environment one, no.

Researcher:

How do you feel compared to the previous session? In your own perception, how are the tool and the workshop different?

Participant 1:

I think the second one, of course, was shorter, but it was much more focused.

Researcher:

Okay, and what would you change if there was another iteration in the interface or in the workshop procedure?

Participant 1:

I think it's perfect as it is, but I would expect to have more suggestions from the model itself for the text input. Of course, I know it's hard to do right now, but that would be really cool to

see because it looks promising. Because if in this state it already helped me have stuff and assumptions emerge, if it gets even more optimized and functional, I think there will be even more assumptions emerging.

Researcher:

Okay. Thank you very much

.

Participant 2

Researcher:

Allora, che cosa ti ha rivelato questo workshop, se ti ha rivelato qualcosa che non ti aspettavi? Puoi anche fare un confronto con la scorsa sessione.

Participant 2:

È strano rifare la risposta, perché se vuoi la rimetto come l'altra volta.

Researcher:

No, no, dimmi, non so se ti ricordi quello che avevi detto.

Participant 2:

Sì, mi ricordo che una riflessione che avevo fatto è che mi sono accorta di quanto anche il design speculativo, che vuole mettere in luce forse dei problemi etici e quindi cerca di essere il più corretto possibile, sebbene nella sua assurdità, allo stesso tempo contenga esso stesso dei bias. Non so se ha senso come elaborazione.

Researcher:

Pensi a qualcosa di nuovo adesso rispetto alla scorsa interazione?

Participant 2:

In questo caso, secondo me, siamo andati parecchio per l'assurdo, però non so se è una grande scoperta. Rispetto all'altra volta siamo andati un po' troppo nell'assurdo, in un futuro molto distopico, e forse l'interazione con la macchina, più che contenerci, ha accentuato questa cosa.

Researcher:

Pensi che in qualche modo lo strumento vi abbia aiutato o vi abbia

bloccato?

Participant 2:

No, penso che ci abbia comunque aiutato sicuramente nell'immaginare lo scenario e a visualizzarlo di sicuro. All'inizio era tutto molto poco elaborato, poi siamo arrivati a definire delle strutture sociali e anche delle routine, ad avere una maggiore consapevolezza di come sarebbe stato il futuro nella sua quotidianità. Abbiamo pensato proprio a delle strutture, anche agli edifici e ai rapporti gerarchici.

Researcher:

Pensi in qualche modo che le tue assunzioni siano diventate più visibili? I tuoi bias presenti, i tuoi stereotipi? Puoi anche fare un confronto con la scorsa volta.

Participant 2:

Ci devo pensare un attimo se i miei bias sono emersi di più. Secondo me, appunto, in questo giro siamo andati molto nell'assurdo e abbiamo trascurato veramente tante cose. Secondo me anche la stessa ipotesi, cioè la stessa affermazione di partenza, era basata su una specie di percezione e noi l'abbiamo data per assodata senza effettivamente esplorarla di più. Non so se è una sorta di bias il fatto di farsi dirigere da... non lo so, non so se capisci quello che voglio dire.

Ad esempio, il fatto che i cani stessero aumentando alla Bovisa: non so se era solo una specie di percezione o se fosse vero o meno, ma noi comunque l'abbiamo dato per assodato. Oppure anche durante le discussioni, anche tutte le altre tematiche che sono uscite, a volte forse non erano così veritiere o comunque ci siamo basati esclusivamente su percezioni e non su dati di fatto. Come l'aumento della gente che va a correre o che fa palestra: forse sì, però il fatto che la gente vada a correre magari è semplicemente perché è cambiata l'età, io ho pensato che possa essere anche una differenza d'età, ecco.

Researcher:

Hai percepito qualche differenza a livello della procedura del workshop e dell'interfaccia che secondo te è rilevante rispetto alla prima interazione?

Participant 2:

Allora, io sinceramente sono riuscita a capirla molto meglio.

Questa volta sì, mi sembrava più chiaro. Forse era tutto meno disturbante, nel senso che era più focalizzato; quindi mi sono distratta meno e sono stata guidata in maniera corretta.

Researcher:

Cosa cambieresti se si dovesse ripetere, sia nell'interfaccia che nel workshop di per sé?

Participant 2:

Non lo so. A noi non è capitato che si rompesse la macchina. Una cosa che mi piacerebbe, se mi si dovesse rompere la macchina, è chiedermi il perché. Non so se tu abbia già pensato, nel caso in cui si rompa la macchina, di spiegare come mai si rompe. Nel nostro caso è sembrato tutto molto lineare. Ritornando alle cose che ci siamo detti l'altra volta, forse farei emergere ancora di più i bias.

Participant 2 (translated)

Researcher:

So, what did this workshop reveal to you, if it revealed anything you hadn't expected? You could also compare it with the last session.

Participant 2:

It's strange to repeat my answer, because if you want, I'll just say the same as last time.

Researcher:

No, no, tell me, I don't know if you remember what you said.

Participant 2:

Yes, I remember that one thought I had was that I realised how even speculative design, which aims to highlight ethical issues and therefore tries to be as correct as possible, albeit in its absurdity, at the same time contains biases of its own. I don't know if that makes sense as a line of thought.

Researcher:

Do you have any new thoughts now compared to the last session?

Participant 2:

In this case, in my opinion, we've gone quite far into the absurd,

though I don't know if that's a major discovery. Compared to last time, we've gone a bit too far into the absurd, into a very dystopian future, and perhaps the interaction with the machine, rather than restraining us, has accentuated this.

Researcher:

Do you think the tool helped you in any way, or did it hold you back?

Participant 2:

No, I think it definitely helped us to imagine the scenario and visualise it for sure. At first, it was all very rough, but then we got to defining social structures and even routines, gaining a greater awareness of what everyday life in the future would be like. We really thought about structures, including buildings and hierarchical relationships.

Researcher:

Do you think in any way that your assumptions have become more apparent? Your existing biases, your stereotypes? You can also compare it to last time.

Participant 2:

I need to think for a moment about whether my biases have emerged more. In my view, actually, in this round we went quite far into the absurd and we really overlooked so many things. I also think the hypothesis itself, that is, the initial statement, was based on a sort of perception, and we took it for granted without really exploring it further. I don't know if it's a sort of bias to be guided by... I don't know, I don't know if you understand what I mean.

For example, the fact that the number of dogs was increasing in Bovisa: I don't know if it was just a sort of perception or whether it was true or not, but we took it for granted anyway. Or even during the discussions, all the other issues that came up, sometimes they perhaps weren't entirely accurate, or at least we based ourselves exclusively on perceptions rather than on facts. Like the increase in people going for a run or going to the gym: perhaps so, but the fact that people go for a run might simply be because the age group has changed; I thought it might also be an age difference, you see.

Researcher:

Did you notice any differences in the workshop procedure and the

interface that you think are relevant compared to the first session?

Participant 2:

Well, to be honest, I managed to understand it much better. This time, yes, it seemed clearer to me. Perhaps it was all less distracting, in the sense that it was more focused; so I was less distracted and was guided properly.

Researcher:

What would you change if it were to be repeated, both in terms of the interface and the workshop itself?

Participant 2:

I don't know. Our machine didn't break down. One thing I'd like, if my machine were to break down, is to be asked why. I don't know if you've already thought about explaining why it breaks down, in case it does. In our case, it all seemed very straightforward. Going back to what we discussed last time, perhaps I'd highlight the biases even more.

Participant 3

Researcher:

What do you think the workshop revealed to you that you did not expect?

Participant 3:

First, starting from the tech point, I was not expecting a functional prototype so visually well-prepared. Also from the tech side, I was not expecting that the AI could be regenerative for me, helping me to think in a different way.

Another thing I noticed is that being by myself would be totally different than working with a group if there is AI included.

Researcher:

Do you think being in a group is better or worse than being by yourself? How do you think that would be different?

Participant 3:

It's less personal, obviously, and much calmer in a group. In a group, you know that you need to respect people and their opinions, so you cannot be super speculative. People's

understanding can be different from one person to another. For example, when we were saying, *"Oh yeah, in this dystopia dogs rule,»* at some point you know that there is not 100% agreement. Sometimes some of them were not commenting because I know they would be able to comment totally differently as well.

Researcher:

Do you think that was a positive or a negative aspect here, working in a group?

Participant 3:

I think with the structure of this platform, it's a positive thing. But if you want to break out completely, then I think it's a negative thing. At least for this group, it worked like that.

Researcher:

Where do you think the tool helped, and where do you think it blocked you?

Participant 3:

It definitely helped by giving you the things that you didn't say. It helped us to think more about a different point of view, not just the four of us in the room, but like millions of people were thinking and giving us concepts that we hadn't thought of before. Also, giving everything in order helped us sometimes to reflect on what we said. Capturing everything from the beginning all together on one page was very helpful.

Researcher:

And do you think it blocked you somehow?

Participant 3:

I expected more questions on different aspects. In a system, you usually have transportation, common areas, the home. But I expected not just the technical idea of this, but also more intimate things, like what is going on with relationships, with family, with love. We didn't actually connect with all of those aspects; we just answered one or two. So I think for the dystopia, there was a lack of information based on what we gave.

Researcher:

Do you think your own assumptions became visible, and if so, how?

Participant 3:

Yeah. I expected something visual, like a generated outcome. This is not a positive or a negative thing, it was just an expectation because we wrote a lot of words. At the beginning, AI worked only with words, but right now you give an input of words and sometimes it gives a visual outcome, like Midjourney, Claude design, Figma, or whatever. So I expected something visual for the dystopia. It doesn't have to be super realistic, by the way; it could be just a fluid image or a video. But I liked it anyway.

Researcher:

If there is anything to be changed, what would you change in the workshop and in the interface?

Participant 3:

The interface, physically and visually, was insanely pretty and clean. It was easy to understand. It had two modes, so if one was not working... for example, in the prototype, there was one part where the written text was super light, but [Participant 1] changed it to dark mode and we could see it immediately. We saw the physical usage of that platform work in real-time. So, UI-based, I liked it a lot. Also, the squares and the layout work perfectly, especially for AI. If you ask Claude to design a page, it comes out very technical, and this layout is super suitable in my opinion.

Nothing more from the functional side. The questions were maybe a bit repetitive on a few things. So instead of giving new information, we were repeating things in a different way. That's it. Except for that, it was perfect.

Participant 4

Researcher:

Allora, prima di tutto, ciao. Spero ti sia piaciuto il workshop.

Participant 4:

Sì, molto.

Researcher:

Pensi che ti abbia evidenziato o ti abbia rivelato qualcosa che non ti aspettavi prima?

Participant 4:

Beh, non avevo aspettative in realtà, quindi direi di no... solo perché non sapevo veramente cosa stavo facendo, cioè a cosa fossi venuta.

Researcher:

Ok. E pensi che nel processo di speculazione lo strumento abbia aiutato in qualcosa o abbia impedito qualcosa?

Participant 4:

Beh, ha evidenziato alcune cose a cui non avevamo pensato, quindi ci ha aiutato magari a riflettere invece su certi aspetti che avevamo tralasciato.

Researcher:

Ok. Pensi che il lavoro di gruppo ti sia servito in questo caso, piuttosto che lavorare da sola?

Participant 4:

Sì, credo che da sola non sarei arrivata né al tema né poi a come è stato sviluppato. Diciamo che ognuno metteva un po' la propria idea e poi si sviluppava partendo un po' da tutti.

Researcher:

Ok. E pensi che i tuoi stereotipi o i tuoi bias siano stati rivelati in qualche modo?

Participant 4:

Non saprei. Alla fine il tema non era un tipo di tema particolarmente orientato sul genere...

Researcher:

Non era molto legato all'etica?

Participant 4:

No, esatto. No.

Researcher:

Ok. E se si dovesse ripetere un workshop di questo tipo. cosa secondo te andrebbe cambiato, sia nella procedura che nell'interfaccia?

Participant 4:

No, allora, è stato un po' dispersivo, ma non è neanche stato un male secondo me. Ho avuto un po' il modo di conoscere anche gli

altri. Mi ha abbastanza affascinata tutta la procedura; cioè, io non saprei come fare una cosa del genere, è stato interessante.

Participant 4 (translated)

Researcher:

Right, first of all, hello. I hope you enjoyed the workshop.

Participant 4:

Yes, very much so.

Researcher:

Do you think it highlighted or revealed anything to you that you hadn't expected beforehand?

Participant 4:

Well, I didn't really have any expectations, so I'd say no... simply because I didn't really know what I was doing, or rather, what I'd come here for.

Researcher:

OK. And do you think that, in the process of speculation, the tool helped in any way or hindered anything?

Participant 4:

Well, it highlighted some things we hadn't thought of, so it perhaps helped us to reflect instead on certain aspects we'd overlooked.

Researcher:

OK. Do you think that working in a group was useful in this case, rather than working on your own?

Participant 4:

Yes, I think that on my own I wouldn't have arrived at the topic or how it was developed. Let's say that everyone contributed their own ideas and then it developed, drawing a bit from everyone.

Researcher:

OK. And do you think your stereotypes or biases were revealed in any way?

Participant 4:

I don't know. In the end, the topic wasn't particularly gender-oriented...

Researcher:

It wasn't very "ethical" connected, right?

Participant 4:

No, exactly. No.

Researcher:

OK. And if a workshop of this kind were to be repeated, what do you think should be changed, both in the procedure and the interface?

Participant 4:

No, well, it was a bit all over the place, but I didn't think it was a bad thing either. I did get to know the others a bit. I found the whole procedure quite fascinating; I mean, I wouldn't know how to do something like that, so it was interesting.

Participant 5

Researcher:

Okay, first of all, what did the workshop reveal to you that you were not expecting before?

Participant 5:

Let me think. I think it's how from one simple idea you can have so many smaller ideas that revolve around it, but they are so disconnected. Like, we talked about the poop on the streets, but we also talked about trying to understand what dogs say. Those are completely different ideas, but they all started from the same starting point. So it was cool to see how much stuff can revolve around a simple idea.

Researcher:

Do you think the tool helped you in the speculation phase, and if so, how? And how did it block you, if at all?

Participant 5:

I think, yeah, the questions were really good. I actually had a university course on speculative design, it was called Design

Futures, and I think I did a much worse job on that course than during this single hour that we spent here today, because I think these questions were better and it was kind of well-structured.

Researcher:

And how did it block you, if it blocked you in anything?

Participant 5:

Ah, let me think. I wouldn't say it blocked me. No, I think it was the opposite. I think maybe it was even like we were forced to exaggerate and go a little bit more crazy before even trying to make sense of a timeline or something.

So yeah, as I was saying, I wouldn't call it blocking because it was the opposite. I think a lot of times we were thinking too much in advance, focusing on the extremes, and not thinking so much about what's the chain of events that caused that or the logic behind it. We just went straight to crazy ideas and asked ourselves what would be "normal" in that scenario.

Researcher:

Do you feel like your assumptions became visible, and how, in building the scenario?

Participant 5:

Yeah, I think so. I think when we exaggerated the idea of dogs so much, it became obvious that we all had this shared vision that people care a lot about dogs, and maybe it's something that's increasing. But also on the other aspects, the ideas that we connected to the dog idea, they also revealed what we think of the world. For example, the fact that we immediately thought about technology for dogs already means something. It means that we all agree on this idea that humans would try to build technology for dogs the exact same way they did for humans. So it reveals kind of this assumption that we have about human intentions.

Researcher:

And one last question: what would you change about this workshop and about the interface?

Participant 5:

Okay, the interface could maybe be smarter. Like, the phrases a lot of times didn't make sense grammatically. It felt like it didn't actually understand the phrase before trying to change it, in a way.

But I think the core idea is very good. The stuff that the interface brought up, even though it was quite badly integrated into the text at times, it still made me think of those themes in some way.

And, just as a gut instinct, I would try to somehow make it more visual. I don't know if making people draw a little bit, or maybe using generative AI or something, but it's something that I missed a little bit.

Researcher:

Okay, that's it.

Participant 5:

Hell yeah! Thank you.

Copyright Statement

Declaration on the Use of Literary, Scientific, or Artistic Works of Others

In compliance with the Italian Law 633/1941 on the "Protection of Copyright and Related Rights"*

Article 70, paragraph 1bis

I declare that the use of photos and images under protection has been for scientific purposes, without profit, and always accompanied by the mention of the title of the work and the names of the author, if such indications appear on the reproduced work.

Article 70, paragraph 1 & paragraph 3

I declare that the full or partial citation or reproduction of excerpts or parts of literary works has been carried out for scientific research purposes, for illustrative purposes, and for non-commercial purposes, accompanied by the mention of the title of the work, the names of the author, the publisher, and, in the case of a translation, the translator, if such indications appear on the reproduced work.

AI Statement

Declaration on the Use of AI Assistants

- I acknowledge using Gemini (https://gemini.google.com) and Claude (https://claude.ai) as brainstorming and writing support assistants during the preparation of this thesis. I used these tools to explore conceptual angles, organize my text, and refine the narrative structure of different chapters.
- I acknowledge using DeepL Translate (https://www.deepl.com/translator) to generate translations of key concepts, quotes, and primary source materials to ensure linguistic accuracy and conceptual clarity across English and Italian text.
- I acknowledge using Elicit (https://elicit.com) as an academic literature assistant to discover, analyze, and sythesize peer-reviewed sources and research relevant to my topic.
- I acknowledge using v0 by Vercel (https://v0.dev) for interface prototyping and “vibe coding”, employing generative code tools to rapidly build, iterate, and visually test the layout and design configurations of the digital interfaces discussed in this research.

I critically reviewed all feedback, suggestions, and outputs provided by these AI technologies, revising and finalizing the layout, code, and text using my own design decisions, words, and expressions.

No content generated by AI technologies has been presented as my own work.

Ackowledgements

The past year of conducting this research project has been humbling for many reasons. This thesis is not the product of my efforts alone, but rather the result of an entire ecosystem that enabled me to pursue this academic path and conduct this research. I am grateful to this ecosystem, of which I am merely the final step.

Firstly, I would like to thank my parents and family for enabling me to attend university on the other side of the country without ever pressuring me to work while studying.

I would like to thank my supervisor, Prof. Giaccardi, who provided me with invaluable academic guidance and encouragement throughout the process. Without her support, I would not have been able to complete this project.

I am also grateful to my friends from back home and, even more so, to the friends I made at Politecnico. Special thanks go to Selene, Giacomo and Alessandra, who supported me throughout these years and were like a family to me far from home. I relied on them more than can possibly be understood.

A huge thanks to Timon, who has put up with me for longer than is humanly tolerable, regardless of the short time we've been in each other's life for.

Thanks to the many people who, implicitly or explicitly, helped this research: from anyone who joined or supported the workshop sessions, to my classmates and professors from the Gender and Diversity class at TU Berlin.

Thanks to la Terna, and everything that it has stood for for the past 30+ years in this university.

Thanks to my fencing teams, Schildwache Potsdam and Compagnia delle Lance Spezzate.

My classmates from my Bachelor's and Master's programmes. I shared the sorrows and hardships of endless study sessions and all-nighters with them. Special thanks go to the classmates I disliked: the obsessively individualistic and egocentric ones who made me despise this university enough to write a thesis about it.

Finally, Politecnico itself: my disdain for this university and for the field of design was the catalyst that made me start writing and the reason I was able to persevere.

Fonts used:
Body: Neue Haas Grotesk Text Pro
Cover and titles: PP Editorial Old
Page numbers, charts, tables: IBM Plex Mono