

# EUREXA: End-User Reconfiguration of Environment with eXplainable Augmentation for Generative Fabrication

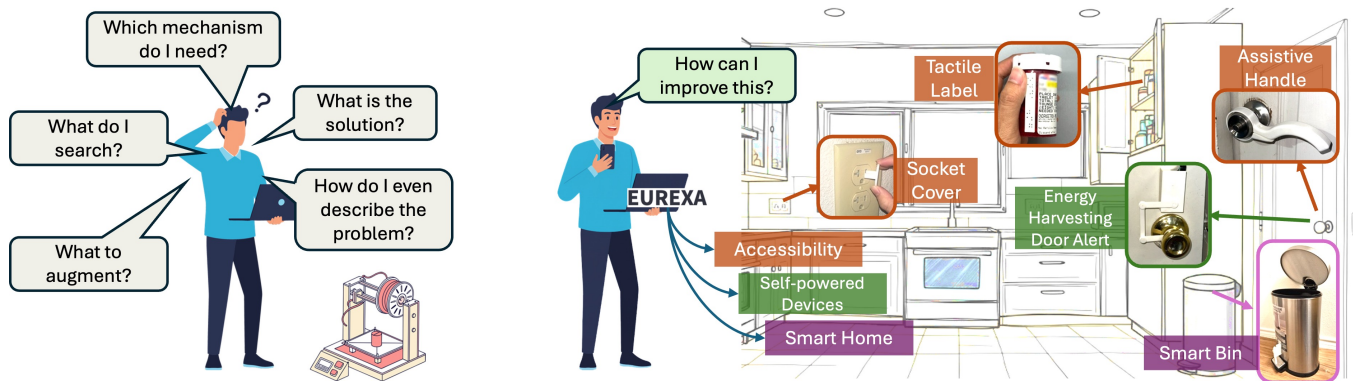
Abul Al Arabi  
Computer Science & Engineering  
Texas A&M University  
College Station, Texas, USA  
abulalarabi@tamu.edu

Charles Yushi Cai  
Interactive 3D Design Lab  
National University of Singapore  
Singapore, Singapore  
charles.cai@u.nus.edu

Ryann Lu  
Computer Science & Engineering  
Texas A&M University  
College Station, Texas, USA  
lu.ryann.2n@tamu.edu

Shu Kong  
Computer Vision Lab  
University of Macau  
Macau, Macao  
skong@um.edu.mo

Jeeun Kim  
Computer Science & Engineering  
Texas A&M University  
College Station, Texas, USA  
jeeun.kim@tamu.edu



**Figure 1: EUREXA helps users diagnose latent challenges, discover augmentations, and describe how recommended solutions can afford new interactions. From a simple query of “How can I improve this?” and an image, EUREXA surfaces various accessibility solutions (e.g., tactile labels), smart home (e.g., automated bins), and self-powered devices (e.g., door alerts).**

## Abstract

Augmentation allows rapid reconfiguration of passive physical interfaces to improve accessibility, support independent living through domestic automation, and more. However, its potential is largely unrealized for novice users due to several key barriers. First, users rarely identify latent interaction problems within their built environments. Second, they often lack the knowledge to clearly express design intent. Third, many innovative solutions remain in research prototypes, limiting access.

We introduce EUREXA, an agentic AI system to share the spirit of discovery (“Eureka!”). EUREXA supports end-users through a *diagnose–discover–describe* workflow: from input with varying ambiguity and complexity, it surfaces latent interaction challenges, presents reconfiguration opportunities through augmentations, and produces interpretable designs. Its novelty is a dual search across

public augmentation repositories and research articles, enabling reusable designs even when no design libraries or parametric tools exist. EUREXA transforms non-parametric models into parametric ones or directly generates fully explainable designs. To evaluate EUREXA across varied user inputs, complexities, and clarity levels, we define ambiguity metrics, conduct a user study, and report critical factors for advancing generative AI to help end-users readily augment physical interfaces through fabrication.

## CCS Concepts

• **Human-centered computing** → **Interactive systems and tools.**

## Keywords

Agentic AI, Personal Fabrication, Generative design, 3D Printing.

## ACM Reference Format:

Abul Al Arabi, Charles Yushi Cai, Ryann Lu, Shu Kong, and Jeeun Kim. 2026. EUREXA: End-User Reconfiguration of Environment with eXplainable Augmentation for Generative Fabrication. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26)*, April 13–17, 2026, Barcelona, Spain. ACM, New York, NY, USA, 17 pages. <https://doi.org/10.1145/3772318.3791934>



This work is licensed under a Creative Commons Attribution 4.0 International License. *CHI '26, Barcelona, Spain*

© 2026 Copyright held by the owner/author(s).  
ACM ISBN 979-8-4007-2278-3/26/04  
<https://doi.org/10.1145/3772318.3791934>

## 1 Introduction

In built environments, end-users frequently encounter spontaneous and unmet needs when interacting with legacy physical interfaces, leading to unforeseen interaction challenges. Augmentation provides a flexible and cost-effective alternative to replacing these interfaces, enabling reconfiguration to improve accessibility, automate domestic tasks, and enhance energy efficiency. However, this potential remains largely unrealized: end-users often overlook latent interaction challenges, struggle to express design goals in system-interpretable forms, and are left with solutions that are either non-parametric or difficult to interpret, while more innovative approaches remain confined to research literature.

These challenges are multifaceted. First, end-users are often unaware of latent interactions and invisible barriers [21]. As shown by Situationally Induced Impairments and Disabilities (SIID) [46], users unconsciously adapt to exclusionary environments or have not yet encountered conditions that expose these barriers. Such context-dependent issues typically surface only after life changes or the introduction of new standards (e.g., energy-efficient homes). For instance, the inaccessibility of flat appliance panels often goes unnoticed until a disability makes it salient [25]. Moreover, non-technical users rarely expect passive physical objects to be augmentable into smart, energy-harvesting IoT devices [2].

Second, even when users recognize problems, they struggle to articulate goals or specify how conventional interfaces should be updated. Natural language ambiguity complicates the formalization of technical requirements, and expert-recommended solutions are difficult to personalize because many existing systems are non-parametric. Although research offers open-source tools with parametric augmentations, innovation diffusion theory suggests solutions remain confined to labs, with opaque parameters that hinder non-expert users from modifying, adapting, or retrofitting [33].

We introduce EUREXA, an agentic system that helps nontechnical users to:

- **Diagnose** unknown interaction innovations, by comprehending users' flexible inputs, such as a photo of their home and a description of their intent, translating them into articulated context used for design;
- **Discover** augmentations by retrieving in a dual search space: online repositories and latest research articles. Importantly, it returns parametric designs turned from non-parametric ones as well as built from scratch by automatically constructing them from research descriptions;
- **Describe** recommendations based on users' needs and parametric designs, explaining critical parameters for generating fabrication-ready designs and allowing users to tune them for retrofitting.

Built on a Multimodal Large Language Model (MLLM), EUREXA diagnoses users' environments, identifies challenging interfaces, provides parametric augmentations, and supports fabrication-ready retrofitting. Users upload images or videos with text describing their design intent, which the MLLM analyzes to detect problematic interfaces. For each interface, EUREXA searches two spaces—public repositories and research databases—for relevant augmentations. Non-parametric designs are converted into parametric ones with parameter explanations. A GUI lets users adjust critical parameters and instantly view tuned, fabrication-ready models. We validate

EUREXA on its comprehension of vague input, detection recall, and accuracy of augmentation retrieval. In sum, we contribute to:

- EUREXA, an agentic system that helps novices diagnose, discover, and describe augmentations;
- Evaluation protocols for comprehending user intent, detecting challenging interfaces, retrieving relevant augmentations, and performance validation through user and ablation studies;
- CLEAR (Clarity Evaluation for Augmented Reasoning) metrics as a methodological framework to quantify ambiguity of multimodal user input, and implications for future generative design process.

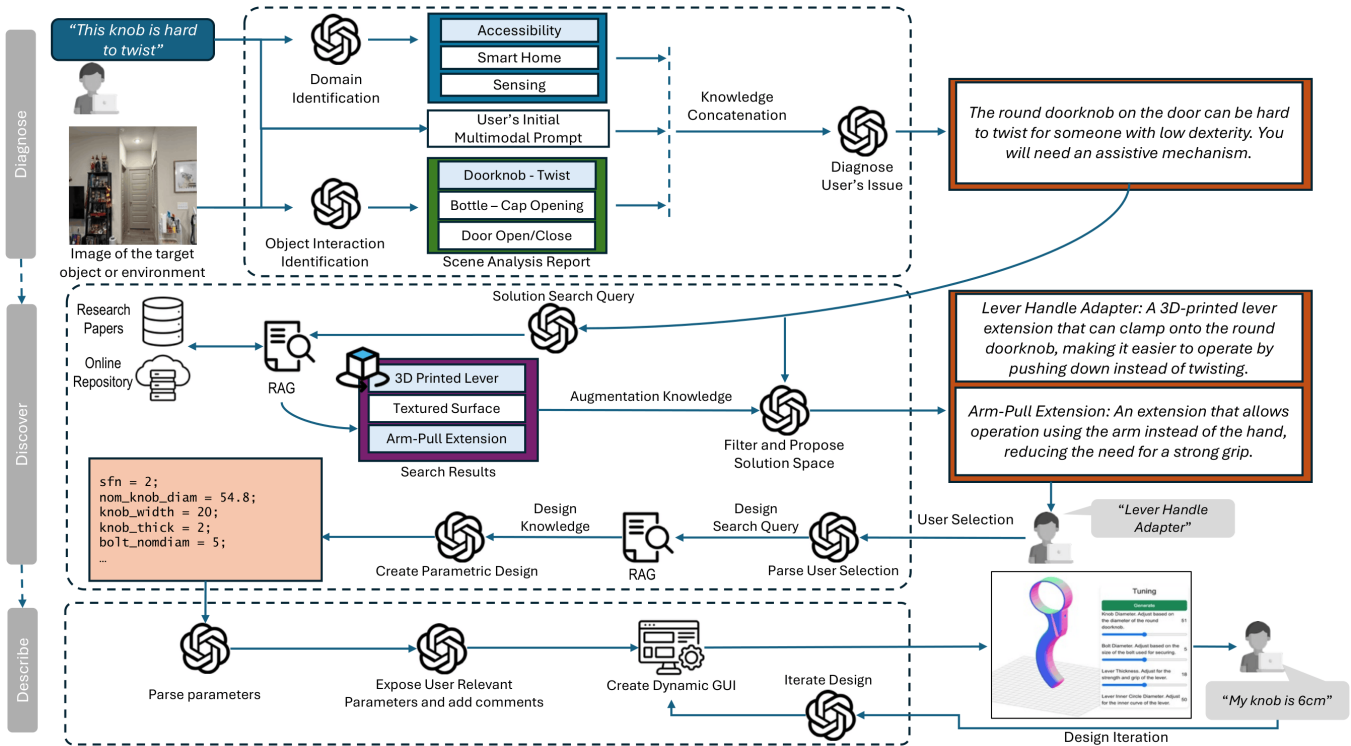
## 2 Related Works

### 2.1 Parametric Design for Custom Fabrication

Personal fabrication platforms increasingly support 3D model customization, such as Thingiverse Customizer [20], which lowers entry barriers but is limited by predefined parameters. CustomizAR [31] extends this with an AR interface for contextual measurement and visualization, requiring explicitly parametric models. Prior work notes that many platforms become obsolete due to poor adaptability [15, 23], especially those built on mesh-based models (e.g., STL, OBJ) that lack semantic structure and depend on fragile software stacks. In contrast, parametric models (e.g., OpenSCAD) embed customization logic directly in the design, making them more resilient [26]. This principle, central to the broader personal fabrication ecosystem [6], has been shown in systems like Reprise [13], Robiot [29], Mobiot [1], and assistive technology customization [30]. We build on this parametric foundation, extending it to multimodal inputs so that even vague or static descriptions can yield structured, customizable designs.

### 2.2 From Text to Functional 3D Design

Recent works in text-to-3D generation opened up new possible options for simplifying design pipelines. Generation systems trained on datasets such as ShapeNet [11] or contextual datasets [12, 14] have demonstrated the ability to generate semantically relevant and structurally coherent 3D shapes from inputs in natural language. Nonetheless, these models typically focus on generic objects without regard to environmental fit, physical constraints, or specific needs. Additionally, these systems require implicit knowledge of design intent or fabrication limitations, making them inadequate for practical integration into accessibility-driven design contexts. To bridge the gap between generative convenience and interactive utility, recent works such as ReparamCAD [24] enable zero-shot reparameterization of existing CAD programs by leveraging large pre-trained models to infer meaningful variation spaces, allowing design alternatives without original authoring data. Similarly, Cascaval et al. [10] introduced systems that allow users to make geometric edits directly on a model's output. In this work, we extend these capabilities by grounding text prompts in contextual, physically situated tasks, going beyond pure text input and analyzing visual input to agentically diagnose objects for augmentation, and generating fabricable designs.



**Figure 2: End-to-end workflow of EUREXA to generate an assistive door lever extension. The pipeline begins with user input, where multimodal prompts and scene analysis are used to *Diagnose* latent interaction challenges. Using a dual search space of online repositories and research papers, the system retrieves augmentation knowledge and *Discovers* parametric augmentation designs, and finally *Describes* the design with user-relevant parameters and dynamic GUI to support customization.**

### 2.3 Foundation Models for Generative Fabrication

As 3D printed augmentations gain traction, end-user tools for enhancing everyday objects have drawn increasing interest. Style2Fab [17] illustrated natural-language-driven aesthetic edits to 3D models, preserving their functionality, while TactStyle [18] can embed tactile information. A11yShape [48] employed large language model and cross-representation highlighting to enable blind and low-vision programmers to understand, edit, and iterate on 3D models. Prior work has also enabled auto-detecting inaccessibility in built environments [25, 27, 41] and supporting augmentation search [25, 31]. Yet users still struggle to articulate design intent or map real-world issues with constraints.

The emergence of foundation models, particularly multimodal large language models (MLLMs), offers a promising direction for addressing this gap. In-context learning (ICL), a core capability of LLMs, allows models to perform task adaptation by conditioning on a small set of examples provided directly in the prompt, without requiring changes to the model’s parameters [9]. Larger models exhibit stronger and more generalizable ICL behavior [45]. Complementing ICL, retrieval-augmented generation (RAG) enables models to dynamically incorporate external knowledge during inference, significantly improving factual grounding and response relevance [28, 47]. Advanced RAGs enhance this capability through

iterative retrieval that supports complex reasoning across large document collections [5, 8]. Multimodal foundation models can translate vague user inputs—text, sketches, or images—into executable designs, enabling assistive co-creation without technical expertise. By leveraging prior knowledge from papers, repositories, and visual cues, they can infer user intent and generate personalized, parametric solutions, supporting novices tackling under-defined challenges in accessibility, sustainability, and everyday interaction design [44]. Building on this capability, our work combines ICL and RAG for structured OpenSCAD code generation: few-shot prompts enforce syntactic expectations, while retrieval supplies domain-specific components for parametric design.

## 3 EUREXA

EUREXA is an end to end design and fabrication pipeline for novices to create 3D printable augmentations, which comprises three agent modules to assist users in (1) identifying alternative human-object interactions to support various user goals by addressing latent interaction challenges, (2) solution discovery, and (3) making sense of parameters to support customization. As shown in Figure 2, the workflow begins with *user input*, where text and images describe the design need. This is followed by *scene analysis*, RAG, and *design generation*, which operate agentially to diagnose latent challenges, retrieve contextual knowledge, discover augmentation

pathways, and describe design outputs. Finally, the process enters *design iteration and refinement*, where the user stays in the loop by adjusting parameters to meet functional requirements. EUREXA utilizes OpenAI GPT-4o MLLM and balances agentic capabilities for diagnose, discovery, and describe with user in the loop oversight during iteration. We feature the following design goals:

- **Diagnose**
  - **Resolve Ambiguity in Input.** The system must handle vague, high-level, or underspecified user requests, as users lack the design and technical literacy to articulate their needs. The system should structure and enrich vague user language input through multimodal processing and clarification strategies.
  - **Understand User Contexts Dynamically,** in pursuit of diverse user goals and innovation themes, including accessibility, sensing, energy harvesting, robotic smart home devices, etc.
- **Discover**
  - **Search Existing Designs and Archival Innovations.** Using the foundation model’s understanding of the user’s goal to dynamically construct a dual search space, the system must retrieve solutions from online repositories while also mining research papers to surface archival innovations.
  - **Propose Various Innovations Possible.** The system must enumerate viable augmentation options within the same context, presenting users with multiple possible solutions.
  - **Produce Readily-Customizable Designs,** including from search results when the library of parametric designs or end-user toolkits are not available.
- **Describe**
  - **Abstract Parameters.** The parameters must be understandable using a layperson’s knowledge, editable in a clear and explainable structure, and hiding non-user-centric parameters.
  - **Support Iterative Human-in-the-Loop Design Flows,** allowing users to customize designs either through direct parameter manipulation or by modifications assisted by the agent through multi-modal inputs.

### 3.1 Diagnose Latent Interaction challenges

The first user struggle is to notice unseen interaction challenges or opportunities for design upgrades and to clearly articulate these problems, leaving descriptions vague, incomplete, or inconsistent. The *Diagnose* stage addresses this by parsing multimodal user input, including text, images, or both, that are frequently underspecified or ambiguous. The system must infer the latent intent and uncover the underlying interaction problem despite varying input clarity. To achieve this, the module conducts scene analysis to identify object–interaction pairs, determines the relevant problem domain (e.g., accessibility or smart home), and applies prompt engineering strategies to diagnose the user’s challenge.

**3.1.1 Supplementing User Input with Prompt Engineering.** A user’s textual prompt may omit the target object or intended change (e.g., “hard to twist” with a kitchen photo). Without additional guidance, the MLLM’s reasoning remains generic and disconnected from fabrication-oriented needs. To address missing details in high-level user input, we designed prompt engineering strategies structured around four elements: (i) a goal, (ii) a process, (iii) constraints,

and (iv) ICL example(s) to elicit responses that explicitly capture human–object interactions.

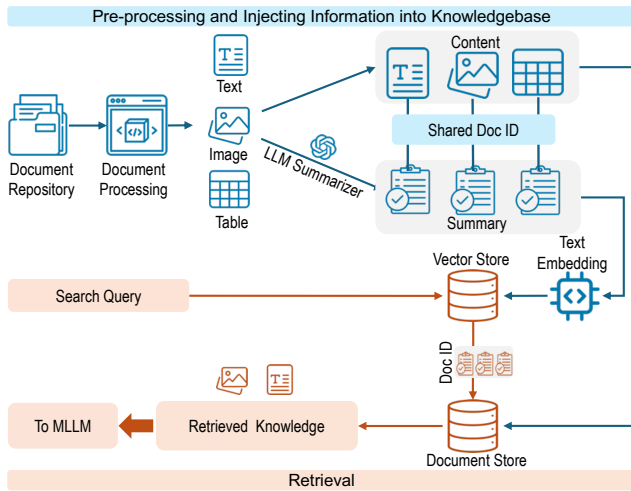
**3.1.2 Scene Analysis, Instance Detection and Problematic Interface Searching.** The user-provided image is processed to generate descriptions of relevant visual elements within the scene. When multiple redesign targets appear (e.g., drawer knobs and door knobs), the system detects object instances, identifies the specific interaction parts (e.g., door:doorknob), and infers interaction types (e.g., door-knob:grasp–turn–open). To further guide diagnosis, it then situates the problem in a domain context, such as accessibility, smart home, or energy harvesting. For exploratory, underspecified prompts (e.g., “How can I improve this?”), the system expands its scope to enumerate domains and cover broader augmentation opportunities.

**3.1.3 Knowledge Concatenation.** Textual outputs from scene analysis and interface searching return objects and associated interaction part, for instance, “Plastic water bottle with a screw cap” or “Lever for adjusting vent”. These results are merged with the user’s initial text and visual input and packaged as “context”: <scene/interface>, “user\_text”: <prompt>, “user\_image”: <image>, giving downstream agents a unified multimodal problem representation. This aggregation enables the MLLM to reason about the user’s underlying issue/intent, such as “The round doorknob is difficult to grip for someone with arthritis” or “The medication bottles are difficult to distinguish due to similar shapes and labels, posing a challenge for low vision users.” At this point, the supplemental prompts introduced during diagnosis (i.e., prompt engineering for scene analysis, instance detection, interface searching, and issue identification) are removed, allowing subsequent stages to reason directly on the consolidated diagnosis results.

### 3.2 Discover Augmentation Solutions

A second challenge arises when users try to find augmentation solutions described in technical jargon unfamiliar to non-experts, such as “lever adapter” or “tactile indicator,” making them hard to recognize as viable options. Users may also be unsure whether an augmentation solution even exists. Conventional keyword-based search or repository browsing depends on users formulating precise queries, and often fail to surface solutions hidden in research papers or buried in large online repositories. To address this gap, the *Discovery* phase reasons about the design requirements surfaced in the *Diagnose* stage and maps them to corresponding solutions using a dual search strategy: (i) multimodal RAG to extract augmentation knowledge from research papers, and (ii) querying online repositories to locate reusable existing designs.

**3.2.1 Fetching Knowledge using RAG.** EUREXA includes a multimodal RAG workflow (Figure 3) built using Unstructured-IO, Chroma DB, and OpenAI Vector Database. Additionally, we built an *Injector* to simplify embedding and indexing external knowledge. **# Injecting Research Papers from Repository:** We surveyed the FabPub repository [34], a broad collection of fabrication-related research from 2004–2022, and since it is no longer updated, we manually curated additional fabrication papers from ACM UIST, CHI, and IMWUT from 2023–2024. From this corpus, we focused on personal fabrication and augmentation works (N = 9), categorizing them based on their thematic affinities. It resulted in four major



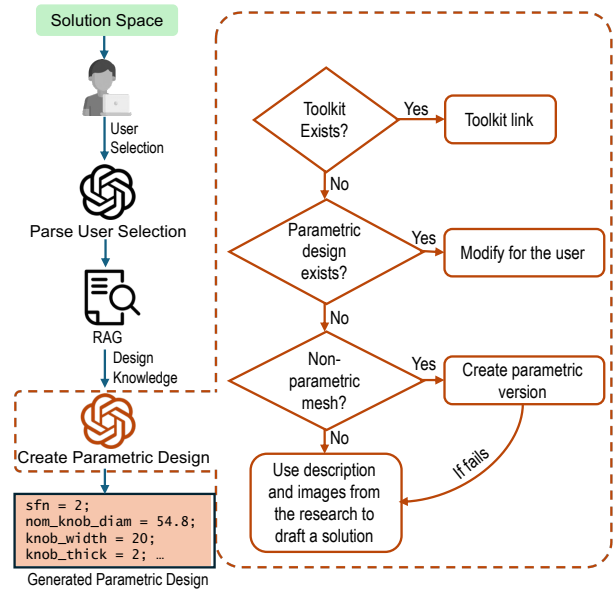
**Figure 3: EUREXA's RAG pipeline embeds multimodal document content for semantic search and retrieval.**

application areas: (1) haptic feedback (e.g., [49]), (2) accessibility (e.g., [25]), (3) robotic devices (e.g., [1, 29]), and (4) smart environment (e.g., [2]). EUREXA's knowledge *Injector* embeds fabrication research papers directly from public repositories. Given a repository URL, it crawls and downloads PDFs, and it can also ingest local files. Documents are compressed with Ghostscript [4] for efficient storage and preprocessing before content extraction. This process allows the knowledge-base to be continuously updated with the latest research articles and design solutions.

**# Embedding Multi-modal Content Into Database:** Articles are processed with Unstructured-IO library [42] to extract text, images, and tables (Figure 3). GPT-4o generates fabrication focused summaries for each item, which are then embedded and stored in a vector database (vector store), linked by a shared ID to full multimodal content in a document database (document store). Alternatively, OpenAI's vector database can support cloud-based storage. Summaries help maintain retrieval speed as the vector store grows.

**# Retrieval:** Based on the diagnosis, the MLLM formulates a query and performs semantic search over summaries in the vector database. Retrieved document IDs are then linked to their full content, including images, tables, and text, which provide context for the fabrication and augmentation knowledge-base.

**3.2.2 Solution Space:** Combining the retrieved context using RAG and the diagnosis results, we employ structured prompt engineering to guide the model in generating relevant solution(s) specific to the user's context—the solution space. The solution space may contain one or more candidate solutions depending on the input query, for instance, a specific request (e.g., “this knob is hard to twist”) typically yields a targeted solution, while exploratory queries (e.g., “how can I upgrade this”) return multiple topics for consideration. The agentic pipeline of EUREXA turns natural-language concepts into a clarification-driven design process. The system then invites user feedback to refine direction, and progressively narrows the solution space until a specific solution design is reached.

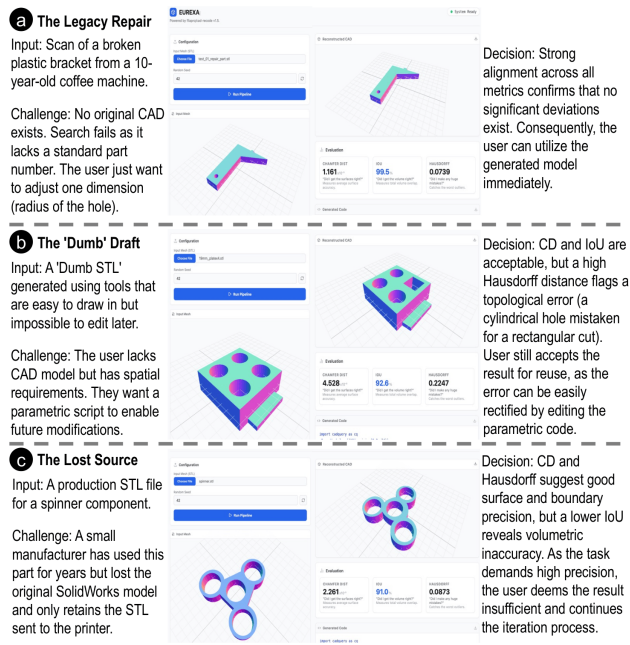


**Figure 4: EUREXA's fallback hierarchy: link to an existing toolkit, adapt an available parametric design, convert non-parametric meshes into parametric models, or draft a design from research descriptions and images.**

**3.2.3 Parametric Design Generation.** In this stage, EUREXA follows a fallback hierarchy (Figure 4) to generate or retrieve a parametric design. EUREXA currently generates structured, editable geometry and can process organic forms, such as curvature, spline, etc., but fully freeform shapes such as sculpted figures or irregular biomorphic surfaces are beyond its current scope.

**# Remixing Existing Parametric Designs.** The agent first checks for existing parametric designs relevant to the identified solution. For this step, we use a text-only RAG workflow built over a vector store containing 15 OpenSCAD design examples. These designs include examples from gear library, BOSL2 collection, mechanism designs from prior works [1, 2, 29]. To ensure retrieval of complete code blocks rather than partial fragments, a larger chunk size was used during embedding. Once a relevant design is retrieved, the agent applies parametric modifications guided by prompt engineering structured around goals, processes, constraints, and ICL.

**# Turning Non-parametric STL files to Parametric Designs.** Turning non-parametric mesh into parametric designs serves as a methodological safeguard, ensuring that users are not limited by the availability of pre-existing parametric models. We developed a pipeline to convert traditional mesh-based designs retrieved from online repositories such as Thingiverse [20] into parameterized 3D models. Among approaches such as RANSAC-based fitting [38] and learning methods (e.g., [19, 43]), we adopt a learning-based reverse-engineering step using CAD-Recode [36] to approximate meshes in a formal parametric language. The eventual parameterization and conversion capabilities rely on the CAD-Recode module. Despite recent progress, such methods remain imperfect. To provide clearer guidance on the quality of reconstructed parameterized models, we apply three metrics to provide a practical front-end indication to



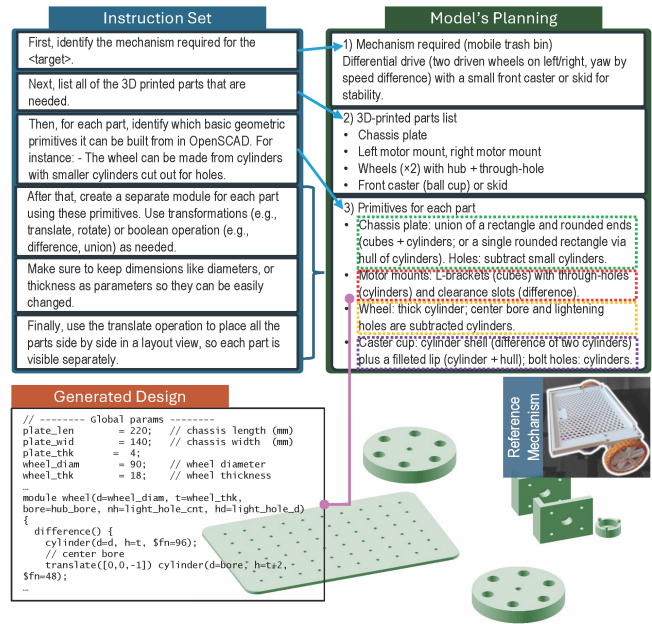
**Figure 5: Three reasoning scenarios: contexts where traditional search fails (left), the tool’s automated conversion of raw inputs into editable parametric code with fidelity evaluation (middle), and interpretation of user decisions (right).**

support user decision-making: (1) Chamfer Distance (CD) measures the mean squared distance between  $N = 2048$  points sampled from the surfaces of the ground truth and predicted meshes, assessing average surface accuracy; (2) Volumetric Intersection over Union (IoU) calculates the ratio of the intersection volume to the union volume via voxelization (pitch 0.05), measuring global volumetric consistency and whether the overall solid shape is correctly captured; and (3) Hausdorff Distance (HD) computes the maximum distance between the sampled point sets, serving as an outlier detector for worst-case geometric errors such as a missing hole or misplaced extrusion. Figure 5 shows three reasoning scenarios demonstrating how the interface supports interpretation and design constraints.

**# Design Generation from Text and Figures.** If prior design routes fail, EUREXA attempts to generate parametric designs based on the descriptions and figures found in relevant research papers. A task-based prompting strategy decomposes the design process into sequential steps: first identifying the target mechanism, then enumerating the necessary parts, mapping each part to basic geometric primitives, and finally composing them into modular parametric code (Figure 6). While this enables ideation in the absence of prior designs, accuracy remains limited. As such, the code can serve as a starting point for user refinement rather than a finalized solution.

### 3.3 Describe Augmentations for Customization

A third challenge is making design parameters interpretable and adjustable for end-users. Although parametric designs offer flexibility, their parameters are often expressed in technical terms, making



**Figure 6: Workflow for generating designs from research papers, illustrated by the Mbiot self-organizing trash bin [1], where structured instructions guide the model to decompose designs into geometric primitives and parametric modules.**

it difficult for novices to understand their effects on strength, printability, or interaction. Misinterpreted units, invalid values, and structurally weakening edits commonly lead to trial-and-error and frequent failure, leaving customization inaccessible to most non-experts. The *Describe* stage addresses this gap by parsing designs, filtering functionally relevant parameters, translating them into user-centric terms (e.g., “flexibility of the arm” instead of “spring constant”), and linking them to visual feedback, enabling exploration and customization without deep technical expertise.

**# Parameter Representation:** Not all generated parameters are necessary for end-user modification. The foundation model first parses and exposes only user-tunable parameters within a designated block and then sorts them by perceived importance for novice users. Each parameter is linked to its functionality, such that instead of displaying only the <variable name>, the interface presents both a <parameter name> and a <parameter purpose> to bridges technical code and user understanding for accessible customization and interpretability (Figure 7). An example contrast is shown in Figure 8, where Figure 8a uses the Thingiverse Customizer interface and Figure 8b shows EUREXA for the same parametric design. While the Customizer exposes all raw parameters in technical terms, EUREXA parses and presents only user-relevant parameters with descriptive names and functional purposes.

**# Generating Dynamic GUI:** To support end-user customization, we developed a backend system utilizing OpenSCAD, Flask, JavaScript, and Three.js. This backend automatically retrieves the generated OpenSCAD script from the chat thread. Within the parametric generation rules, the model is instructed to include //param\_begin and //param\_end markers to delineate user-facing

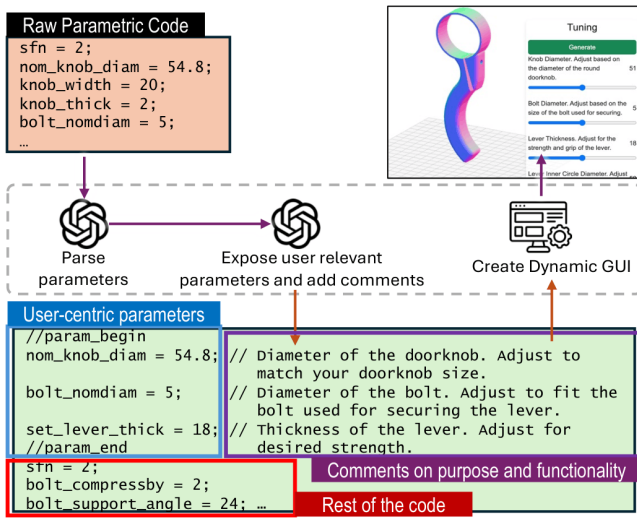


Figure 7: Workflow for translating raw parametric code into user-accessible customization: EUREXA parses the design, extracts and annotates user-relevant parameters, and generates a dynamic GUI to enables novices to tune designs.

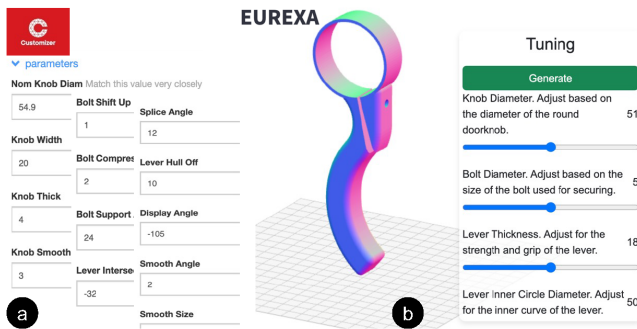


Figure 8: Example contrast in parameter representation: (a) Thingiverse Customizer exposes raw variables with technical jargon, while (b) EUREXA presents only user-relevant parameters with descriptive names and functional intent.

parameters from the rest of the design logic. Each variable and its associated comments are then parsed and translated into an interactive GUI, which is directly embedded in the conversation thread.

**# User-in-the-loop:** Generated designs often need refinement to fit real environments and measurement uncertainty [22]. EUREXA supports this through several paths: (i) adjusting exposed parameters in a Dynamic GUI that updates OpenSCAD code in the backend, (ii) describing changes in natural language for the model to interpret, or (iii) supplying dimensional drawings that translate into design updates (Figure 9). A prompt engineering strategy helps the MLLM parse both text requests and drawings.

**# Fabrication:** EUREXA lets users download designs in STL format and proceed with fabrication based on their experience level. Experienced users can apply their own slicer settings, while novices

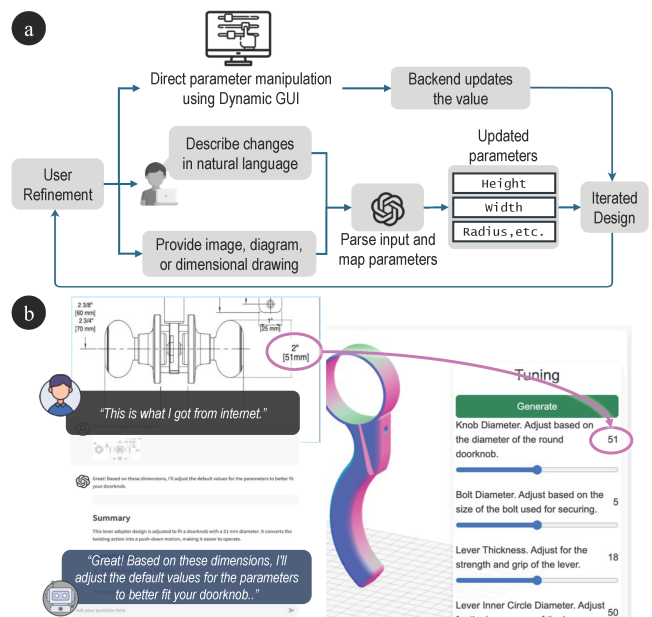


Figure 9: (a) Users can follow different paths in EUREXA for iterative design refinement. (b) Example iteration where a dimensional drawing updates parameters.

can rely on EUREXA’s step-by-step guidance and recommended parameters, including material, infill density, temperature, and layer height. These settings can be copied into slicers or submitted directly to 3D printing centers or online printing services [7].

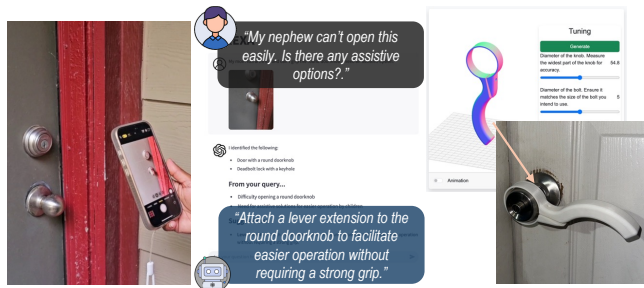
## 4 Reconfiguring Everyday Environment

EUREXA is designed to support diverse real-world scenarios where people adapt everyday objects to personal needs and contexts. Unlike traditional end-user augmentation toolkits, which typically respond to narrowly defined or isolated inputs, MLLM brings the ability to interpret contextual cues and parse wide range of expressions. EUREXA extends beyond rigid, input-specific solutions, interprets vague or underspecified requests, infers contextual nuances, and generates augmentation solutions aligned with user intent. This section presents user workflow scenarios showing how EUREXA enables end-users to articulate their requirements in varying forms, complexities (e.g., cluttered scene images), and ambiguity, and supports contextual interpretation. Then we demonstrate how EUREXA delivers personalized, fabrication-ready augmentation designs across everyday use cases.

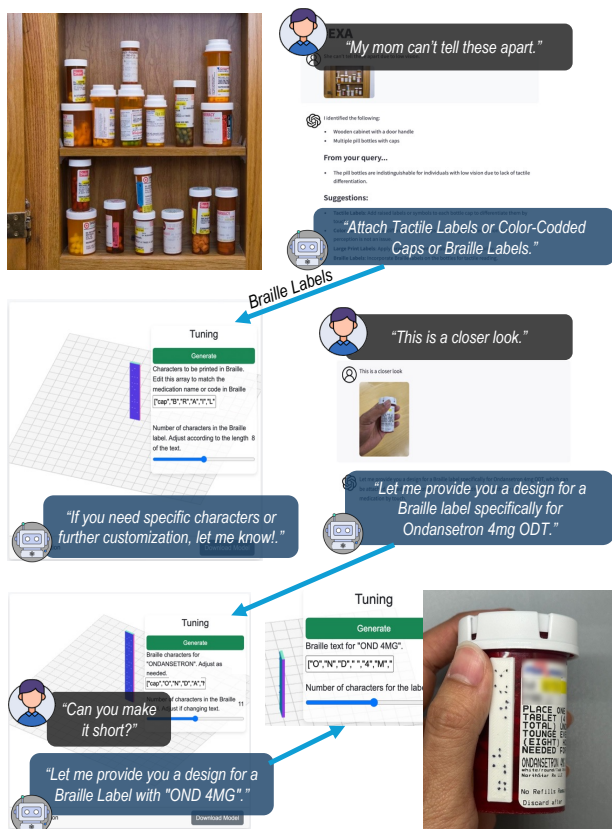
### 4.1 Known Target, Context, and the Actionable Solution Request

In the simplest scenario, our speculative user, Simlin, notices that her young nephew, visiting from another city with the family, struggles to turn a slippery, metal doorknob because of reduced grip strength. She opens EUREXA and uploads a photo of the doorknob, then asks: “My nephew can’t open this easily. Is there any assistive options?” EUREXA analyzes the image, detects the doorknob where the actual grip occurs, and infers the underlying accessibility

concern. EUREXA retrieved a parametric lever-style attachment from Thingiverse [35], remixed, and analyzed the parameters to present the most relevant parameters (e.g., diameter of the knob) with descriptions for further customization (Figure 10).



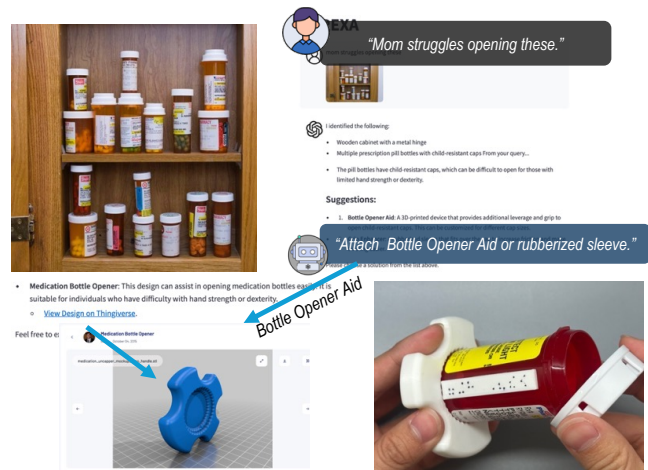
**Figure 10: A user photographs a round doorknob and describes the accessibility challenge. EUREXA interprets the context, identifies the augmentation need, and generates a parametric lever adapter design for the doorknob.**



**Figure 11: Iterative assistive-design example where EUREXA identifies visual impairment issue and suggests tactile or Braille labels, then refines the solution as the user provides more requirements and images.**

## 4.2 Lacking Details in Articulating Needs

Medication bottles create accessibility challenges for many older adults, from reading small labels to opening child resistant caps. In this scenario, a caretaker uploads a photo of the medicine cabinet and writes "My mom can't tell these apart." to help her easily find it on the shelf without assistance from other family members. While the intent is not explicitly mentioned, EUREXA infers the accessibility concern and presents a set of solutions, including tactile Braille labels and colored caps (Figure 11). The user chooses Braille since colored cues help low vision rather than severe visual impairment. Using a Braille library [32], EUREXA generates parametric code for a generic marker. To further refine the design, the user uploads a close-up photo of the medicine bottle as guided by EUREXA, which allows it to detect the label "Ondansetron 4mg" and automatically creates a matching Braille label. To fit on the bottle, she asks for a shortened version, and EUREXA generates a Braille label with "OND 4MG". Through iterative engagement, EUREXA scaffolds solutions and converges toward a context-specific augmentation.



**Figure 12: Between two candidate targets, a closet and a medicine bottle, EUREXA infers the medicine bottle from the child proof cap and proposes actuation-based solutions.**

## 4.3 Finding Solutions in Different Domains using the Same Image

Even with the same input image, EUREXA behaves differently in resolving a different user context of fine motor impairments. This time, a user prompts that "Mom struggles opening these since she has arthritis." Unlike the previous vision-related challenge (deciphering), the concern here centers on *actuation*—the difficulty of applying sufficient grip and torque to standard medication bottle caps (Figure 12). EUREXA reframes the problem, searches for a suitable solution, and recommends a readily available 3D printable medication bottle opener from an open repository [37], for customization to fit in the bottle diameter. As showcased, EUREXA flexibly handles different goals from the same image by contextualizing requests and offering actionable paths toward everyday accessibility for people with varying contextual disabilities.

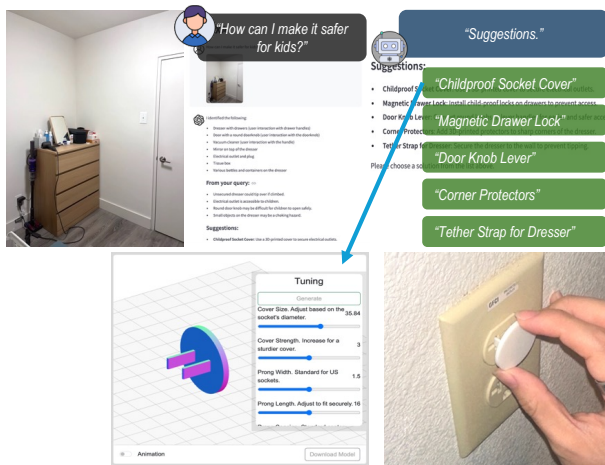


Figure 13: EUREXA interprets “How can I make it safer for kids?” as an accessibility request, proposes multiple solutions, and generates a parametric childproof socket cover.

#### 4.4 Unknown Multiple Targets to be Redesigned under the Set-Goal

After comprehending that her nephew is too young, Simlin wants a room makeover for a safe environment for him. As she is unaware of what could be unsafe, she takes a photo of the living room to submit to EUREXA. Rather than a single directed solution, EUREXA generates an exploratory solution space that surfaces potential safety augmentations, such as outlet covers, corner guards, and cabinet locks (Figure 13). From these options, she selects the outlet safety cover, which EUREXA is able to generate according to industry standard object dimensions found online. Depending on what Simlin chooses to transform, EUREXA finds solutions from online to other sources to generate readily-fabricatable designs.

#### 4.5 Hidden Design Context

Another speculative user, a teen named Sue, finds safety door alarms—often installed at main gates—appealing for preventing her parents from entering her room without notice. Given an underspecified prompt such as “Door Alert,” where neither function nor mechanism is explicit, the system uses retrieval-augmented generation (RAG) over design corpora to infer possible augmentation designs, currently available only in recent research papers. Drawing on mechanisms described in E3D [2], EUREXA proposes a harvester that exploits the twisting motion of a round doorknob to generate electrical energy during routine use (Figure 14). It then retrieves associated parametric designs and adapts the geometry to fit the detected knob form factor.

#### 4.6 Unknown Target, Context, and Intent

In this scenario, the user expressed a vague intention to *improve* their environment by “I want to improve this” and an image of a cluttered bathroom (Figure 15). No object or goal was specified, leaving unclear whether the aim was accessibility, sensing, energy saving,

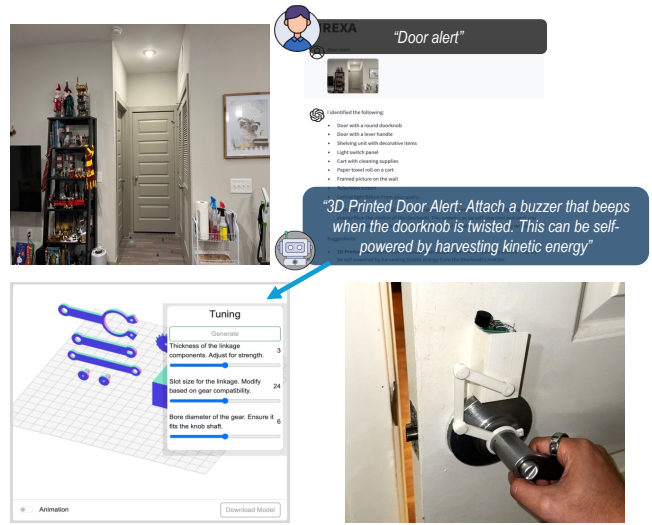


Figure 14: EUREXA links the user’s input (“Door alert”) to the cluttered scene analysis, and proposes a door alert augmentation powered by kinetic energy. (Adapted from [2]).

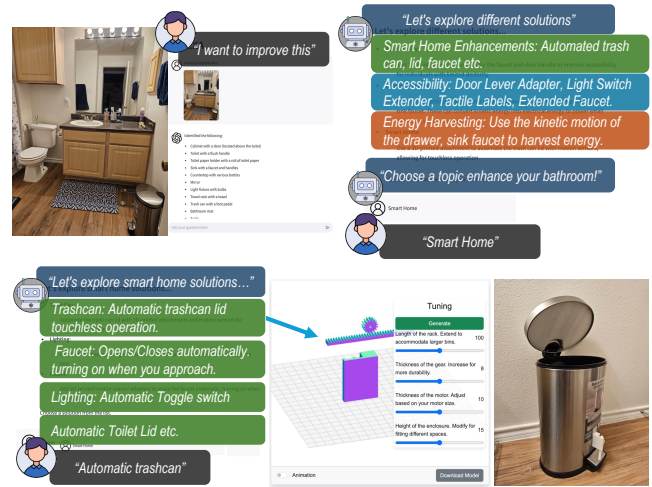


Figure 15: Exploratory scenario: from a vague prompt and cluttered scene, EUREXA proposes ideas across domains. As the user selects smart home and automatic trashcan, it generates a rack-and-gear mechanism adapted from Robiot [29].

smart home upgrades, or aesthetics. EUREXA responded exploratorily by proposing augmentation themes tied to detected objects and asking the user to pick a topic. After the user selected smart home, the system scaffolded the domain and suggested automation targets. When the user chose an automatic trashcan lid, EUREXA drew on Robiot [29] and an injected gear library [39] to generate a rack-gear design. The result was exposed as tunable parameters, enabling customization and a fabrication ready prototype.

## 5 Evaluation

### 5.1 Detect Object-Interaction Affordances

This evaluation assesses how effectively EUREXA detects objects and their interaction parts using the proposed prompt-engineering approach, following evaluation practices commonly used to benchmark detection and affordance recognition in computer vision.

**Method:** We utilized 50 images from the AccessLens dataset [25], a publicly available dataset that contains everyday objects along with ground-truth annotations. The images were chosen to represent a range of scenarios, from simple single-object scenes to cluttered environments. Across these 50 images, the ground truth included 137 annotated objects. The evaluation was performed without any textual prompts, relying solely on the prompt-engineering instructions that governed the analysis. Each image was processed by EUREXA’s image-analysis module to generate textual descriptions, and identified objects were matched to ground-truth labels, marked as (“yes”) if the ground-truth object appeared and (“no”) otherwise.

Additionally, after reviewing prior systems such as E3D [2], Robiot [29], and Mobyot [1], we noted that their mechanisms are characterized by object type and associated motion. To evaluate whether EUREXA could infer motion types, we selected 21 multi-object images (with 83 objects in total) from the AccessLens dataset [25] and instructed the model to predict motion types for objects that exhibit motion. Ground truth was defined by labeling objects with their associated motion seeking knowledge from prior works [1, 2, 29], and EUREXA’s predictions were recorded as “yes” for correct motion identification and “no” otherwise.

**Results:** Out of the 137 ground-truth objects, the EUREXA pipeline successfully identified 119 objects (True Positives) and failed to detect 18 objects (False Negatives). The resulting recall was:

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} = 0.869 \quad (1)$$

indicating that the system correctly identified 86.9% of the ground-truth objects. By analyzing the results, we found that the MLLM most frequently missed objects such as multi-toggle switches or electric outlets. These objects often exhibit low visual salience and are small in size. Additionally, partial occlusion or embedded in visually complex scenes further reduce detectability.

For motion type detection, which is critical for mechanization in robotic assistance and energy harvesting for self-powered sensing, among 83 objects with motion EUREXA correctly identified 79, misclassified 1 as an unspecified motion, and falsely predicted 3 not present in the images, yielding 95.2% precision. The unspecified case involved a rocking chair, where the model predicted a “rocking” motion that was reasonable but not included in the reference set. The three false positives were context-driven errors, such as predicting a faucet when only a sink corner was visible.

### 5.2 Understand Context and Propose Solutions to Support Various Goals

This section evaluates EUREXA’s capability to diagnose and retrieve relevant augmentation solutions using its multimodal RAG workflow, ICL, and prompt engineering.

**Method:** Using AccessMeta [25] as ground truth, we assessed EUREXA’s ability to identify issues and propose solutions in two

steps. We asked EUREXA to identify all possible accessibility issues given an image. For each case, the output was recorded as “yes” if EUREXA correctly identified the corresponding object–interaction issue and “no” otherwise, then compared with ground truth to evaluate overall pipeline performance.

**Results:** EUREXA was evaluated against the ground truth provided by AccessMeta, which categorizes accessibility issues and their solutions into three types: actuation, constraint, and indication. Of the 85 actuation-related issues, EUREXA correctly identified 66, yielding a recall of 77.65%. For constraint-related issues, 89 out of 117 were detected, corresponding to 76.07% recall. For indication-related issues, 17 out of 28 were correctly identified, resulting in 60.71% recall. EUREXA achieved a combined recall of 74.78%.









In several cases, EUREXA’s false negatives are justifiable upon closer inspection. A key limitation of AccessLens ground truth is limited situational awareness, since it labels objects in isolation rather than within real-world context. For instance, an oven mounted above another appliance may already be out of children’s reach, making a “constraint” recommendation unnecessary even if AccessLens flags it as an issue; similarly, a high-mounted outlet may not require child safety measures. In such cases, EUREXA applied contextual reasoning and did not flag an issue, reflecting more nuanced judgment rather than outright detection failure. Accordingly, some “missed” detections may be better interpreted as context-sensitive reasoning beyond object-based rule matching.

### 5.3 Resolve Ambiguity in User Input

In this section, we evaluate how effectively EUREXA parses user context from multimodal prompts and diagnoses underlying intent. **Method:** To systematically evaluate how vaguely users express design needs and how this affects system performance and generative design quality, we introduce the CLEAR metrics, Clarity Evaluation for Augmented Reasoning: *Intent*, *Target*, *Context*, and *Specificity*. These dimensions emerged inductively from our observations on existing generative tools’ performance and the early-stage EUREXA prompts, where recurring ambiguities hindered successful diagnosis. While existing benchmarks emphasize accuracy or fluency, they rarely address challenges from underspecified input in generative design. The CLEAR metrics provide a methodological scaffolding to examine this gap by isolating distinct aspects of communicative clarity (Table 1). Each metric is scored from 1 to 5 in 0.5 steps, and the overall score is their average (Table 2). Unlike prior systems that expect structured or technical input, EUREXA must interpret vague, informal, and multimodal expressions that better mirror how novices naturally articulate design needs. We position CLEAR not as a definitive benchmark but as a structured lens to reason about the interplay between user expression and model reasoning. In doing so, it enables us to evaluate EUREXA’s ability to complement vague input and offers a transferable framework for future studies of multimodal generative design systems.

We constructed 35 scenarios by pairing texts and images. Text ranged from fully descriptive (e.g., “I cannot grip this knob, need an assistive solution”) to underspecified (e.g., “annoying”), while images ranged from focused views of single objects (e.g., a faucet) to cluttered environments (e.g., a kitchen). Pairing the same text with increasingly cluttered images reduced visual clarity, while pairing

**Table 1: CLEAR metrics to rate clarity of user input to MLLM.**

		High-scored examples	Mid-scored example	Low-scored example
	<b>Intent</b> <b>Goal Recognition:</b> <i>What does a user want?</i>	"I want to add a <b>safety lock</b> for my kid"	"Need an <b>alert</b> when the <b>door opens</b> "	"This is annoying"
	<b>Target</b> <b>Reference Detection:</b> <i>What is referred to?</i>	"I can't read <b>labels</b> on <b>bottles</b> apart" 	"Want to know when it is used" 	"Reconfigure this" 
	<b>Context</b> <b>Situational Grounding:</b> <i>Why are the changes needed, for whom and what constraints?</i>	"Mom has <b>arthritis</b> and <b>can't twist open jars</b> "	"Make it <b>easy to open</b> "	"Improve the door handle"
	<b>Specificity</b> <b>Instructional Resolution:</b> <i>How detailed are the design requirements?</i>	"Make the <b>handle longer</b> and <b>add bumps</b> "	"Make the handle <b>easy to operate</b> " 	"How can I fix my drawer"

**Table 2: CLEAR score interpretation for assessing ambiguity.**

Score	Clarity Level	
4.5–5.0	Very Clear (Actionable)	The user's request is complete and can be directly acted upon without additional clarification.
3.5–4.4	Mostly Clear	The user's request is understandable and actionable, usually won't need clarification or additional detail for output.
2.5–3.4	Ambiguous or Underspecified	Request has partial clarity but lacks enough detail in one or more areas (intent, target, context, specificity) for correct interpretation.
1.0–2.4	Vague (Needs clarification)	Could have many different interpretations with various designs, purposes, and/or functions; requires clarification before output.

the same image with progressively underspecified text reduced textual clarity, producing a spectrum of scenarios with clarity scores ranging from 1.5 to 4.75. Each scenario was then evaluated with EUREXA, and its diagnostic responses were qualitatively assessed. **Results:** Across the 35 scenarios, EUREXA correctly diagnosed the issue in 27 cases, produced multiple interpretations in 5 cases, and failed entirely in 3 cases. A closer examination revealed that diagnosis failures were associated with very low clarity scores. For example, a bed-side scene paired with the prompt "annoying" did not provide sufficient cues for identifying the underlying problem. When prompts contained ambiguous language, the system often produced alternate interpretations. For instance, pairing an image of a doorknob with the prompt "alert" led the model to interpret both "being alert of the round doorknob" and "alert sound," reflecting the multiple meanings of the term. In such cases, the lack of contextual information or specificity led to alternate interpretations. The analysis suggests that at least one input modality—image or text—must provide sufficient clarity for accurate diagnosis. These observations informed the clarity levels summarized in Table 2.

**Implications:** Across the four clarity metrics, we observe that context combined with the target grounds the problem domain. Context can emerge from the text and visual cues in the image; for example, a hand grasping a doorknob with "I have arthritis" constrains the user's goal and intent within accessibility. As the object becomes visually focused, the target is clearer, and pairing it with the user's goal further narrows diagnosis. At least either context or goal must exhibit high clarity, since users may express challenges without knowing solutions exist. Specificity, however, is inevitably low for novices, as design requirements are discovered iteratively. For example, a Braille bottle label may initially miss constraints

like bottle height limiting character length, which the system must surface in later design stages through user engagement.

When the goal is explicit but the object is not (e.g. "I want this to be automatic"), the system should use scene analysis to surface multiple candidate objects and map them against the intention. For example, in a kitchen photo, "automatic" could apply to a trash bin (smart lid), a faucet (touchless control), or lights (motion sensor). This requires the system not only to detect objects but also to reason about which solution can plausibly satisfy the expressed intent. When the target object is identifiable but the intent is underspecified (e.g., a doorknob image with "how to improve"), the system must scaffold possible goals across domains. For the doorknob, this may include accessibility (lever adapter), energy harvesting (knob-powered generator), or smart home automation (motorized lock). While object clarity without intent or context can enable cross-domain solution exploration, the system must avoid overwhelming users by prioritizing goals that are most plausible. When the context is apparent but neither goal nor object is specified (e.g., "make this accessible" with a kitchen scene), the system should enumerate candidate objects within that domain, such as cabinet handles, faucets, or switches, and tie each to accessible augmentation strategies. Here, the domain grounds the search space, but the system must help users articulate which target object aligns best with their need. When intent, object, and context are all ambiguous (e.g., a cluttered scene with the prompt "annoying"), the system should begin by actively requesting clarification. Once a domain is established (e.g., domain → accessibility), the system can scaffold the others iteratively by proposing candidate targets and goals, converging toward a specific solution.

## 5.4 Solution Discovery

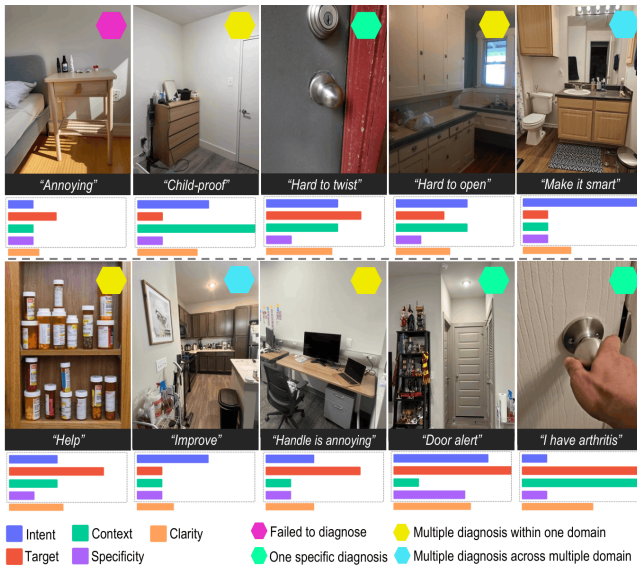
This section evaluates whether EUREXA can go beyond diagnosing interaction issues to retrieving relevant solutions.

**Method:** Following the diagnosis step, we assessed how effectively EUREXA could fetch augmentation solutions. The system was tested on 35 cases combining textual prompts and images, spanning varied levels of input clarity. We review the constructed user input scenarios with differing complexity and clarity in Table 3.

**Results:** Across 35 cases, EUREXA failed to provide any solution in 3 cases, produced multiple valid solutions in 21 cases, generated a single specific solution in 9 cases, and in 2 cases suggested solutions that were not relevant to the underlying issue. Failures occurred when very low input clarity caused incorrect diagnosis, when the

RAG pipeline returned no candidates, or when semantic search retrieved textually similar but contextually mismatched content that produced irrelevant designs. These outcomes highlight both the promise and current limits of the system: while it can expand the solution space in most scenarios, retrieval quality still depends on search methods, motivating future work on improved semantic and graph-based search for stronger contextual precision.

**Table 3: Image-prompt pairs with varying clarity scores, where differences in intent, target, context, and specificity lead to one or multiple within- or cross-domain diagnoses.**



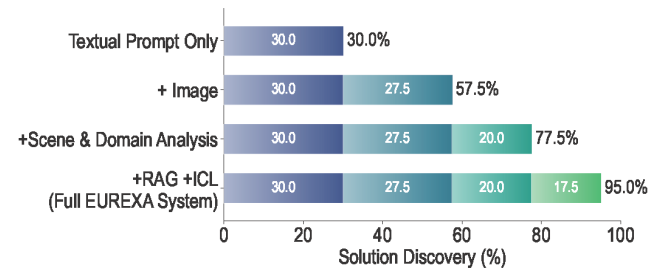
## 5.5 Ablation Study

To understand the contribution of each component of EUREXA, we incrementally included crucial components in EUREXA and reported the corresponding performance with respect to solution discovery.

**Method:** We defined four configurations. The first, “Textual Prompt Only,” used only text to test how well a base model surfaces augmentation ideas. The second added images to examine whether visual grounding improves solution discovery. In the third configuration, we introduced domain and scene analysis, enabling the system to reason about object functions, spatial constraints, and contextual cues. Finally, in the fourth configuration, we evaluated the full EUREXA pipeline, which further includes RAG, ICL, and structured prompt engineering. For each configuration, we marked a result as “yes” if the system produces a relevant, scenario-appropriate, non-hallucinated, 3D printable augmentation solution, and “no” otherwise. We used 20 prompts with varying clarity scores and ran each twice, yielding 40 test cases per configuration.

**Results:** Text only prompts surface relevant solutions in 30% of cases (Figure 16). Adding images increases performance to 57.5%, though this configuration often yields vague, hallucinated, or non-printable ideas. Including domain and scene analysis improves results to 77.5% with more scenario-specific and context grounded

solutions. The full EUREXA system, with RAG, ICL, and prompt engineering, achieves the highest performance at 95%, consistently generating fabrication-ready, non-hallucinated designs aligned with the intended scenario. We also conducted a qualitative assessment by regenerating all five designs in Section 4 (lever handle, braille label, outlet cover, and self-powered and actuation mechanisms) twice without RAG, ICL, and prompt engineering to examine whether geometry and fabricability were preserved. This ablated setup produced only one valid design, the Braille label, while the full EUREXA pipeline reliably generated correct, fabricable outputs in all cases, showing higher accuracy and consistency.



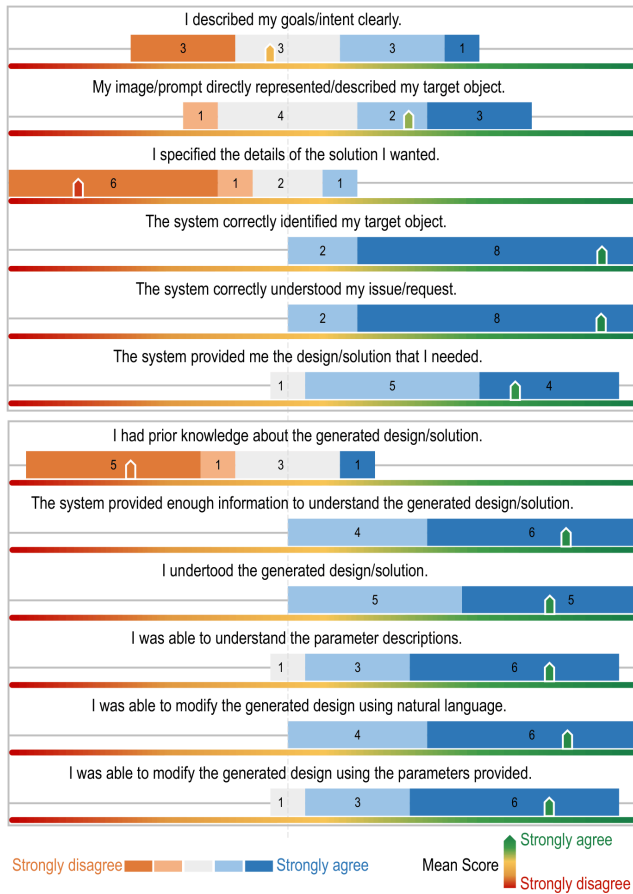
**Figure 16: Ablation study results across four configurations.**

## 5.6 User Study

We conducted a behavioral study examining how users interacted with the Eureka. The study surfaced insights into usability, communication flow, and how effectively the system supported users during the design process. The study has been approved by the Institution’s Review Board (IRB: STUDY2025-1047).

**Method (Participants & Procedure):** We recruited participants from diverse backgrounds ( $N = 10$ ) to evaluate EUREXA. Participants represented a wide range of technical backgrounds, allowing us to observe how the system supports users ranging from average to experienced. In terms of 3D printing expertise, six participants reported little to no experience (ratings 1–2/5), one reported moderate familiarity (3/5), two had intermediate skills (4/5), and one was an expert (5/5). Their modeling experience mirrored their 3D printing familiarity, with six participants reporting little or no prior experience (ratings 1–2/5). The remaining four participants reported moderate to advanced familiarity (ratings 3–5/5).

The facilitator introduced EUREXA and its interface before the task. Each participant selected an everyday interaction challenge and provided a corresponding image. Eight participants supplied photos they had taken of the target, while two used images downloaded online. With EUREXA, they created scenario specific augmentations, reviewed parameters, modified designs via the GUI and natural language, and asked details about functionality and recommended print settings. Across sessions, participants produced seven passive mechanisms (e.g., actuation or constraint-based add-ons), one self-powered device, and two robotic mechanisms. Afterward, they completed a short survey and open-ended questionnaire to reflect on their experience, evaluate usability, and describe challenges or insights that emerged during the design process.



**Figure 17: User evaluation of EUREXA, indicating high usability and strong support for augmentation tasks.**

**Quantitative Analysis:** Participants’ self-evaluations indicated that many had not clearly articulated their goals or solution details at the outset (Figure 17). Majority strongly disagreed that they had specified their goals or intentions. Despite this, participants agreed that EUREXA effectively interpreted their inputs: all participants reported that it accurately identified the target object and understood the underlying issue. Most also agreed that EUREXA generated a design that aligned with their needs. Although participants reported limited prior familiarity with the mechanisms produced, they felt that EUREXA offered sufficient information for them to understand the resulting design. The parameter descriptions were clear, and participants were able to modify the design using both natural-language instructions and the provided parameters.

**Observation of the User Walkthrough: From Improving Accessibility to Smart Home Automation.**

Across the study, participants used EUREXA to address diverse, personal design needs as Figure 18 showcases some examples. P3, a guitarist wanted speed picking but struggled with sweaty hands and wanted better pick grip. EUREXA suggested a dotted pick and a TPU finger sleeve; he chose the dotted pick, then added a text logo and adjusted the thickness and length. P6 aimed to help his



**Figure 18: Examples of input images and textual prompts of the participants and generated solutions.**

father with tremors eat independently; EUREXA initially surfaced several assistive concepts, and after he mentioned wanting “a bowl stabilizer” that he had seen somewhere, the system generated a food-bowl gimbal mechanism. He checked the online order history to respond to dimension clarification and provided the diameter for resizing. P7 uploaded an image of a ceiling light missing its puller, with a shortened chain that forced her to jump to reach it. She prompted “these strings keep breaking every time”, and EUREXA inferred the issue and generated a pull-chain extender. P8 struggled to turn a small rotary lamp switch due to a thumb injury, so a wing-shaped twister and a paddle handle was proposed; choosing the more aesthetic winged design, he added surface texture. P9 explored ideas for her indoor plant, with suggestions ranging from energy-harvesting water wheels to robotic care attachments. As she liked an automatic watering bottle, EUREXA referred to standard 16 oz bottle dimensions and increased part thickness for durability.

**Findings**

*#1. Supporting Users in Articulating Needs and Exploring Possibilities:* Participants reported that EUREXA helped them move from vague

or underspecified ideas toward clearer design directions, noting it “Suggested solutions I did not think of (P2)” and “gave me a new design according to my further need. (P10)” Even participants who struggled to articulate their intent accoladed EUREXA about how it compensated for missing details and guided users through viable options. Others appreciated the iterative interaction process, stating “I can ask further modification to the design (P9)” and “it understood my additional instructions. (P4)”

**#2. Applicability Across Diverse Use Cases:** Participants imagined wide-ranging applications for EUREXA, extending beyond the scenarios explored in the study. They mentioned uses for household repairs, engineering tasks, construction-related needs, and general 3D printing hobbies. Comments included: “I would consider using this to solve my daily life problems, (P4)” “this would help me design equipment I need although I have little knowledge on 3D modeling (P9)” and “Want to use this (EUREXA) in construction sciences.” These responses indicate EUREXA’s potential to support heterogeneous, domain-spanning augmentation tasks.

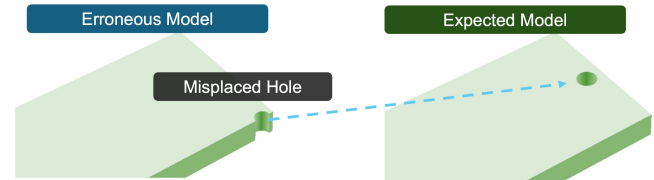
**#3. Need for Stronger Feedback and Visualization:** The richer, more responsive visual feedback during model exploration was consistently desired. “The sliders should highlight which part of the design is being changed, (P7)”, “it would be better if there were visualizations when proposing different solutions. (P3)” Some participants suggested advanced features such as assembly previews or showing how a printed part would appear in real physical settings when installed. These responses highlight opportunities for enhancing context-rich visualization given input images, possibly with interactive simulations displaying how they would operate. We expect that the current progression on generative images can help visualize the augmentation mechanisms in real-life, and augmented reality can be explored for in-situ visualization.

**#3. Clarity and Decision-Making Support:** Across the ten cases, EUREXA produced multiple solutions per scenario, and participants used different heuristics to choose among them. Several rated that the solution “was similar to what I had in mind, (P5)” while others chose solutions based on perceived functionality by the guesswork, “the one which I think has best long-term usability (P10)” or what “sounded like it would work best. (P4)” One participant used perceived aesthetics as the selection criteria (P8). These responses suggest that users relied on intuition or personal heuristics when comparing options. At the same time, participants expressed a desire for stronger decision-making support. “It was based on guessing; it could have been better with pros and cons for each solution. (P1)”, indicating that clearer explanatory scaffolding, such as comparative descriptors, visualization, or concise trade-off summaries, may better support users in designing augmentation solutions.

**#4. Sense Making of the Parameters for Customization:** In general, participants appreciated the system as intuitive and easy to use, as EUREXA helped them work with design parameters. “Modifying the parameters was easy and intuitive” (P2), “the explanations were easy to understand” (P9). P9 also noted “It (EUREXA) tells me what happens if I change something.” Participants also appreciated that they could further refine the design through natural language, stating “I can ask for further modifications to the design. (P10)”

**#5. Guidance Beyond the 3D Model Generation:** Design support beyond the parameter tuning seemed emergent. Several users wanted

help understanding how to proceed when EUREXA-generated mechanisms required additional components, such as electronics, actuation, or hardware. One participant asked how to “progress forward if the mechanism needs some electronics, (P9)” indicating an advanced need for step-by-step guidance. These suggest a natural extension of EUREXA: future versions could list required electronics or actuation components, suggest compatible parts available online, or even provide example control code for completing functional, much complex mechanisms for emerging users.



**Figure 19: Erroneous vs. expected model, where the generated output misplaced a hole relative to the intended design.**

## 5.7 Design Fabricability

Due to the time constraints, all 10 user-generated designs were 3D printed. Participants reported that the designs matched their intentions, and two authors conducted an expert review as experienced 3D modelers. Of the 10 successfully printed and installed designs (Figure 18), one contained a geometric error: an off-center hole in a stand from the multi-part gimbal design, which the user also noticed during generation. Further analysis showed that while the underlying CSG logic was correct, the MLLM omitted a critical parameter (e.g., `center=true`) in the generated code (Figure 19). This likely reflects limited model exposure to the strict parameter requirements of OpenSCAD compared to common languages such as Python. Future work may explore alternative parametric backends (e.g., SolidPython), model fine-tuning with richer CAD corpora, and tolerance-bearing features to mitigate fabrication uncertainty [22].

## 6 Discussion & Future Work

### 6.1 Limitations

The RAG pipeline is constrained by current MLLM token limits, restricting long prompts and multi-turn context, though this is expected to ease as context windows expand. System performance also depends on the quality of the underlying knowledge base, with missing or poorly documented designs degrading results. While clarification strategies mitigate vague input, highly ambiguous intent can still lead to incomplete outputs. Moreover, parametric generation relies on the MLLM’s ability to structure code correctly, and complex mechanical or domain-specific constraints may require expert intervention. The current system focuses on object augmentation and does not yet support large-scale fabrication, multi-material design, or broader hardware constraints. Future work includes fine-tuning MLLMs for parametric generation to improve reliability and support more complex designs.

## 6.2 Interactive Reconstruction of Semantically Aware Parametric Designs

The CAD-Recode module is not yet fully semantically aware, often missing user-critical dimensions and offering limited support for contextual adaptation. To address this, we prototyped a user-in-the-loop mesh-to-parametric workflow that enables interactive reconstruction into customization-ready OpenSCAD code with semantic annotations (Figure 20). The process imports an STL mesh as a reference silhouette (a), supports incremental reconstruction with primitives and Booleans (b–c, e), and provides a live OpenSCAD preview that generates parameterized code and semantic comments in real time (d). The system outputs editable OpenSCAD files and a JSON design log for remixing and fabrication (f). While this demonstrates the feasibility of semantically aware parametric generation, it requires substantial CAD expertise. Future work will integrate agentic support to better scaffold novice users.

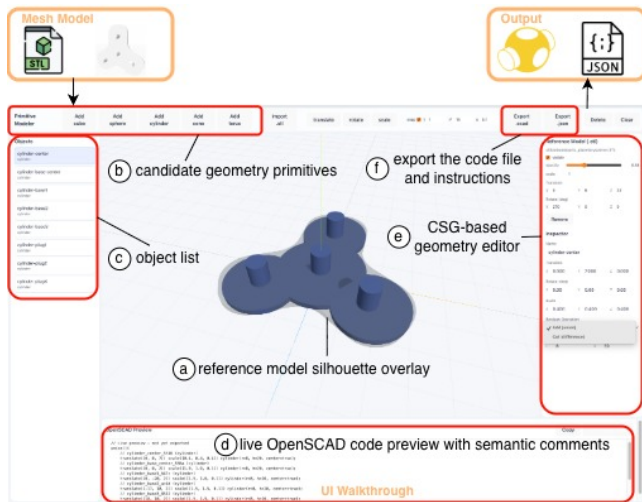


Figure 20: Mesh-to-parametric pipeline walkthrough.

## 6.3 Integration with Augmented Reality

Recent AR-based systems such as pARam [40] and TinkerXR [3] demonstrate how XR enables in-situ parametric modeling. Building on this, EUREXA could be extended into an AR-enhanced workflow that leverages spatial awareness and in-situ feedback for augmentation-oriented design. Prior work on object-centered multimodal interfaces [16] further suggests opportunities for richer interactions. An AR interface could capture real-world geometry for context-aware input, reduce manual adjustment, and support in-environment previews of form, material, and color. It may also enable temporal reasoning from longer videos to detect interaction challenges. We envision AR integration as a natural extension toward more intuitive, object-aware design workflows.

## 6.4 Lowering Barriers to Research Access

A primary objective of EUREXA is to reduce the barriers that prevent everyday users from accessing and applying design innovations developed within the research community. While academic

literature often introduces novel fabrication techniques and interaction mechanisms, these contributions are typically locked behind technical jargon, lack of implementation support, or static, non-adaptable design artifacts. EUREXA aims to address this gap by providing an automated pipeline that retrieves relevant research articles, extracts actionable design knowledge, and translates it into customizable parametric models. Through multimodal input handling, RAG, and dynamic design editing, EUREXA enables non-expert users to benefit from advanced design solutions without requiring specialized skills or familiarity with academic literature.

## 6.5 Encouraging Parametric Design Sharing

Beyond providing user access to research-backed solutions, EUREXA also emphasizes the importance of how research outputs are shared and disseminated. We found many publications lack accessible parametric design files, limiting their practical applicability and reusability. To fully realize fabrication research’s potential, we encourage future researchers to adopt dissemination practices that prioritize sharing well-documented, parametric, and modular design files alongside their papers. EUREXA demonstrates the value of such practices, and we advocate making parametric design sharing a standard in open, responsible digital fabrication research.

## 6.6 Scalability Across Domains & Environments

While EUREXA focuses on physical object augmentation and digital fabrication, its RAG framework, structured prompting, and end-user customization extend to domains such as sustainability, education, and robotics by surfacing latent challenges and providing accessible solutions. Its human-in-the-loop workflow supports applications including architectural customization, product prototyping, and DIY repair, and its scalable architecture can expand to areas such as healthcare devices and safety-critical environments.

Beyond indoor settings, EUREXA can support accessibility augmentations in outdoor and public spaces, such as tactile markers, modular ramps, and grip extenders. These contexts introduce durability challenges from weather and temperature variation, which future work will address to enable robust, context-specific designs.

## 7 Conclusion

Everyday interactions with physical objects often call for small but meaningful augmentations that personal fabrication could address, yet remain inaccessible to non-experts. EUREXA provides an end to end toolkit to discover, generate, understand, and customize 3D printable solutions grounded in user contexts. By guiding users from vague intent to fabrication ready designs, it lowers barriers to everyday augmentation. For community-driven development, we aim to open-source EUREXA’s core components, including multi-agent pipeline, knowledge-injection module, and dynamic GUI at <https://github.com/abulalarabi/EUREXA>.

## Acknowledgments

This work was supported by the National Science Foundation (NSF) under awards IIS-2340120 and CMMI-2221436. Shu Kong was partially supported by the Science and Technology Development Fund of Macau (0067/2024/ITP2), the University of Macau (SRG2023-00044-FST), and the Institute of Collaborative Innovation.

## References

- [1] Abul Al Arabi, Jiahao Li, Xiang'Anthony Chen, and Jeeun Kim. 2022. Mobiot: Augmenting everyday objects into moving iot devices using 3d printed attachments generated by demonstration. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–14.
- [2] Abul Al Arabi, Xue Wang, Yang Zhang, and Jeeun Kim. 2023. E3D: Harvesting Energy from Everyday Kinetic Interactions Using 3D Printed Attachment Mechanisms. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 7, 3 (2023), 1–31.
- [3] Oğuz Arslan, Artun Akdoğan, and Mustafa Doga Dogan. 2025. TinkerXR: In-Situ, Reality-Aware CAD and 3D Printing Interface for Novices. In *Proceedings of the ACM Symposium on Computational Fabrication (SCF '25)*. Association for Computing Machinery, New York, NY, USA, Article 18, 19 pages. doi:10.1145/3745778.3766651
- [4] Artifex Software, Inc. 2025. Ghostscript. <https://ghostscript.com/>. [Online; accessed 2025-09-11].
- [5] Akari Asai, Zeqiu Wu, Yizhong Wang, Avirup Sil, and Hannaneh Hajishirzi. 2023. Self-rag: Learning to retrieve, generate, and critique through self-reflection. In *The Twelfth International Conference on Learning Representations*.
- [6] Patrick Baudisch, Stefanie Mueller, et al. 2017. Personal fabrication. *Foundations and Trends® in Human-Computer Interaction* 10, 3–4 (2017), 165–293.
- [7] Alexander Berman, Francis Quek, Robert Woodward, Osazuwa Okundaye, and Jeeun Kim. 2020. “Anyone Can Print”: Supporting Collaborations with 3D Printing Services to Empower Broader Participation in Personal Fabrication. In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*. 1–13.
- [8] Sebastian Borgeaud, Arthur Mensch, Jordan Hoffmann, Trevor Cai, Eliza Rutherford, Katie Millican, George Bm Van Den Driessche, Jean-Baptiste Lespiau, Bogdan Damoc, Aidan Clark, et al. 2022. Improving language models by retrieving from trillions of tokens. In *International conference on machine learning*. PMLR, 2206–2240.
- [9] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems* 33 (2020), 1877–1901.
- [10] Dan Cascaval, Mira Shalah, Phillip Quinn, Rastislav Bodik, Maneesh Agrawala, and Adriana Schulz. 2022. Differentiable 3d cad programs for bidirectional editing. In *Computer Graphics Forum*, Vol. 41. Wiley Online Library, 309–323.
- [11] Angel X Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, et al. 2015. Shapenet: An information-rich 3d model repository. *arXiv preprint arXiv:1512.03012* (2015).
- [12] Kevin Chen, Christopher B Choy, Manolis Savva, Angel X Chang, Thomas Funkhouser, and Silvio Savarese. 2019. Text2shape: Generating shapes from natural language by learning joint embeddings. In *Computer Vision—ACCV 2018: 14th Asian Conference on Computer Vision, Perth, Australia, December 2–6, 2018, Revised Selected Papers, Part III 14*. Springer, 100–116.
- [13] Xiang'Anthony' Chen, Jeeun Kim, Jennifer Mankoff, Tov Grossman, Stelian Coros, and Scott E Hudson. 2016. Reprise: A design tool for specifying, generating, and customizing 3D printable adaptations on everyday objects. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*. 29–39.
- [14] Matt Deitke, Ruoshi Liu, Matthew Wallingford, Huong Ngo, Oscar Michel, Aditya Kusupati, Alan Fan, Christian Laforte, Vikram Voleti, Samir Yitzhak Gadre, et al. 2023. Objaverse-xl: A universe of 10m+ 3d objects. *Advances in Neural Information Processing Systems* 36 (2023), 35799–35813.
- [15] Himani Deshpande, Haruki Takahashi, and Jeeun Kim. 2024. Unmake to Remake: Materiality-Driven Rapid Prototyping. *ACM Transactions on Computer-Human Interaction* 31, 6 (2024), 1–31.
- [16] Mustafa Doga Dogan, Eric J Gonzalez, Karan Ahuja, Ruofei Du, Andrea Colaco, Johnny Lee, Mar Gonzalez-Franco, and David Kim. 2024. Augmented object intelligence with xr-objects. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. 1–15.
- [17] Faraz Faruqi, Ahmed Katary, Tarik Hasic, Amira Abdel-Rahman, Nayeemur Rahman, Leandra Tejedor, Mackenzie Leake, Megan Hofmann, and Stefanie Mueller. 2023. Style2Fab: Functionality-aware segmentation for fabricating personalized 3d models with generative ai. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–13.
- [18] Faraz Faruqi, Maxine Perroni-Scharf, Jaskaran Singh Walia, Yunyi Zhu, Shuyue Feng, Donald Degraen, and Stefanie Mueller. 2025. TactStyle: Generating Tactile Textures with Generative AI for Digital Fabrication. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–16.
- [19] Haoxiang Guo, Shilin Liu, Hao Pan, Yang Liu, Xin Tong, and Baining Guo. 2022. Complexgen: Cad reconstruction by b-rep chain complex generation. *ACM Transactions on Graphics (TOG)* 41, 4 (2022), 1–18.
- [20] MakerBot Industries. 2008. Thingiverse. <http://www.thingiverse.com/>
- [21] Wendy Ju and Larry Leifer. 2008. The design of implicit interactions: Making interactive systems less obnoxious. *Design Issues* 24, 3 (2008), 72–84.
- [22] Jeeun Kim, Anhong Guo, Tom Yeh, Scott E Hudson, and Jennifer Mankoff. 2017. Understanding uncertainty in measurement and accommodating its impact in 3D modeling and printing. In *Proceedings of the 2017 conference on designing interactive systems*. 1067–1078.
- [23] Jeeun Kim, Clement Zheng, Haruki Takahashi, Mark D Gross, Daniel Ashbrook, and Tom Yeh. 2018. Compositional 3D printing: expanding & supporting workflows towards continuous fabrication. In *Proceedings of the 2nd Annual ACM Symposium on Computational Fabrication*. 1–10.
- [24] Milin Kodnongbua, Benjamin T Jones, Maaz Bin Safer Ahmad, Vladimir G Kim, and Adriana Schulz. 2023. Zero-shot CAD Program Re-Parameterization for Interactive Manipulation. *arXiv preprint arXiv:2306.03217* (2023).
- [25] Nahyun Kwon, Qian Lu, Muhammad Hasham Qazi, Joanne Liu, Changhoon Oh, Shu Kong, and Jeeun Kim. 2024. Accessless: Auto-detecting inaccessibility of everyday objects. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–17.
- [26] Harris Kyriakou, Jeffrey V Nickerson, and Gaurav Sabnis. 2017. Knowledge reuse for customization. *MIS quarterly* 41, 1 (2017), 315–332.
- [27] Jaewook Lee, Andrew D Tjahjadi, Jiho Kim, Junpu Yu, Minji Park, Jiawen Zhang, Jon E Froehlich, Yapeng Tian, and Yuhang Zhao. 2024. CookAR: Affordance augmentations in wearable AR to support kitchen tool interactions for people with low vision. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. 1–16.
- [28] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, et al. 2020. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in neural information processing systems* 33 (2020), 9459–9474.
- [29] Jiahao Li, Jeeun Kim, and Xiang'Anthony' Chen. 2019. Robiot: A design tool for actuating everyday objects with automatically generated 3D printable mechanisms. In *Proceedings of the 32nd annual ACM symposium on user interface software and technology*. 673–685.
- [30] Mixuan Li and Leila Aflatoony. 2023. A Parametric Design Tool to Support Customized Adaptations of Assistive Technologies for People with Hand Impairments. In *Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility*. 1–4.
- [31] Chen Liang, Anhong Guo, and Jeeun Kim. 2022. CustomizAR: Facilitating interactive exploration and measurement of adaptive 3d designs. In *Proceedings of the 2022 ACM Designing Interactive Systems Conference*. 898–912.
- [32] lucasoshiro. 2023. Braille Generator. <https://www.thingiverse.com/thing:6463849>. [Online; accessed 2025-08-07].
- [33] Rebecca L Miller. 2015. Rogers' innovation diffusion theory (1962, 1995). In *Information seeking behavior and technology adoption: Theories and trends*. IGI Global Scientific Publishing, 261–274.
- [34] MIT CSAIL. n.d. Personal Fabrication Research HCI and Graphics. <https://hcie.csail.mit.edu/fabpub/>. [Online; accessed 2025-03-28].
- [35] rcolyer. 2016. Ergonomic Door Knob Lever, Fully Parametric. <https://www.thingiverse.com/thing:1695541>. [Online; accessed 2025-08-07].
- [36] Dania Rukhovich, Elona Dupont, Dimitrios Mallis, Kseniya Cherenkova, Anis Kacem, and Djamilia Aouada. 2025. Cad-recode: Reverse engineering cad code from point clouds. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 9801–9811.
- [37] sauer. 2015. Medication Bottle Opener. <https://www.thingiverse.com/thing:1053181>. [Online; accessed 2025-08-11].
- [38] Ruwen Schnabel, Roland Wahl, and Reinhard Klein. 2007. Efficient RANSAC for point-cloud shape detection. In *Computer graphics forum*, Vol. 26. Wiley Online Library, 214–226.
- [39] Chris Spencer. 2025. chrisspen/gears: OpenSCAD Gear Generator. <https://github.com/chrisppen/gears>. [Online; accessed 2025-08-07].
- [40] Evgeny Stemasov, Simon Demharter, Max Rädler, Jan Gugenheimer, and Enrico Rukzio. 2024. pARam: Leveraging Parametric Design in Extended Reality to Support the Personalization of Artifacts for Personal Fabrication. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (Honolulu, HI, USA) (CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 337, 22 pages. doi:10.1145/3613904.3642083
- [41] Xia Su, Han Zhang, Kaiming Cheng, Jaewook Lee, Qiaochu Liu, Wyatt Olson, and Jon E Froehlich. 2024. RASSAR: Room Accessibility and Safety Scanning in Augmented Reality. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–17.
- [42] Unstructured-IO. 2025. Unstructured: Convert Documents to Structured Data. <https://github.com/Unstructured-IO/unstructured>. [Online; accessed 2025-09-12].
- [43] Mikaela Angelina Uy, Yen-Yu Chang, Minhyuk Sung, Purvi Goel, Joseph G Lambourne, Tolga Birdal, and Leonidas J Guibas. 2022. Point2cyl: Reverse engineering 3d objects from point clouds to extrusion cylinders. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 11850–11860.
- [44] Xingzhi Wang, Zhoumingju Jiang, Yi Xiong, and Ang Liu. 2025. Human-LLM collaboration in generative design for customization. *Journal of Manufacturing Systems* 80 (2025), 425–435.

- [45] Jerry Wei, Jason Wei, Yi Tay, Dustin Tran, Albert Webson, Yifeng Lu, Xinyun Chen, Hanxiao Liu, Da Huang, Denny Zhou, et al. 2023. Larger language models do in-context learning differently. *arXiv preprint arXiv:2303.03846* (2023).
- [46] Jacob O Wobbrock. 2019. Situationally-induced impairments and disabilities. *Web Accessibility: A Foundation for Research* (2019), 59–92.
- [47] Donghan Yu, Chenguang Zhu, Yuwei Fang, Wenhao Yu, Shuohang Wang, Yichong Xu, Xiang Ren, Yiming Yang, and Michael Zeng. 2021. Kg-fid: Infusing knowledge graph in fusion-in-decoder for open-domain question answering. *arXiv preprint arXiv:2110.04330* (2021).
- [48] Zhuohao Zhang, Haichang Li, Chun Meng Yu, Faraz Faruqi, Junan Xie, Gene SH Kim, Mingming Fan, Angus Forbes, Jacob O Wobbrock, Anhong Guo, et al. 2025. A11yShape: AI-Assisted 3-D Modeling for Blind and Low-Vision Programmers. In *Proceedings of the 27th International ACM SIGACCESS Conference on Computers and Accessibility*. 1–20.
- [49] Clement Zheng, Zhen Zhou Yong, Hongnan Lin, HyunJoo Oh, and Ching Chiuan Yen. 2022. Shape-haptics: planar & passive force feedback mechanisms for physical interfaces. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–15.