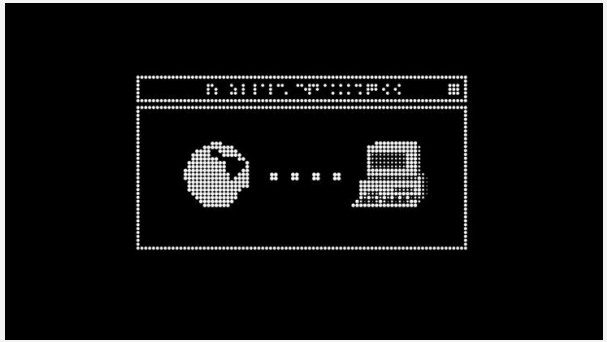


Capable but Not Scalable

Requirements and Interaction Costs of Building Multi-View CAD Datasets for Design Research

Danyal Sarfraz
Researcher, HCDT, UNIST

SeEun Lee
Ulugbek Ismatullaev
JoungHyun Kim
Abdurakhimov Donierbek
Graduate Students, HCDT, UNIST

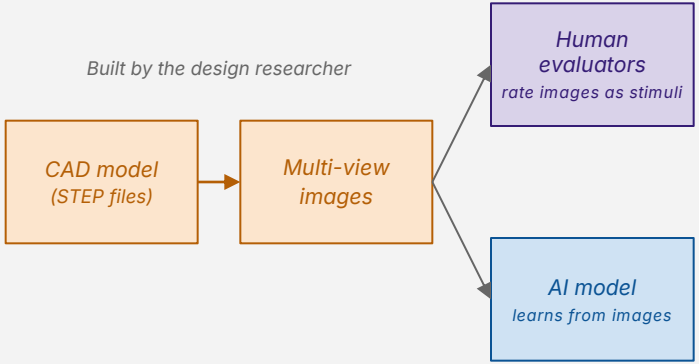
KwanMyung Kim
Professor, Department of Human-Centered Design & Technology, UNIST

From CAD Models to Image Datasets

Image-based AI is now a practical instrument in product design research: models predict aesthetic appeal, recognize 3D shapes from rendered views, and generate CAD from text prompts.

These models depend on datasets, and CAD models are an attractive source: design teams already produce them.

*Turning CAD into multi-view images is **data work**, often undervalued, yet model quality depends on it (Sambasivan et al., 2021).*



*Two consumers, one rendering pipeline: every choice made while rendering reaches **both**.*

Rendering Is Not Neutral

Every rendering choice reaches both the model and the human rater. Uncontrolled choices become hidden variables in the dataset.

Our argument: *for design researchers, the problem lies less in what tools can do than in* ***how the work scales****.*

Almost every function a controlled dataset needs exists, but it is offered either through code or as a manual step repeated for every product.

1

Sizes not normalized

A model may learn that small parts look small, rather than learning their form.

2

“Front” points at a different side

Neither a model nor a rater compares like with like across products.

3

Appearance tracks the file source

Shortcut learning (Geirhos et al., 2020); ratings mix form with lighting and material.

Research Questions

	Question	Contribution	
RQ1	*How do current studies convert 3D and CAD models into multi-view images for AI, and what does each approach demand of the person operating it?*	**10** sources	*Analysis of reported rendering practice*
RQ2	*What requirements should a rendering pipeline satisfy so that design researchers can produce datasets that are both AI-ready and meaningful for design research?*	**14** requirements	*Requirements for the dataset and the interaction*
RQ3	*To what extent can design researchers meet these requirements with existing tools without writing code, and which steps remain manual for each product?*	**11** tools	*Analytic comparison through cognitive dimensions*

Three Lines of Prior Work

View-based learning

*Su et al. (2015): a network trained on **12 or 80 rendered views** recognizes 3D shapes better than 3D descriptors.*

*Design adds a need: corresponding views must have **consistent semantic meaning** across products, so front and side stay comparable.*

Why multi-view images

CAD data & dataset quality

*Arjomandi Rad et al. (2024): **quality assurance and harmonization** are recurring barriers to dataset creation in design research.*

*Datasheets for datasets (Gebru et al., 2021): documentation criteria that apply to the **CAD-to-image stage**.*

What a dataset must be

Analyzing tools without users

Cognitive dimensions (Green & Petre, 1996; Green & Blackwell, 1998).

Threshold and ceiling (Myers et al., 2000).

Six learning barriers of end-user programming (Ko et al., 2004).

Our analytic lens

An HCI Lens for Rendering Tools

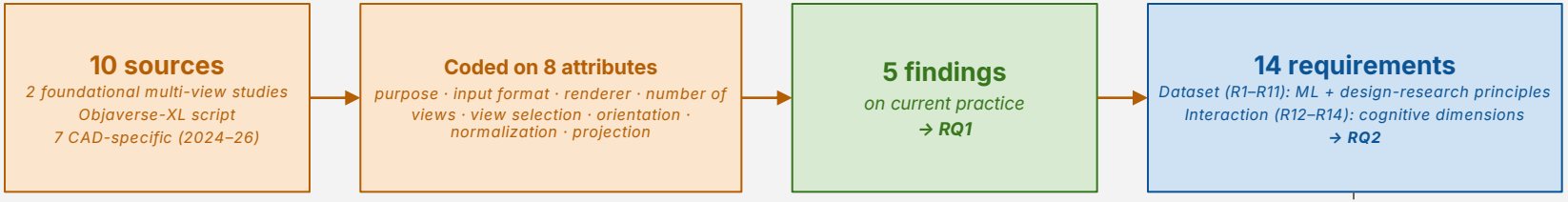
HCI offers analytic methods for judging tools before users are studied. We use them to read each tool against one reference task.

Note: *Myers et al. (2000) observe that successful tools have tended to be* ***low-threshold and low-ceiling, or high in both****.*

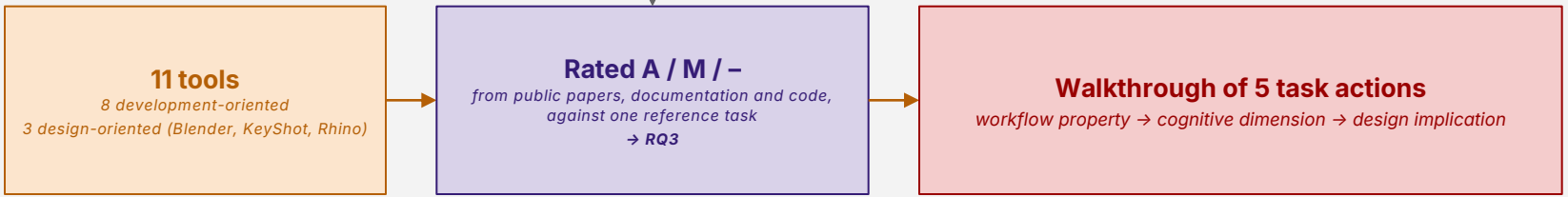
Concept	Meaning
Threshold / ceiling	*How hard a tool is to learn / how much can be done with it*
Repetition viscosity	*Resistance to change when one change must be repeated in many places*
Hidden dependencies	*Relationships the tool does not show*
Visibility & progressive evaluation	*Seeing the work, and checking it as it develops*
Closeness of mapping	*How directly the tool's terms match the task*
Error-proneness	*How easily the tool invites slips*
Learning barriers	*Six barriers people face when programming (Ko et al., 2004)*

Research Methods

Step 1 — Practice analysis & requirements



Step 2 — Tool comparison & interaction cost



Requirements are candidate criteria for the reference task, not an exhaustive account of researchers' needs.

Reference Task & Rating Scheme

Assumed user: *can configure documented interfaces and run an existing script, but does not write or modify code, including visual programs.*

Ratings describe accessibility for this user*, not the tool's overall technical capability. A multi-part requirement is rated by its least supported part.*

A	*Automatic for every product once configured*
M	*Manual step repeated per product, or a workaround*
–	*Needs writing / modifying code, or unsupported*

REFERENCE TASK

300 *STEP files of consumer products*

1. **Fixed view set**
   same named views for every product
2. **Clay baseline**
   material and color cues removed
3. **Normalize**
   uniform scale to the largest dimension
4. **Group by product**
   with provenance for every sample

Current Rendering Practice

Five findings from ten sources that render 3D or CAD models for AI

Finding 1	Finding 2	Finding 3	Finding 4	Finding 5
Code-only	**1 → 80**	**1 of 7**	**3 of 7**	**Per task**
Rendering is fragmented	**View protocols vary widely**	**Orientation is rarely reported**	**Normalization is rarely stated**	**Projection & appearance vary**
5 of 7 CAD sources start from STEP, yet renderers range from Blender and pythonOCC to PyVista; every inspected implementation is code.	*Views per object range from a single view or four-view composite to 80; selection from drawing views to random cameras.*	*Only Ortho2CAD reports a view convention (drafting), not a semantic front; public code renders models in their native CAD frame.*	*CADBench, CADFS and Ortho2CAD state a scaling rule; the other CAD-specific sources state none.*	*Orthographic for drawing-like tasks, perspective for photo-like ones; perspective distortion degraded VLM annotations.*

Consequence: *datasets rendered under different protocols are hard to pool or compare view by view, and reconciling them may require re-rendering.*

Fourteen Requirements

Dataset requirements (R1–R11): what the images must be

Interaction (R12–R14)

Input	Geometry	View protocol	Observation	Output	Interaction
R1 *Native STEP, size-relative meshing*	**R2** *Uniform scaling to a reference size*	**R5** *Fixed view set for all products*	**R7** *Orthographic and perspective*	**R10** *Product-grouped output and splits*	**R12** *Collection-level (batch) operation*
	R3 *Centering, framing, clipping checks*	**R6** *Shared framing within a product*	**R8** *Clay baseline, matched conditions*	**R11** *Per-sample provenance and scale*	**R13** *Reusable, shareable protocol preset*
	R4 *Semantic front and view labels*		**R9** *Aligned mask, depth, normal passes*		**R14** *Collection-wide checks and correction*

R4 *For a designer, "front" is a semantic property of a product, not a coordinate axis.*

R10 *If views of one product fall in both train and test sets, a model scores well by recognizing products it has already seen.*

Which Side Is the “Front”?

For a designer, the front is the face people look at: a semantic property of the product, not a fixed axis (R4).

	Product	Fixed “Front” camera (+X)	Designer’s front	Result
Speaker ***front face perpendicular to the ground*** *seen by a horizontal camera: the renderer’s “Front”*				**✓ Match** *The fixed Front camera already shows the face.*
Mouse ***front face parallel to the ground*** *seen from above: the renderer’s “Top”*				**× Mismatch** *The Front camera shows only the nose; the researcher relabels the Top view as the front.*

Same rule, different answer: *one fixed front axis cannot fit every product. The tool proposes a front;* ***the researcher confirms the view names.***

Consistency & Proportion

Three levels of consistency

Across products: *common scale (R2), centering (R3), semantic views (R4).*

Within a product: *shared framing preserves the relation between views (R6).*

Across appearance conditions: *geometry and cameras reused; lighting or material changed deliberately (R8).*

Proportion, not absolute size

A 5 mm fillet reads as soft on a 150 mm remote but almost sharp on an 800 mm cabinet. ***Scale uniformly*** *to a declared reference dimension;* ***store original size and scale factor*** *(R11).*



(a) *True size, one front camera: the mini speaker almost disappears; per-view framing makes the boombox look taller from the side.*

(b) *Uniform scaling to the same largest dimension, one framing for all views.*

Adapted from Figure 1, with renders from our dataset (sizes approximate)

Tool Comparison

31 *automatic* **27** *manual* **96** *code / –*

	Development-oriented								Design-oriented		
Requirement	**Blender-Proc2**	**Kubric**	**python-OCC**	**UnrealCV**	**Objaverse-XL**	**Omni-verse**	**CAD2-Render**	**Ortho2-CAD**	**Blender**	**KeyShot**	**Rhino 8**
Interface needed	*Py*	*Py*	*Py*	*UE+Py*	*Py*	*Py*	*Unity*	*CLI*	*GUI*	*GUI*	*GUI*
R1 *Native STEP, size-relative meshing*	–	–	–	M	–	A	–	A	–	A	A
R2 *Uniform scaling to a reference size*	–	–	–	–	A	–	M	A	M	M	M
R3 *Centering, framing, clipping checks*	–	–	–	–	–	–	M	A	M	M	M
R4 *Semantic front and view labels*	–	–	–	–	–	–	M	–	M	M	M
R5 *Fixed view set for all products*	–	–	–	–	–	–	–	A	A	A	A
R6 *Shared framing within a product*	–	–	–	–	–	–	A	A	A	A	A
R7 *Orthographic and perspective*	–	–	–	–	–	–	–	–	A	A	A
R8 *Clay baseline, matched conditions*	–	–	–	–	–	–	–	–	A	M	A
R9 *Aligned mask, depth, normal passes*	–	–	–	–	–	–	A	–	A	A	A
R10 *Product-grouped output and splits*	–	–	–	–	–	–	–	–	M	M	M
R11 *Per-sample provenance and scale*	–	–	–	–	–	–	–	–	M	M	M
R12 *Collection-level (batch) operation*	–	–	–	–	–	–	M	A	M	A	M
R13 *Reusable, shareable protocol preset*	–	–	–	–	–	–	A	–	A	A	A
R14 *Collection-wide checks and correction*	–	–	–	–	–	–	M	–	M	M	M
Count A / M	**0/0**	**0/0**	**0/0**	**0/1**	**1/0**	**1/0**	**3/5**	**6/0**	**6/7**	**7/7**	**7/7**

Highlighted rows: not automatic in any assessed workflow. *Ratings reflect a non-programming user, not overall tool capability. Adapted from Table 1.*

Three Routes, One Gap

Threshold and viscosity are different properties. Each route lacks at least one of the three properties a researcher needs.

Gap: *no assessed workflow combines a **low threshold** with **low repetition viscosity** while also allowing the **protocol to change without code**.*

Not automatic in any assessed workflow (non-programming user):

- R4 Front
- R10 Splits
- R11 Provenance
- R14 Checks

	Low threshold	Low repetition viscosity	Protocol changeable without code
Ready-made scripts *e.g., Ortho2CAD generator*	✓	✓	✗
Frameworks *BlenderProc2, Kubric, Omniverse, UnrealCV, pythonOCC*	✗	✓*	✗
Design tools *Blender, KeyShot, Rhino*	✓	✗	✓
What design research needs	✓	✓	✓

** Once a script exists, a change is a small edit and a rerun; writing that script is the barrier.*

Walkthrough of the Reference Task

Task action	Documented workflow property	HCI concept	Design implication
Set up views, passes and normalization	*Frameworks: Python calls. Design tools: a template scene.*	***Threshold; learning barriers***	*Preset with defaults; code access kept for extension (R13)*
Change the normalization rule	*Design tools: rescale each product again. Code: a small edit and rerun.*	***Repetition viscosity***	*Re-apply the revised preset to all products at once (R12, R13)*
Assign the front of each product	*A per-product judgment, entered as rotations in code or by hand.*	***Closeness of mapping***	*Proposed fronts confirmed by direct manipulation in a collection grid (R4, R14)*
Check framing and clipping	*One product at a time in design tools; after the run in scripts.*	***Visibility; progressive evaluation***	*Collection-wide preview that flags clipping and failed conversions (R3, R14)*
Record original size and scale	*Design tools: manual rescaling leaves no scale factor with the images.*	***Hidden dependencies***	*Write dimensions, scale factor, front and split to every sample by default (R10, R11)*

Adapted from Table 2. The walkthrough identifies sources of potential interaction cost; it does not measure effort.

Toward a Designer-Facing Tool

1. **Preset with defaults, code kept**
   Low threshold with a high ceiling (Myers et al., 2000).

2. **Proposed front, confirmed by the researcher**
   Mixed-initiative (Horvitz, 1999); "no meaningful front" allowed.

3. **Collection-wide checks**
   Flag clipping and failed conversions before rendering.

4. **Provenance by default**
   Source, units, transforms, cameras, protocol version, size, scale, front, split.

Note: *design hypotheses; effects on time, errors and workload remain to be evaluated.*

Our prototype: *a Blender add-on (work in progress)*

1. **Import** — R1
   STEP / IGES and mesh files, product by product

2. **Standardize** — R2 R3
   Recenter; scale the largest dimension to a fixed size

3. **Fixed view set** — R5 R6 R7
   6 orthographic + 3 perspective cameras, one framing

4. **Confirm the front** — R4
   Tool proposes; researcher confirms each product's view names

5. **Render & name** — R8 R10 R14
   Clay renders, batch pre-check, one folder per product

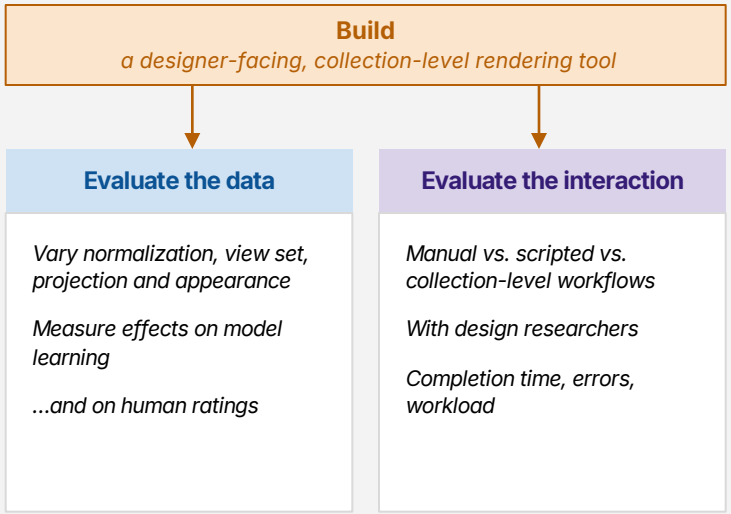
```
AA_01_Product/
AA_01_Product_Front_clay.png
```

Built on Blender, one of the assessed design tools: the add-on targets the steps it leaves manual for each product.

Limitations & Future Work

Limitations

- *Analytical study: ratings and walkthrough by the authors, on one reference task, without participants.*
- *M counts show where manual work remains, not measured effort.*
- *Ten practice sources, not drawn from a systematic search; ratings reflect public documentation only.*

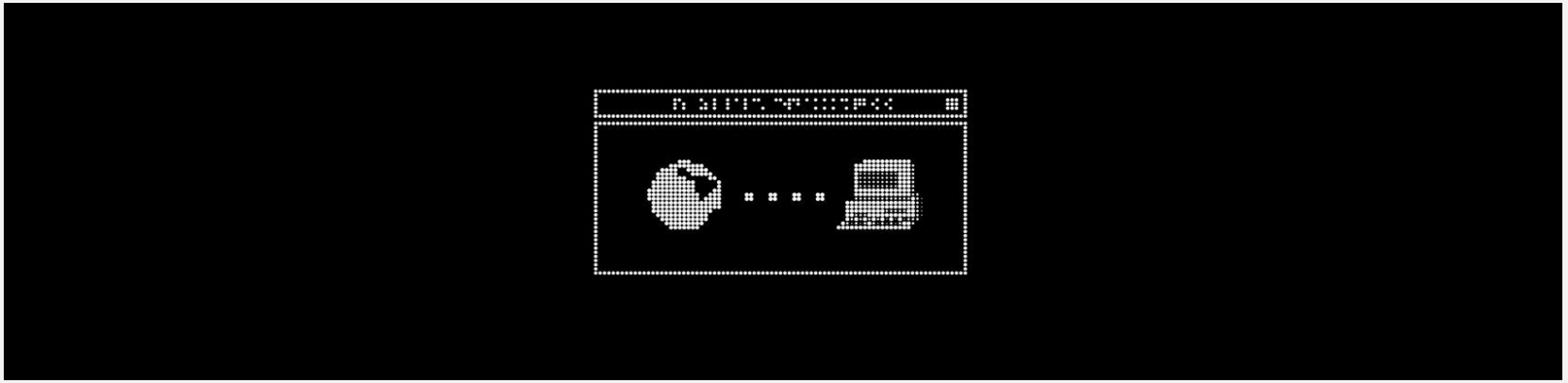


Capable but Not Scalable

Rendering CAD models into multi-view images shapes what AI models learn and what evaluators judge, yet it is usually handled by one-off scripts.

RQ1	RQ2	RQ3
Practice is fragmented	**14 requirements**	**Code or repetition**
Rendering is code-based and task-specific; view protocols vary widely, while orientation and normalization are often underreported.	*11 for the dataset (input to output) and 3 for the interaction that produces it.*	*Almost every requirement can be met, but through code or manual steps repeated for every product.*

Takeaway: *these findings motivate* ***collection-level workflows*** *for non-programming design researchers.*



Thank You.

Open for Questions.

Supported by the Technology Innovation Program (RS-2025-04752997), funded by Ministry of Trade, Industry & Energy (MOTIE), Korea.

Danyal Sarfraz
Researcher, HCDT, UNIST

KwanMyung Kim
Professor, HCDT, UNIST

한국어로 보기

LinkedIn