



Elegance Automated

Final Technical Report

New Venture Design: Team 001-1

Joshua Himmens	54456710
Dawson March	25907593
Sloan Sobie	12094512
Genevieve Merz	10268696
Cameron Powell	61395679
Jaden Legate	63330716

April 8, 2026

Change Log

Date	Author	Change	Reason
2026-04-08	Team 001-1	Initial technical report compiled	Final deliverable

Contents

1	Product Brief	5
2	Requirements	6
2.1	Context	6
2.2	Product Vision	6
2.3	Requirements and Evaluation Criteria	8
2.3.1	Evaluation Criteria	9
3	Design	11
3.1	Product Architecture	11
3.1.1	Structural Frame & Workspace Fixture	11
3.1.2	Workspace Base Surface	11
3.1.3	Aluminium Extrusion Gantry	12
3.2	Robotic Arm Configuration & Workspace Geometry	12
3.2.1	Arm Spacing Trade-off	12
3.3	System-Level Concept Generation	13
3.3.1	Concepts Considered	14
3.3.2	Concept Selection Rationale	14
3.4	End-Effector Design	14
3.4.1	Down-Selected Prototype Concepts	15
3.5	Machine Learning	15
3.5.1	Policy Selection	15
3.5.2	Software Architecture	17
3.5.3	Data Collection	17
3.5.4	System Architecture	18
3.5.5	Policy Training	20
3.5.6	Results Summary	21
3.5.7	Future Directions	22
4	Verification & Validation	23
4.1	Overview	23
4.2	Verification	23
4.2.1	Fold Accuracy and Throughput (REQ-01, REQ-02)	23
4.2.2	Policy Results and Analysis	25
4.2.3	End-Effector Selection (FR-02, REQ-02)	25
4.2.4	Additional Verification Status	27
4.3	Validation	28
4.3.1	Concierge Simulation at Wildlight Kitchen + Bar	28
4.3.2	Letter of Intent Validation	29
4.3.3	Deferred Validation: Manager's Blind Lineup (VAL-01)	29
4.4	Summary of Verification and Validation Status	30
4.5	Prototyping Risks and Mitigation	30
4.6	Future Verification and Validation Roadmap	31
5	IP Assessment	33
5.1	Related IP (Prior Art)	33
5.2	Freedom to Operate	33
5.3	Patentability	34

References	35
6 Appendices	36
Document List	36
In the Google Drive Archive	36
Externally Hosted	37
Appendix A — USB-UART Bridge Hardware Notes	39
Cable Strain Relief	39
Schottky Diode Voltage Drop	40
Appendix B — Concept Visualisation	42
Appendix C — End-Effector Concept Generation	44
Appendix D — Function Structure Diagram	46
Appendix E — Morphological Chart	47
Appendix F — Weighted Decision Matrix	48
Appendix G — Physical System Construction	49
Initial Assembly: Optical Breadboard + Planar Lights	49
Secondary Assembly: Continuous Hardboard + Ring Lights	50
Appendix H — Annual Napkin Folding Cost per Restaurant	51
Stakeholder Evidence	51
Hand Calculations	51
Appendix I — Satisfaction Curves	52
Appendix J — Hugging Face Datasets	54

1 Product Brief

Premium hospitality venues dedicate 2–3 hours of paid server labour per day to decorative napkin folding, a task that generates no revenue, degrades staff morale, and produces inconsistent results. Across fine dining, catering, and event preparation, this translates to \$15,000–\$30,000 in annual labour cost per venue, with roughly 30% of hand-folded napkins failing to meet manager presentation standards. The hospitality industry’s worsening labour shortage makes this allocation increasingly untenable.

Linique is a compact, dual-arm robotic system purpose-built for autonomous linen napkin folding. Two SO-101 6-DOF manipulators, guided by a multi-camera RGB perception pipeline and a learned visuomotor policy, execute presentation-grade decorative folds without human intervention. Custom end-effectors enable reliable cloth capture, while a 450M-parameter vision-language-action model (SmolVLA) composes grasp and fold behaviours from real-world demonstration data, eliminating the sim-to-real gap that makes deformable-object manipulation intractable in simulation. The system fits on a standard prep table ($\leq 60 \times 60$ cm), runs on 120 V AC, and is operable within two minutes of setup.



Figure 1: Product rendering of the Linique system.

A week-long concierge deployment at Wildlight Kitchen and Bar confirmed all five operational hypotheses: folding consumed ~2 hrs/day of labour, its removal freed staff for neglected maintenance tasks, closing shifts shortened by 30–60 minutes, staff sentiment was unanimously positive, and napkin usage tracked guest count at 1.08 napkins/guest. Management and staff independently expressed willingness to pay for the service.

The serviceable addressable market in North America is approximately \$700M across 130,000 premium venues. Linique targets a \$7,000 sale price with a \$200/month lease alternative, yielding 78% gross margins and sub-one-year customer payback. Letters of intent have been secured from early partner venues.

The prototype is alpha-ready with all subsystems integrated. SmolVLA achieves 70% end-to-end fold success; closing the gap to 80%+ requires resolving dataset merging limitations, implementing stage-aware reward modelling for RL fine-tuning, and evaluating diffusion-based action generation for the pickup-to-fold transition. A second concierge trial at an unaffiliated venue and pursuit of Wildlight as a paying beta customer are the immediate commercial priorities.

2 Requirements

2.1 Context

Linique is motivated by a repetitive but operationally significant back-of-house task in hospitality environments: decorative napkin folding. Across fine dining restaurants, wineries, catering teams, and convention centres, staff currently spend paid preparation time manually folding napkins before service. From both stakeholder interviews and our team's combined 13 years of hospitality experience, this task consistently emerged as a high-friction labour sink that adds no direct revenue while still having a visible impact on guest experience.

This motivation was strongly reinforced by stakeholder feedback. Interviews with hospitality operators showed that venues commonly dedicate 1–3 hours of server labour per day to folding hundreds of napkins, which translates to approximately \$15,000–\$30,000 in annual labour cost for luxury restaurants and substantially more for larger venues. See [Appendix H](#) for a full breakdown of how these values are calculated. Stakeholder comments made the pain point especially clear, with one operator stating, “If I didn’t have to pay people to fold napkins, I wouldn’t,” while others emphasized that labour was the largest bottleneck during catering and event preparation. These interviews also highlighted that inconsistent folds negatively affect fine-dining presentation standards, making the problem both an economic and brand-consistency issue.

For the final MVP, the project scope was intentionally narrowed to isolate the most technically uncertain portion of this broader workflow: autonomous robotic cloth folding using dual-arm manipulation and machine learning. While earlier user-interaction scenarios included feed trays, output stacking, storage, and simple restaurant-facing controls, the current prototype focuses exclusively on the core tabletop folding task. Specifically, the system accepts a manually placed square cloth napkin and executes a single decorative fold style using two SO-101 robotic arms, RGB vision, and an ML policy. This scoping decision was made to maximise technical depth within the short New Venture Design timeline while directly addressing the hardest engineering challenge in the workflow: reliable deformable-object manipulation.

The primary external stakeholders remain hospitality operators, including restaurant managers, catering supervisors, and event-preparation staff, whose needs centre on labour reduction, fold consistency, and smooth workflow integration. Full interview notes and user scenarios are archived in the project Drive — see [Table 17](#). However, for this MVP, an equally important internal stakeholder group is the technical development team responsible for robotic integration, perception, and policy training. Their constraints — limited SO-101 workspace overlap, camera frame rate, controlled lighting, tabletop dimensions, budget, and the absence of tactile sensing — directly shape the requirements for this prototype generation. These prototype-specific stakeholders justify why the MVP prioritises repeatable autonomous fold execution in a controlled environment over deployment-level features such as automated feeding and output handling.

2.2 Product Vision

The primary operational stakeholders are restaurant managers, supervisors, and prep staff responsible for back-of-house opening workflows. Across multiple hospitality interviews, these users consistently identified manual napkin folding as a repetitive side duty that consumes approximately

1–3 hours of paid labour per day, depending on venue volume and staff efficiency. These stakeholders drive the most critical product attributes of speed, accuracy, and ease of use, since the system must reduce labour burden without adding operational complexity. A breakdown of product attributes appears in [Table 1](#).

#	Design Issue	Product Attribute Statement	Stakeholder Source
DI-01	Speed	The system folds napkins at a rate that meaningfully reduces manual labour time.	Sarah W., Michael C.
DI-02	Fold Accuracy	The final fold alignment is precise enough for fine-dining aesthetic standards.	Natalie Fraser, Karen Jones
DI-03	Versatility	The machine handles multiple fold types and varying fabric materials.	Needs Assessment 6
DI-04	Footprint	The device fits within space-constrained back-of-house prep areas.	Fleur Adele, Karen Jones
DI-05	Ease of Use	The machine is intuitive to operate, requiring no specialised training.	Technical Plan
DI-06	Hygiene	All surfaces exposed to linens are non-porous and easily cleaned.	NSF/ANSI 2 [1]
DI-07	Reliability	The system operates continuously through a full batch without failure.	Mission Hill Winery
DI-08	Autonomy	The system operates with minimal human supervision once started.	Sarah W., Catering Mgrs.
DI-09	Noise	The machine operates at a noise level acceptable in a hospitality environment.	Restaurant Managers
DI-10	Cost	The system is priced to deliver a payback period of less than one year.	Multiple stakeholders
DI-11	Durability	The system withstands years of daily commercial use.	Catering Managers
DI-12	Serviceability	Faults can be diagnosed and repaired without specialised technicians.	Restaurant Managers

Table 1: Design issues and product attribute statements derived from stakeholder consultations.

One stakeholder not encapsulated in the product attributes above is the technical development team, a key stakeholder for the MVP stage itself. Unlike hospitality users, their primary concern is reducing technical uncertainty within the hardware constraints of the current prototype. These constraints are driven by the fixed SO-101 workspace, camera frame rate, lighting assumptions, table dimensions, budget limitations, and the absence of tactile sensing. They explain why the current system is intentionally limited to one fold style, one manually placed napkin, and a controlled tabletop environment. This perspective is important because it justifies the narrowed MVP scope as a deliberate engineering decision to isolate the hardest unsolved challenge: autonomous dual-arm deformable cloth manipulation.

Taken together, hospitality operators prioritise labour reduction, presentation consistency, and workflow simplicity, while the development team prioritises technical feasibility under strict hardware constraints. This dual perspective ensures the MVP remains grounded in real end-user value while still serving as a technically credible foundation for future prototypes with automated feeding, stacking, faster throughput, and multiple fold styles.

2.3 Requirements and Evaluation Criteria

Each design issue from [Table 1](#) was classified as a requirement (must-have pass/fail threshold), an evaluation criterion (nice-to-have scored on a satisfaction curve), or both. Speed (DI-01) appears as both: the requirement (REQ-01) sets the minimum acceptable cycle time while the evaluation criterion (EC-02) rewards designs that exceed it. Autonomy (DI-08) is purely an evaluation criterion because any level of reduced supervision adds value without a hard threshold. Cost (DI-10) is also purely an evaluation criterion; no hard ceiling was imposed because stakeholders across different venue sizes indicated different willingness-to-pay thresholds. Ease of Use (DI-05) and Hygiene (DI-06) are absorbed into specific measurable requirements (REQ-13, REQ-15, REQ-16).

Functional requirements ([Table 2](#)) define the essential services Linique must provide. Engineering requirements ([Table 3](#)) establish measurable performance thresholds, safety limits, and physical constraints that filter out ineligible solutions.

ID	Requirement	Description	Source
FR-01	Napkin Intake	The system shall accept clean square cloth napkins into the folding workspace, either by manual placement (MVP) or automated feed in future prototypes.	BOH prep staff workflow
FR-02	Fold Execution	The system shall autonomously execute at least one complete decorative fold program from a flat initial napkin state.	Labour reduction
FR-03	Fold Style Selection	The system shall allow the user to select from available fold programs. MVP supports one style; future versions support multiple.	Versatility
FR-04	Completion & Reset	The system shall terminate in a completed fold state and return to a safe neutral pose for the next cycle.	Ease of use, repeatability
FR-05	Output Handling	The system shall present the folded napkin in a way that preserves geometry for immediate table placement or downstream stacking.	Fine-dining presentation

Table 2: Functional requirements (FR-01 through FR-05).

ID	Requirement	Target	Justification
REQ-01	Cycle Time	≤ 90 s / napkin	Human benchmark 7–40 s; 300 napkins in < 8 hr shift
REQ-02	Fold Accuracy	≤ 10 mm edge dev.	Fine-dining “table ready” presentation standard
REQ-03	Footprint	$\leq 36 \times 36$ in	Standard commercial prep table
REQ-04	Supply Voltage	120 V AC	North American restaurant infrastructure
REQ-05	Emergency Stop	Present (≤ 0.5 s)	ISO 10218-1 [2]; ISO/TS 15066 [3]
REQ-06	Pinch-Point Safety	≤ 150 N	ISO/TS 15066 [3] quasi-static contact limit
REQ-07	Operating Temp.	0–40 °C	Unheated BOH to kitchen-adjacent range
REQ-08	Recalibration Interval	≥ 8 hr	Full shift without operator recalibration
REQ-09	1st-Attempt Success	≥ 90 %	Minimise rework; Dyna benchmark ~99 %
REQ-10	Fold Recognition	≥ 99 %	Prevent defective output reaching tray
REQ-11	Position Resolution	≤ 0.5 mm	Multi-step fold alignment after error propagation
REQ-12	Mechanical Drift	≤ 1 mm / hr	Cumulative error within REQ-02 over 4 hr
REQ-13	Calibration Time	≤ 5 min	Pre-service setup; daily use target < 2 min
REQ-14	Noise Level	≤ 70 dB(A)	FOH ambience / BOH communication threshold
REQ-15	Water Resistance	Splash-proof	NSF/ANSI 2 [1] cleanability
REQ-16	Cleanability	Wipe-clean	No disassembly; NSF/ANSI 2 [1]
REQ-17	Design Lifetime	≥ 5 yr (438k cycles)	Commercial equipment amortisation period
REQ-18	Spare Part Lead	≤ 1 month	Remote venue downtime minimisation
REQ-19	Repairability	No specialist req.	General technician or operator serviceable

Table 3: Engineering requirements (REQ-01 through REQ-19).

2.3.1 Evaluation Criteria

While the preceding requirements define go/no-go conditions for an acceptable product, the following evaluation criteria capture the performance dimensions that most strongly increase customer value and guide future prototype development. These criteria intentionally extend beyond the MVP scope to preserve alignment with the long-term commercial vision of autonomous restaurant deployment.

ID	Criterion	Metric	Target	Weight	Why it adds value
EC-01	Versatility	Fold types	5+	15 ± 5 %	Adaptability to seasonal menus; secondary differentiator
EC-02	Throughput	napkins / hour	500+	20 ± 5 %	Directly scales financial savings for high-volume venues
EC-03	Autonomy	supervision min / batch	≤ 1 min	40 ± 10 %	Unanimous stakeholder consensus; primary value driver
EC-04	Unit cost	Retail price (CAD)	\$3k–\$10k	25 ± 5 %	Lower cost improves payback and broadens market access (cost factor)

Table 4: Evaluation criteria with nominal weights and uncertainty ranges.

Benefit weights (EC-01 through EC-03) sum to 0.75; cost (EC-04) is 0.25. Autonomy receives the highest weight based on unanimous stakeholder consensus that “set-and-forget” operation is the primary value driver. Throughput is weighted above versatility because speed directly scales financial savings, while fold variety shows diminishing returns beyond three to five types. Satisfaction curves mapping each criterion to a 0–1.0 score appear in [Appendix I](#) a formal value equation with uncertainty propagation and sensitivity analysis is documented in the Specification (see [Table 17](#)). These criteria also establish the primary verification priorities for future prototype generations, particularly feed automation reliability, throughput scaling, and fold-style generalisation.

3 Design

3.1 Product Architecture

The Linique napkin-folding system is a dual-arm robotic platform designed to autonomously fold napkins for fine-dining restaurants. The system couples a physical hardware layer with a vision-guided perception subsystem and a machine-learning software stack to achieve autonomy. [Appendix G](#) shows the physical system as constructed; the authoritative CAD is hosted on Onshape and linked in [Table 18](#).

The system is decomposed into five major subsystems:

- **Robotic manipulation & alignment:** dual SO-101 6-DOF follower arms with custom end-effectors that perform napkin pick-up, corner capture, fold initiation, and transfer.
- **Leader arm interface:** a matched pair of SO-101 leader arms used by a human operator to demonstrate folding trajectories during teleoperation-based data collection.
- **Sensing & perception:** two overhead RGB cameras mounted on aluminium extrusion rails and two end-effector-mounted cameras, providing a fixed top-down view for napkin pose estimation and fold quality verification.
- **Control & verification:** a learned controller that estimates the commanded state for all six actuators per arm simultaneously.
- **Structural frame & workspace fixture:** aluminium-extrusion gantry supporting the camera rail above the workspace, with arm-mounting brackets bolted to the breadboard and subsequent wood base.

[Appendix D](#) shows the full system architecture diagram, breaking the end product into material, energy, and information flows.

3.1.1 Structural Frame & Workspace Fixture

The structural subsystem provides a stable, repeatable testing environment. It consists of a workspace base surface, an aluminium-extrusion gantry, and custom arm-mounting brackets. A key evolution during development was the transition from an optical breadboard to a wood-panel base, described below.

3.1.2 Workspace Base Surface

Initial prototyping used an optical breadboard, provided by the Engineering Physics department, as the arm-mounting and napkin workspace surface. The breadboard provided a rigid, flat base with a dense grid of mounting points for repositionable arm brackets. This helped in early-stage prototyping when arm positions were not yet fixed. However, its high cost and significant weight made it impractical as a long-term platform for a deployable system. The workspace was subsequently transferred to a custom wood-panel base, which preserved the flat, stable working surface while substantially reducing cost and improving manoeuvrability. Arm brackets were redesigned to mount to the wood surface using standard hardware.

3.1.3 Aluminium Extrusion Gantry

The overhead camera and lighting system is supported by a gantry constructed from aluminium-extrusion profiles with L-bracket corner fittings. Two vertical uprights rise from the rear corners of the base and are connected by a horizontal cross-member. The cameras and LED lighting strips are clamped to this rail on adjustable slide mounts, allowing the camera height and fore-aft position to be adjusted without structural modification. The gantry ensures a fixed, vibration-resistant camera perspective.

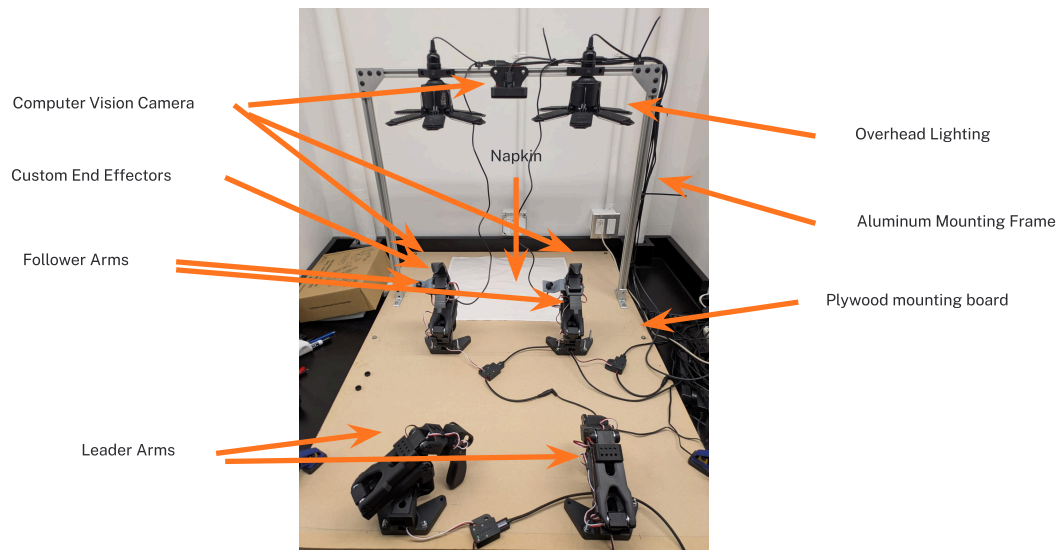


Figure 2: Linique prototype showing the aluminium-extrusion gantry, overhead camera mounts, and dual SO-101 arms on the wood-panel base.

3.2 Robotic Arm Configuration & Workspace Geometry

3.2.1 Arm Spacing Trade-off

Two arm-spacing configurations were evaluated before committing to a final mounting layout (Figure 3). The first placed both arms centrally, producing overlapping reach envelopes and a shared manipulation zone. The second separated the arms to the outer edges of the base, maximising each arm's individual reachable workspace while centring the fold zone between them.

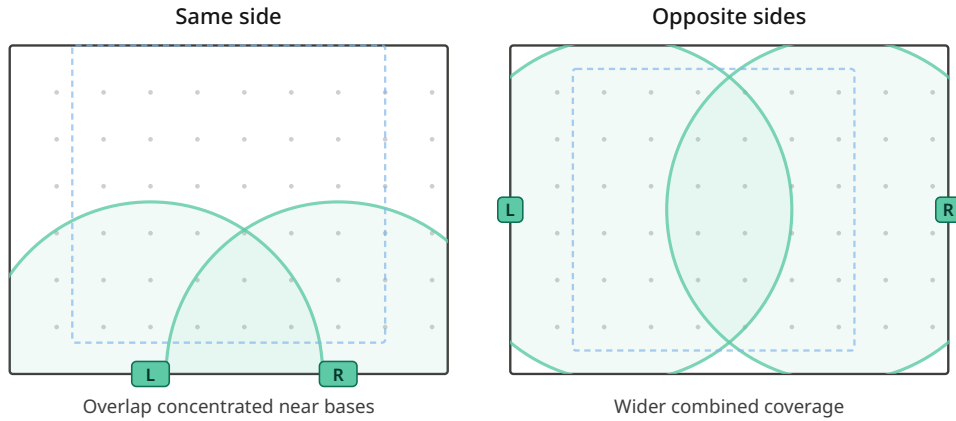


Figure 3: Arm mounting geometry trade-off. Left: centred, overlapping reach envelopes. Right: separated outer-edge layout with the fold zone centred between the arms. The right layout was selected to reduce inter-arm collision risk during fold motions.

The separated outer-edge layout was selected. It places the napkin fold workspace centrally between the two arms, reduces inter-arm collision risk during fold-over motions, and provides each arm with sufficient reach to independently capture a napkin corner. The trade-off is that cooperative two-arm tasks requiring simultaneous grip on a single corner must use extended arm poses; folding trajectories are designed to account for this.

3.3 System-Level Concept Generation

Prior to arriving at the current system architecture, four top-level concept directions were generated from a morphological matrix spanning the key design-driving functions: napkin pick-up and isolation, cloth alignment, corner manipulation, fold creation, fold verification, misfold recovery, output stacking, and control strategy. Combinations of sub-concepts across each function were synthesised into four coherent system architectures. The full morphological chart can be found in [Appendix E](#).

3.3.1 Concepts Considered

Concept	Name	Summary
A	Dual Robotic Arm Folding	Two robotic manipulators actively isolate, reposition, and fold napkin corners with maximum flexibility across fold styles. High adaptability but larger footprint and greater control complexity.
B	Conveyor-Guided Folding	Passive linear guides and conveyor-driven transport fold napkins through a controlled sequence of geometric constraints. High throughput and reduced actuation complexity, but limited fold-style flexibility.
C	Rotary Folding Drum	A rotating drum with compliant fingers and crease guides folds napkins in a continuous cyclic process. Considered for scalability, but eliminated due to footprint and limited fold-style range.
D (selected)	Hybrid Robot + Passive Guides	Robotic corner manipulation combined with passive alignment and crease-forming guides. Balances fold precision with reduced mechanical complexity. Selected as the leading direction.

Table 5: System-level concepts generated from the morphological matrix. Concept C was eliminated at initial screening; Concept D was selected as the refined direction.

Visualisations of each concept can be found in [Appendix B](#).

3.3.2 Concept Selection Rationale

Concept C (Rotary Drum) was eliminated at initial screening as it could not satisfy the back-of-house footprint constraint and was limited to a narrow range of fold styles. Concepts A, B, and D proceeded to weighted scoring against the evaluation criteria, with fold quality, repeatability, and robustness to linen variation carrying the highest weights based on stakeholder interview findings. The Hybrid Robot + Passive Guide architecture (Concept D) ranked highest, scoring 4.15 out of 5.00, by providing the strongest balance of fold consistency, manageable control complexity, and robustness to different linen conditions. This was not the final concept but it provided a general direction for us to follow. The full weighted score table can be found in [Appendix F](#).

3.4 End-Effector Design

The end-effector is the highest-impact mechanical component in the system. It is the sole point of physical contact between the robot and the napkin, and its geometry directly determines pick-up reliability, cloth deformation, and the downstream complexity of the control stack. End-effector optimisation was therefore the primary focus of the mechanical design phase.

Preliminary end-effector concepts were sketched to explore the cloth-manipulation design space before committing to physical prototypes. These sketches can be found in [Appendix C](#).

3.4.1 Down-Selected Prototype Concepts

Four end-effector concepts were down-selected from the broader ideation phase for physical prototyping and testing. Early sketched concepts were eliminated during concept screening against requirements for reliable corner capture, low cloth deformation, and integration with the SO-101 wrist geometry. The four retained concepts each embody a distinct and meaningfully different combination of the design parameters above, bracketing the realistic ways a single-DOF rotary gripper can engage a flat cloth edge. All prototypes were 3D-printed in PLA+.

Code	Name	DP Combination	Rationale for inclusion
OG	Narrow-tip pinch	DP-1: Force-closure DP-3: Top-down DP-4: Parallel	Establishes a performance baseline. The simplest and fastest force-closure geometry; retains a narrow tip to target napkin corners precisely. Included to characterise the upper bound of throughput at the cost of cloth compliance.
EEJ	Thick-jaw pinch	DP-1: Force-closure DP-3: Lateral DP-4: Parallel	Lateral engagement direction differentiates this from OG. A thicker jaw body was hypothesised to improve grip area and reduce localised stress on cloth fibres compared to the narrow-tip baseline.
EED	Inverse close	DP-1: Force-closure DP-3: Top-down DP-4: Normal	Wide parallel pads distribute normal force over a larger contact area. Included to test whether spreading compressive load reduces cloth crumpling compared to concentrated-contact force-closure designs.
EES	Tangential scoop	DP-1: Geometry-entrapment DP-3: Underside-lift DP-4: Tangential	The only geometry-entrapment concept. A curved scoop slides beneath the napkin edge and lifts it before closing. Included as the principal departure from force-closure — the only design that eliminates compressive cloth loading by design.

Table 6: Down-selected end-effector prototypes and rationale for inclusion. Testing and performance results are discussed in [Section 4.2](#).

The final end-effector design selected was EES because of its versatility in the range of angles with respect to the napkin pick-up plane, and its surface contact ensured the napkin would not be dropped. A more detailed analysis of the experimental validation conducted on the end-effector design can be found in [Section 4.2.3](#).

3.5 Machine Learning

3.5.1 Policy Selection

The central design decision in the software stack is the choice of learned control policy. Because the system manipulates deformable linen — a material whose dynamics are highly non-linear and

difficult to simulate with fidelity — the standard reinforcement-learning pipeline of training in simulation and transferring to hardware (sim-to-real) is not viable. The sim-to-real gap for cloth is prohibitively large: contact mechanics, self-occlusion, and drape behaviour vary with fabric weight, weave, and starch level in ways that current physics engines do not capture. This constraint eliminates all policy families that depend on a simulator reward signal and restricts the design space to methods that learn directly from real-world human demonstrations — collectively termed imitation learning or behaviour cloning.

Within imitation learning, three policy families were evaluated:

Action Chunking with Transformers (ACT) [4]. ACT encodes short temporal “chunks” of demonstrated actions and decodes them autoregressively at inference time. Its key advantage for this project is data efficiency: published results show competent bimanual manipulation from as few as 30 demonstration episodes, with training completing in under 24 hours on a single consumer GPU. ACT was selected as the initial policy architecture for these reasons, and was augmented with temporal ensembling (smoothing action predictions across overlapping chunks to reduce jitter) and learning-rate scheduling (cosine annealing to improve convergence stability). These modifications improved trajectory smoothness and training reliability relative to the baseline ACT implementation.

Vision-Language-Action (VLA) models. VLA policies couple a pretrained vision-language backbone with an action head, giving them strong generalisation and the ability to transfer across tasks and embodiments. Two VLA architectures were evaluated during development:

- **xVLA [5]:** An early experiment with the xVLA architecture produced poor results — the policy commanded trajectories that drove the arms into the table surface, indicating a fundamental mismatch in action-space calibration or observation processing. Due to time constraints, debugging was not pursued further. xVLA may warrant revisiting with more careful action-space configuration, as its architecture is theoretically well-suited to bimanual manipulation.
- **SmolVLA [6]:** At 450 million parameters and approximately 4 GB of VRAM, SmolVLA is a compact VLA model that leverages pretrained visual and linguistic representations for better generalisation than ACT. SmolVLA was successfully trained and produced the most performant policy in our evaluation, as discussed in [Section 4.2.2](#).

Diffusion Policy [7]. Diffusion-based action generation naturally represents multimodal action distributions by modelling the full distribution over possible actions rather than predicting a single point estimate. This property makes diffusion policies a strong candidate for tasks with multimodal action distributions, such as the pickup-to-fold transition. Diffusion Policy remains a candidate for future work; flow matching [8] — a related generative framework — would also allow the temporal ensembling and learning-rate scheduling improvements applied to ACT to transfer naturally to the diffusion setting.

A key constraint on policy selection is the target inference platform. For a commercially deployable system, the policy must run on an edge compute module such as the NVIDIA Jetson Orin, which provides only 8 GB of VRAM. This rules out larger VLA models (which often require 16+ GB) and motivated the focus on SmolVLA’s compact 4 GB footprint and ACT’s even smaller memory requirements. If additional GPU resources were available at inference time, larger and potentially more capable VLA models could be explored, but the Jetson Orin constraint reflects the practical reality of deploying a robot in a restaurant back-of-house environment.

3.5.2 Software Architecture

3.5.2.1 LeRobot Framework

The software stack is built on LeRobot [9], an open-source framework from Hugging Face for robotic imitation learning. LeRobot provides integrated tooling for data collection, dataset management, policy training, and inference. However, several modifications were necessary to adapt the framework to this project's requirements.

3.5.2.2 Custom Inference Pipeline

The default LeRobot inference pipeline performs image encoding (resizing, normalisation, and colour-space conversion) on the CPU as a preprocessing step before passing observations to the policy on the GPU. On the inference workstation — initially equipped with an RTX 2070 and a mid-range CPU — this image processing created a CPU bottleneck that starved the GPU of data, resulting in unacceptably low control-loop frequencies. At low frequencies, the temporal gap between observation and action grows, and the robot's motion becomes jerky and unresponsive.

To resolve this, custom inference scripts were written that bypass LeRobot's CPU-side image encoding entirely. Raw camera frames are passed directly to the GPU, where the policy's vision encoder handles all necessary preprocessing as part of the forward pass. This eliminated the CPU bottleneck and restored the control loop to its target frequency. The CPU starvation issue on the original inference machine was also a motivating factor in requesting an upgraded workstation (RTX 4070) from the Engineering Physics department, which further improved both inference and training throughput. [Appendix A](#) contains detailed hardware notes on the USB-UART bridge used in this pipeline.

3.5.2.3 Dataset Management

Data is stored in the LeRobot dataset format, which packages synchronised image frames and joint-angle trajectories into a structured layout. Two dataset categories were collected: full-task episodes (pickup through fold) and pickup-only episodes. Merging these datasets into a single combined training set proved non-trivial due to compatibility issues in the LeRobot dataset schema — specifically, differences in episode metadata and frame indexing between the two collection runs caused intermittent crashes during training when the dataloader attempted to sample across the boundary between the two dataset partitions. Due to time constraints, this merging issue was not fully resolved, which meant that the pickup-only data could not be incorporated into SmoIVLA training. The SmoIVLA policy was therefore trained exclusively on the full-task dataset. Resolving the dataset merging pipeline is identified as a prerequisite for future training runs that combine stage-specific and full-task data.

All collected data is saved locally during collection, backed up to the UBC Sockeye research computing cluster, and has been made publicly available through Hugging Face to support reproducibility.

3.5.3 Data Collection

3.5.3.1 Collection Procedure

Training data was gathered through kinesthetic teleoperation. A human operator physically manipulates a pair of SO-101 leader arms; the follower arms mirror the leader poses in real time, and the system records synchronised joint-angle trajectories and RGB image streams from all three

cameras at 30 frames per second. Each demonstration — one complete pickup-and-fold sequence — constitutes a single episode.

Two categories of episodes were collected:

- **Full-task episodes (400 episodes, $\approx$ 4 hours of demonstration):** The operator picks up a flat napkin, aligns the corners, and executes a complete fold.
- **Pickup-only episodes (99 episodes, $\approx$ 1 hour of demonstration):** The operator performs only the napkin pickup and corner-capture phase. These were added after early ACT training runs revealed that the pickup phase was the dominant failure mode; the additional pickup data was intended to increase the density of pickup-state coverage in the training distribution.

In total, the dataset comprises 499 episodes and over 550,000 synchronised image frames.

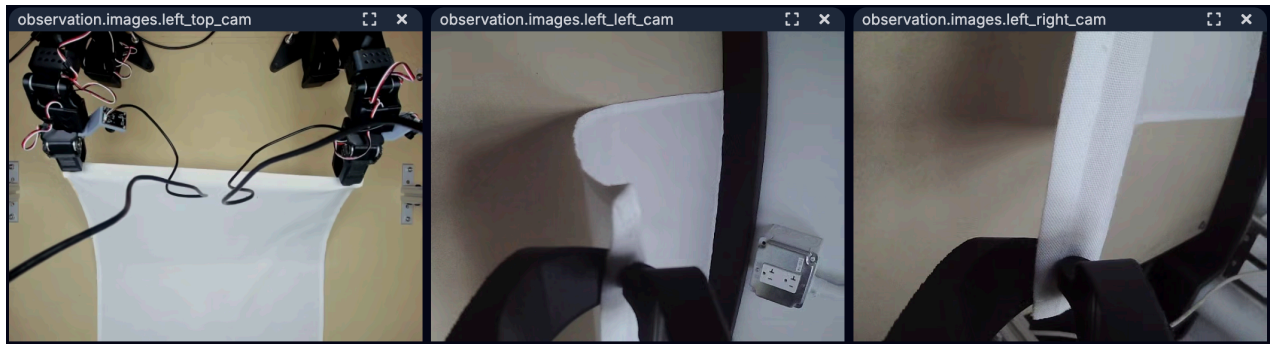


Figure 4: Example visual information collected during a demonstration episode: synchronised sample frames from the overhead and wrist cameras showing the pickup and fold stages.

3.5.3.2 Minimising Distribution Shift

A persistent challenge in imitation learning is distribution shift: if the states the policy encounters at inference time differ from those seen during training, performance degrades rapidly. Several design decisions in the data-collection setup were made specifically to minimise this effect:

1. **Fixed jig geometry.** The aluminium-extrusion gantry and wood-panel base described in [Section 3.1.1](#) ensure that camera viewpoints, lighting conditions, and arm-base positions are identical between data collection and inference. No component is repositioned between sessions.
2. **Consistent napkin placement.** The operator places napkins in a standardised region of the workspace at the start of each episode, bounding the initial-state distribution.
3. **Overhead and wrist cameras.** The three-camera arrangement (one overhead, two wrist-mounted) provides complementary viewpoints. The overhead camera captures global napkin pose; the wrist cameras capture local gripper-to-cloth geometry. This redundancy reduces the likelihood that a critical visual feature is occluded in all views simultaneously.

3.5.4 System Architecture

The system architecture decomposes the information flow from raw sensor data to motor commands into three stages: perception, policy inference, and actuation.

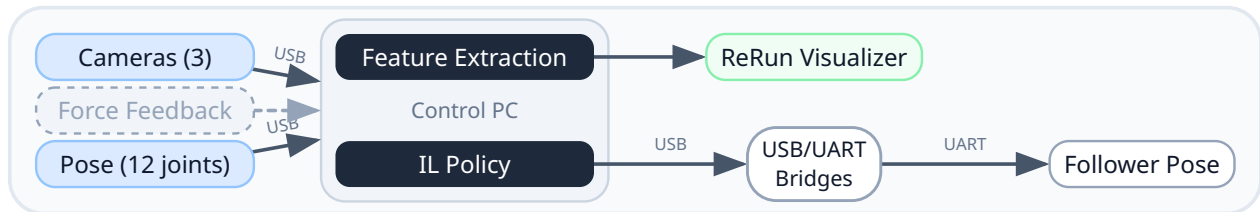
Perception. At each control timestep, synchronised frames are captured from the overhead and two wrist-mounted RGB cameras. These images, along with the current joint-angle readings from all twelve servos (six per arm), are concatenated into a single observation vector and passed to the policy.

Policy inference. The policy receives the observation vector and outputs a chunk of future joint-angle targets — a short temporal sequence of commanded positions for all twelve actuators. For the ACT policy, temporal ensembling is applied across overlapping chunks to smooth the commanded trajectory. The chunk length is a hyperparameter set during training; longer chunks smooth trajectories but reduce the policy’s ability to react to unexpected states.

Actuation. The commanded joint angles are sent to the follower-arm servo controllers, which execute the trajectory chunk. At the end of the chunk, a new observation is captured and the cycle repeats.

This architecture satisfies the control and verification functional requirement: the learned controller estimates commanded states for all actuators simultaneously, and the overhead camera provides a fixed top-down view that can be used for fold-quality verification.

Inference



Data Collection

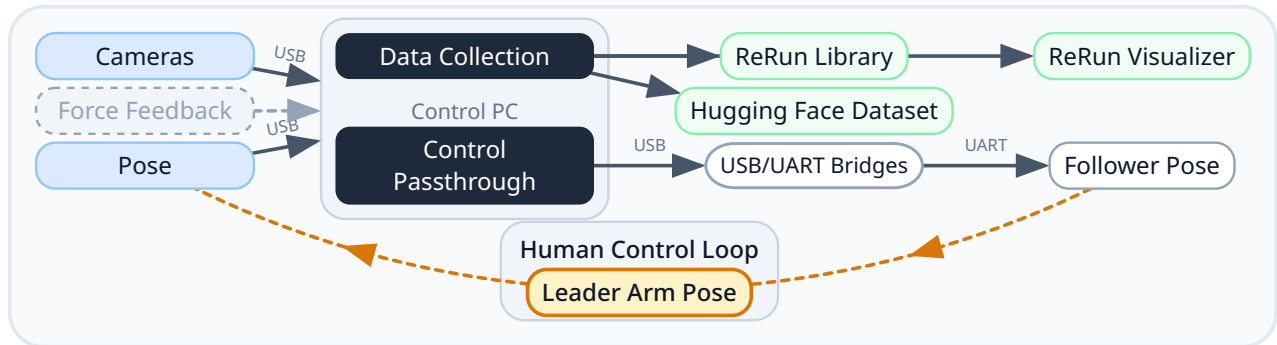


Figure 5: Linique system architecture. Top: inference dataflow from cameras and pose sensors through the control PC into the follower arms. Bottom: data-collection dataflow, with a human-in-the-loop control of the leader arms writing synchronised frames and poses to a Hugging Face dataset.

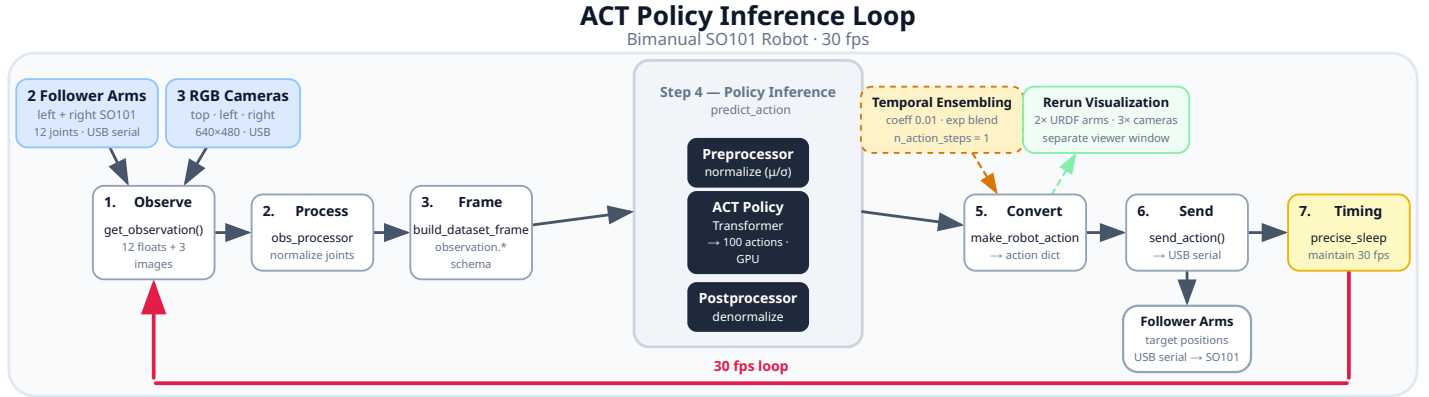


Figure 6: ACT policy inference loop: seven pipeline steps running at 30 fps on the inference workstation, showing temporal ensembling, Rerun visualisation, and the feedback loop back to the follower arms.

3.5.5 Policy Training

3.5.5.1 Training Infrastructure

Training was conducted across multiple GPU configurations spanning local, cloud, and institutional HPC environments (Table 7).

Environment	GPU	Notes
Local (initial)	NVIDIA RTX 2070	First development machine; also used for inference, where CPU starvation during LeRobot’s image preprocessing motivated the hardware upgrade below.
Local (upgraded)	NVIDIA RTX 4070	Requested from Engineering Physics department to address the CPU/GPU bottleneck; became the primary training and inference workstation.
Cloud (RunPod)	NVIDIA H100 / RTX 5090	Highest peak throughput; used for large-scale and SmoVLA training runs. Data-transfer overhead (> 550k frames) reduced effective iteration speed for short experiments.
UBC Sockeye HPC	NVIDIA A40	Institutional cluster; job-queue scheduling latency and data-staging overhead made it less practical for rapid iteration.

Table 7: Training environments used across the project.

Despite the higher raw compute available on cloud and HPC platforms, local training on the RTX 4070 proved to have the fastest end-to-end development cycle for ACT. The overhead of transferring the large image dataset and the latency of job-queue scheduling on Sockeye negated the compute advantage for iterative experimentation. Cloud GPUs (particularly the H100) were reserved for longer training runs, including SmoVLA, where the per-step speedup justified the data transfer cost.

3.5.5.2 ACT Training

The ACT policy was trained for 100,000 gradient steps, requiring approximately 15 hours on the local workstation. Learning-rate scheduling (cosine annealing) was applied to stabilise convergence, and

temporal ensembling was used at inference time to smooth action predictions across overlapping chunks. The training loss curve decreased steadily over the run and showed no signs of divergence or catastrophic forgetting. While the loss value is difficult to interpret mechanistically — it reflects the policy’s prediction error on demonstrated action chunks, not a direct measure of task success — the smooth, monotonically decreasing curve provides confidence that there are no critical errors in the data pipeline, preprocessing, or training script that would prevent the model from navigating the loss landscape.

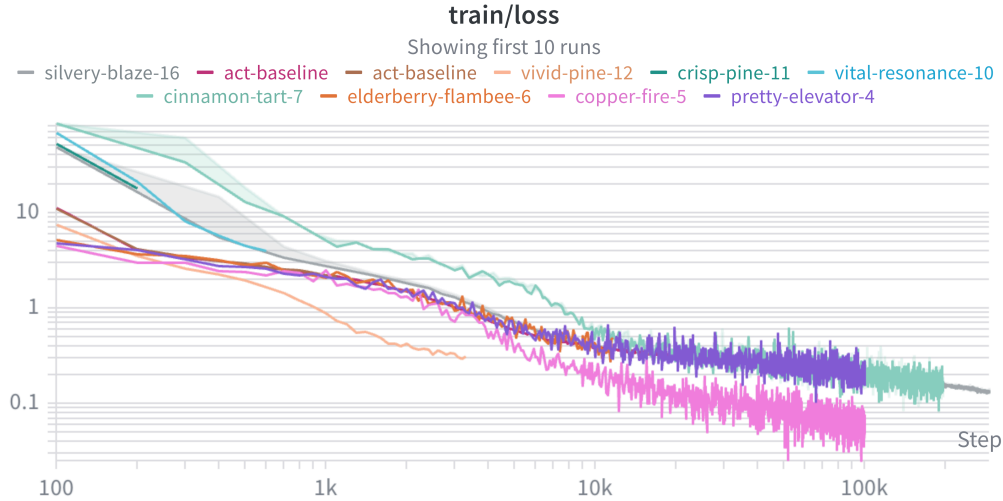


Figure 7: ACT training loss curve over 100,000 gradient steps. The smooth, monotonically decreasing curve provides confidence in the data and training pipeline; additional training beyond this point produced no meaningful further reduction in loss.

Beyond 100,000 steps, additional training runs extending to 200,000 steps showed no appreciable further decrease in loss, suggesting that the model had reached the representational capacity of the ACT architecture given the current dataset.

3.5.5.3 SmolVLA Training

Following the ACT experiments, a SmolVLA policy was trained on the full-task dataset (400 episodes). Due to the dataset-merging issues described in [Section 3.5.2](#), the pickup-only episodes could not be incorporated. SmolVLA was trained without the temporal ensembling and learning-rate scheduling modifications applied to ACT, as time constraints prevented adapting these techniques to the VLA training loop. Flow matching — a generative modelling framework compatible with VLA architectures — would enable these improvements to be applied naturally in future training runs. Despite the absence of these refinements, the SmolVLA policy produced the most performant results in our evaluation, as discussed in [Section 4.2.2](#) and detailed further in [Section 4.2](#).

3.5.6 Results Summary

The SmolVLA policy achieved the highest task success rate of any policy evaluated, while the ACT full-task policy failed to compose the pickup and fold stages into a coherent end-to-end behaviour. Stage-isolated ACT variants (pickup-only and fold-only) each performed reliably in isolation, pointing to a distributional complexity problem at the phase boundary rather than a data quality issue. Full

quantitative results, stage-isolated diagnostics, failure-mode analysis, and fold quality assessment are presented in [Section 4.2.2](#).

3.5.7 Future Directions

Three directions are planned to address the remaining limitations:

1. **Stage-Aware Reward Modelling (SARM) and reinforcement-learning fine-tuning.** Rather than training a single monolithic policy, SARM explicitly decomposes the task into stages (pickup, align, fold, release) and trains a separate reward model for each stage. A meta-controller sequences the stage policies and triggers transitions based on the reward model's assessment of stage completion. Critically, SARM transforms the sparse, binary task reward (fold complete / fold incomplete) into a dense, differentiable reward signal by providing continuous feedback at each stage boundary. This dense reward is a prerequisite for reinforcement-learning fine-tuning: standard RL algorithms require a reward signal that varies smoothly with policy quality, whereas binary success/failure provides almost no gradient information for policy improvement. SARM would therefore enable an RL fine-tuning loop in which the imitation-learned policy (ACT or SmolVLA) is used as an initialisation and the stage-aware reward models provide the training signal for further optimisation. This combination of imitation-learning initialisation and RL fine-tuning is a promising path toward robust end-to-end performance.
2. **Dataset merging and combined training.** Resolving the LeRobot dataset compatibility issues that prevented merging the pickup-only and full-task datasets would allow SmolVLA (and other policies) to train on the full 499-episode corpus. The additional pickup coverage may improve grasp reliability, which remains the dominant failure mode even for SmolVLA.
3. **Diffusion Policy [7] / flow matching [8].** Diffusion-based or flow-matching action generation naturally represents multimodal action distributions. Flow matching in particular would allow the temporal ensembling and learning-rate scheduling improvements already validated on ACT to transfer to the generative policy setting. These approaches remain candidates for addressing the pickup-to-fold transition without requiring explicit task decomposition.

4 Verification & Validation

4.1 Overview

Verification and validation address two distinct questions: *did we build the product right?* and *did we build the right product?* [Table 8](#) summarises the key tests conducted, their associated requirements, and outcomes.

Test	Type	Requirement	Question addressed	Result
20-Cycle Bench Test	Verification	REQ-01, REQ-02	Does the system meet fold accuracy and throughput specs?	Partial — SmolVLA 14/20, below 80% threshold
End-Effector Comparative Trial	Verification	FR-02, REQ-02	Which end-effector geometry maximises grasp reliability?	EES (tangential scoop) selected
Concierge Simulation (Wildlight)	Validation	VAL-01, EC-03	Is napkin folding a pain point worth solving?	All 5 hypotheses confirmed
Manager’s Blind Lineup	Validation	VAL-01	Are robot folds indistinguishable from human folds?	Deferred — MVP not yet at required quality
LOI Customer Discovery	Validation	EC-04	Will operators commit to purchasing before prototype completion?	2 LOIs in progress (22.2% conversion)

Table 8: Summary of verification and validation tests performed.

4.2 Verification

4.2.1 Fold Accuracy and Throughput (REQ-01, REQ-02)

4.2.1.1 Test Definition

The most critical verification requirement is VR-01: the system must complete a waterfall fold on a standard 20" × 20" linen napkin within 90 seconds, with final edge-alignment error ≤ 10 mm in at least 80% of trials (16/20). This requirement derives directly from REQ-01 (cycle time) and REQ-02 (fold accuracy), which together define the minimum performance for a commercially viable MVP. The 90-second target reflects the human folding benchmark under overnight batch-processing assumptions; the 10 mm accuracy threshold was established through stakeholder interviews identifying the tolerance below which folds are considered “table-ready” in fine-dining settings.

4.2.1.2 Procedure

The test was conducted on the integrated Linique platform under standard controlled lighting conditions. Twenty identical, freshly ironed 20" × 20" linen napkins were used. For each trial, a napkin was placed flat in a randomised position within the workspace (no automated feed, consistent with the MVP scope defined in FR-01). The operator issued a start command, and the system executed its full autonomous pipeline: perception, policy inference, and dual-arm actuation. Two metrics were recorded per trial:

- **Cycle time:** Wall-clock duration from start command to fold-complete signal, measured by stopwatch.
- **Edge deviation:** Maximum misalignment of the final fold edges against a reference template, measured with calipers to ± 1 mm precision.

A trial was scored as a success if both conditions were met (cycle time ≤ 90 s *and* edge deviation < 10 mm). The test was repeated identically for each of the three trained policy variants.

4.2.1.3 Results

Policy	Successes / 20	Pass rate	Mean cycle time (s)	Mean edge deviation (mm)	Notes
ACT (full-task)	0	0%	N/A	N/A	Failed at grasp stage; never initiated fold
SmolVLA	14	70%	40.2	6	Best performer; 6 failures traced to pickup

Table 9: Twenty-cycle bench test results by policy.

None of the tested policies met the 80% success criterion. ACT's total failure on the end-to-end task is consistent with the distributional complexity analysis presented in [Section 4.2.2](#): the unified pickup-fold action space exceeds ACT's representational capacity at the phase boundary. SmolVLA achieved the highest success rate at 70%, with its pretrained vision-language backbone providing stronger observation representations that partially disambiguate the pickup-to-fold transition. Of SmolVLA's six failures, five were traced to unsuccessful initial grasps (the napkin was not securely captured by the EES end-effector), and one to a mid-fold misalignment where the napkin slipped during the crease phase.

4.2.1.4 Interpretation and Gap Analysis

The 10-percentage-point gap between SmolVLA's 70% and the 80% target is attributable to two factors. First, the pickup-only dataset (99 episodes) could not be merged into SmolVLA training due to the LeRobot [\[9\]](#) schema incompatibility described in [Section 3.5.2](#); resolving this would increase pickup-state coverage and is expected to improve grasp reliability. Second, SmolVLA was trained without the temporal ensembling and learning-rate scheduling refinements validated on ACT, because time constraints prevented adapting these techniques to the VLA training loop. Flow matching [\[8\]](#) — compatible with VLA architectures — would enable both improvements. Closing this

gap is the critical technical milestone before the system can proceed to the Manager’s Blind Lineup validation test.

4.2.2 Policy Results and Analysis

To diagnose the ACT [4] full-task policy’s failure mode, stage-isolated variants were evaluated alongside the SmoIVLA [6] results reported above. A pickup-only ACT policy (99 episodes) achieved reliable grasps in nearly all trials; a fold-only ACT policy (given a pre-grasped napkin) completed folds reliably with consistent corner alignment. The full-task ACT policy (400 episodes) failed to compose the two stages, either missing the grasp or transitioning erratically between pickup-like and fold-like motions.

This pattern points to a distributional complexity problem at the phase boundary: the pickup and fold phases occupy distinct regions of the joint observation-action space, and the ACT architecture, which predicts a single action chunk per timestep, cannot resolve the multimodal action distribution where both pickup-like and fold-like actions are plausible given similar observations. SmoIVLA’s pretrained vision encoder provides richer representations that partially disambiguate this transition, explaining its superior end-to-end performance — though its 70% success rate indicates the stage-composition challenge is not fully resolved by architectural improvements alone.

When the pickup bottleneck is bypassed (fold-only policy or human-assisted grasp), fold quality is acceptable: the napkin is folded along the intended diagonal with aligned corners and a well-defined crease (Figure 8). No continuous quantitative metric for fold quality has yet been implemented; current evaluation relies on binary success/failure labelling. A vision-based fold quality metric is planned for the next prototype generation.

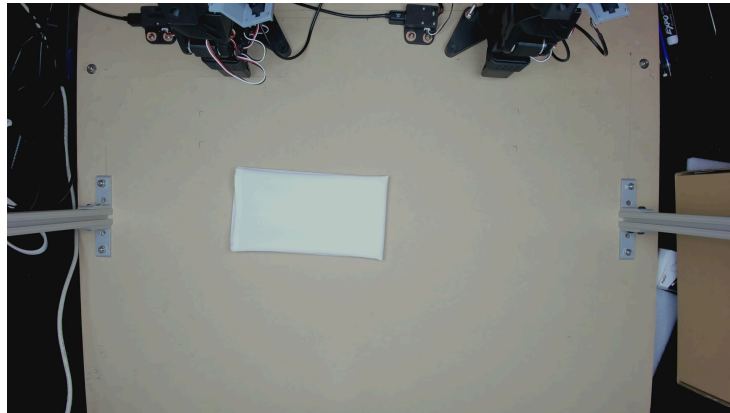


Figure 8: Completed fold produced by the fold-only policy: diagonal crease, aligned corners, and a well-defined edge.

4.2.3 End-Effector Selection (FR-02, REQ-02)

Four 3D-printed (PLA+) end-effector prototypes were evaluated through teleoperated pickup trials to characterise grasp speed, cloth deformation, and robustness across napkin states. Each design was tested on both flat and pre-folded napkins, from both the first-encounter (F.E.) and second-encounter (S.E.) approach sides, to capture the full range of grasp conditions the learned policy would face during autonomous operation.

End-Effector	Flat avg	Flat F.E.	Flat S.E.	Folded avg	Folded F.E.	Folded S.E.
OG (narrow-tip pinch)	4.03	4.79	3.26	3.60	2.26	4.94
EES (tangential scoop)	5.74	7.13	4.21	4.57	5.70	3.33
EED (inverse close)	5.27	5.32	5.22	4.17	4.11	4.24
EEJ (thick-jaw pinch)	12.68	7.32	18.04	9.44	6.78	12.11

Table 10: End-effector pickup time (seconds) across flat / folded napkins and first- / second-encounter approaches.

End-Effector	Flat avg	Flat F.E.	Flat S.E.	Folded avg	Folded F.E.	Folded S.E.
OG	100%	50%	50%	100%	50%	50%
EES	0%	0%	0%	0%	0%	0%
EED	100%	50%	50%	100%	50%	50%
EEJ	77%	40%	37%	47%	13%	33%

Table 11: End-effector crumple rate (% of trials producing visible cloth deformation).

EES was selected for integration. Despite slightly slower average pickup times than OG (5.74 s vs. 4.03 s flat; 4.57 s vs. 3.60 s folded), EES achieved a 0% crumple rate across all conditions — every other design produced deformation in at least 47% of trials. This is the decisive metric: post-grasp cloth deformation propagates downstream through the entire fold sequence, causing corner misalignment and crease defects that cannot be corrected by the policy. The 1–2 second pickup penalty is negligible relative to the 90-second cycle time budget (REQ-01) and is far outweighed by eliminating the dominant source of fold-quality degradation.

EEJ was the weakest candidate on both axes — its thick-jaw geometry required lengthy repositioning (18.04 s worst case) and still crumpled napkins in the majority of flat trials. OG and EED, while fast, crumpled the cloth in 100% of trials due to compressive force-closure loading, confirming the hypothesis that geometry-entrapment (sliding beneath the cloth edge before closing) is fundamentally better suited to deformable linen than force-closure for this application.

4.2.4 Additional Verification Status

Requirement	Status	Evidence
REQ-13 (Calibration ≤ 5 min)	Validated	Setup time is effectively irrelevant: trained policies are loaded from disk on startup and, once loaded, the robot can initiate a fold within 5 seconds of receiving a start command.
REQ-15, REQ-16 (Hygiene)	Met	All contact surfaces are 3D-printed PLA+; non-porous and wipe-clean
REQ-03 (Footprint $\leq 36 \times 36$ in)	Met (theoretical)	The SO-101 arms only require reachable workspace <i>in front of</i> their base, so the mounted footprint is bounded by the base plate itself. Bench testing has confirmed that no additional clearance is required behind or to the sides of the arms during fold cycles.
REQ-04 (Supply Voltage 120 V AC)	Met in lab	The prototype runs end-to-end off a standard 120 V AC outlet and functions correctly under lab conditions. The current power chain relies on a series of low-quality 5 V AC-DC converters that have caused intermittent low-voltage errors in the past; a cleaner regulated supply is an acknowledged follow-up.
REQ-05 (Emergency Stop)	Met	Emergency stop button wired; tested and functional
REQ-08 (Recalibration Interval)	Partial	The 20-cycle bench test ran without system crashes, servo disconnects, or power failures, providing baseline evidence of batch stability. Formal stress testing over extended runs (100+ napkins) has not been conducted; intermittent low-voltage errors from the Schottky diode drop (Appendix A) remain a risk during prolonged operation.
Portability	Met	The assembled prototype weighs approximately 12 kg and is routinely moved by a single person between the lab and demonstration venues.
FR-04 (Completion & Reset)	Met	System returns to neutral pose after fold-complete signal in all SmoIVLA trials

Table 12: Additional verification status for requirements not covered by the dedicated tests above.

4.3 Validation

4.3.1 Concierge Simulation at Wildlight Kitchen + Bar

4.3.1.1 Objective and Rationale

The original validation plan specified a Manager's Blind Lineup test (VAL-01), in which five robot-folded and five human-folded napkins are presented in randomised order to a restaurant manager for blind quality assessment. This test was deferred because SmoVLA's 70% fold success rate does not yet produce the five consecutive high-quality folds required for a meaningful comparison. Instead, a concierge simulation was conducted to validate the upstream question: whether the problem Linique addresses — manual napkin folding — represents a real, measurable operational pain point that hospitality operators would pay to eliminate.

4.3.1.2 Method

A team member employed at Wildlight Kitchen + Bar (Vancouver) manually folded all napkins required for service each day from February 22–28, 2026, simulating the output of the future Linique system. A contract covering liabilities and terms was signed with the General Manager. Each morning, the procedure was: count leftover napkins from the prior evening, check reservation count, and fold sufficient napkins for the day plus a 70–100 unit buffer. Six metrics were tracked daily: leftover count, reservations, napkins folded, fold time, final guest count, and napkins used. Following the test week, structured debrief interviews were conducted with one manager and two servers, and a sentiment survey was distributed to the full front-of-house team.

4.3.1.3 Results

Labour quantification (Hypotheses 1–3). Total folding time across seven days was 13.5 hours, averaging 1.93 hours/day — confirming the 2–3 hour per day estimate from stakeholder interviews. With folding removed, staff redirected time to neglected maintenance (window cleaning, fridge organisation, station restocking). The lead closing server reported releasing staff 30–60 minutes earlier each night, as napkin folding had been the final bottleneck in closing duties. Payroll data for the test week (Feb 22–28) and control week (Mar 1–7) has been requested from management but was not available at time of writing; once received, this will allow quantification of the exact labour-cost differential.

Staff sentiment (Hypothesis 4). All three interviewees gave unequivocally positive feedback. The assistant manager highlighted improved fold consistency and increased floor availability during high-turnover periods. When regular folding resumed the following week, servers expressed visible disappointment and the lead server reported spending additional time correcting inconsistent folds during section checks — a direct quality regression that reinforces the value proposition of automated, consistent output.

Napkin logistics (Hypothesis 5). Napkin usage tracked final guest count at 1.08 napkins/guest across the week, confirming that demand is predictable enough to plan batch automation around reservation data. Friday's 600-napkin volume required additional storage space, suggesting that high-volume days may need dedicated staging — an operational detail relevant to future deployment planning.

Day	Leftover	Folded	Fold time	Guests	Used	Napkins/guest
Sun	55	295	2h 32m	174	199	1.14
Mon	151	45	0h 40m	89	99	1.11
Tue	97	159	1h 40m	158	166	1.05
Wed	90	199	1h 40m	211	215	1.02
Thu	74	194	1h 30m	147	174	1.18
Fri	94	503	1h 45m	220	224	1.02
Sat	373	0	—	195	210	1.08
Total		1,395	13.5 h	1,194	1,287	1.08 avg

Table 13: Concierge simulation daily metrics, Feb 22–28, 2026, Wildlight Kitchen + Bar.

4.3.1.4 Limitations

The primary limitation is selection bias: the team member who conducted the simulation is a Wildlight employee, so staff feedback may skew favourable. The sentiment survey has not yet reached majority completion. Labour-cost data for the control week remains outstanding. A second concierge trial at an unaffiliated venue is planned to address the bias concern and strengthen external validity.

4.3.2 Letter of Intent Validation

To validate willingness-to-pay independently of the concierge trial, a structured LOI campaign was conducted across 19 hospitality businesses. Nine completed interviews (47% response rate). All nine confirmed manual folding; 88% reported increased folding volume before events; 66% stated they would prefer staff focused elsewhere. Five expressed strong purchase interest, three requested follow-up, and two are in the process of signing non-binding LOIs — a 22% conversion rate against a 20% target. The strongest interest came from banquet halls and high-volume event venues rather than standalone restaurants, refining the go-to-market segmentation toward event-driven operators.

4.3.3 Deferred Validation: Manager's Blind Lineup (VAL-01)

The Blind Lineup test remains the definitive validation of commercial fold quality. It will be executed once SmoVLA (or its successor policy) achieves $\geq 80\%$ success on the 20-Cycle Bench Test, ensuring that five consecutive table-ready folds can be reliably produced. The test protocol is unchanged: ten napkins (five robot, five human) presented in randomised order to a restaurant manager for blind quality assessment and source identification. Success requires all five robot folds marked “Table Ready” and the evaluator unable to reliably distinguish robot from human output.

4.4 Summary of Verification and Validation Status

Requirement	Target	Current status	Gap
REQ-01 Cycle Time (≤ 90 s)	≤ 90 s	40 s (SmolVLA avg)	Satisfied
REQ-02 Fold Accuracy	16/20 < 10 mm	14/20 (70%)	10 pp — dataset merge + training refinements
VAL-01 Commercial Quality	Blind lineup pass	Deferred	Requires REQ-02 closure first
Labour pain validation	Confirmed	5/5 hypotheses validated	Pending: control-week payroll, second venue
Market commitment	≥ 2 LOIs	2 in progress	On track

Table 14: Summary of V&V status across all key requirements.

The system is not yet MVP-ready on the technical verification axis: SmolVLA's 70% success rate must reach 80%+ through dataset merging, training refinements, and potentially diffusion-based action generation [7]. However, the validation evidence — confirmed labour savings, strong staff sentiment, predictable napkin logistics, and early LOI commitments — establishes that the problem is commercially real and the product direction is correct. Closing the verification gap is the single remaining gate to pilot deployment.

4.5 Prototyping Risks and Mitigation

Table 15 summarises the most significant prototyping risks identified during this development cycle, their assessed severity, and the mitigation steps taken or planned.

Risk	Likelihood	Impact	Mitigation
Pickup grasp failure rate limits end-to-end success below 80% threshold	High	High	Resolve LeRobot [9] dataset merging to incorporate the 99 pickup-only episodes into SmolVLA [6] training; apply temporal ensembling and learning-rate scheduling refinements already validated on ACT [4]; evaluate diffusion-based action generation [7] for the pickup-to-fold transition.
Power supply instability causes servo disconnects during extended operation	Medium	High	Schottky diode voltage drop on the XIAO USB-UART bridge [10] has caused intermittent low-voltage alarms. Short-term: monitor voltage levels during bench tests. Medium-term: bypass the protection diode or replace with a dedicated regulated 7.4 V supply.
Selection bias in concierge validation overstates labour pain point	Medium	Medium	A second concierge trial at an unaffiliated venue is planned to establish external validity. The sentiment survey will be extended to reach majority completion.
Single-operator demonstration data limits policy generalisation	Medium	Medium	All 499 episodes were collected by a single human operator, which may embed operator-specific motion habits. Future data collection will include demonstrations from at least two additional operators to broaden the training distribution.
Single fold style limits commercial viability	Low	High	The satisfaction curve analysis (Appendix I) confirms that a single reliable fold captures the majority of stakeholder value. Multi-style support is an evaluation criterion (EC-01) and is planned for the next prototype generation via SARM and additional demonstration data.
Cable strain failures at USB-C connectors	Low	Medium	Custom 3D-printed strain-relief housings (Appendix A) have been fitted and have eliminated connector failures since installation. Risk is considered mitigated.

Table 15: Prototyping risks, assessed severity, and mitigation status.

4.6 Future Verification and Validation Roadmap

The following verification and validation activities are planned for subsequent prototype generations, mapped to the evaluation criteria defined in Section 2.3.1.

- **EC-01 — Fold versatility (5+ fold types):** Each new fold style will undergo the same 20-cycle bench test protocol used for the waterfall fold. A fold is considered verified when it meets the REQ-01 cycle time and REQ-02 accuracy thresholds independently. Cross-fold regression testing (re-running

verified folds after adding a new style) will ensure that multi-style training does not degrade performance on previously learned folds.

- **EC-02 — Throughput (500+ napkins/hour):** Throughput verification will require extended batch runs of 100+ napkins to measure sustained cycle times under continuous operation. The test will also monitor for thermal drift in servos, power supply stability, and end-effector wear over extended duty cycles.
- **EC-03 — Autonomy (≤ 1 min supervision / batch):** An operator time-and-motion study will be conducted during batch runs to quantify the total human intervention time per batch. This includes napkin loading, system monitoring, and any manual recovery steps. The target is a fully autonomous batch cycle requiring only initial napkin staging and a start command.
- **EC-04 — Unit cost (\$3k–\$10k):** The current BOM will be reviewed against production-volume pricing. Key cost drivers (SO-101 arms, compute module, camera system) will be benchmarked against alternative components. The \$7,000 target retail price assumes 78% gross margin; this will be validated against the finalised production BOM.
- **VAL-01 — Manager’s Blind Lineup:** As described in [Section 4.3.3](#), this test is gated on achieving $\geq 80\%$ on the 20-cycle bench test. Additional blind lineup tests at two or more venues are planned to strengthen external validity beyond a single evaluator.

5 IP Assessment

5.1 Related IP (Prior Art)

Prior art in robotic textile manipulation spans four loose categories relevant to Linique’s system: dual-arm deformable-object control, airflow-assisted textile manipulation, household garment folding, and fabric handling for garment manufacturing. [Table 16](#) summarises the most relevant granted and published applications in each category, the scope of their technical claims, their overlap with the Linique approach, and the assessed infringement risk.

Patent (assignee)	Technical focus	Overlap with Linique	Risk
US 10,875,189 B2 — City U. of Hong Kong [11]	Dual-arm deformable-material control using real-time perception.	Broad claims over textile handling; no hospitality-fold or symbolic-program specificity.	Moderate
US 11,628,566 B2 — Draper Laboratory [12]	Coordinated manipulation of deformable and fracturable materials.	Method-level dual-arm coordination overlap; no restaurant-fold or fold-program claims.	Moderate
US 2023/0256608 A1 — Toyota Research Institute [13]	Airflow-assisted dual-arm garment unfolding in an open workspace.	Garment preparation rather than decorative folding; methodologically distinct.	Low–Moderate
US 2019/0390396 A1 — LG Electronics [14]	Household laundry garment folding.	Flat garment folding within a laundry workflow; no staged napkin manipulation.	Low
WO 2024/077436 A1 — Center for Garment Production [15]	Dual-manipulator textile handling for garment manufacturing.	Method-level dual-arm overlap; different end-effector geometry and workflow.	Moderate

Table 16: Consolidated prior-art landscape: cited patents, technical scope, overlap with the Linique system, and assessed infringement risk.

Commercial landscape: Dyna Robotics represents the most relevant emerging competitor, having demonstrated autonomous napkin folding with high reliability. However, no clearly published patents are assigned to the company; the primary risk is commercial rather than IP-based.

5.2 Freedom to Operate

The prior art surveyed above does not constrain Linique’s ability to deploy its system. The existing patents focus on general-purpose deformable-object manipulation, garment folding in household or industrial contexts, and airflow-assisted unfolding — none of which target decorative napkin folding in hospitality settings. Their claims are oriented toward broad textile handling rather than the specific combination of symbolic fold programs, imitation-learning-driven motion planning, and multi-camera state estimation for thin linens that define Linique’s approach.

No identified patent covers the integrated system architecture Linique employs: staged fold execution with coordinated dual-arm control, perception optimised for presentation-grade napkin geometry, and end-effectors tailored for precise cloth handling. While broad claims in patents such as US 10,875,189 B2 [11] and US 11,628,566 B2 [12] create some overlap at the method level (dual-arm coordination, deformable-object control), Linique's narrow application domain and distinct technical pipeline provide sufficient differentiation to operate without infringement.

Dyna Robotics poses no known IP barrier at this time, though their patent activity should be monitored as the competitive landscape evolves.

Open-source dependencies. Linique's software stack rests on a small set of permissive open-source dependencies — LeRobot (Apache 2.0), the SmolVLA backbone and released weights (Apache 2.0), and the underlying PyTorch (BSD) and Hugging Face `transformers` (Apache 2.0) infrastructure. None of these carry a copyleft obligation, so a closed-source commercial release of the integrated Linique system is compatible with every upstream licence; the only downstream obligation is standard attribution bundled with the distributed binary.

Assessment: Linique has freedom to operate under the current IP landscape.

5.3 Patentability

While the general concept of a napkin-folding robot is not independently patentable due to prior art, Linique's specific implementation contains several novel, patentable elements:

- **Imitation-learning-driven fold representation system** that maps symbolic fold programs to dual-arm motion plans, enabling repeatable execution of complex decorative folds.
- **Multi-camera state estimation pipeline** optimised for thin linens, providing accurate pose estimation within a controlled workspace where conventional textile perception approaches struggle.
- **Coordinated dual-arm execution architecture** purpose-built for presentation-grade folds, where visual consistency and speed are primary design constraints.
- **Specialised end-effectors** designed for precise cloth handling in a hospitality context.

These elements shift Linique's innovation from generic cloth manipulation toward a system-level approach with clear novelty in its integration and application domain.

Linique should pursue patent protection on the integrated system design, with independent claims covering the fold representation and perception pipeline. Filing early is advisable given the absence of published IP from direct competitors such as Dyna Robotics, establishing a first-mover advantage in a narrowly defined but commercially valuable space. From an economic standpoint, a granted patent in this area would create a meaningful barrier to entry for competitors targeting the hospitality napkin-folding niche, strengthening Linique's position in licensing discussions and protecting margins as the market develops.

References

- [1] NSF International, "Food Equipment," NSF International, NSF/ANSI 2–2019, 2019.
- [2] International Organization for Standardization, "Robots and Robotic Devices --- Safety Requirements for Industrial Robots --- Part 1: Robots," ISO, ISO 10218-1:2011, 2011.
- [3] International Organization for Standardization, "Robots and Robotic Devices --- Collaborative Robots," ISO, ISO/TS 15066:2016, 2016.
- [4] T. Z. Zhao, V. Kumar, S. Levine, and C. Finn, "Learning Fine-Grained Bimanual Manipulation with Low-Cost Hardware," in *Robotics: Science and Systems*, 2023. [Online]. Available: <https://tonyzhaozh.github.io/aloha/>
- [5] J. Zheng *et al.*, "X-VLA: Soft-Prompted Transformer as Scalable Cross-Embodiment Vision-Language-Action Model," 2025. [Online]. Available: <https://arxiv.org/abs/2510.10274>
- [6] M. Shukor *et al.*, "SmoIVLA: A Vision-Language-Action Model for Affordable and Efficient Robotics," 2025. [Online]. Available: <https://arxiv.org/abs/2506.01844>
- [7] C. Chi *et al.*, "Diffusion Policy: Visuomotor Policy Learning via Action Diffusion," in *Robotics: Science and Systems*, 2023. [Online]. Available: <https://diffusion-policy.cs.columbia.edu/>
- [8] Y. Lipman, R. T. Q. Chen, H. Ben-Hamu, M. Nickel, and M. Le, "Flow Matching for Generative Modeling," in *The Eleventh International Conference on Learning Representations*, 2023. [Online]. Available: <https://openreview.net/forum?id=PqvMRDCJT9t>
- [9] R. Cadene, S. Alibert, A. Soare, Q. Gallouedec, and T. Wolf, *LeRobot: State-of-the-Art Machine Learning for Real-World Robotics*. (2024). Hugging Face. [Online]. Available: <https://github.com/huggingface/lerobot>
- [10] Seeed Studio, "Bus Servo Driver Board for Seeed Studio XIAO Datasheet." [Online]. Available: https://files.seeedstudio.com/wiki/reComputer-Jetson/lerobot/bus_servo_xiao.pdf
- [11] J. Pan, Z. Hu, and others, "System and Method for Manipulating Deformable Objects," US 10,875,189 B2, Jan. 2021
- [12] D. M. S. Johnson, S. Wagner, S. Lines, and M. Hebert, "Manipulating Fracturable and Deformable Materials Using Articulated Manipulators," US 11,628,566 B2, Apr. 2023
- [13] R. Tamura and others, "Systems and Methods for Deformable Object Manipulation Using Air," US 2023/0256608 A1, Aug. 2023
- [14] S. Park and others, "Household Laundry Garment Folding," US 2019/0390396 A1, Dec. 2019
- [15] Z. Li and others, "Fabric-Handling Dual-Manipulator System with Fabric Rolling-Up End-Effectors," WO 2024/077436 A1, Apr. 2024

6 Appendices

Document List

Documents supporting this report split into three groups: appendices bound into this PDF (A-J, below), supplementary documents in the project’s Google Drive archive, and externally hosted living artifacts

In the Google Drive Archive

All supplementary documents are collected in a single shared folder: [Linique Project Drive Archive](#).

Type	Link	Name	Note
CAD drawing	Drive	Assembly Drawing	Static SVG export of the V1 full assembly — referenced from Section — Product Architecture and Appendix G . The authoritative live model is on Onshape (see Table 18).
CAD drawing	Drive	UART Bridge Mount Drawing	SVG drawing of the custom 3D-printed strain-relief housing for the XIAO USB-UART bridge — matches the CAD shown in Appendix A .
Prototype photos	Drive	Prototype Images	Photo set documenting the most up-to-date hardware build of the Linique system — gantry, arm mounts, end-effectors, and custom housings as physically assembled.
COTS datasheet	Drive	XIAO USB-UART Bridge	Vendor PDF for the Seeed Studio XIAO Servo Driver Board used as the USB-UART bridge — the source for the schematic figure in Appendix A .
Requirements matrix	Drive	Engineering Requirements	Master engineering-requirements spreadsheet backing Section — Requirements .
Specification	Drive	Specification	Engineering specification defining design issues, requirements (REQ-01–REQ-19), evaluation criteria, satisfaction curves, and value equation — backing Section — Requirements .
BOM / costing	Drive	BOM and Costs	Bill of materials with unit costs and project-budget tracking.
Test data	Drive	End Effector Testing	Raw comparative-trial data for the four end-effector variants evaluated in Section — End-Effector Design and Appendix C .

Type	Link	Name	Note
Interview bundle	Drive	Stakeholder Interviews	Ten interview transcripts covering both discovery and post-concierge debriefs. Aman Nijar and Damian Luong are the two post-concierge debriefs referenced in Section — Concierge Simulation the remaining eight (Elise Hall, Fleur Adele, Karen Jones, Margot Baloro, Michael Cooke, Sarah Watson, Sean Worthington, Stephanie Couture) are discovery / LOI interviews.
Concierge notes	Drive	Concierge Interview Notes	Running notes from the Wildlight Kitchen + Bar concierge simulation — see Section — Concierge Simulation .
Validation report	Drive	Customer Validation Report	Written summary of the customer-discovery process and the LOI conversion funnel described in Section — Concierge Simulation .
Rollout video	Drive	Inference Rollout Videos	Four-camera capture (top, angle, left, right) of a SmolVLA policy rollout on hardware — supporting evidence for the results reported in Section — Fold Accuracy and Throughput .
Conference poster	Drive	IEEE ML Poster	Poster summarising the ML / policy work prepared for IEEE submission.

Table 17: Supporting documents hosted in the Linique Google Drive archive.

Externally Hosted

CAD that is still being revised, code under active development, and public datasets live at their authoritative URLs rather than as frozen copies in the archive.

Type	Link	Name	Note
CAD (Onshape)	Onshape	Linique Full Assembly	Authoritative live CAD for the complete rig — gantry, arm mounts, end-effectors, and camera mounts.
CAD (Onshape)	Onshape	Custom End-Effector (EES Tangential Scoop)	Live CAD for the selected EES tangential-scoop end-effector described in Section — End-Effector Design and Appendix C .
CAD (Onshape)	Onshape	USB-UART Bridge Strain-Relief Casing	Live CAD for the custom 3D-printed housing described in Appendix A .

Type	Link	Name	Note
Code repo (GitHub)	GitHub	linique-napkins/robo-ops	Main Linique monorepo: training, inference, teleoperation, and data-ingest tooling.
Code repo (GitHub)	GitHub	linique-napkins/nvd-utils	NVD project-management utilities: Gantt charts, milestone planning, and report-tooling scripts.
Dataset collection (HF)	Hugging Face	jhimmens/linique-v2-*	Public teleoperation datasets used for policy training. ¹
Project website	Website	Linique Robotics	Public-facing project website — overview, demo footage, and contact information.

Table 18: Externally hosted Linique artefacts referenced in this report.

¹See [Appendix J](#) for the full breakdown.

Appendix A — USB-UART Bridge Hardware Notes

Communication between the host computer and each arm's servo chain uses Seeed Studio XIAO Servo Driver Boards acting as USB-UART bridges. Two hardware issues were identified and partially resolved during integration.

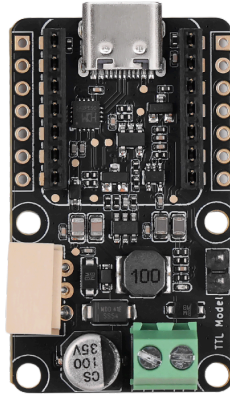


Figure 9: As-installed USB-UART bridge with the custom 3D-printed strain-relief housing fitted over the USB-C connector.

Cable Strain Relief

USB-C cable failures were observed at the bridge connector due to repeated bending during arm motion. A custom 3D-printed PLA+ strain-relief housing ([Figure 10](#)) was designed and fitted to clamp the cable at the connector body, transferring bending loads to the housing rather than the solder joint.

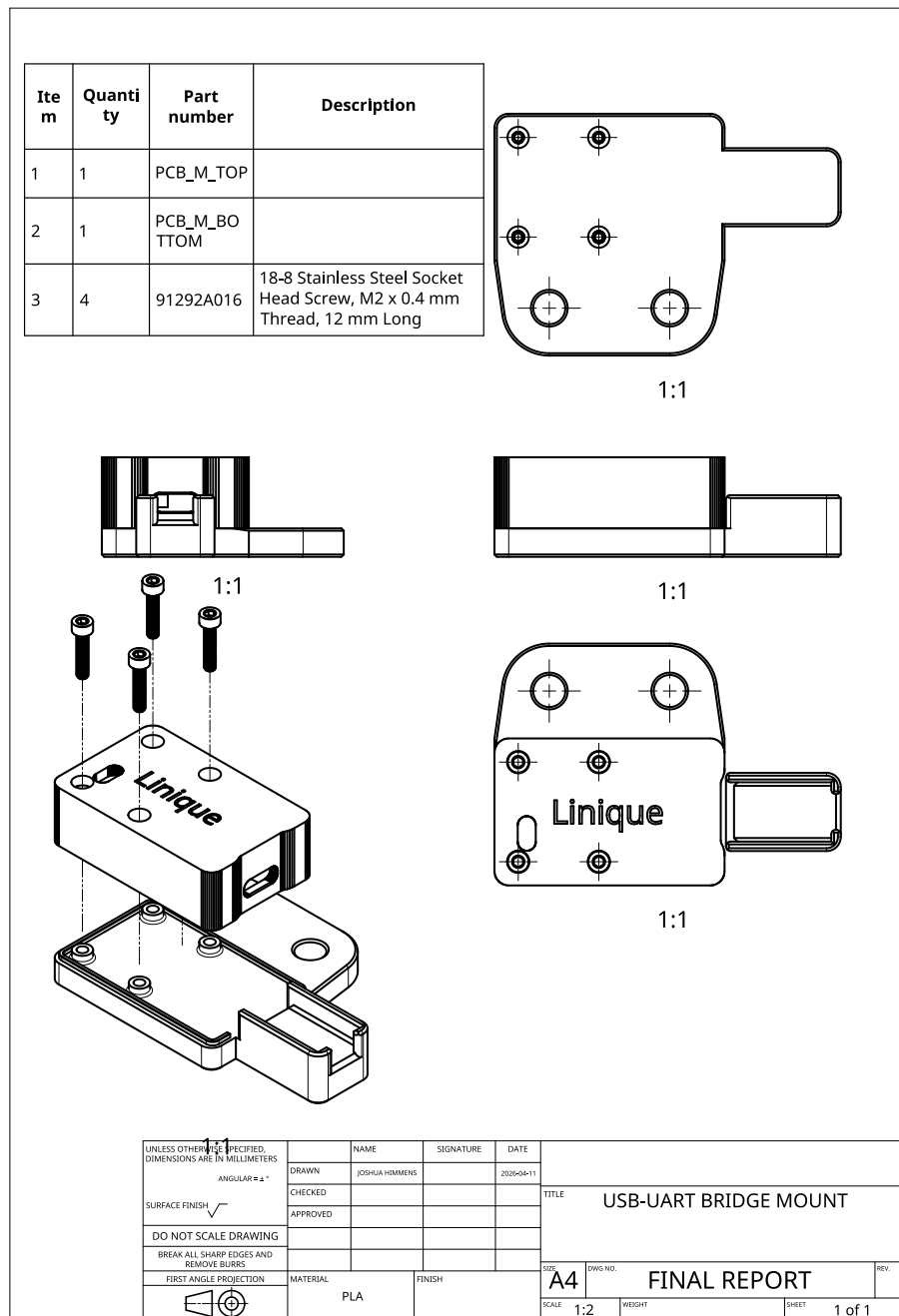


Figure 10: CAD drawing of the custom strain-relief housing. The housing clamps around both the bridge PCB and the USB-C cable so that any bending moment is carried by the plastic shell rather than the solder joint at the connector.

Schottky Diode Voltage Drop

The bridge board's internal Schottky protection diode reduces the delivered bus voltage from the rated 7.4 V to approximately 4.8–4.9 V. The FeeTech motors require 5–8.4 V (rated 7.4 V), so this drop causes intermittent low-voltage alarms and LeRobot framework disconnects. The full vendor

schematic for the XIAO Servo Driver Board (Figure 11) shows the B5819WS Schottky on the 5 V rail feeding the SCS servo chain; this is the component responsible for the observed drop. A bypass or dedicated regulated supply is under evaluation for future revisions.

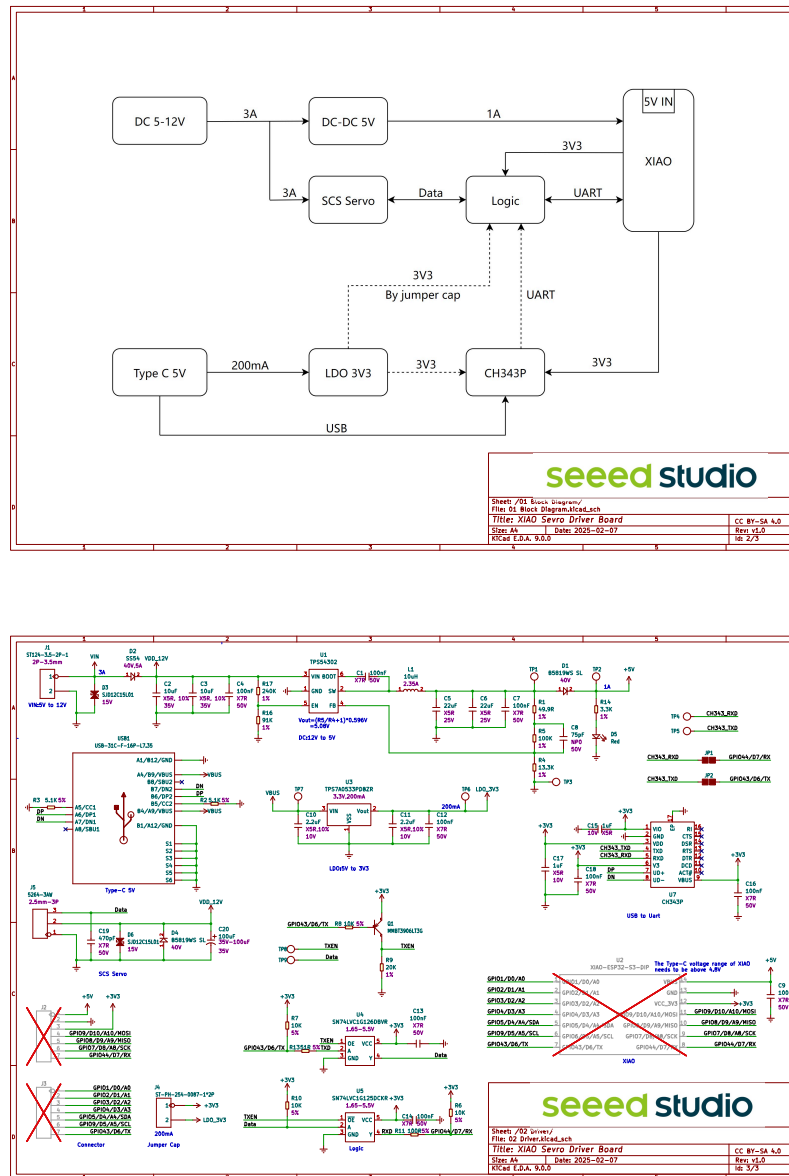


Figure 11: Sseed Studio XIAO Servo Driver Board vendor schematic (CC BY-SA 4.0), reproduced here as the authoritative reference for the bridge hardware. Top: block diagram showing the bridge's power rails and data paths — a DC 5–12 V input is stepped down to 5 V for the SCS servo chain and the XIAO, while the USB-C port feeds a separate 3.3 V rail via an LDO into the CH343P USB-to-UART controller. Bottom: full driver schematic. The B5819WS Schottky on the 5 V output rail (between the TPS54302 buck converter and the servo supply connector) is the protection diode responsible for the 0.5 V drop discussed above. The XIAO footprint is crossed out because the board revision in use does not populate it.

Appendix B — Concept Visualisation

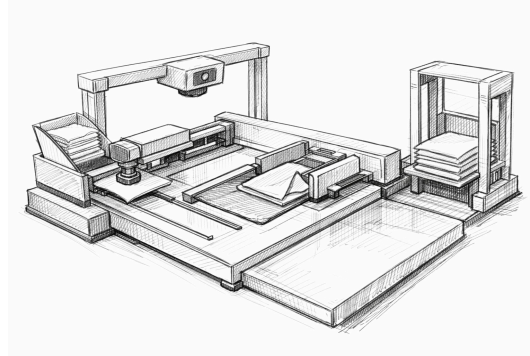


Figure 12: Classical fixture concept — napkins arrive in a known orientation on a feed table and are folded through a fixed sequence of timed actuators (lifting table, clamps, suction cups) that keep the cloth fully constrained at every step. Corresponds to Concept B in Table 5.

The top sketch represents the classical end of the concept space and corresponds most closely to **Concept B (Conveyor-Guided Folding)** in Table 5. Napkins enter in a defined orientation on a feed table, and a fixed sequence of timed actuators, a lifting table, compliant clamps, and suction cups, carries the cloth through each fold stage, so the napkin is geometrically constrained at every intermediate state. The hardware is purpose-built for one fold style and one napkin geometry, but within that narrow operating point it is fast, highly repeatable, and inexpensive once tuned, because every motion is driven by a designed mechanism rather than a learned controller. The cost is flexibility: changing the fold, the cloth weight, or the napkin dimension requires re-tooling the actuators and guides rather than retraining a model.

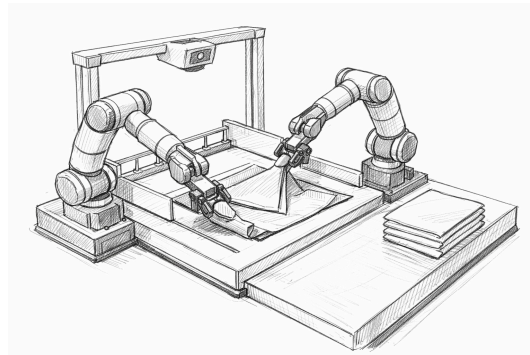


Figure 13: Dual-arm learned concept — two 6-DOF manipulators draw napkins from an unstructured input pile, compose the fold in a shared workspace, and place the folded cloth on an output stack, with no hard mechanical constraints on intermediate napkin pose. Corresponds to Concept A in Table 5.

The second sketch represents the general-purpose end of the concept space and corresponds to **Concept A (Dual Robotic Arm Folding)** in Table 5. Two 6-DOF arms reach into an unstructured input pile, lift an arbitrary napkin, and compose the fold in a shared workspace before returning the folded cloth to an output region. Because the napkin is never mechanically constrained, every motion has to be driven by a closed-loop controller that reacts to perceived cloth pose, which, in practice, means a learned policy rather than a hand-tuned trajectory. The problem is materially harder to engineer, imitation-learning data collection, vision, and dual-arm coordination all become first-class concerns,

but the resulting system can be retargeted to new folds, new cloth types, and new napkin sizes without changing hardware.

Together, these two sketches define the trade-off underpinning the concept-selection decision documented in [Section — Concept Selection Rationale](#): the classical fixture approach trades flexibility for tuned precision and cost, while the dual-arm learned approach trades engineering complexity for breadth of capability. Linique’s final direction (Concept D, Hybrid Robot + Passive Guides) sits between them, retaining the learned dual-arm manipulator from Concept A for the flexible corner-handling work while borrowing the idea of passive alignment guides from Concept B wherever a mechanical constraint is cheaper than a learned one.

Appendix C — End-Effector Concept Generation

Early end-effector ideation explored six distinct geometries before the down-selection described in [Section — Down-Selected Prototype Concepts](#). [Figure 14](#) is a composite sketch sheet capturing all six geometries; [Table 19](#) summarises the rationale for each.

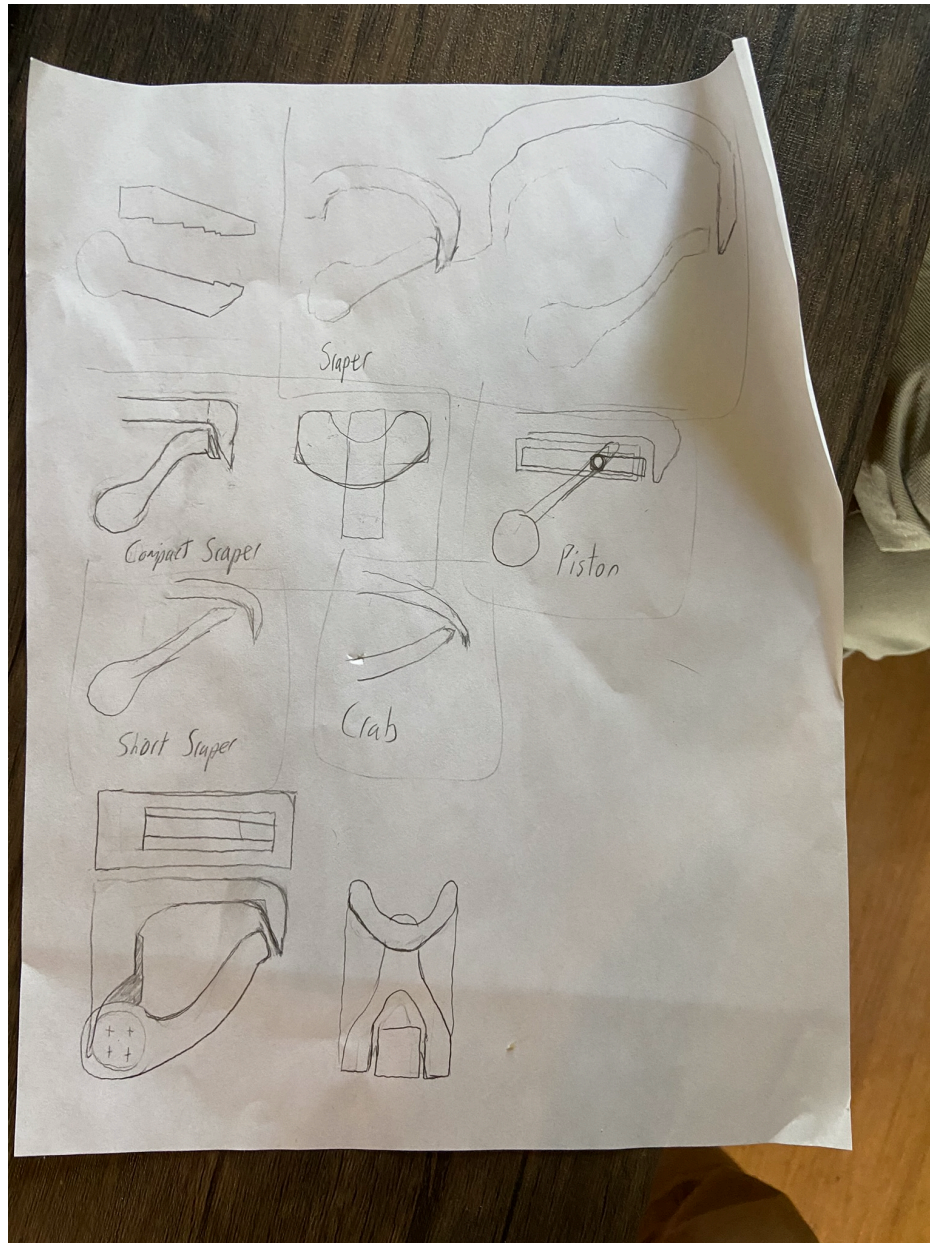


Figure 14: Composite end-effector concept sheet: Piston, Crab, Scaper, Compact Scaper, Short Scaper, and Hybrid Scaper geometries explored during early ideation.

Concept	Rationale
Piston	Linear plunging mechanism intended to create a controlled downward tuck by pressing the napkin corner into a fixed guide path. Prioritised for simple actuation and repeatable vertical motion.
Crab	Dual-hook lateral grabbing concept designed to pull and reposition napkin corners inward from two sides. Explored for improved corner isolation during fold initiation.
Scraper	Curved sweeping end effector intended to drag fabric edges into alignment through a single continuous motion. Considered for its low mechanical complexity and broad cloth contact.
Compact Scraper	Short-radius scraper geometry developed to maintain the sweeping alignment function while reducing footprint and improving integration with the robot wrist workspace.
Short Scraper	Reduced-length scraper variant optimised for tighter turning radius and faster localised cloth redirection near fold corners.
Hybrid Scraper	Combined scraping and guided support concept integrating broad contact alignment with localised corner capture features.

Table 19: Early end-effector concepts and their design rationale. These preceded the four physical-prototype down-selects (Table 6) that were ultimately tested on hardware.

Appendix D — Function Structure Diagram

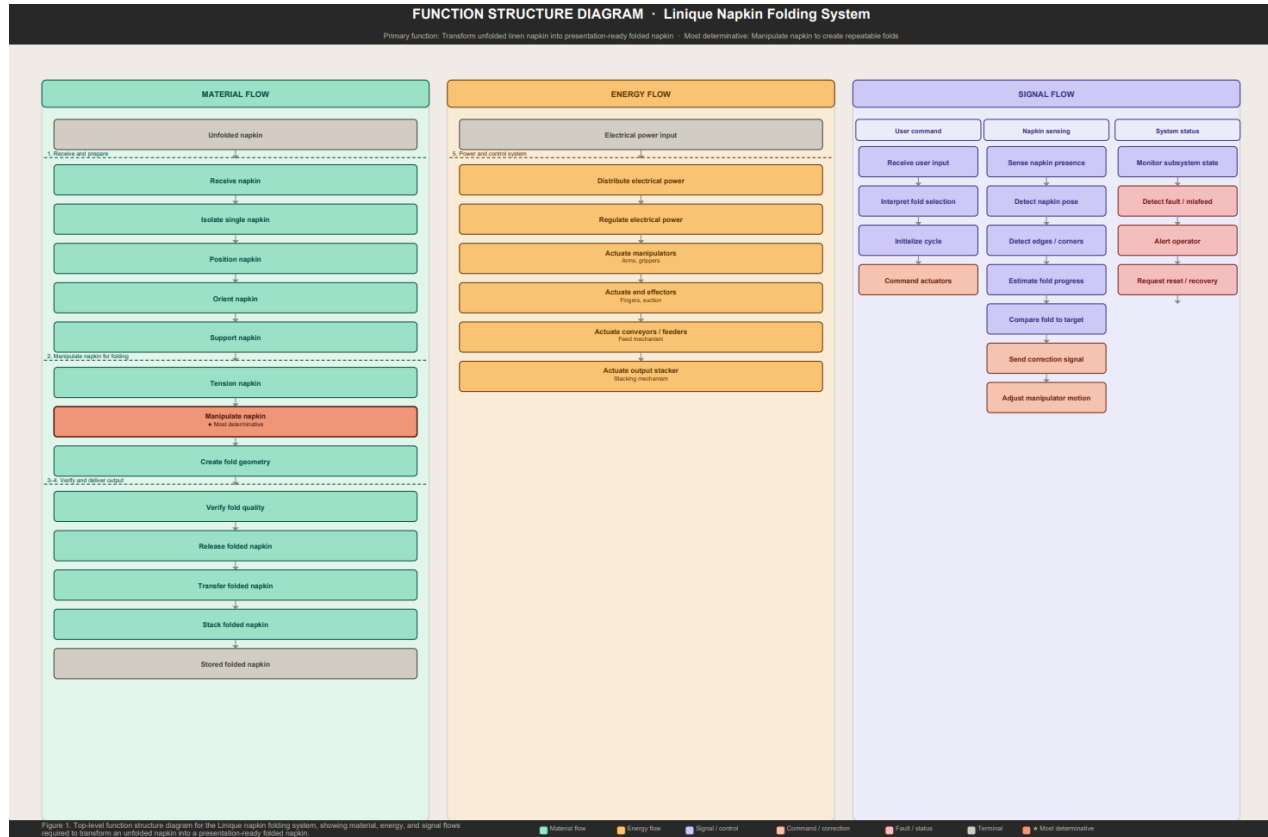


Figure 15: Function structure diagram for the Linique napkin-folding system, decomposed into material (linen), energy (electrical and mechanical), and information (vision, control, user command) flows through the major subsystems. Referenced from [Section — Product Architecture](#).

Appendix E — Morphological Chart

The morphological chart spans the eight design-driving functions of the system and four sub-concepts per function. System-level concepts A–D discussed in [Section — System-Level Concept Generation](#) were composed by selecting one sub-concept per row.

Function	Option A	Option B	Option C	Option D
Pick / isolate napkin	Friction roller singulation	Vacuum pickup head	Compliant finger pinch	Sliding separator plate
Align cloth	Fixed alignment fence	Vision-guided alignment	Passive guide rails	Tension against reference edge
Manipulate corners / edges	Dual robot grippers	Suction cups	Soft compliant fingers	Paddle / flipper arms
Create fold geometry	Robot fold-over motion	Passive folding plates	Hinged crease guides	Roller-assisted crease
Verify fold geometry	Overhead RGB camera	Wrist-mounted cameras	Template-matching vision	Edge / corner detection
Recover from misfold	Regrasp and retry	Reject to discard bin	Operator reset prompt	Pause and alarm
Stack completed napkin	Passive stack tray	Guided vertical stacker	Conveyor outfeed	Robot place-down
Control strategy	ML + hybrid scripted	Vision-assisted closed loop	Rule-based correction	Open-loop control

Table 20: Morphological chart used to generate system-level concepts A–D. Each concept in [Table 5](#) was assembled by picking one sub-concept from each row.

Appendix F — Weighted Decision Matrix

Weighted scoring of the three surviving system-level concepts (A, B, D; Concept C was eliminated at initial screening) against eight evaluation criteria. Scoring scale: 1 = poor, 5 = excellent. Concept D — Hybrid Robot + Passive Guides — ranked highest with a weighted total of 4.15, as discussed in [Section — Concept Selection Rationale](#).

Criterion	Weight	A	A·w	B	B·w	D	D·w
Fold quality	0.20	5	1.00	3	0.60	4	0.80
Repeatability	0.15	4	0.60	3	0.45	4	0.60
Cycle time	0.10	3	0.30	5	0.50	4	0.40
Footprint	0.10	3	0.30	4	0.40	4	0.40
Cost	0.10	2	0.20	4	0.40	4	0.40
Ease of user interaction	0.10	3	0.30	4	0.40	4	0.40
Robustness to linen variation	0.15	4	0.60	2	0.30	5	0.75
Scalability	0.10	3	0.30	4	0.40	4	0.40
Total	1.00		3.60		3.45		4.15

Table 21: Weighted decision matrix for system-level concepts A, B, and D. Concept D (Hybrid Robot + Passive Guides) ranked highest with 4.15 / 5.00.

Appendix G — Physical System Construction

The physical build progressed through two distinct hardware revisions. The first focused on getting a teleoperated data-collection rig working at all; the second incorporated the lessons learned during the first round of dataset capture.

Initial Assembly: Optical Breadboard + Planar Lights

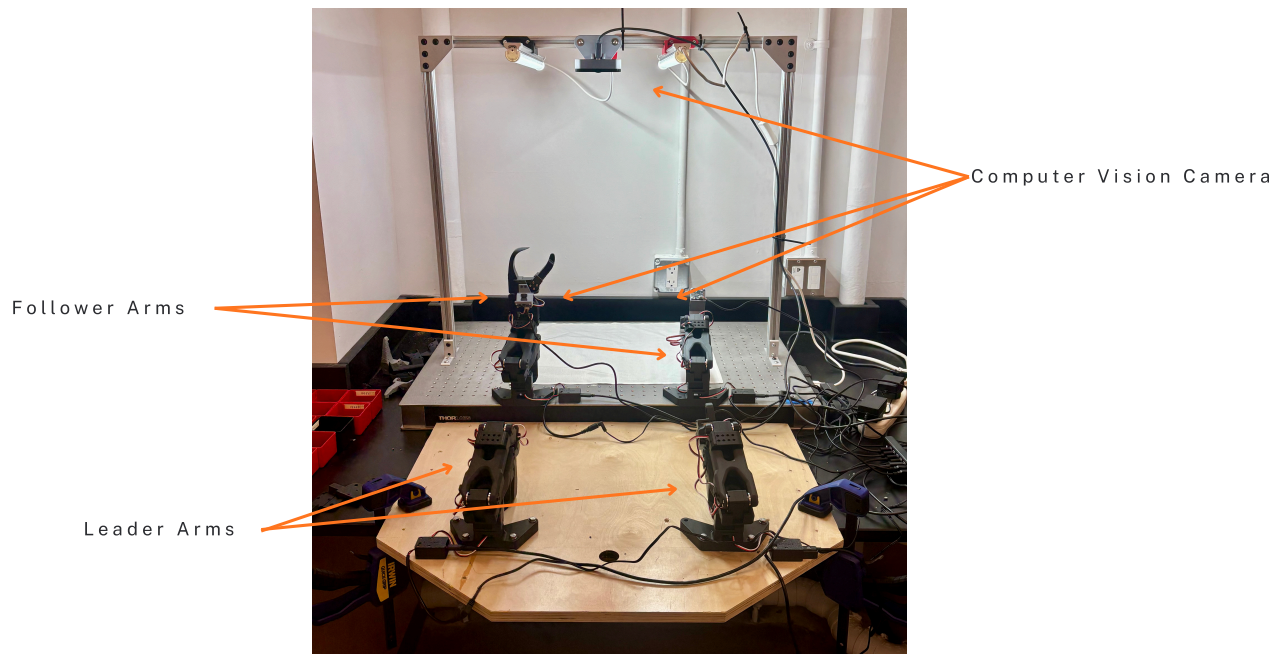


Figure 16: As-built Linique hardware (secondary assembly): continuous hardboard base, aluminium-extrusion gantry, dual SO-101 leader and follower arms, overhead ring light, wrist cameras, and the EES tangential-scoop end-effectors.

The first prototype was built on a full aluminium optical breadboard with the SO-101 follower arms mounted directly to the breadboard surface and a pair of planar LED light bars mounted above on an extrusion gantry for controlled illumination. The optical breadboard was selected for its precision hole pattern and rigidity — every component could be anchored at exact, repeatable positions, which simplified arm calibration against the overhead camera.

In practice the breadboard was extraordinarily expensive and almost completely immobile — transporting, demoing, or reconfiguring the rig was impractical, and the rigidity it provided turned out to be far in excess of what the task actually required.

Secondary Assembly: Continuous Hardboard + Ring Lights

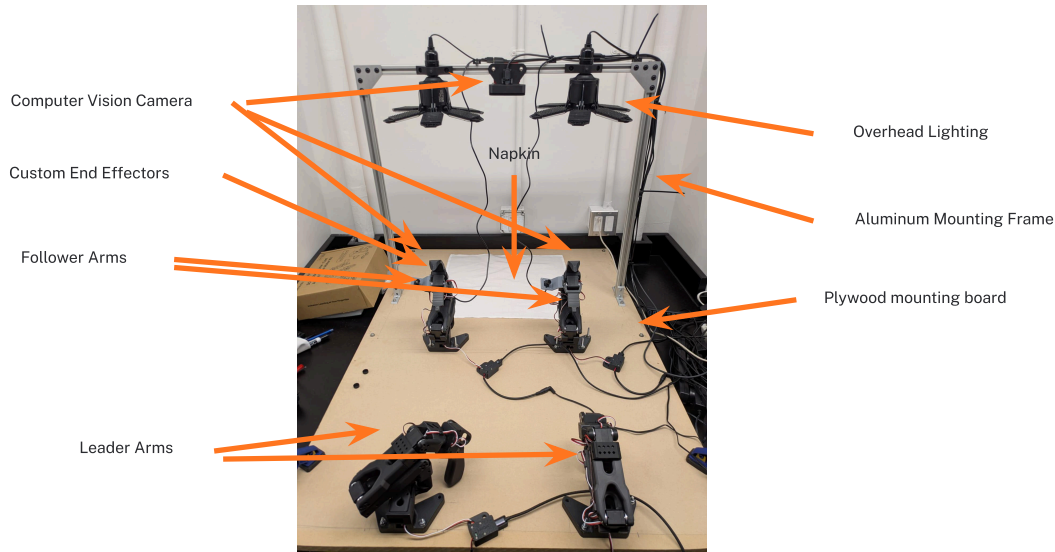


Figure 17: Linique prototype showing the aluminium-extrusion gantry, overhead camera mounts, and dual SO-101 arms on the wood-panel base.

The second-revision assembly (Figure 16) retains the overall geometry of the first build but replaces the optical breadboard with a single continuous hardboard panel, which was empirically stiff enough to hold the arms square during folding while being cheap, portable, and easy to modify. Both the leader and follower SO-101 arms are mounted to this single panel in fixed positions, keeping the teleoperation-to-autonomy pipeline aligned without any re-calibration between modes.

The planar light bars were also replaced in this revision. The planar bars' AC-driven rectification circuit created visible flicker that beat against the camera exposure, producing aliasing artifacts in the captured frames. They were swapped for DC ring lights, which eliminated the aliasing and also happened to give more uniform illumination on the napkin surface.

Appendix H — Annual Napkin Folding Cost per Restaurant

Restaurants spend meaningful paid time on decorative napkin folding every day. Across our stakeholder interviews, managers consistently described folding as a repetitive side duty that takes staff away from more valuable work such as table resets, polishing, guest prep, and event setup.

Referenced from [Section — Context](#).

Stakeholder Evidence

- Sarah Watson (Mission Hill Winery) reported 200–500 napkins per day and stated: *“If I didn’t have to pay people to fold napkins, I wouldn’t.”*
- Sean Worthington (Mission Hill Winery) estimated folding speed ranges from 7–10 s per napkin for strong staff to 30–40 s per napkin for slower workers.
- Michael Cooke and Margot Baloro both reinforced that side-duty labour is a major cost concern in fine dining.

These interviews support using a simple 1–3 hours per day estimate for premium hospitality venues.

Hand Calculations

Time estimate (middle case). Using 300 napkins/day at 20 s per napkin average:

$$\text{Time} = 300 \times 20 \text{ s} = 6000 \text{ s/day} = 1.67 \text{ h/day}$$

which fits comfortably inside the 1–3 h/day stakeholder claim.

Annual cost estimate (typical venue). Using 2 h/day, \$25/h, and 300 service days/year:

$$\text{Cost} = 2 \times 25 \times 300 = \$15,000 \text{ /year}$$

Annual cost estimate (upper-end venue). Using 3 h/day, \$30/h, and 330 service days/year:

$$\text{Cost} = 3 \times 30 \times 330 = \$29,700 \text{ /year}$$

The \$15–30k annual range cited in [Section — Context](#) is the simple, defensible estimate produced by these two bracketing calculations against the stakeholder-reported labour rates.

Appendix I — Satisfaction Curves

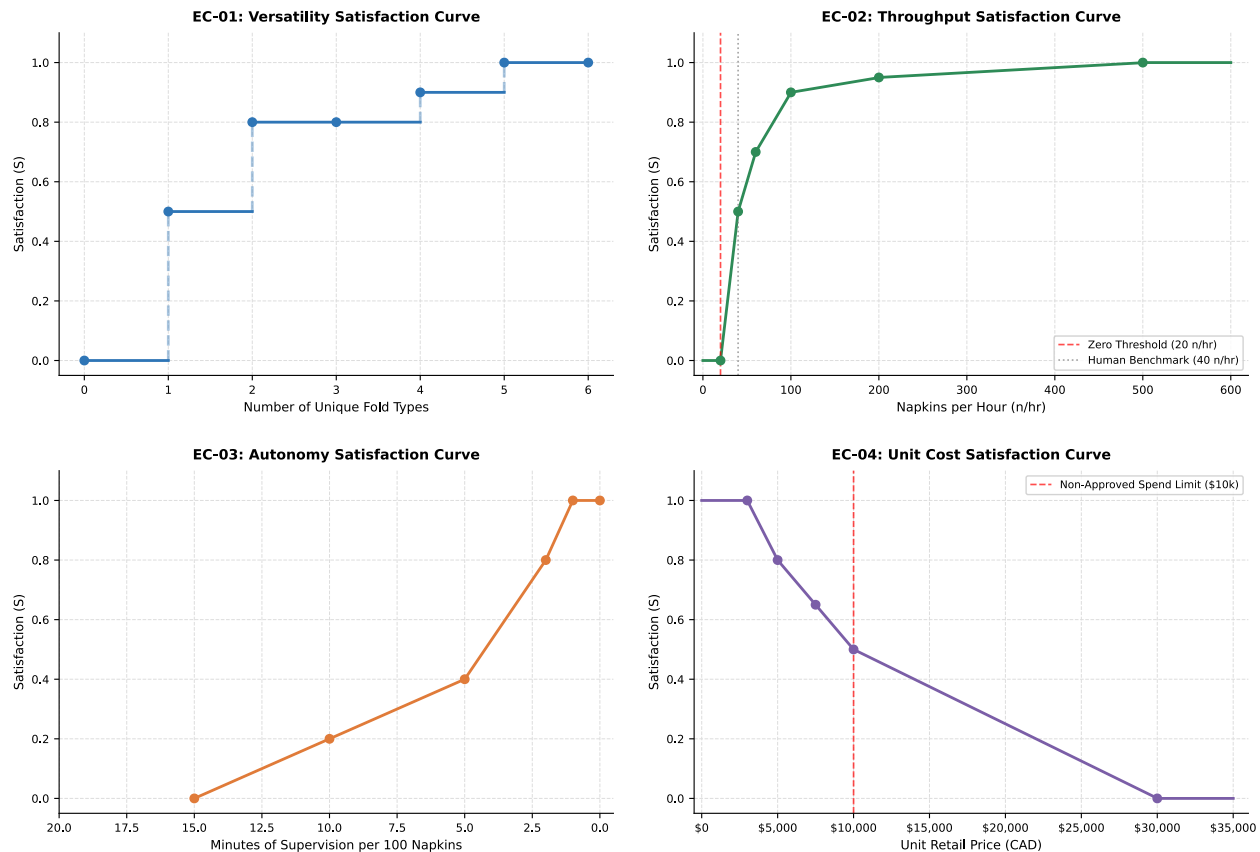


Figure 18: Stakeholder satisfaction curves across the primary evaluation criteria (versatility, throughput, autonomy, unit cost). Regenerated from `diagrams/satisfaction_curves.py`. Referenced from [Section — Evaluation Criteria](#).

Versatility (number of supported folds). The steepest jump in satisfaction comes from moving between zero and one fold; each additional fold style contributes progressively less. Operators consistently told us that having *any* reliable automated fold was vastly more valuable than having none, and the marginal value of a second and third fold style, while real, was much smaller. This implies that an MVP which performs a single fold well is already an operable product — a useful framing because it let us focus engineering effort on one decorative fold without apologising for a narrow scope.

Throughput (napkins per hour). The throughput curve is also quasi-logarithmic, with its knee at roughly 40 napkins per hour — the pace a trained human operator can sustain. Below 40/hr the system fails to displace any meaningful labour; above it, each additional napkin per hour contributes a smaller satisfaction increment. The SmoVLA policy evaluated in [Section 4.2](#) achieves a mean cycle time that comfortably exceeds the human-parity threshold, so the Linique MVP sits well past the throughput knee and into the diminishing-returns regime — i.e., it is not merely acceptable on this axis, it is clearly better than the human baseline.

Unit cost. The cost curve is the mirror image: satisfaction starts high at low prices and falls off as prices rise. The midpoint of the curve sits at roughly \$10,000 per unit. The Linique MVP's BOM comes

in meaningfully under that figure, which places the product on the favourable side of the knee and — combined with the throughput and versatility results above — makes the economic case for the MVP reasonably strong.

Taken together, these four curves say something specific about the MVP: it is not the most capable napkin-folding product an operator could eventually want, but it is already past the knee on every satisfaction axis that matters — one operable fold style, throughput that exceeds a human operator, and a unit cost well below the midpoint of the price-sensitivity curve. That is a meaningful outcome, especially given how tight the project budget was.

Appendix J — Hugging Face Datasets

All data collected during the project is hosted publicly on Hugging Face under the `jhimmens/` namespace. The primary training corpus is `linique-v2-fold-pickup`, which combines the full-task fold episodes (`linique-v2`) with the additional pickup-only episodes (`linique-v2-pickup`) into a single merged dataset. [Figure 19](#) shows a representative observation window from this dataset: synchronised frames from the overhead and wrist cameras paired with the joint-angle trajectory driving the arms.

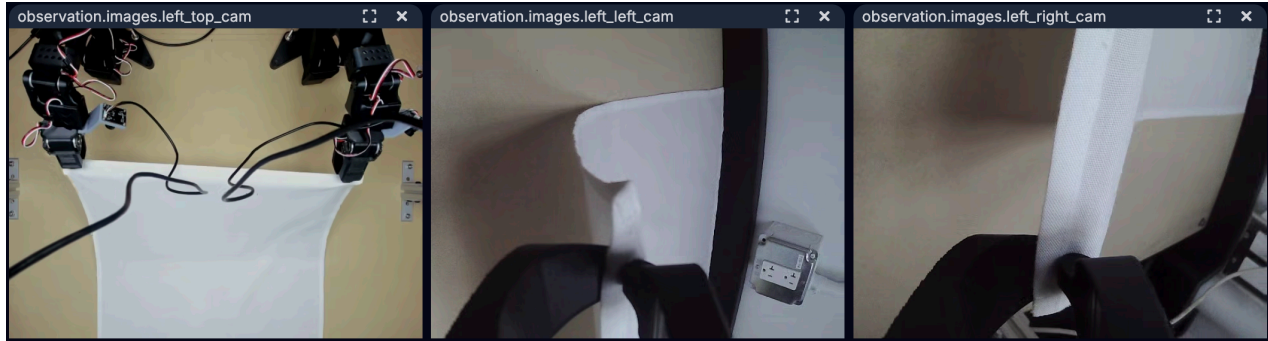


Figure 19: Example observation window from the primary training dataset `jhimmens/linique-v2-fold-pickup`: synchronised overhead and wrist camera frames alongside the commanded joint-angle trajectory. A single episode contains roughly 1,100 such frames captured at 30 Hz across the pickup and fold stages.

Dataset	Episodes	Description
<code>linique-v2-fold-pickup</code> (primary)	499	Merged full-task <i>and</i> pickup-only episodes with no force-feedback channels. This is the best training corpus and the one used for the final SmolVLA runs. Three data-cleaning passes were applied during construction: (1) any episode shorter than 2 seconds is dropped, (2) the last 0.7 seconds of every episode is trimmed to remove terminal hesitation and save-initiation artefacts, and (3) a handful of obviously-failed runs are excluded.
<code>linique-v2</code>	314	The original end-to-end dataset: full pickup-through-fold teleoperated episodes captured on the secondary-assembly rig. All subsequent datasets derive from or extend this corpus.
<code>linique-v2-pickup</code>	185	A smaller companion corpus consisting only of pickup episodes, the napkin is picked up but the fold is not executed. Added to increase pickup-state coverage after early ACT / SmolVLA runs showed the pickup-to-fold transition was the dominant failure mode.
<code>linique-v2-pickup-force</code>	189	The pickup-only dataset augmented with servo force-feedback channels. Each frame carries the FeeTech motors' reported current / load alongside the standard joint-angle observations.

Dataset	Episodes	Description
<code>linique-v2-combined-force</code>	503	<code>linique-v2-pickup-force</code> combined with <code>linique-v2-load-padded</code> into a single schema-compatible corpus — i.e., the full pickup+fold pipeline with force-feedback channels present on every frame.
<code>linique-v2-load-padded</code>	314	The original end-to-end <code>linique-v2</code> data with zero-padded force / load channels appended so that its schema matches the force-augmented pickup datasets. Produced during the schema compatibility work required for LeRobot dataset merging.
<code>eval_lerobot-inference-test</code>	0 ²	Recorded inference rollouts of trained policies running on hardware. Kept for traceability and after-the-fact analysis of the custom GPU-side inference pipeline rather than for training.

Table 22: Public Hugging Face datasets supporting the Linique project. All datasets are hosted under huggingface.co/datasets/jhimmens/. Episode counts are taken directly from each dataset’s `meta/info.json`.

The full-task dataset `linique-v2` backs the ACT and SmolVLA training runs reported in [Section — Policy Training](#). Attempts to merge the pickup-only corpus into SmolVLA training surfaced LeRobot schema incompatibilities discussed in [Section — Software Architecture](#) the force- and load-padded variants above were produced during that investigation and are preserved for reproducibility.

²The episodes in the dataset are recorded to a scratch which we purge due to space constraints; the raw video data is not pushed to the Hugging Face Hub.