Turn Operational Data Into AI Training Assets

Date: 2026-04-27
Audience: academics, SBIR/STTR researchers, mechanics, field technicians, lab operators, small business owners, professional specialists, and technical people sitting on hard-won economic data.

Start Here

If you do economically valuable work, the useful data is probably not the file cabinet. It is the workflow:

- What did the expert see?
- What options were available?
- What action did they take?
- Why was the obvious answer wrong?
- What happened later?
- How would you score a good answer?

That is what AI labs and frontier-data vendors increasingly need: task worlds, rubrics, traces, verifiers, edge cases, and outcome-linked examples. Raw data gets treated like a commodity. Scoreable workflow data gets treated like an asset.

The move is simple:

1. Inventory the operational data.
2. Map the rights before sharing anything.
3. Convert messy records into task units.
4. Attach expert rubrics and outcomes.
5. Sell paid pilots, licenses, environments, or quarterly update contracts.

Do not give away raw samples that contain the hard part. Package the judgment.

Market Signals Worth Citing

Signal	Why it matters	Source
Mercor says it builds benchmarks, evaluation environments, RL environments, and large-scale human datasets for frontier AI, and says it is used by the top 5 AI labs and 6 of the Mag 7.	The buyer category is real and explicitly tied to expert data.	Mercor Research

Signal	Why it matters	Source
Mercor's public marketplace shows `Average Pay $101/hr`, `Roles Created 186.4K`, and `Daily Payouts $2M+`, with visible expert roles such as equity research at `$120/hr`, biology PhDs at `$105/hr`, Python SWE at `$100/hr`, and physics PhDs at `$70-$90/hr`.	Domain expertise is already being priced openly.	Mercor
Business Insider reported Mercor was paying more than `$1.5M/day` to over `30,000` contractors training AI for major companies including OpenAI and Anthropic.	This is not a niche side market. It is already a large labor and data supply chain.	Business Insider
TechCrunch reported Mercor was approaching `$450M` annualized run-rate revenue by September 2025, with an outsized portion from a small set of major AI customers, including OpenAI.	A small number of frontier buyers can create huge vendor revenue.	TechCrunch
Business Insider reported a Meta-related Mercor project moving from `$21/hr` to `$16/hr`, with more than `5,000` workers at peak.	Commodity labeling is under pressure. The money is not in low-end generic work.	Business Insider
Epoch AI reports RL environment contracts are often six to seven figures per quarter, task costs often sit around `$200-$2,000/task`, and exclusivity can cost `4-5x` non-exclusive pricing.	This gives a public pricing spine for serious environment/data assets.	Epoch AI
Mechanize estimates around `$2,400` lifetime compute cost per RL task and argues labs will pay more for high-quality tasks to avoid wasting compute.	Good tasks are valuable because they protect expensive training runs.	Mechanize
SBA says SBIR/STTR data are protected from disclosure by participating agencies for not less than 20 years if properly handled.	SBIR data can be a commercial asset, but markings and rights hygiene matter.	SBIR FAQ

The Useful Vocabulary

Use these terms when talking to buyers. They turn “I have files” into “I have an asset.”

Term	What it means	What an operator can provide
Case	One real work instance	Repair order, failed test, quote, inspection, claim response
Trace	The path from input to decision to outcome	Symptoms, diagnosis, action, result, comeback/no comeback
Rubric	How an expert scores quality	What is correct, risky, incomplete, noncompliant, unsafe
Verifier	A way to check success	Test result, database state, pass/fail rule, expert adjudication
Eval	A test set for measuring model performance	Cases plus hidden answers and scoring
Task	One unit the model attempts	Initial state, goal, allowed actions, answer format, score
Environment	A task world where the model can act and be scored	Mock workflow, files, tools, rules, hidden state, scoring function
Benchmark	A standardized eval that compares models	Public or private leaderboard-ready task set
Synthetic variant	A generated task derived from real patterns	Same failure mode, changed facts, controlled difficulty
Holdout set	Tasks the buyer cannot train on	Evaluation-only cases for measuring improvement

What Is an Environment?

An environment is a place where an AI agent can try to do a task, take actions, receive feedback, and be scored.

The minimum pieces are:

1. **State:** what the agent can see.
2. **Actions:** what the agent can do.
3. **Goal:** what success means.
4. **Feedback:** how the result is scored.
5. **Hidden truth:** the answer, outcome, or constraint the agent cannot simply read.

For software, the environment might be a browser, spreadsheet, codebase, database, ERP screen, ticket queue, CAD file, lab information system, or folder of PDFs. For physical work, it is usually a digital reconstruction of the work: measurements, photos, notes, logs, constraints, choices, and outcome.

Examples:

Domain	Environment example	Scoring
Auto repair	Symptoms, scan codes, photos, prior repairs, part prices, labor constraints	Actual fix, comeback rate, unnecessary parts avoided
Machine shop	Drawing, tolerance stack, material, tooling, vendor lead times, prior quotes	Expert quote, margin, scrap risk, delivery outcome
Lab operations	Protocol, instrument logs, sample metadata, reagent lots, QC thresholds	Correct diagnosis, next experiment, replication result
Patent prosecution	Claims, prior art, office action, examiner history, client constraints	Expert review, amendment quality, allowance/prosecution outcome
Audit/compliance	Access logs, invoices, approvals, org chart, policy text, exceptions	Control finding, evidence quality, standard alignment

An environment is more valuable than a static dataset because the model has to act. It is cheaper than a full simulator because it only needs enough reality to make the task meaningful and hard to game.

The Data Asset Scorecard

Score your data before pitching it. A high score means you may have an asset, not just work-for-hire.

Factor	0 points	1 point	2 points
Real workflow	Toy or hypothetical	Based on real work but cleaned heavily	Directly tied to real work
Outcome signal	No known outcome	Partial or delayed outcome	Clear success/failure or downstream result
Expert decision	No decision captured	Decision captured but little reasoning	Decision plus reasoning and alternatives
Hidden state	Everything obvious from input	Some hidden constraints	Important hidden truth or context
Edge cases	Mostly routine	Some unusual cases	Many rare, ambiguous, failure-rich cases
Scorability	Subjective only	Expert can score manually	Rule, test, outcome, or strong rubric exists
Proprietary access	Public or easy to scrape	Hard to gather but not unique	Privileged workflow or longitudinal record
Rights clarity	Unknown ownership	Some restrictions mapped	Clear owner and license path

Factor	0 points	1 point	2 points
De-identification path	Sensitive and hard to clean	Cleanable with work	Low privacy burden or strong redaction path
Buyer relevance	Narrow curiosity	Useful to one buyer type	Useful to labs, vendors, or enterprises

Interpretation:

Score	Meaning	Best move
0–7	Mostly raw material	Sell expertise or improve collection
8–13	Pilotable dataset	Build 25-task paid pilot
14–17	Strong data product	Price as eval/environment asset
18–20	Strategic asset	Consider exclusivity, licensing, or recurring SOW

What Technical Operators Actually Have

The good stuff is usually buried in boring systems:

Operator type	High-value data	Why buyers care
Mechanic / repair shop	Diagnostic notes, codes, parts replaced, repeat visits, photos	Teaches troubleshooting under uncertainty
Machine shop	Quote history, tolerance problems, scrap causes, fixture notes	Teaches manufacturability and cost/risk reasoning
Calibration lab	Failed tests, drift patterns, corrective actions, environmental conditions	Teaches measurement reliability and root-cause analysis
University lab	Failed protocols, modified methods, replication notes, instrument quirks	Teaches experimental judgment not visible in papers
SBIR company	Test data, design iterations, drawings, software, reports	Teaches frontier technical workflows with protected rights
Patent practice	Office actions, amendments, arguments, outcomes, examiner patterns	Teaches legal-technical reasoning with outcomes
Compliance/audit shop	Evidence sets, exception findings, control failures, remediation	Teaches standard application and judgment under ambiguity

Operator type	High-value data	Why buyers care
HVAC/electrical/plumbing	Site photos, symptoms, repair path, callback data, part substitutions	Teaches field diagnostics and real-world constraints
Logistics operator	Exceptions, substitutions, carrier performance, delays, penalties	Teaches operational recovery and tradeoffs
Specialty distributor	Quote history, substitutions, lead times, customer objections	Teaches procurement and supply-chain judgment

The high-value pattern is: **input plus action plus reason plus outcome.**

Product Ladder

Do not jump straight to "sell my dataset." Package the asset at the right level.

Product	What it contains	Buyer	Typical pricing logic
Expert hours	SME reviews cases or writes answers	Mercor-like platforms, labs, vendors	$60–$300/hr worker-side for skilled experts
Case pack	De-identified real cases with outcomes	Vendors, applied AI teams	Per-case or pilot fee
Rubric pack	Scoring guides, failure taxonomy, examples	Eval teams, AI product teams	Fixed fee, often more valuable than raw cases
Private eval	Hidden task set with answers and scoring	Labs, model developers, enterprises	Per-task, license, or annual update fee
RL task set	Tasks with state, actions, scoring, variants	Labs, environment vendors	$200–$2,000/task public benchmark range from Epoch
Workflow environment	Mock app/files/tools plus scoring	Labs, agent builders, enterprises	Six to seven figures per quarter for serious SOWs
Recurring data license	Updated cases, outcomes, rubrics, drift reports	Labs and vendors	Quarterly or annual license
Exclusive domain environment	Same as above, but one buyer gets holdback/exclusivity	Frontier labs	4–5x non-exclusive pricing anchor from Epoch

Practical Shadow Price Card

These are not guarantees. They are negotiation anchors.

Asset	Plausible market band	Comment
Commodity labeling labor	$16–$45/hr worker-side	Weak leverage, high churn
Skilled evaluator / analyst	$60–$120/hr worker-side	Good for income, weaker as asset
Senior domain expert	$100–$300/hr worker-side	Stronger if tied to proprietary cases
Client-side expert labor line	Roughly $80–$500/hr	Depends on vendor layer and expertise
Expert comparison / annotation	Often $100–$400/item when packaged	Use when task requires true judgment
Serious benchmark task	Low thousands per task	Time reported Mercor spent >$500K on 200 APEX tasks, implying >$2.5K/task; see Time
RL task	$200–$2,000/task	Epoch's public range
Rare complex SWE task	Up to $20K/task	Rare, but reported by Epoch interviewees
UI / website environment clone	Around $20K/site in reported examples	Epoch citing SemiAnalysis
Complex workflow environment	Hundreds of thousands	Slack-like app clone example in Epoch
Quarterly eval/RL environment SOW	$300K–$1M+	Serious buyer, recurring improvement loop
Exclusivity	4–5x non-exclusive	Do not give this away casually

Rule of thumb: if you sell only your time, you are in the labor market. If you sell a reusable eval, rubric, verifier, or environment, you are in the asset market.

Offer Menu

These are seller-side package shapes, not guaranteed market-clearing prices. Use them to avoid inventing a quote from scratch.

Offer	Deliverable	Good for	Indicative ask
Expert discovery call	2–4 hours explaining workflow, failure modes, and data availability	Qualifying buyer seriousness	Paid hourly or $1K–$5K
Synthetic sample task	One fake-but-realistic task card with rubric	Showing structure without leaking raw data	Free only after NDA, or bundled into paid pilot
Micro pilot	10–15 tasks, rubric, rights memo, debrief	Proving signal fast	$10K–$25K

Offer	Deliverable	Good for	Indicative ask
Paid pilot	25–50 tasks, hidden answers, scoring, failure taxonomy	First serious buyer engagement	$25K–$100K
Private eval	100–500 tasks, holdout set, rubric, scorer/adjudication	Model comparison and regression testing	$100K–$300K+
Domain task set	500–2,000 tasks with variants and QA	Training/eval vendor pipeline	Per-task or quarterly license
Workflow environment	Mock files/tools/state, tasks, grader, documentation	Agent training and RL	$300K–$1M+ SOW range
Quarterly refresh	New cases, updated rubrics, drift/failure report	Ongoing model improvement	Quarterly license
Exclusive holdback	Buyer blocks competitors from same tasks/domain for a period	Frontier lab differentiation	4–5x non-exclusive anchor

Pricing formula:

```
Minimum ask =
expert time
+ data cleanup / redaction
+ rubric and scoring work
+ rights and legal friction
+ scarcity premium
+ reuse value
+ exclusivity premium, if any
```

The key negotiation move: separate **access**, **training rights**, **resale rights**, and **exclusivity**. If those are bundled, the buyer will try to pay once for four different assets.

Minimum Viable Data Product

Build this before serious buyer conversations.

Component	What to prepare	Rent it pays
Workflow one-pager	The work, who does it, why it is hard, what decisions matter	Lets buyers understand the domain fast
25 case pilot	De-identified inputs, expert action, reasoning, outcome	Proves the data exists
Rubric	How to score answers from 0–5 or pass/fail	Converts judgment into training signal

Component	What to prepare	Rent it pays
Failure taxonomy	10–20 common ways a model/human gets it wrong	Shows expert depth
Rights memo	Ownership, restrictions, sensitive fields, buyer permissions	Prevents accidental giveaway
Datasheet	Provenance, collection period, fields, exclusions, intended/prohibited uses	Makes the asset legible
Sample task card	One complete task with hidden answer and scoring	Gives the buyer something concrete
Pilot offer	Price, scope, timeline, deliverables, use limits	Forces paid engagement

Sample Task Card

Use this format for any domain.

```
Task ID:
Domain:
Workflow:
Initial state:
Available materials:
Allowed actions:
Goal:
Answer format:
Hidden ground truth:
Expert answer:
Expert reasoning:
Common wrong answers:
Scoring rubric:
Outcome evidence:
Sensitive fields removed:
Use restrictions:
```

Example: Repair Diagnostic Task

```
Task ID: AUTO-DIAG-042
Domain: independent auto repair
Workflow: intermittent no-start diagnosis
Initial state: customer complaint, scan codes, battery age, weather, prior repair history
Available materials: photos, test results, part prices, labor constraints
Allowed actions: choose diagnostic tests, recommend repair, estimate cost
Goal: identify likely root cause and avoid unnecessary part replacement
Answer format: diagnosis, confidence, tests, repair plan, cost range
Hidden ground truth: final repair, comeback/no comeback, actual labor time
Expert answer: replace corroded ground strap after voltage-drop test
```

```
Common wrong answers: replace starter, replace battery, clear code and release vehicle
Scoring rubric: root cause 40%, diagnostic path 25%, cost/risk 20%, customer explanation 15%
Outcome evidence: invoice, technician note, no comeback within 90 days
Sensitive fields removed: customer name, VIN, phone, plate
Use restrictions: evaluation and internal training only; no resale
```

Rights Map Before You Share Anything

This is the section that saves the asset from becoming a mess.

Risk	Ask	Useful reference
Ownership	Who owns the data: you, employer, customer, university, sponsor, subcontractor?	Contract, employment agreement, grant terms
SBIR/STTR	Was it generated under SBIR/STTR? Is it marked correctly?	SBIR Data Rights FAQ, SBIR Data Marking
Privacy	Does it include customers, patients, employees, students, or household data?	NIST Privacy Framework, FTC personal information guide
Health data	Is there PHI or re-identification risk?	HHS HIPAA de-identification guidance
CUI	Is it controlled federal information?	NARA CUI Program, NIST SP 800-171 Rev. 3
Export control	Does it involve defense, aerospace, nuclear, encryption, dual-use, or controlled technical data?	Counsel, EAR/ITAR review
Trade secrets	Would disclosure destroy your own moat?	NDA, clean-room packaging, synthetic variants
Human subjects	Was research subject to IRB or consent limits?	IRB protocol, consent forms
Customer contracts	Do contracts restrict secondary use or model training?	MSA, DPA, confidentiality clause

Default posture: if rights are messy, sell expertise, rubrics, synthetic tasks, or on-prem evaluation services before selling raw records.

SBIR/STTR: Do Not Sleep On This

SBIR/STTR firms should treat data as a protected commercial asset.

Useful facts:

- SBA says SBIR/STTR data are protected from disclosure by participating agencies for not less than 20 years beginning at award. Source: SBIR FAQ.

- SBA's tutorial explains that the government generally cannot disclose SBIR data outside the government during the protection period. Source: SBIR Data Rights Tutorial.
- SBA emphasizes that marking matters; failure to mark SBIR data properly can lead to loss of protection. Source: SBIR Data Marking Tutorial.

Practical implication:

- Do not upload SBIR technical data into random AI tools.
- Do not send unmarked reports as "sample data."
- Do not let a buyer turn protected test data into their model asset without a license.
- Consider selling derived evals, redacted traces, synthetic variants, or hosted scoring instead of raw SBIR data.

Deal Structures

Pick the structure that matches asset maturity.

Structure	Use when	Watch out for
Paid expert work	You have expertise but no packaged data yet	Buyer owns outputs unless contract says otherwise
Paid pilot	You can deliver 25–100 tasks or cases	Avoid broad training rights in pilot
Eval license	You have hidden scoring tasks	Keep holdout rights and update fees
Data license	You have clean reusable records	Limit resale, retention, and derivative rights
Environment build	You can reconstruct a workflow	Price setup plus recurring maintenance
Quarterly SOW	Buyer needs continuous task refresh	Include volume, QA, and update cadence
Exclusivity	Buyer wants competitors blocked	Charge a large premium and limit duration/domain
On-prem/private eval	Data cannot leave your control	Charge for access, scoring, and expert adjudication
Revenue share	Buyer commercializes your domain product	Audit rights and minimum guarantees matter

Clause Checklist

Ask counsel to translate these into actual terms.

- **Use scope:** evaluation only, training allowed, fine-tuning allowed, internal use only, no re-sale.
- **Model rights:** whether buyer may train foundation models, adapters, reward models, or classifiers.
- **Derivative works:** who owns transformed tasks, synthetic variants, rubrics, embeddings, traces, and graders.
- **Exclusivity:** none, limited holdback, buyer-exclusive, domain-exclusive, time-limited.

- **Retention:** how long buyer may keep raw data and derived data.
- **Deletion:** what must be deleted at end of term and what may remain in trained models.
- **Audit:** whether you can verify access, deletion, and downstream use.
- **Attribution:** public reference, private reference, anonymous, or prohibited.
- **Confidentiality:** trade secrets, SBIR data, customer records, sensitive operations.
- **Data security:** encryption, access controls, logging, subcontractor limits.
- **Payment:** setup fee, per-task fee, quarterly license, success fee, royalty, minimum guarantee.

Buyer Map

You do not need to sell directly to OpenAI on day one.

Buyer type	What they buy	Best first offer
Frontier AI labs	Novel tasks, evals, environments, expert traces	Private pilot eval or environment SOW
Data vendors	Expert labor, rubrics, task sets, domain SMEs	Paid expert/rubric package
RL environment startups	Workflow reconstruction, domain edge cases, verifiers	`25–100` task seed set plus consulting
Enterprise AI teams	Internal evals, guardrails, workflow traces	Private eval and failure taxonomy
Vertical SaaS vendors	Domain-specific agent evals and training data	Benchmark for their customer workflow
Insurers / auditors / regulators	Risk taxonomies, compliance evals, evidence review	Scored case library
Universities / labs	Research datasets, reproducible benchmarks	Dataset card plus benchmark tasks

Routes to market:

- Start with a paid expert project to learn buyer vocabulary.
- Convert expert work into reusable rubrics.
- Convert rubrics into private evals.
- Convert private evals into recurring environment updates.
- Charge exclusivity only when the buyer pays for it.

Buyer Red Flags

Do not be naive. These requests often extract the value without paying for it.

Red flag	Better response
“Send us a few representative raw files.”	Send a synthetic or redacted task card under NDA, or offer a paid pilot.
“We just need to evaluate fit.”	Define a paid evaluation package with limited use rights.

Red flag	Better response
“Our standard terms own all work product.”	Carve out pre-existing data, rubrics, methods, and domain taxonomy.
“We need exclusivity for free.”	Price exclusivity separately and limit duration/domain.
“We cannot say whether it will train models.”	No raw data until use scope is explicit.
“We will anonymize it later.”	You anonymize or host; do not rely on buyer cleanup.
“This is just metadata.”	Metadata can still reveal customers, trade secrets, or workflow advantage.
“We need perpetual rights.”	Charge accordingly or offer renewable term rights.

What To Build This Week

Day 1: Data Inventory

Create a spreadsheet with:

- Source system.
- Data type.
- Date range.
- Number of cases.
- Outcome field.
- Expert decision field.
- Sensitive fields.
- Owner.
- Contract restrictions.
- Example edge cases.

Day 2: Rights Triage

Mark each data source:

- `Green`: owned, low sensitivity, clean to license.
- `Yellow`: usable after redaction, consent review, or contract review.
- `Red`: do not share raw; use synthetic, hosted, or expert-only path.

Day 3: Task Extraction

Pick `25` cases. For each:

- Input.
- Decision.
- Reasoning.
- Outcome.
- Common wrong answer.
- Scoring rule.

Day 4: Rubric and Failure Taxonomy

Write:

- 5 score dimensions.
- 10 common failure modes.
- 5 cases where naive models will look plausible but fail.

Day 5: Buyer One-Pager

Make one page:

- Domain.
- Why public data is insufficient.
- What workflow is captured.
- What outcomes exist.
- What a pilot includes.
- What use rights are offered.
- Price range and timeline.

One-Page Buyer Pitch Template

```
Title:
Private eval / RL task set for [workflow]

Problem:
Current AI agents fail at [domain] because public data does not capture [hidden constraints, ec

Asset:
We have [N] real cases from [workflow], covering [date range / equipment / customer class / tec

Signal:
Each case includes [inputs], [expert action], [reasoning], and [outcome].

Pilot:
We can deliver [25-50] de-identified tasks with rubrics, hidden answers, and scoring in [timeli

Rights:
Pilot is for evaluation only. Training, model improvement, resale, or exclusivity require separ

Expansion:
After validation, we can provide [quarterly refresh / environment build / synthetic variants /

Price:
Pilot: [$X]. Production: per-task, quarterly license, or environment SOW.
```

Email Template

```
Subject: Private eval data for [technical workflow]

Hi [Name],

We operate in [domain] and have outcome-linked workflow data that public web data does not capt

The asset is not raw documents. It is scoreable task data: inputs, expert decisions, reasoning

This may be useful for evaluating or training agents on [capability]. We are not sharing raw da

Open to a short call next week?
```

Resource Map

Market Proof

- Mercor Research for frontier-data buyer language.
- Mercor for public expert rates and role categories.
- Mercor Enterprise AI for "agentic workflow data."
- Epoch AI RL environments FAQ for contract sizes, task prices, and exclusivity premiums.
- Mechanize RL task economics for why task quality protects compute spend.
- TechCrunch on Mercor's run rate for buyer-market scale.
- Business Insider on Mercor daily payouts for human-expert labor scale.
- Business Insider on low-end Mercor project rates for commodity-rate pressure.

Evals and Environment Tooling

- OpenAI Evals for model evaluation structure.
- OpenAI Agent Evals for trace grading and workflow-level agent evaluation.
- Inspect AI for evals with datasets, solvers, scorers, tools, agents, and sandboxes.
- METR Task Standard for portable agent task structure.
- SWE-bench for real work issues plus reproducible scoring.
- SWE-bench GitHub for benchmark packaging.
- OSWorld for real computer-use environments.
- OSWorld GitHub for task setup and evaluation scripts.
- WorkArena for enterprise workflow tasks.
- WorkArena GitHub for browser-based knowledge-work task packaging.
- tau-bench for tool-agent-user interaction with state-based grading.
- tau-bench GitHub for policies, APIs, user simulation, and database-state scoring.
- BrowserGym / Foundry for browser-agent simulation and telemetry.

Dataset Packaging

- Datasheets for Datasets for provenance, composition, collection, intended use, and limitations.
- Hugging Face Dataset Cards for a buyer-recognizable dataset documentation template.

- FAIR Principles for research-data hygiene.
- DataCite Metadata Schema for citable dataset metadata.
- MLCommons Croissant for ML-ready dataset metadata.
- Croissant 1.1 Specification for machine-readable dataset packaging.

Rights and Safety

- SBIR Data Rights FAQ for SBIR/STTR ownership and protection basics.
- SBIR Data Rights Tutorial for the 20-year protection period.
- SBIR Data Marking Tutorial for marking requirements.
- NIST Privacy Framework for privacy risk management.
- HHS HIPAA De-identification Guidance for health-data de-identification.
- FTC Protecting Personal Information for small-business data security.
- NARA CUI Program for Controlled Unclassified Information.
- NIST SP 800-171 Rev. 3 for protecting CUI in nonfederal systems.
- NIST AI Risk Management Framework for AI risk governance.

What To Copy From Existing Environments

Do not copy the domains. Copy the structure.

Resource	What to copy
SWE-bench	Real work item plus hidden success test
OSWorld	Initial state, available actions, end-state checker
WorkArena	Enterprise UI tasks and compositional workflows
tau-bench	Policy document plus tools plus final database-state grading
METR Task Standard	Portable task format
Inspect AI	Datasets, solvers, scorers, tools, sandboxes
Datasheets for Datasets	Provenance, limitations, intended use
Hugging Face Dataset Cards	Clean documentation template

Final Operating Principle

Operational data is valuable when it captures a decision under constraints and proves what happened afterward.

The highest-value asset is not a database dump. It is a rights-clean workflow package with cases, outcomes, rubrics, hidden answers, failure modes, and scoring.

You made the data. You understand the domain. Package it like an asset, price it like an asset, and do not give away the learning signal for free.

Work With Ashiba

Ashiba is looking for researchers, operators, SBIR/STTR teams, skilled practitioners, labs, shops, and technical businesses with real workflow data that could become AI training, evaluation, or environment assets.

The best fit is not "I have a pile of files."

The best fit is:

- You have repeated technical work.
- You know what good and bad answers look like.
- You have outcomes, failures, callbacks, test results, inspection results, quote history, repairs, experiments, or decisions under constraints.
- Public web data does not capture your edge cases.
- A buyer would waste time or compute without your domain signal.
- You care about not giving away the raw learning signal for free.

Ashiba can help turn that into:

- A 25–50 task paid pilot.
- A private eval.
- A workflow environment.
- A rubric and verifier package.
- A quarterly data/update license.
- A cleanroom-cleared procedural dataset.
- A rights map and buyer-facing data room.
- A pricing and exclusivity strategy.

We are also developing **LADDER**, a standard data contract and clearance stack for operational and procedural AI data. LADDER is meant to solve the part that kills these deals: unclear rights, messy provenance, trade secrets, privacy issues, buyer overreach, raw-sample leakage, and underpriced exclusivity.

The goal is simple:

> Turn hard-won operational knowledge into licensable AI data assets without handing buyers the raw moat for free.

The simplest way to understand LADDER:

```
workflow data
-> rights map
-> cleanroom review
-> reserved know-how
-> clean skill episodes
-> LADDER Passport
-> buyer license
```

It separates access, evaluation rights, training rights, resale, retention, updates, and exclusivity so buyers cannot quietly collapse seven assets into one vague pilot.

This can be real money.

Not because every dataset is valuable. Most are not. But if you have the right workflow, outcome signal, scarcity, rights position, and buyer relevance, you should not be selling it like spare consulting time. You should be packaging it like an asset.

If you think you have one, Ashiba wants to talk.

Bring:

- One workflow.
- Ten representative cases.
- The outcome or scoring signal.
- Any known rights restrictions.
- The buyer type you think would care.

We will help answer the only question that matters:

> Is this just operational exhaust, or is it an AI data asset?