We turn the real world into a training environment for AI.

We give away the software for free.
We make money selling the data to AI labs.
Businesses that put AI to work grow revenue and profit 100% to 300%.

ALEX DEBELOV · VLAD CHISTIAKOV · KHODOR SALAM

The pitch, in one deck.

TL;DR

The whole pitch, in six lines.

The problem	Frontier labs are out of good data. Text is done; sweatshop annotation is over.
The bottleneck	The next unlock is expert judgment AND outcome — real operators, real stakes.
Our offer	Free agents to real businesses. Owners approve; the world scores. Labs pay for the record.
Where we start	Go X: 100 tasks, 41 businesses ranked. We already run 12 alone and propose on 33 more.
What we sell	Environment suite, launch tasks, live-fire evals, OpsBench free.
The bet	Nobody at scale, with consent, packaged for labs. Whoever runs the pipes wins.

SECTION 01

THE FRAME

Sweatshop data is over. Real operators are the bottleneck.

Mechanize, July 2025: 'Full-time contributors, not contractors. Deep expertise. Their tacit knowledge, skills, and experience are now the bottleneck to further AI progress.'

01

THE FRAME

They named the bottleneck. We are the pipe.

> *"Sweatshop data is over. Good RL environments are the bottleneck. The tacit knowledge, skills, and experience of real operators are now the bottleneck to further AI progress."*
>
> — Mechanize, July 2025 · Ege Beşiroğlu, Tamay Beşiroğlu, Matthew Barnett

Sources: mechanize.work/blog/sweatshop-data-is-over (July 10, 2025).

THE THREE SHIFTS

The data pipeline reshaped itself in twelve months.

SOFTWARE, NOT DATASETS

Labs buy live environments and evaluations — code that can be run, not files of text.

Static labels used to be enough. Now the model needs to try, fail, and be scored.

FULL-TIME EXPERTS, NOT CONTRACTORS

The scarce input is deep operator expertise, not marginal labor.

Legal, medical, biotech, ops. Judgment AND outcome — not throughput.

MEASURED IN THE WORLD, NOT THE LAB

Ground truth comes from the field: money moved, jobs delivered, customers won.

Live-fire evals replace static benchmarks. The world is now the grader.

Sources: Mechanize (2025); Anthropic Project Vend (2025); Kalshi + Robinhood (Nov 2025) — real-world payoffs standardizing outcome-labeled data.

SECTION 02

THE DEAL

Three parties, one loop.

Business gets free agents. Midas runs them and owns the record. Lab pays for the record and evaluates on it.

02

THE DEAL

Everyone gets one thing, and it fits together.

BUSINESS

Free agents that run the daily work.

- Free software from day one
- Owner approves; agents propose
- New revenue lines that were too small to run before
- Data license: paid quarterly

MIDAS / GO X

The instrumentation layer and the record.

- Software vendor to operators
- Data vendor to labs
- Owns 100-task ledger + provenance
- Two revenue lines

AI LAB

Ground-truth data + live-fire evals.

- State / action / override / outcome / reward
- Consent + provenance built in
- Live evals against reality
- Multi-vertical coverage

One transaction, three winners. The loop runs itself.

SECTION 03

WHY NOW

$10M in Spirit's inbox. 10 of 100 automated. $0 spent.

03

WHY NOW

Every operator is Spirit Airlines waiting to happen.

$10M

In Spirit Airlines' customer-service inbox this year.

Revenue moving through a channel nobody staffs.

10 of 100

Ops tasks a good agent can already handle.

The other 90 still need a human — for now.

$0

What labs spend to buy this data today.

Because nobody built the pipe.

Frontier labs are chasing expert operators. We already run them.

SECTION 04

HOW IT WORKS

One loop. Five steps. Ground truth every time.

04

HOW IT WORKS

The loop that makes everyone money.

01

PLUG IN

OpsBench connects to the business's systems: CRM, dispatch, inbox, phones.

02

AGENTS PROPOSE

Every task on the 100-list gets an agent action, drafted in seconds.

03

HUMANS APPROVE

Owner approves, edits, or overrides — with a reason.

04

WORLD RESPONDS

Money moves. Jobs get done. Customers pay or complain.

05

LABS TRAIN

Every decision becomes state / action / override / outcome / reward.

The lab gets the training set. The business gets free labor. Midas gets both.

SECTION 05

WHAT THE BUSINESS GETS

Free agents. You approve. You get paid.

05

WHAT THE BUSINESS GETS

Six things Go X could not do a year ago.

QUOTES

Agents draft every quote from the enquiry.

Owner approves in one click.

DISPATCH

Agents match jobs to crews based on constraints.

Owner sees the reason, ships it.

SUPPORT

Every inbound answered in under two minutes.

Human takes the hard ones.

PRICING

Prices set daily by demand, not by gut.

Owner overrides if needed.

HIRING

Every candidate scored against the last 500.

Owner interviews the top three.

REPORTS

Weekly ops review, generated by the agent.

Owner signs off in an hour.

TODAY AT GO X

Agents run about 12 of the 100 tasks alone and propose on about 33 more. The other 55 still need a human.

COVERAGE TODAY

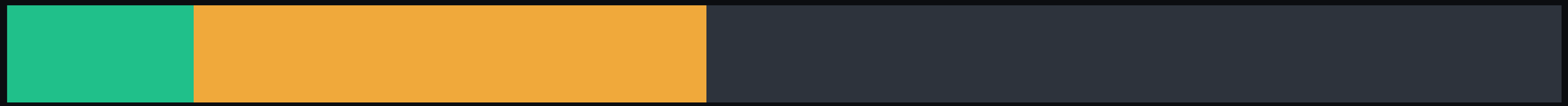
Where the frontier actually is.

This is what a live task ledger looks like. It moves every quarter. Labs will pay to watch it move.

AGENT ALONE
12

AGENT PROPOSES, HUMAN APPROVES
33

HUMAN STILL
55



100 TASKS

The ledger every operator must run

Sources: Go X internal task audit, Q3 2026.

SECTION 06

WHAT THE LAB GETS

Every decision, on the record.

06

WHAT THE LAB GETS

The decision record. Five fields. Every task. Every day.

STATE	What the world looked like: inputs, prices, calendar, weather, inventory.
AGENT ACTION	What the model proposed and why. Draft, price, script, dispatch.
HUMAN OVERRIDE	What the operator changed. Free-text reason attached.
OUTCOME	What actually happened. Money moved, job done, customer response.
REWARD	How the operator or the market scored it. Numeric or graded.

The state-action-outcome-reward tuple. RLHF made real.

SECTION 07

THE 100-TASK LEDGER

Ten areas. Every operator runs them.

07

THE 100-TASK LEDGER

The list, so a lab can point at it.

01 SUPPLY
Sign, price, activate

02 DEMAND
Ads, funnels, matching

03 QUOTING
Draft, price, follow-up

04 DISPATCH
Assign, route, exceptions

05 SERVICE
Physical delivery — humans still
PHYSICAL / HUMAN

06 CUSTOMERS
Support, retention, upsell

07 MONEY
Invoice, collect, reconcile

08 PEOPLE
Hire, onboard, review

09 COMPLIANCE
Regulator, license, tax

10 GROWTH
New geo, new SKU, new channel

Nine areas are software-shaped. One area is not — and that's a moat.

SECTION 08

THE MAP

41 businesses on one 2×2.

08

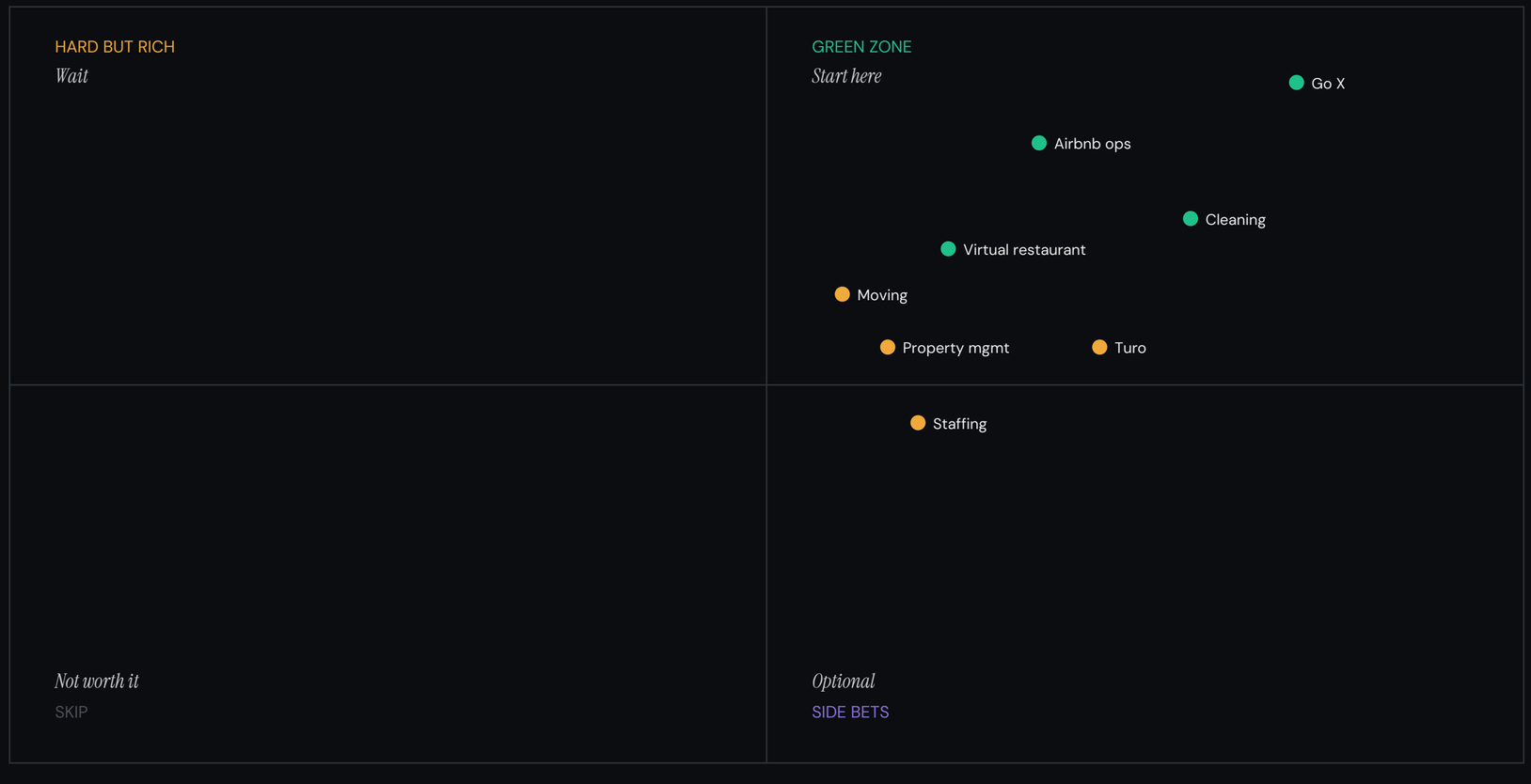
THE MAP

Easy to capture × Valuable to labs.

41

businesses ranked

We plotted every service business we could plausibly run. The green quadrant is the only zone with software leverage AND lab-grade decisions.



Sources: Midas internal ranking, Sept 2026. Full list on the next slide.

SECTION 09

SIX DECISIONS

Six patterns. Every business runs them.

09

SIX DECISIONS

What labs actually want to score.

NEGOTIATION

Price a job. Hold a customer. Close a deal.

Money on the line, judgment matters.

ALLOCATION

Who gets what. Crew to job. Dollar to channel.

Constrained optimization with soft costs.

LAUNCH

New product, new city, new SKU.

Multi-step, uncertain, expensive to reverse.

REGULATION

License, complaint, audit.

Judgment AND outcome — one mistake compounds.

PEOPLE

Hire, fire, promote, discipline.

High cost, few labels, deep expertise.

RECOVERY

Refund. Redo. Rescue. Save the account.

The hardest tasks. Rarest data.

Every operator runs all six. Every lab wants all six. Nobody's built the pipe.

SECTION 10

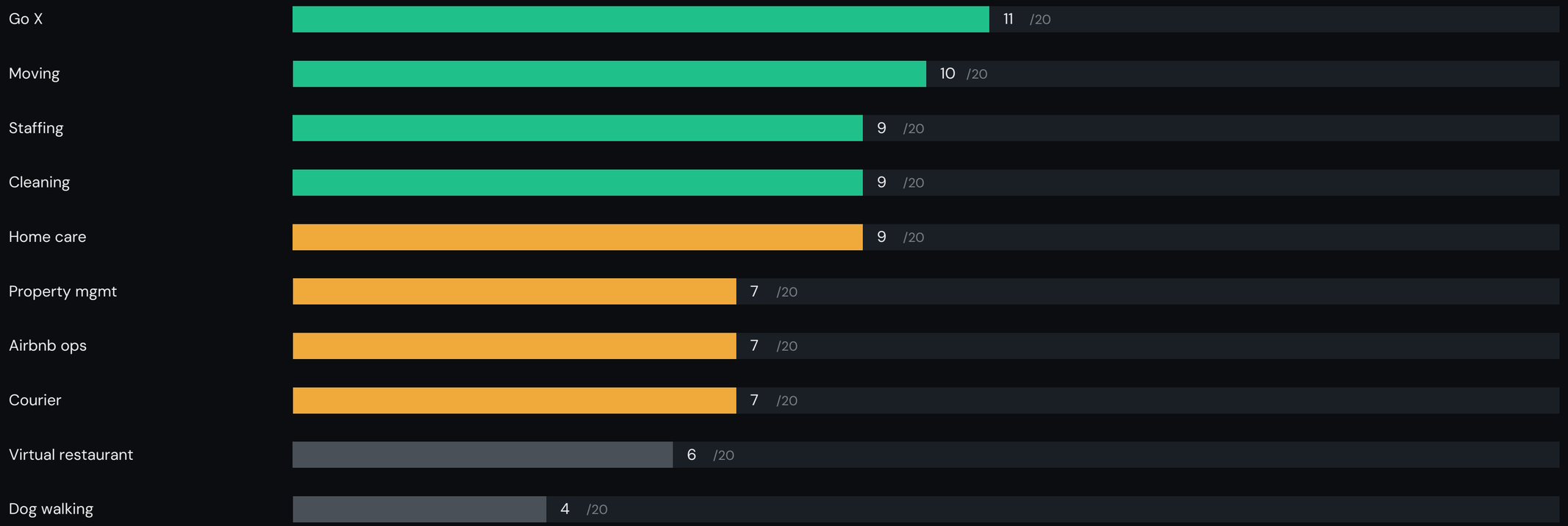
RANKING

Two scores per business. Ten questions each.

10

RANKING

Top 10 of 41 businesses, scored 0–20.



SCORE (0–20) · 10 questions on capture-ability + 10 questions on lab value

SECTION 12

FIRST EIGHT

The first eight. In order.

12

FIRST EIGHT

The order matters. Own, integrate, license.

OWN

Go X

The flagship. Ours end-to-end.

OWN

Cleaning

High volume, tight loop, clean labels.

OWN

Airbnb ops

Millions of hosts, one workflow.

OWN

Virtual restaurant

Pure software, per-order margin.

INTEGRATE

Moving

Existing operators, we plug in.

INTEGRATE

Turo

Consumer marketplace + host tools.

INTEGRATE

Property mgmt

PMS integrations exist. Data flows.

INTEGRATE

Staffing

Recruiter tools + candidate matching.

Wave one is where we prove the loop. Then license the ledger to the rest.

SECTION 13

ROLLOUT

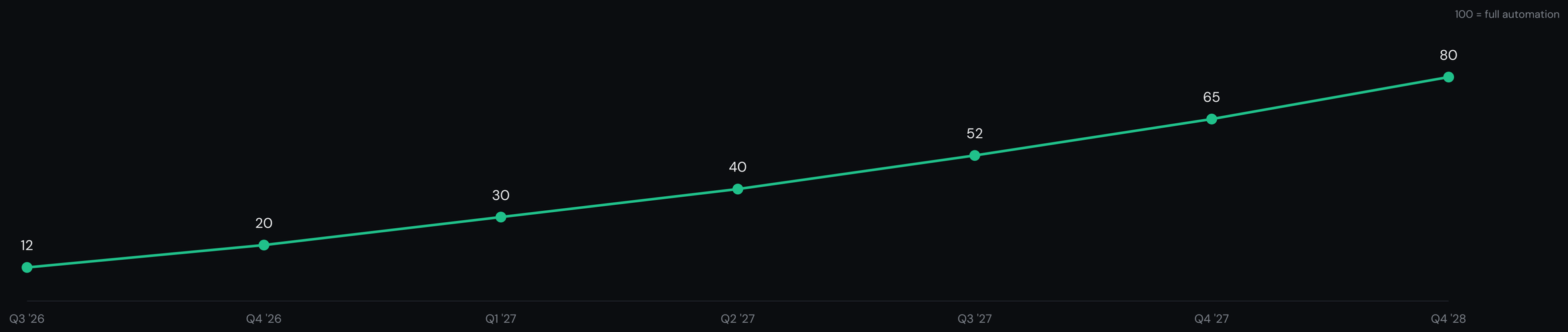
Three waves. Sequenced. Compounding.

13

ROLLOUT

Coverage grows every quarter. That's the thing labs pay for.

AGENT-ALONE % OF THE 100 TASKS



WAVE 1
Own 0–3

WAVE 2
Integrate 4–9

WAVE 3
License 10–18

Sources: Midas internal roadmap, Sept 2026. Trajectory assumes labs co-fund coverage build.

SECTION 14

SELL

Two pitches. Same loop.

14

SELL TO THE OWNER

"Free agents. You approve. You get paid."

Free software	Zero upfront. Zero SaaS. Agents run 12 of 100 tasks day one.
You approve	Every action, one click. Override anytime. Reason attached.
You keep the upside	Your revenue is yours. Your customers are yours. We touch nothing.
You get paid	Data license, quarterly. We split what labs pay for your ledger.

The pitch fits on one napkin. That's the point.

SELL TO THE LAB

Four SKUs. Priced against alternatives.

ENVIRONMENT SUITE

Live operator loops the lab can train on.

$1M – $3M / lab / year

LAUNCH TASKS

Specific ops tasks generated on demand.

$5K – $20K per task

LIVE-FIRE EVAL

Test the model against real Go X operations.

$250K – $1M per eval

OPSBENCH

Open coverage dashboard, updated weekly.

Free — the wedge

The environment suite pays the electric bill. Everything else is upside.

SECTION 15

WHY WE WIN

Four things. Nobody has all four.

15

WHY WE WIN

The moat is the loop, not the model.

01

SUPPLY NOBODY CAN SIGN

Real operators, on the tool, with consent. Not survey panels. Not annotators.

02

TASKS GENERATED, NOT WRITTEN

The 100-task ledger is discovered from live operations, not hand-authored by ex-consultants.

03

FRONTIER MEASURED DAILY

OpsBench moves every week. Labs see coverage go from 12 → 20 → 40 in public.

04

PROVENANCE BUILT IN

Every row: who, when, what override, what outcome. Legally clean. Audit-ready.

Nobody at scale, with consent, packaged for labs. That's the whole thesis.

SECTION 16

THE FIELD

Everyone's building pieces. We're building the pipe.

16

THE FIELD

Everyone is playing one square. We're playing the board.

PLAYER	WHAT THEY DO	WHAT THEY CAN'T
Thrive / Andon	Ops copilots for restaurants	One vertical. No lab pipe.
Mercor / Scale / Surge	Annotation, expert contractors	Judgment AND outcome — missing.
HUD / Prime Intellect	RL environments as software	No live operators. No consent.
Wonder Create	Restaurant ops brand	Vertical player, not data vendor.
Uber / DoorDash	Real operators at scale	Not selling the data. Won't ever.
Midas	Ops copilots + data pipe + provenance	That's the play.

Sources: Company websites, TechCrunch coverage, Sept 2026.

SECTION 17

THE MODEL

Two revenue lines. Neither cannibalizes the other.

17

THE MODEL

Line 1 pays the bills. Line 2 is the fortune.

LINE 1

LABS PAY FOR THE DATA

Environment suite subscriptions. Task generation. Live-fire evals. Sold to 5-10 frontier labs.

LINE 2

OPS COMPOUND

Take rate on business the agents generate. Volume × margin. Every business runs 24/7 with 12+ tasks automated.

	REVENUE	WHAT IT LOOKS LIKE
YEAR 1	$1M – $3M	First 2-3 lab contracts, 4 owned businesses
YEAR 2	$10M – $30M	5-8 lab contracts, 8 businesses live
YEAR 3	$100M+	Full lab roster, 18-business coverage, ops line kicks in

Sources: Midas financial plan, Sept 2026. Anchored to Mercor's revenue ramp — but we own supply.

SECTION 18

RISKS

Five risks. How we handle them.

18

RISKS

Five things that break the plan.

CONSENT	Operators say no when they hear 'labs'.	*Fixed price per row. Contract templates. Data trust.*
LABS GO IN-HOUSE	Frontier lab builds its own operator pipeline.	*We're already there. Consent takes years, not weeks.*
HUD / MERCOR PHYSICAL	One competitor sits on physical tasks too.	*Multi-vertical + owned businesses = defensible flywheel.*
OWNERS CHURN	Operator finds a better co-pilot elsewhere.	*Free software + rev share + integrated dispatch = high switching cost.*
SMALL SAMPLES	Not enough decisions per task per week.	*Coverage strategy: 8 businesses × 100 tasks = 800 loops running daily.*

Nothing is fatal. All are known. All have countermoves shipping now.

SECTION 19

NEXT 90 DAYS

Instrument. Publish. Sell.

19

NEXT 90 DAYS

Prove the loop on four businesses.

DAYS 1–30

INSTRUMENT

- Score the Go X ledger with the ops team
- Ship the approval app: propose, approve, override with reason
- Launch cleaning + virtual restaurant with agents from day one
- Fleet Ops Gym v0 on HUD

DAYS 31–60

PUBLISH

- Airbnb co-hosting live with three hosts
- OpsBench v0: coverage per vertical, public
- Sample export and a Colab a researcher can open
- First ten Go X franchisees and venues on the platform

DAYS 61–90

SELL

- Mercor: the live version of the archive you bid on
- Anthropic: the team that ran Project Vend
- HUD vendor listing
- One paid pilot, $250K to $500K

SECTION 20

CLOSE

The one line to remember.

20

"Give the software away.
Sell what the businesses teach it."

— Alex Debelov, CEO Midas / Go X

alex@midas.com · midaspitch.pplx.app