Here’s a straight-shooting senior-dev review of the repo + your sample leads, with hard fixes so this thing stops hallucinating and starts feeding you only qualified targets.

**What I see (repo & behavior)**

**Repo summary.** The README claims a full “production-ready” pipeline with strict target profile enforcement, validators, classification, scoring, exports, etc. (SQLite, Pydantic, aiohttp, orchestrator, scheduler) ([GitHub](https://github.com/markfilliph/business-acquisition-mvp)). In practice, your current lead list shows multiple misses on category filters, geography hygiene, and evidence capture—exactly the failure modes you called out (making things up; not writing to the single source of truth; weak cross-validation).

**Symptoms in the leads you shared (quick spot-check):**

* **Retail / convenience (should be excluded):** Eastgate Variety (1505 Main St E) is a variety store/convenience, not a target professional service. ([MapQuest](https://www.mapquest.com/ca/ontario/eastgate-variety-gifts-455922240?utm_source=chatgpt.com))
* **Gas/retail (should be excluded):** Pioneer at 603 King St E is a gas station & c-store. ([MapQuest](https://www.mapquest.com/ca/ontario/pioneer-359355342?utm_source=chatgpt.com))
* **Automotive retail/aftermarket (likely excluded):** Performance Improvements is a retail speed-shop. ([Yelp](https://www.yelp.ca/biz/performance-improvements-hamilton?utm_source=chatgpt.com))
* **Pure retail (should be excluded):** Worldwide Mattress Outlet is mattress retail. ([MapQuest](https://www.mapquest.com/ca/ontario/worldwide-mattress-outlet-456037655?utm_source=chatgpt.com))
* **Large auto dealership (retail):** Mountain Hyundai is a car dealership. ([Mountain Hyundai](https://www.mountainhyundai.com/?utm_source=chatgpt.com))
* **Funeral home (borderline/professional service but typically owner-operator model; validate per your thesis):** Truscott, Brown & Dwyer. ([MapQuest](https://www.mapquest.com/ca/ontario/truscott-brown-dwyer-funeral-281716436?utm_source=chatgpt.com))
* **Signage (likely in-scope “printing / professional services”):** Dee Signs has a valid site and address (but note **Burlington**, not Hamilton). ([Dee Signs](https://deesigns.ca/?utm_source=chatgpt.com))
* **Engineering (likely in-scope professional services):** Landtek Limited (geotechnical/environmental; valid site). ([landtek.ca](https://landtek.ca/?utm_source=chatgpt.com))
* **Lavender/apiary (retail/agr-tourism; excluded):** Weir’s Lane Lavender moved from farm address; retail shop now on King St W, Dundas. ([Weir's Lane Lavender & Apiary](https://www.weirslanelavender.ca/?srsltid=AfmBOorV8pbECLCSbMn-0qXcXAoFGrJTN5I2MVzemZLyhgbHxrTPSWAD&utm_source=chatgpt.com))
* **Edward Jones / State Farm (franchise/agent offices; typically owner-operator with small SDE and high owner reliance; likely exclude):** (EDJ Hamilton listing noise; State Farm similar). ([Shoogle](https://ca.shoogle.net/ontario/hamilton/edward-jones-investments/?utm_source=chatgpt.com))
* **Pardon Applications of Canada:** address line shows a **postal code error** in your lead (L4P vs **L8P**); their own site/BBB show 21 King St W, 5th Fl, **L8P 4W7** in Hamilton. That’s a data-quality red flag. ([pardonapplications.ca](https://www.pardonapplications.ca/contact/?utm_source=chatgpt.com))

Bottom line: the pipeline is not reliably enforcing **industry/category + geography + evidence** gates before anything hits your “qualified” list.

**Why it’s happening (engineering diagnosis)**

1. **Discovery → Enrichment happens before persistence.** You’re likely enriching transient objects and only persisting “keepers”; this hides where/why items failed and invites silent fall-through to exports. You need **idempotent upserts of every raw lead**, plus an **evidence ledger**.
2. **Classifier & exclusion rules are too weak.** Names like “Eastgate Variety”, “Pioneer”, “Mattress Outlet” should be trivially blocked by a **category blacklist** (convenience, gas, mattress, auto dealer, nail salon, hair salon, etc.), ideally grounded by external place types.
3. **LLM enrichment is allowed to “guess.”** If prompts don’t strictly disallow fabrication (and you don’t require citations), models will confidently invent revenue, staff counts, and history.
4. **No multi-source corroboration.** You’re accepting a single directory hit (or an LLM assertion) as truth. That’s how the PAC postal code drifted to **L4P** (Keswick area) instead of **L8P** (Hamilton core). ([pardonapplications.ca](https://www.pardonapplications.ca/contact/?utm_source=chatgpt.com))
5. **Radius/geofence not enforced with coordinates.** Dee Signs is **Burlington**—close, but if your rule is Hamilton/Dundas/Ancaster etc., you need a **distance check** from a defined centroid, not just “contains ‘Hamilton’ in the address.” ([Dee Signs](https://deesigns.ca/contact-us/?utm_source=chatgpt.com))

**Hard fixes (code & data)—do these next**

**1) Data model: make “evidence” first-class and persist EVERYTHING**

Add normalized tables (or keep SQLite for now, same idea):

* **businesses** (canonical record; unique on normalized\_name + street + city + postal).
* **observations** (one row per source check; fields: business\_id, source\_url, field, value, confidence, observed\_at).
* **validations** (gate results per rule with pass/fail, reason).
* **exclusions** (rule\_id, reason, evidence\_ids[]).
* **exports** (snapshot of what was sent to CSV/CRM, with the validation version).

Make the **discovery stage write raw candidates immediately** (status = NEW), then enrichment/validation update status to QUALIFIED or EXCLUDED with explicit reasons. This single change stops “ghost leads” that never land in your source of truth.

**2) Deterministic category gating (no LLM here)**

Create a **two-layer filter** before enrichment:

* **Places/Directory types** (Google Places/Yelp/OpenStreetMap tags) with a **positive whitelist**: {manufacturing, printing, sign shop, equipment rental, engineering/consulting, commercial services, B2B wholesale} and a **negative blacklist**: {convenience\_store, gas\_station, mattress\_store, car\_dealer, barber, salon, nail\_salon, vape, liquor, funeral\_home (if you choose), pawn, payday, cannabis, restaurant, retail\_general}.  
  (Your own README says “EXCLUDED: skilled trades,” add those too: {welding, HVAC, roofing, plumbing, electric, machining, construction}) ([GitHub](https://github.com/markfilliph/business-acquisition-mvp))
* **Name & URL heuristics**: regexes that instantly reject obvious retail (“variety”, “mart”, “mattress”, “hyundai”, “nails”, “salon”, “barbershop”, “gas”, “convenience”, “tobacco”, “pawn”, “boutique”, “spa”). These catch cases where directories mis-type.

**Only if a candidate passes both** do you proceed to enrichment.

**3) Multi-source corroboration (2-of-N rule)**

Define a **corroboration policy** per field:

* **Address & phone**: require **2 matching sources** (e.g., website + Yelp/MapQuest/BBB).
* **Website**: require HTTP **200**, non-parked content, and **Wayback** snapshot (age signal).
* **Category**: one authoritative type (Google Places/Yelp) **plus** site content keyword match.
* **Postal code**: must match in two sources (PAC shows how this breaks). ([pardonapplications.ca](https://www.pardonapplications.ca/contact/?utm_source=chatgpt.com))

If a field fails corroboration, **leave it null** and block export (do not “estimate”).

**4) Geography: coordinate every record**

Geocode each candidate (Nominatim or Google). Store lat,lng and **compute distance** from your **Ancaster/Hamilton centroid**. Block anything outside your **km radius** (you specified 15–20km in past work) from being “Qualified” without manual override.

**5) Revenue & staff: ranges with confidence, never single-point “guesses”**

* Pull **company size bands** from LinkedIn/Facebook/BBB if available; parse **Careers/About** pages for signals.
* Map category → **Ontario/Canada industry benchmarks** (StatsCan, industry reports) to generate **range estimates** (e.g., P50/P75) with a **confidence score** (low if staff unknown).
* **Reject** any record missing both (a) staff signal and (b) category benchmark.

**6) LLM usage: extraction only, with citations**

Where you still want LLMs:

* Use **strict JSON schemas** (Pydantic-validated) for **extraction only** (e.g., “extract founding year if present on the About page; else return null”).
* **Disallow fabrication** in the prompt (“If not present verbatim, return null”).
* Require the LLM to include a source\_url for every extracted field.
* **Never** let the LLM compute revenue/SDE or “decide category” on its own.

**7) Tests you’re missing (write these now)**

* **Classifier regression tests:** feed in “Eastgate Variety”, “Pioneer Gas”, “Mountain Hyundai”, “Worldwide Mattress Outlet” and assert **EXCLUDED** with the right rule id. (Lock these as golden tests with snapshots.) ([MapQuest](https://www.mapquest.com/ca/ontario/eastgate-variety-gifts-455922240?utm_source=chatgpt.com))
* **Geofence tests:** “Dee Signs, Burlington” should fail Hamilton-only profile unless you accept “within X km” (then pass with distance\_km < threshold). ([Dee Signs](https://deesigns.ca/contact-us/?utm_source=chatgpt.com))
* **Corroboration tests:** PAC postal fix—ensure postal mismatches fail export until reconciled. ([pardonapplications.ca](https://www.pardonapplications.ca/contact/?utm_source=chatgpt.com))
* **No-guess policy tests:** when website lacks founding year, the extractor must return null, not 2012 “because vibes”.

**8) Orchestrator & export “gates”**

In agents/orchestrator.py (per README), add a **single function** that decides if a record is eligible to export:

def is\_exportable(b):

return all([

b.status == "QUALIFIED",

b.validation.pass\_rate >= 1.0, # all hard gates passed

b.category in TARGET\_WHITELIST,

b.distance\_km <= RADIUS\_KM,

b.website.ok and b.website.age\_years >= 3,

b.corroboration.score >= 2, # 2+ independent sources

b.revenue.estimate\_confidence >= 0.6

])

Make your export script **only** pull rows where is\_exportable was True at a recorded validation version. This prevents “oops” records from slipping out.

**9) Logging: store *why* something was excluded**

Every exclusion should write: rule\_id, reason, evidence\_ids[]. This is gold for audits and rapid rule tuning.

**10) CLI hygiene (reflects your “main source of data” problem)**

* generate should **always** push into businesses + observations before any filtering.
* scripts/test\_validation.py should output a **confusion matrix** of passes/fails by rule to a report in /output/.

**Quick re-take on your sample list (evidence-based)**

| **Lead** | **Verdict** | **Why (rule)** | **Evidence** |
| --- | --- | --- | --- |
| Eastgate Variety | **EXCLUDE** | Retail/Convenience | Map/Yelp list as variety store; tobacco sale violation logged → not target. ([MapQuest](https://www.mapquest.com/ca/ontario/eastgate-variety-gifts-455922240?utm_source=chatgpt.com)) |
| Pioneer (603 King E) | **EXCLUDE** | Gas/C-store retail | Multiple sources show gas station chain. ([MapQuest](https://www.mapquest.com/ca/ontario/pioneer-359355342?utm_source=chatgpt.com)) |
| Performance Improvements | **EXCLUDE** | Retail/aftermarket | Yelp/MapQuest show retail speed-shop. ([Yelp](https://www.yelp.ca/biz/performance-improvements-hamilton?utm_source=chatgpt.com)) |
| Worldwide Mattress Outlet | **EXCLUDE** | Retail | Storefront mattress retail. ([MapQuest](https://www.mapquest.com/ca/ontario/worldwide-mattress-outlet-456037655?utm_source=chatgpt.com)) |
| Mountain Hyundai | **EXCLUDE** | Auto dealer (retail) | Dealer site & directories. ([Mountain Hyundai](https://www.mountainhyundai.com/?utm_source=chatgpt.com)) |
| Truscott, Brown & Dwyer | **REVIEW** | Funeral home (borderline) | Professional service, but model fit? (owner reliance). ([MapQuest](https://www.mapquest.com/ca/ontario/truscott-brown-dwyer-funeral-281716436?utm_source=chatgpt.com)) |
| Dee Signs | **POTENTIAL** | Signage/printing (OK), **but Burlington** | Valid site/address; accept only if within radius rule. ([Dee Signs](https://deesigns.ca/?utm_source=chatgpt.com)) |
| Landtek Limited | **POTENTIAL** | Engineering consultancy (OK) | Valid site; professional services. ([landtek.ca](https://landtek.ca/?utm_source=chatgpt.com)) |
| Weir’s Lane Lavender | **EXCLUDE** | Retail/agr-tourism | Retail shop now in Dundas. ([Weir's Lane Lavender & Apiary](https://www.weirslanelavender.ca/?srsltid=AfmBOorV8pbECLCSbMn-0qXcXAoFGrJTN5I2MVzemZLyhgbHxrTPSWAD&utm_source=chatgpt.com)) |
| PAC (Pardon Applications) | **FIX DATA then REVIEW** | Postal code mismatch | Use L8P 4W7, not L4P; confirm team size, SDE. ([pardonapplications.ca](https://www.pardonapplications.ca/contact/?utm_source=chatgpt.com)) |

**Concrete repo changes I’d make this week**

1. **Add a sources/places.py** with deterministic lookups (Google Places/Yelp/OSM).
2. **Build services/validation\_service.py** housing: category\_gate, geo\_gate, website\_gate, corroboration\_gate, revenue\_gate.
3. **Create src/core/rules.py** with explicit **BLACKLIST**/**WHITELIST** (unit-tested). Seed blacklist terms from this review.
4. **Add src/core/evidence.py** (helpers to create observation & validation rows).
5. **Rewrite the orchestrator** so discover → persist(raw) → enrich(extract only) → validate(gates) → score → qualify/export.
6. **Tests:** Add golden tests for the 10 leads above; add fixtures for common sources (fake Places/Yelp JSON).
7. **Reporting:** scripts/validation\_report.py generating a weekly HTML with fails by rule and top reasons (to tune rules fast).
8. **Prompt hardening** (if you keep LLM): “If field not present verbatim, return null; include source\_url; do not infer or estimate.”

**What “done” looks like (for you)**

* Every exported lead has: (a) **two corroborating sources** for address/phone, (b) **verified category in whitelist**, (c) **distance** within your radius, (d) **working website** with age signal, (e) **revenue range + confidence** grounded in benchmark + staff signal, and (f) **zero fabricated fields**.
* Anything that fails sits in the DB with a **clear exclusion reason** so we can tune rules or override case-by-case.

If you want, I can convert the above into issues/PR stubs against the repo structure shown in the README and drop in a ready-to-run **validation gate** module + unit tests. Otherwise, you can hand this to your dev and tell them: “no more vibes—**evidence or it doesn’t ship**.”