



Factorize Vision

what I need from you

Complete · tested standalone and inside the hub (Sept 5, 2026)

What it does for you

Film the shelf with a phone and get product rows: whole-video coverage, several products per frame, Catalog mode (one row per product, inventory duplicate check, CSV or draft products) or Count mode (everything on the shelf with units that become stock quantity), intensity 1–3 consensus voting, a live token meter, run analytics, and a filming coach — all on your own Ollama box, nothing leaves the network.

What I need from you

Everything below is something only the store owner can supply or decide — the code side is finished and waiting for these values.

1. The Ollama box, reachable and GPU-resident

Today: 192.168.1.6:11434. qwen2.5vl:7b needs ~15 GB of VRAM at the 4k context the plugin requests; on CPU it is 3× slower with minute-long cold loads.

Keep that machine on the LAN the WordPress server can reach; confirm the GPU has ≥ 16 GB. On the box: `OLLAMA_HOST=0.0.0.0` and `OLLAMA_KEEP_ALIVE=30m`. If the address changes:

```
wp option update factorize-vision_options '{"ollama_host":"192.168.1.6","ollama_port":11434}' -format=json
```

2. More test videos with a hand count

Count mode is calibrated on one 58 s video (owner count ≈ 42 ; result 53 products / 165 units). Two or three more videos with your own count let me tune tiles and pooling.

Film 3 shelves: slow pan, pause 2–3 s per item, portrait, good light; send the files and how many products are on each.

3. Workflow decisions

Who reviews before anything is published, and which mode is the default.

Say: Count for stock-taking (units $\rightarrow$ stock) or Catalog for listing; default intensity (1 sweep / 3 before adding); who approves drafts.

4. Pricing rule

Model prices are placeholders.

Your rule (e.g. 40% below retail) so rows are pre-filled correctly.

5. **Accuracy upgrade (optional hardware)**

A 24 GB GPU (used RTX 3090 class) runs qwen2.5vl:32b, which reads fine print the 7B model misses — the biggest single accuracy lever.

Say go and I switch the model name in the config once the card is in the box.

6. **UPC lookup API (optional, the 99% route)**

Barcodes give exact catalog identity.

A UPCitemdb or Barcode Lookup API key.

Optional / later

- Live camera mode.
- Background removal for cleaner product photos.
- Auto-publish pipeline with an approval queue.

How you will know it is working

- Factorize → Vision Intake → pick a video → Analyze: rows appear, the token meter runs, and the Run analytics panel logs the run.
- Count mode on a shelf: rows carry a units badge; Add to WooCommerce sets that as stock.

Factorize — Empire Tools build · generated Sept 6, 2026 · technical details: wp-content/plugins/factorize-vision/README.md, QUICKSTART.md, readme.txt