

Formul8 Product Roadmap and Scope of Work

Document Purpose: This document defines the full scope of work for the development of *Formul8*, a modular AI-powered consultant platform designed to serve cannabis operators across cultivation, processing, manufacturing, and retail. It is intended for internal developer guidance and will later be adapted for investors and enterprise clients.

Core Product Vision: Formul8 leverages multi-agent AI orchestration with internal proprietary datasets and expert system logic. A key differentiator is its **Agent-to-Agent Verification** layer, where multiple specialized agents interact, validate each other's outputs, and generate refined, consensus-based answers for the end user.

Phase 1 — Core Functional Agents (Modules)

All agents below are to be developed simultaneously as part of the MVP. There is no sequential priority in Phase 1.

Kevin's note from Vrsion call- Check for Discrepancies. Make sure the AI notifies the user when discrepancies are noted. Consider human intervention. Another selling point to ensure hallucinations are minimized and AI answers are production-ready.

1. Compliance Agent

Primary Responsibilities:

- Provide up-to-date regulatory guidance across jurisdictions for various purposes to the user
- Interacts with all agents
- Cross-check SOPs, product formulations, and operational practices against internal regulatory databases
- Alert users to compliance risks and upcoming regulatory changes, or even upcoming comments, live sessions, or other similar events by regulators. Potential to be tied into existing calendars or event indicators
 - Can be defined based on users interests or location
- Primary KPI- accuracy of answer.

Core Functionalities:

- **Formula & Process Regulatory Check**

- Internal DB check for pre-vetted formulations
- External source scan for jurisdiction-specific compliance
- Ability to get human verified answers or guidance
- **SOP Verification**
 - Compares user SOPs against trained internal templates and external resources
 - Adding proprietary Formula8 data and templates will be key here
 - Including jurisdiction specific SOP requirements
 - Assesses jurisdictional validity before external lookup
 - Lookup additional data that is checked for how up to date it is, that we scrub for and keep updated
- **Facility Compliance & AHJ Code Checks**
 - Search internal code references and verify with latest codes
 - API references for ICC, OSHA, and local safety code access
- **Licensing Compliance (e.g. CRA)**
 - Extracts and parses PDF documents from municipal or state sites
 - Very important that it has an accurate and solid PDF reader
 - County-level focus in Phase 1; link to local resources
 - Phase 2 + expands to more city and local level (depending on the ability to gather and keep updated such data)
 - May include Simplifya scraping or similar service integration, we would like to identify the strongest solution in this category.
- **Testing Requirements**
 - Region-specific required test panels, thresholds
 - Testing labeling requirements
- **Marketing Compliance**
 - Flags child-appealing language or design
 - Reviews structure-function claims against FDA or similar guidelines
 - Work directly and importantly with the FTO agent, amongst others
- **Regulator Assistant Sub-Agent**
 - Explains complex cannabis processes (e.g. terpene chemistry)
 - Helps regulators improve rule creation and understanding
 - Help highlighting important factors related to creating SOPs
- **Auto-classifies expenses into COGs categories for minimizing 280e losses.**

Internal Database Examples:

- SOP templates (state-specific)
- Regional compliance guides
- Regulatory FAQs and municipal PDFs
- Testing panel summaries
- Manufacturer Compliance Documents
- And more...

Developer Consideration:

- Add a confidence scoring layer for compliance suggestions
 - Enable agent response memory for multi-step verifications
-

2. Patent/Trademark Agent (Freedom to Operate)

Primary Responsibilities:

- Search patent databases to identify existing intellectual property
- Guide users to avoid infringement
- Highlight potential “close calls” as well
- Interacts with Compliance and Marketing Agents

Core Functionalities:

- **Patent & Trademark Search**
 - Links to USPTO, TSDR, PatentsView APIs
 - Optionally link to CrossRef and Semantic Scholar
- **Infringement Risk Assessment**
 - Uses prompt pattern recognition to flag risks
 - “I am not a lawyer” disclaimer
- **Human-Verified Answer Upsell**

Internal Database Examples:

- Public patent summaries
- Internal logs of filed trademarks by state
- Public infringement cases and landmark cases

Developer Consideration:

- Allow flagged queries to queue for legal review or upsell to expert
-

3. Operations & Equipment Agent

Primary Responsibilities:

- Support cannabis operators with day-to-day and strategic operations
- Calculate product outputs, plan expansions, and maintain equipment
- Find solutions to equipment problems
- Find recommendations for most efficient use for various equipment and manufacturers
- Interacts with compliance agent

Core Functionalities:

- **SOP Creation Assistant**
 - Cross-checks against compliance rules
 - Can be repurposed for license expansions
- **Operations Calculators**
 - Yield, loss, concentration
 - Terpene blending, infusion ratio estimates
 - Highlight potential efficiency improvements
- **Financial Projection Engine**
 - Projected cost vs margin analysis (detailed)
 - Ability to put into multiple financial formats
- **Equipment Maintenance & Troubleshooting**
 - Pulls structured equipment manuals and alert histories

Internal Database Examples:

- Yield calculators, conversion formulas
- Equipment maintenance logs
- Licensing expansion samples
- Manufacturer specific equipment documents, some not readily available to the public

Developer Consideration:

- Integrate equipment prompt responses with internal error resolution dataset
 - Some of these error codes can be pulled from existing documents, but serve as an easier way to diagnose and correct them
-

4. Formulation Agent

Primary Responsibilities:

- Assist in cannabis product development and refinement
- Provide chemistry-backed formulation recommendations
 - Another huge area for upsell, human verified answers and more
- Interacts with Compliance, Operations, Marketing agents

Core Functionalities:

- **"Your PhD Formulator" Assistant**
 - Guides THC/CBD ratios, terpene profiles, excipient selection
 - Based on internal formulation science and academic sources
- **Human-Verified Answer Upsell**

Internal Database Examples:

- In-house formulation guides and testing results
- Ingredient compatibility rules
- White papers from universities and other reputable organizations

Developer Consideration:

- Integrate PubMed, PubChem, and Semantic Scholar APIs
 - Tie formulation agent to operations agent for yield or cost estimates
-

5. Sourcing Agent (HUGE REVENUE GENERATOR - VERY IMPORTANT AGENT)

Primary Responsibilities:

- Source materials, equipment, services, and suppliers
- Providing accurate and vetted recommendations
- Provide a way to contact a recommended vendor
- Recommendations on what to avoid
- Ability to view a snapshot of a companies reputation score, which we could train on how to calculate.
- Interacts with all agents

Core Functionalities:

- **Equipment Sourcing**
 - Connect to trusted sources defined by us
 - Possible connection to Supply the Brand excel database
 - Possible connection to EAG used repository site
 - Manually loaded database entries
 - Links to MFG websites
- **Consumables & Services Sourcing**
 - Link to internal preferred vendors and contact forms
- **Packaging Guidance (Compliance + Sourcing)**

Internal Database Examples:

- Affiliate equipment spreadsheet
- Consumables cost lists
- Contact form templates
- Links to EAG and STB product databases

Developer Consideration:

- Auto-submit RFQs or contact requests where allowed
 - Needs to be tagged that it originated from Formula8
-

6. Marketing Agent

Primary Responsibilities:

- Guide compliant marketing content and validate feasibility of new product strategies
- Interacts with Compliance, FTO

Core Functionalities:

- **Copywriting Engine**
 - Works with Compliance and FTO Agent to validate content
- **Market Feasibility Review**
 - Based on pricing, state data, BDSA, Headset integration
 - Would be good for Formula8 to develop a special relationship where we can access snippets of this data for free to entice the user, then sell a full report via referral, for another revenue generating area.
- **Mockup Images**
 - Uses prompt-to-image pipeline (e.g., Unsplash API)
 - MUST have a high-end image creator, this is extremely important and must be unique.
- **Role-Based Answer Formatting**
 - Tailors response for C-Suite, R&D, Compliance, Legal, etc.
 - Special Note: Should we have role-based answer formatting as a unique feature for all our agents? This would avoid complex prompting on audience or purpose.

Clarification Needed:

- Confirm how user role would be detected: dropdown, login setting, or prompt inference?

Internal Database Examples:

- Product category-specific marketing benchmarks
 - Compliance limits on claims or visuals by region
-

7. Internal Agent: Process Optimization

Primary Responsibilities:

- Is this possible for phase 1? Need to discuss with dev team
- Monitor app usage, user intent, error rate, hallucination detection
- Continuous improvement and optimization of agent logic

Core Functionalities:

- **User & Agent Analytics Dashboard**
- **Model Confidence Score Feedback Loop**

Developer Consideration:

- Connect to backend logs for performance monitoring
 - Add admin-side tooling to flag bad outputs for model fine-tuning
-

Phase 1 — Platform-Wide Features

- **Native App (High Priority)**
 - **Remove Auto Delivery of Sources** (user chooses)
 - **Unique Logins** for users and teams
 - **Data Functions**
 - Save results in various formats
 - Export to third-party apps (CSV, PDF, Asana, Slack)
 - **Persistent Memory per Project**
 - **Vector Database Maintenance**
 - Continuously ingest SOPs, formulations, documents from clients and partners
-

Phase 2 & 3 — Future Functionality Roadmap

- **Hosted Projects & Handoff Flow**
 - Users collaborate across departments
 - E.g., formulation → marketing → production
- **Content Creator Marketplace**
 - Users can browse, preview, and purchase SOPs, guides, etc. from experts
- **Backend Automation Agents** (Internal)
 - CRM, Sales, Marketing, Research
 - Competitor surveillance (ISR)
- **Expanded Compliance Depth**
 - Move from county to city level data (SOPs, licenses, testing)
- **Enterprise Integration API Suite**
 - BioTrack, METRC, openFDA, USPTO, etc.
- **Project Performance Metrics** → Add KPIs like:

- Avg. time to response of cross-agent verifications
 - Dataset expansion milestones
 - Internal- SEO and marketing agent
-

Data Acquisition Strategy (Ongoing)

- Partnerships with (or a trusted database who already syndicates this)
 - Universities
 - Equipment manufacturers
 - Cannabis Training Academies (CTAs)
 - Consultants
 - Content creators
 - AHJs
 - Who else?
-

Long-Term Vision (Formul8 Ecosystem)

- Platform becomes a branded ecosystem offering:
 - Proprietary offerings across all buckets and agents
 - Licensed SOPs
 - Equipment reselling (beefed up)
 - End-to-end support for cannabis businesses from R&D to shelf
-

Additional Agent- DreaMS Agent- GCMS.

-

Additional Agent- Sales / Customer Service Agent (after initial design gets ironed out)

- Basic support / Q&A then escalate to a human for complex inquiries.

Road Map V2

Formul8 Product Roadmap & Scope of Work

Core Differentiator – Agent-to-Agent Verification

How it works & what the dev team must track

Layer	What happens under the hood	Metrics to surface
Retriever	Pulls top-N law or data chunks from vector DB.	<code>docs_retrieved</code> , <code>source_coverage %</code>
Expert Generator (GPT-o3)	Drafts answer from those chunks.	<code>gen_confidence</code> (model log-prob)
Verifier	Second LLM plus rule-engine re-reads sources, scores factual match, flags hallucination.	<code>verifier_score</code> , <code>disagree?</code>
Consensus	If verifier score < 0.7 or conflict > 15 % → answer shown as “Needs human verify.”	<code>needs_human %</code> , <code>avg_time_to_consensus</code>

Back-end stores every layer’s JSON in the audit table so QA or regulators can replay how an answer formed.

Phase 1 Functionality Buckets (“Agents”)

Goal: replace 10-15 full-time roles worth of manual work.

Compliance Agent (first-pick model = GPT-o3)

What the user gets

1. **Ask-a-rule chat box** – instant YES / NO, legal clause, risk color.
2. **SOP drop-zone** – redline PDF: missing steps clickable or highlighted.
3. **Formula checker** – pastes recipe → flags banned substances, wrong THC limits, ingredient regs etc..
4. **Facility map wizard** – upload address → returns zoning notes & ICC/OSHA highlights.
5. **Ahead-of-the-curve feed** – weekly e-mail: “CA DCC just scheduled a public hearing on inhalable hemp.”

Sub-feature	Data path & UI deliverable	First pick	3 alt. tools
Formula & process check	Internal vector DB → popup pass/fail & link to rule	GPT-o3 + Sentence-T	OpenAI Emb., Haystack, Pinecone
SOP verification	Compare TXT vs. template, show diff lines	GPT-o3	Claude 3, Mistral-Large, Cohere
AHJ code fetch	Crawl ICC/OSHA PDFs → snippet cards	Scrapy + OCR	Firecrawl, Puppeteer, Colly
Licensing lookup	Auto-parse municipal PDFs → JSON table	PyPDF / Tika	AWS Textract, Form Recog., GVision

Compliance
calendar

Cron job → email + app
banner

Airflow cron

Prefect, Dagster,
GitHub Cron

Operations & Equipment Agent

User deliverables

- **Run Calculator** – enter flower weight → potency yield, per-gram cost, gross margin.
- **280E wizard** – paste QuickBooks CSV → deductible vs. non-deductible chart.
- **Error-code doctor** – enter “F4-3” from extractor → steps, part link, downtime estimate.

Module	Tech path	GPT-o3 backup
Yield / loss calc	pandas, saved formulas in DB	Polars, Excel-JS, DuckDB
Equipment docs	Scrape manuals → vector DB	DeepFinder, PDFPlugin, Azure Search
280E logic	rule-engine + IRS memos	TaxJar API, Avalara

Formulation & Processing Agent

User deliverables

- **Recipe Builder UI** – sliders for THC/CBD ratio, terpene target; outputs full GRAS ingredient list, weights in g/L, pH window.
- **Stability hints** – predicted shelf-life curve & “add citric acid” tip.
- **Cost dashboard** – per-unit COGS with links to Sourcing Agent SKUs.

Engine	First pick	Alt options
Chemistry calc	RDKit + GPT-o3	DeepChem, ChemAxon, MolSimplify
Excipient safety	PubChem API	EPA SRS, USDA FoodData, OpenFDA

Sourcing Agent (“big revenue”)

User deliverables

- **Marketplace table** – category, MOQ, lead time, rep score (1-5).
- **Auto-RFQ** – click vendor → sends templated e-mail with specs & attaches NDA.
- **Affiliate linked** - Commissions from recommendation engine for ingredients, equipment, services

Data feed

First pick

Alt

Vendor DB Our CSV + Stripe affiliate tag LeafLink API, Alibaba, Piloterr

Rep score GPT-o3 summarises FDA 483s ChatGPT-BCA, HuggingFace sentiment, RavenPack

Spectra Agent

User deliverables

- **File drag-drop** – 30 MB mzML → in <120 s returns potency, terpene %, pass/fail table.
- **Unknown alert** – clusters novel peaks, links to Formulation agent for R&D.
- **ISO-ready PDF** – auto-formatted COA with inline chromatograms.

Quick summary

1. **Prepare an open file** – vendor raw data are first converted to *.mzML* (or *.mgf*) so any program can read the spectra [ccms-ucsd.github.io](https://github.com/ccms-ucsd).
2. **Drag-and-drop or POST** – the user drags that file onto the Spectra Agent's upload panel (built with a React "drop-zone") or sends it via an HTTPS API [react-dropzone.js.org](https://medium.com/blog/alexdevero.com/react-dropzone.js.org)[medium.com](https://medium.com/blog/alexdevero.com)[blog.alexdevero.com](https://medium.com/blog/alexdevero.com).
3. **Backend unpacks each spectrum** – Python readers stream every MS/MS scan and its metadata pyteomics.readthedocs.io.
4. **Convert to DreaMS tensor** – intensities + metadata are stored in a compact HDF5 dataset for fast math drea.ms/docs.readthedocs.io[hdfgroup.org](https://hdfgroup.org/support)support.hdfgroup.orgneonscience.org.
5. **DreaMS fingerprints & matches** – the model turns each spectrum into a 1 024-number "fingerprint," then looks for matches or contaminants in milliseconds nature.com.
6. **COA panel appears** – the user sees potency, terpene %, any red-flag contaminants, plus a PDF they can hand regulators.

Tool summary for devs

Layer	First-choice library	Backup options
Front drag-drop	<code>react-dropzone</code> hook	HTML5 <code><input type=file multiple></code> + custom JS
Spectrum parser	Pyteomics.mzML	pymzML, psims
HDF5 write	<code>h5py</code>	PyTables, HDF5 C API
Fingerprint model	DreaMS (PyTorch)	msFineTune, MetaboAnalyst
Vector search	FAISS - FlatL2	ScaNN, Annoy

Sub-task	First pick	Alt
Fingerprint	DreaMS PyTorch	msFineTune, MetaboAnalyst

Few-shot contam	GPT-o3 head	LightGBM, SVM, CatBoost
--------------------	-------------	-------------------------

Report PDF	ReportLab	WeasyPrint, VivaCharts
------------	-----------	------------------------

DreaMS never guesses your method parameters—it reads them directly from the spectrum file’s metadata and bakes them into the vector it produces. The process happens in three spots:

1 – File format You (or the LIMS) export data as mzML / mzXML / MGF. Those open formats already carry instrument settings such as ion mode, collision energy, adduct, precursor m/z, scan polarity, retention time and even the chromatographic column if the vendor wrote it.

2 – Tensor build The loader (pyteomics.mzML or pymzML) streams every scan, grabs those header fields, and writes them alongside m/z–intensity pairs into DreaMS’s ML-friendly HDF5 tensor. Each spectrum row has fixed columns: precursor_mz, charge, adduct, collision_E, ion_mode, instrument, rt, column_id, etc.

3 – Model embedding During pre-training the transformer learns two tasks—(a) predict masked peaks and (b) predict which of two spectra elutes first (retention-order objective). To solve that, it must pay attention to the metadata tokens it’s given; those tokens are concatenated to the peak list, so the final 1 024-number fingerprint already “knows” the method context.

Why this works in practice

If you change collision energy from 20 eV to 35 eV, that token flips, so fingerprints from the two methods don’t alias each other.

When you fine-tune DreaMS for a new panel, you include the method tag (e.g., “HILIC_7 min”) in the few-shot spectra, so the small output head learns to pay attention only to data from that method—no cross-talk.

Because the metadata are stored in the same HDF5 row, downstream agents (QC, Formulation) can filter spectra by instrument or column without extra look-ups.

Bottom line: DreaMS “knows” your method because the spectrum file tells it, the tensor keeps it, and the transformer is trained to notice it. No hand-coding, no manual form filling—just export, upload, and the parameters ride along automatically.

Marketing Agent

User deliverables

- **Copy Studio** – product name → 3 taglines, 100-word blurb, 30-char SMS, all compliance-checked.
- **Hero image** – 1024×1024 PNG, brand colours, no cartoon kids.
- **Market sanity** – table: competitor SKU, average price, recommended MSRP.

Component	First pick	Other
Text gen + compliance	GPT-o3	Bard, Jasper BrandGuard, Cohere
Image gen	SDXL	DALL·E 3, Midjourney, Imagen 2
Price feed	Headset snapshot	BDSA, LeafLink data, Weedmaps scrape

Additional INTERNAL agents (Phase 1)

Agent	What it does for the team	Main tool	3 alternates
Process-Optimization	Dashboards: most-used agent, avg. answer time, “weird answer” log.	Kafka + Grafana	Segment, Mixpanel, Amplitude

CRM Intelligence	Auto-creates deal in HubSpot when new company signs up; emails CSM if usage drops 50 %.	HubSpot API	Salesforce, Zoho, Freshsales
Billing & Usage	Reads API count, writes Stripe invoice nightly; shuts off over-quota.	Stripe Usage	Chargebee, Zuora, Recurly
Knowledge Governance	Tags PII, blocks copyrighted uploads.	Amazon Macie	Azure Purview, BigID, Collibra
Incident Bot	Spike in errors → Jira ticket + Slack ping + log link.	Opsgenie	PagerDuty, VictorOps, ServiceNow

Phase 2 / 3 (Most Likely) – Hosted Projects & Marketplace

- Project hand-off flow (formulation → marketing → finance)
- Content marketplace (buy vetted SOPs & templates)
- Deeper city-level compliance, mobile app, back-office automation bots (Sales, Marketing, Competitor ISR).

Must-Have APIs (Phase 1)

Metrc, BioTrack, openFDA, PubChem, PatentsView, OPC-UA, Stripe Usage, Pinecone, AWS Textract, Google Vision OCR.

Should-Have APIs (Phase 2 +)

Headset, BDSA, LeafLink, Confident Cannabis, CrossRef, Semantic Scholar, Google Ads
PolicyCheck, Unsplash, EPA SRS, USDA FoodData, USPTO TSDR, Alibaba supplier, NIOSH
Pocket Guide, ICC Code Connect.

Completeness score 93 / 100

Still open: screen wireframes, GPU cost line for Spectra, final legal-review SLA.