

### **Thematic Roadmap:**

<b>High-Level Work</b>	<b>Year 1 2025-26</b>	<b>Year 2 2026-27</b>	<b>Year 3 2027-28</b>	<b>Year 4 2028-29</b>	<b>Year 5 2029-30</b>
<b>Ease of Use</b>	Install, bug issues, GUI ease of use	Streamline workflows, menus	Drag-and-drop tools	Personalized User Settings	Voice and Mobile Interaction
<b>Lightweight Hardware</b>	Optimize scripts	Resource-efficient scaling	Code Optimization for Cost	Run on Low-Power Devices	Energy Consumption Reporting
<b>Strategic Partnerships / External Funding</b>	Identify partners	Outreach and interviews (what are the needs?)	Early user testing?	Early user testing?	Final Handoff Presentation and Case Study
<b>Flexible Architecture</b>	Modular components	Standardize Feature APIs	Easy Code Updates and Upgrades	Support for New Frameworks	No-Downtime System Maintenance
<b>Process Quality / Automation</b>	Document workflows	Automated testing and error checks	Consistent Coding Standards	Proactive Bug Fixing	Full Codebase Review and Audit

## Feature-Based Roadmap:

High-Level Work	Year 1 2025-26	Year 2 2026-27	Year 3 2027-28	Year 4 2028-29	Year 5 2029-30
<b>Web Scraping</b>	Establish safe practices around API usage.  Test with Gemini and ChatGPT models for internal use only.	Integrate More Scraping	Implement other file format conversions	Robust Error Handling	Ethical and Legal Scraping Framework
<b>Data Cleaning</b>	Create basic cleaning scripts and processes.  Benchmark against industry tools and identify missing features.  Begin building a lightweight, custom cleaning tool.	Full integration with Scraping Output	Advanced Metadata Extraction (Attribution)	Industry Tool Benchmarking Features	Deep Integration with Fine-Tuning Module
<b>Fine-Tune LM</b>	Set up LMForge and explore model training.  Compare approaches used by tools like N8N and Langflow.  Run a small-scale test model.	Implement Model Quantization (LoRA/QLoRA) - <b>Might already be integrated.</b>	Database for Model Hyperparameters & Search	Advanced Hyperparameter Optimization Tool	Model Deployment and API Management

<b>RAG Database</b>	Vector DB integration MySQL->PostgreSQL migration script (if necessary)	Vector DB scaling Graph DB architecture for integration and scaling (may require \$ for relationship extraction with LLM)	Semantic Search and Document Versioning	Multi-Modal RAG (Image/Video Support)	Explore backing up of datasets (flesh out this idea with tools like Dolthub, etc.)
<b>Agentic Systems</b>	Prototype a basic agent creation tool  Create demo workflows in Langflow for internal review.	Basic Agent UI Implementation & Tool Selection	Custom Tool Upload (Python Scripting)	Multi-Agent Collaboration and Coordination	Agent Performance Monitoring and Auditing
<b>Governance &amp; Handoff</b>	Other: Containerization with Docker or Podman	Living Documentation and GitHub Issues	User Interviews & Feedback Cycle	Security and Ethical AI Compliance	Final Project Handoff and IP Review

To ensure long-term stability and reduce technical debt, **15% of total development time** each year will be dedicated to:

- Ongoing bug fixes
- Code refactoring
- Reducing technical debt buildup
- Improving maintainability and overall code quality

This allocation applies consistently across all roadmap years.