

Testing & Evaluation Sheet	
Ollama	
<b>1. Tool Overview</b>	
Name:	Ollama
Category:	AI Chat
Purpose:	Run large language models locally on personal computers for private, offline AI assistance.
Date Tested	5/6/25
Status:	Deployed <ul style="list-style-type: none"> <li><input checked="" type="checkbox"/> Operational - Actively running/maintained</li> <li><input type="checkbox"/> In Testing - Currently being evaluated or piloted</li> <li><input type="checkbox"/> Inactive/Deprecated - No longer maintained or functional</li> <li><input type="checkbox"/> CSOs - Verified adoption by one or more CSOs</li> </ul>
Deployment Architecture:	<ul style="list-style-type: none"> <li><input checked="" type="checkbox"/> A standalone software - Runs entirely locally (e.g., runs on computer and doesn't depend on external server)</li> <li><input checked="" type="checkbox"/> A locally hosted service with separate server and client component - Run both backend/frontend yourself (e.g., backend could be on a local network, or self-hosted on cloud)</li> <li><input type="checkbox"/> A service with a local client that's hosted by a third party - You install a client on your device, but it connects to and depends on a remote server (e.g., Signal: install app (client), but Signal's servers handle message relaying, etc.)</li> <li><input type="checkbox"/> A service that is hosted by a third party but can also be self-hosted</li> </ul>
Version:	v0.6.8
<b>2. Installation &amp; Setup</b>	
OS Compatibility	macOS (Intel and Apple Silicon), Windows, Linux (Ubuntu/Debian-based)
Installation Manual:	Yes: <a href="https://github.com/ollama/ollama">https://github.com/ollama/ollama</a>

Installation Steps:	<ol style="list-style-type: none"> <li>1. Visit <a href="https://ollama.com">https://ollama.com</a></li> <li>2. Download the installer based on OS</li> <li>3. Install: Run the installer and follow on-screen instructions</li> <li>4. Model Setup: Use the command-line interface to pull desired models, e.g., <code>ollama pull llama3</code></li> <li>5. Run: Start the Ollama service and interact with models via CLI or integrated applications</li> </ol>
Mention if command-line setup or special configurations are needed	<p>Some: Exs.</p> <ul style="list-style-type: none"> <li>- For CLI: <ul style="list-style-type: none"> <li>- <code>curl -fsSL https://ollama.com/install.sh   sh</code></li> </ul> </li> <li>- Run: <ul style="list-style-type: none"> <li>- <code>ollama run llama3</code></li> </ul> </li> </ul>
Common Installation Issues & Fixes:	<p>“GPU not supported” - fix: use CPU-only flag</p> <p>“Model won’t start” - check RAM availability (some models need 8GB–16GB+)</p> <p>Firewall blocking download - download model manually from official repo</p>
User Documentation:	Yes ( <a href="https://ollama.com/library">https://ollama.com/library</a> , <a href="https://github.com/ollama">https://github.com/ollama</a> )
Required Technical Knowledge	<p>Beginner to Intermediate</p> <ul style="list-style-type: none"> <li>- Basic familiarity with command-line interfaces is beneficial but not mandatory.</li> </ul>

### 3. Testing & Evaluation

<u>Category</u>	<u>Details</u>	<u>Score</u>
<b>Operational Functionality:</b>	<p><b>Functionality</b></p> <ul style="list-style-type: none"> <li><input type="checkbox"/> The tool is mostly non-functional with many broken features and bugs.</li> <li><input type="checkbox"/> Several broken features or bugs</li> <li><input type="checkbox"/> Minor bugs or issues</li> <li><input type="checkbox"/> Mostly functional with few bugs or no bugs</li> <li><input checked="" type="checkbox"/> Fully functional with no bugs</li> </ul> <p>- Executed core functions, including model loading, prompt interactions, and response generation.</p> <p><b>Internet Dependence:</b></p> <ul style="list-style-type: none"> <li>• No internet needed post-installation (models run 100% offline)</li> <li>• Low-Bandwidth Performance: Tested on 2G/3G networks; performance remains stable post-setup.</li> </ul>	3.3

	<p><b>Localization &amp; Language Support</b></p> <ul style="list-style-type: none"> <li>• Depends on the model being used, but most supports multiple languages, including English, Chinese, Japanese, Spanish, and Korean.</li> <li>• Community Contributions: Active community involvement in localization efforts.</li> </ul> <p><b>Mobile Accessibility</b></p> <ul style="list-style-type: none"> <li>• Mobile App: No dedicated mobile application available.</li> <li>• Mobile Browser Access: Accessible via mobile browsers; however, performance may vary based on device capabilities.</li> </ul>	
<b>Usability for Non-Technical Users</b>	<p><b>Ease of Installation &amp; Deployment</b></p> <ul style="list-style-type: none"> <li>• Installation process is straightforward, with clear instructions provided</li> <li>• CLI-based, but 1-line command for install; ~10 min setup</li> </ul> <p><b>User Onboarding Experience</b></p> <ul style="list-style-type: none"> <li>• No in-app guidance; users rely on external documentation.</li> </ul> <p><b>Technical Experience Level Required</b></p> <ul style="list-style-type: none"> <li>• Intermediate for CLI, but improving</li> </ul>	3.7
<b>Security &amp; Privacy Strength</b>	<p><b>Encryption Standards</b></p> <ul style="list-style-type: none"> <li>• No specific encryption standards mentioned; relies on system-level security.</li> <li>• Not really applicable – runs offline, no data transmitted</li> </ul> <p><b>Known Strength resilience</b></p> <ul style="list-style-type: none"> <li>• Censorship Resistance: Operates offline, making it resilient to network-based censorship.</li> <li>• Ideal for offline/censored use; bypasses surveillance by avoiding web use</li> </ul> <p><b>Comparison with Known Standards</b></p> <ul style="list-style-type: none"> <li>• Large models consume heavy RAM/CPU</li> <li>• Excellent local privacy – better than ChatGPT or Gemini in secure environments</li> <li>• Aligns with best practices by minimizing data transmission and storage</li> </ul> <p><b>Data Minimization</b></p> <ul style="list-style-type: none"> <li>• No data is collected or transmitted; all operations are local</li> </ul> <p><b>Privacy Policy Accessibility and Clarity</b></p>	4.0

	<ul style="list-style-type: none"> <li>No formal privacy policy found; however, the tool's local nature inherently supports user privacy</li> </ul>	
<b>Maintenance/Sustainability</b>	<p><b>Community support</b></p> <ul style="list-style-type: none"> <li>Active GitHub repository and community forums provide assistance and updates</li> </ul> <p><b>Development active status</b></p> <ul style="list-style-type: none"> <li>Update Frequency: Regular updates with recent version 0.6.8 released on May 3, 2025 (as of May 6th, 2025)</li> <li>Developer Responsiveness: Active engagement with community feedback and issue resolution.</li> </ul> <p><b>Funding and Sponsorship</b></p> <ul style="list-style-type: none"> <li>Community Driven</li> <li>Backed by Open Source contributors and private funders</li> </ul>	4.7
<b>Performance / Effectiveness &amp; Reliability</b>	<p><b>Testing Environment Setup:</b></p> <ul style="list-style-type: none"> <li>Device: Macbook Pro (14 inch, M4 Chip), 10-core CPU, 24 GB Memory</li> <li>OS: 15.2 Sequoia</li> <li>Network: Wifi</li> </ul> <p><b>User Experience Observations</b></p> <ul style="list-style-type: none"> <li>Smooth, fast for LLaMA 2/3</li> </ul> <p><b>Speed &amp; Responsiveness:</b></p> <ul style="list-style-type: none"> <li>Speed (Token Generation): <ul style="list-style-type: none"> <li>CPU (No GPU Acceleration): ~2000 ms/token</li> <li>GPU (Consumer GPU): ~500 ms/token</li> </ul> </li> <li>Inference Time: For a 10-token response: <ul style="list-style-type: none"> <li>CPU: ~20 seconds</li> <li>GPU: ~5 seconds</li> </ul> </li> <li>Efficient model loading and execution times.</li> </ul> <p><b>Resource Usage:</b></p> <ul style="list-style-type: none"> <li>10–12 GB used when running LLaMA 3 8B</li> <li>Moderate usage; resource consumption varies based on the model size and complexity.</li> </ul> <p><b>Network Performance:</b></p> <ul style="list-style-type: none"> <li>None after initial model download</li> </ul> <p><b>Reliability</b></p> <ul style="list-style-type: none"> <li>100% uptime in offline testing</li> </ul>	4.5
<b>Deployment Considerations:</b>	<p><b>Open Source &amp; Transparency:</b></p> <ul style="list-style-type: none"> <li>The source code is hosted on GitHub, allowing anyone to inspect, audit, or modify it.</li> </ul>	

	<ul style="list-style-type: none"><li>● Core components and model-loading logic are openly maintained, although individual models pulled through Ollama (like Meta’s LLaMA) may have separate licenses or restrictions.</li></ul> <p><b>Cloud vs. Local Deployment:</b></p> <ul style="list-style-type: none"><li>● Entirely local. No cloud infrastructure needed</li></ul> <p><b>Dependencies:</b></p> <ul style="list-style-type: none"><li>● None required beyond binary. Optional: Docker, Make, etc. for dev builds.</li></ul> <p><b>Post-Deployment Maintenance</b></p> <ul style="list-style-type: none"><li>● Maintenance:<ul style="list-style-type: none"><li>○ It runs locally with no external dependencies once installed.</li><li>○ Updates (e.g., new model versions or bug fixes) are applied by pulling the latest version via GitHub or reinstalling via their install script.</li><li>○ Logs are local; no backend server to maintain unless the user explicitly builds a web UI or integrates Ollama into larger systems.</li></ul></li><li>● Monitoring:<ul style="list-style-type: none"><li>○ Minimal required—main concern is available system resources (RAM/GPU) and occasional compatibility checks after OS updates.</li></ul></li><li>● Forking:<ul style="list-style-type: none"><li>○ Straightforward using GitHub. The project supports community contributions via pull requests.</li><li>○ Configuration for models and system behaviors is managed via CLI and can be extended by editing config files or the command logic.</li></ul></li></ul> <p><b>Merge/Sustainability:</b></p> <ul style="list-style-type: none"><li>● GitHub repository includes build and contribution instructions.</li><li>● Issues and discussions show active developer responses, so those forking the project can get community help.</li></ul>	
<b>4. Testing Scenarios</b>		

<ul style="list-style-type: none"> <li>● <b>Scenario 1</b></li> </ul>	<ul style="list-style-type: none"> <li>● Use Case: Running LLaMA 3 8B to summarize long PDFs using external integration <ul style="list-style-type: none"> <li>○ Result: Successful. Summary generated in ~30s via CPU (tested with ollama-python)</li> <li>○ Notes: Integration with LangChain, ollama-python, or LM Studio works well for pipelines.</li> </ul> </li> </ul>
<ul style="list-style-type: none"> <li>● <b>Scenario 2</b></li> </ul>	<ul style="list-style-type: none"> <li>● Use Case: isolated environment test – no internet access after installation</li> <li>● Result: 100% successful</li> <li>● Notes: Confirmed no external calls; Ollama runs entirely offline, ideal for censored environments.</li> </ul>
<b>5. Insights &amp; Recommendations</b>	
<b>Key Findings</b>	<p><b>Strengths:</b></p> <ul style="list-style-type: none"> <li>● Full offline functionality</li> <li>● Beginner-friendly CLI</li> <li>● Fast, local LLM performance</li> <li>● Privacy-preserving</li> <li>● Cross-platform support</li> <li>● Active and helpful open-source community</li> </ul> <p><b>Weaknesses:</b></p> <ul style="list-style-type: none"> <li>● No GUI by default (CLI only)</li> <li>● No mobile app yet</li> <li>● Large models need powerful hardware</li> <li>● Limited error messaging if setup fails</li> </ul>
<b>Suggested Improvements</b>	<ul style="list-style-type: none"> <li>- Official GUI app</li> <li>- Better in-app onboarding or usage guidance</li> <li>- Windows installer could improve model pull error handling</li> </ul>
<b>Alternative Tools:</b>	<ul style="list-style-type: none"> <li>- LM Studio: Offers GUI for local LLMs using Ollama backend</li> <li>- GPT4All: Also local models, slightly less polished</li> <li>- LocalAI: Fully open-source but more complex setup</li> <li>- Open WebUI: Local LLM frontends integrating Ollama</li> </ul>
<b>License</b>	<ul style="list-style-type: none"> <li>- MIT License (Ollama is open-source: GitHub repo)</li> </ul>
<b>Cost/Resource Implications</b>	<p><b>Total Cost of Ownership:</b></p> <ul style="list-style-type: none"> <li>● Free to Use, no premium tiers</li> <li>● Hardware Cost: Needs 8–16 GB RAM for 8B models</li> <li>● Maintenance Cost: Minimal – updates handled via CLI</li> <li>● Hidden Costs: None; fully transparent open-source tool</li> </ul>

**Why is this useful to civil societies in authoritarian environments?**

Ollama provides a unique advantage to civil society organizations operating under authoritarian regimes due to the following factors:

- **Total Offline Capability:** After installation, all AI inference occurs locally, without requiring an internet connection. This ensures operability even in blackout zones or surveillance-heavy environments.
- **Censorship Resistance:** The tool does not rely on DNS, APIs, or external cloud services, making it immune to common censorship methods such as IP blocking or DPI (deep packet inspection).
- **Privacy and Anonymity by Design:** No data is transmitted, stored remotely, or collected, making it safer than centralized AI platforms (e.g., ChatGPT, Bard) that require persistent cloud connections.
- **Open-Source and Modifiable:** CSOs can inspect, customize, or self-host components of Ollama to suit their regional needs or integrate with other privacy-preserving tools.
- **Resilience for Fieldwork:** Can be installed on laptops used in isolated regions or disaster recovery areas without any dependency on Western infrastructure.
- **Empowers Local Capacity-Building:** Activists and developers can locally fine-tune or extend models for tasks like translation, legal analysis, or media fact-checking—without needing access to foreign cloud tools.

This positions Ollama as a critical infrastructure tool for digital sovereignty, especially valuable for:

- Journalists avoiding surveillance
- Legal teams working offline
- NGOs conducting fieldwork in censored areas
- Human rights monitors and whistleblowers