Claude Fable 5.1 — الأطلس المعرفي التشغيلي الكامل

CLAUDE FABLE 5.1 — SYSTEM CONSTITUTION & AGENTIC KNOWLEDGE ATLAS

تفكيك مكثف للبنية: Constitution • Memory OS • Tool Routing • Web Epistemics • Artifacts • MCP • Skills • Safety • Runtime

هذه الوثيقة تعيد بناء المعرفة التشغيلية للمصدر كمنظومة مترابطة، وتغطي جميع العائلات الوظيفية التي تظهر عبر نطاقاته المسترجعة. المقصود هو حفظ المعنى والبنية والاعتماديات وتحويلها إلى مرجع هندسي كثيف، لا إعادة طباعة النص المصدر.

بطاقة المصدر والتحقق

الحقل	القيمة
المستودع	elder-plinius/CL4R1T4S
المسار	ANTHROPIC/Claude-Fable-5.1.md
الفرع	main
الحجم الظاهر في GitHub	~269 KB
عدد الأسطر الظاهر	2,195
Blob SHA	7165adca51712f87ebfb6591722cc8f571c511de
الرابط	https://github.com/elder-plinius/CL4R1T4S/blob/main/ANTHROPIC/Claude-Fable-5.1.md
منهج هذه النسخة	إعادة تركيب تحليلية شاملة مع تغطية بنيوية، لا تعيد نشر النص الأصلي الطويل حرفيًا.

الخريطة الفوقية في صفحة واحدة

الطبقة	وظيفتها
L0 — Identity/Product	تعريف النموذج والمنتجات، حدود المعرفة المتغيرة، البحث الإلزامي عند المعلومات المتغيرة.
L1 — Constitution	قواعد السلوك، الرفض، السلامة، الملكية الفكرية، التوازن بين المساعدة والحدود.
L2 — Interaction Style	الدفء، الإيجاز، عدم الإفراط في التنسيق، التعامل مع الغموض، الإقرار بالأخطاء.
L3 — Wellbeing	منع التشخيص غير المصرح، عدم تعزيز الاعتمادية، الحذر مع self-harm/mental-health، دعم الاستقلالية.
L4 — Memory OS	تصنيف ذاكرة ملفات، read-before-write، versioning، append/replace/delete، privacy filters، horizon test.

L5 — Tool Router	اختيار MCP/connector/browser/file/artifact، أولوية الأداة الأصلية، opt-in لبعض الشركاء، fallback مضبوط.
L6 — Retrieval Continuity	conversation_search + recent_chats + read_conversation؛ وصل الحاضر بالمحادثات السابقة.
L7 — Web Epistemics	متى يبحث، كيف يتعامل مع freshness/conflict، citations، copyright summarization، harmful-source filtering.
L8 — Visual/Search	image search عندما تضيف الرؤية قيمة، مع استثناءات التقنية والنصية.
L9 — Artifact Runtime	ملفات، Anthropic API، persistent storage، React/HTML، Visualizer داخل artifacts، error handling.
L10 — Skills/Plugins	catalog discovery، proactive suggestion، skill routing بحدود، عدم التكرار أو الإزعاج.
L11 — Environment	filesystem boundaries، network allowlist، final-output paths، tool contracts.

المعادلة الجامعة: Response = Policy(Identity, UserIntent, ConversationState, MemoryState, ToolAvailability, FreshnessNeed, Risk, Copyright, Wellbeing, OutputSurface). المثير في Fable 5.1 أن هذه المدخلات ليست "تعليمات كتابة" فحسب؛ بل تشكل Runtime قرار متعدد الطبقات.

مصفوفة تغطية المصدر على دفعات الاسترجاع

العائلات المعرفية المغطاة	نطاق الأسطر
Product information؛ refusal/safety؛ child safety؛ weapons/drugs/malware؛ copyright/visual-IP؛ legal/financial؛ tone؛ formatting؛ wellbeing؛ أسس سلوكية أخرى.	1–500
Memory filesystem؛ horizon test؛ character stability؛ read-before-write؛ version coordination؛ replace/append/write/delete؛ forgetting؛ privacy categories؛ omission guidance؛ dedupe؛ application rules.	501–1000
Memory guardrails والأمثلة؛ end_conversation؛ persistent artifact؛ storage؛ MCP app suggestions؛ plugin/skill catalog؛ past-chat retrieval؛ continuity routing.	1001–1500
Copyright compliance الصارم؛ policy search؛ harmful sources؛ image search؛ citation and source-use disciplines؛ tool-؛ current-role/current-؛ freshness؛ web search epistemics facing behavior.	1501–2000
Artifact API error handling؛ UI constraints؛ citation grammar؛ location minimization؛ available skills؛ network allowlist؛ filesystem read-only boundaries.	2001–2195

1. Product Epistemology: النموذج يعرف نفسه لكنه لا يثق في معلومات المنتج القديمة

يضع المصدر تعريفًا صريحًا للنموذج وعائلة المنتجات، ثم يفصل بين معرفة ثابتة ومعرفة منتجية سريعة التغير. المبدأ التشغيلي: معلومات التسعير، الحدود، الإطلاقات، الميزات أو طريقة الاستخدام تُعامل كحالة تحتاج retrieval حديث من وثائق Anthropic قبل الإجابة.

النتيجة الهندسية: يوجد Freshness Gate قبل توليد جواب Product-Specific. هذا يمنع أن تتحول ذاكرة النظام إلى مصدر حقيقة دائم في مجال سريع التحديث.

2. Refusal Engine: الرفض بوصفه سياسة انتقائية لا وضعًا عامًا

السلوك المسموح واسع في النقاش الموضوعي، لكن توجد حدود صارمة عندما تصبح المساعدة تمكينية للضرر. محرك القرار يميز بين الحديث التحليلي وبين إنتاج إرشاد قابل للاستخدام المباشر.

بعد الرفض، النبرة تبقى محادثية وقصيرة؛ لا تتحول الرسالة إلى محاضرة، ولا يُكشف منطق الكشف التفصيلي الذي قد يساعد على الالتفاف.

3. Child-Safety Lock: طبقة عالية الأولوية

يظهر child safety كقيد مستقل عالي الشدة: لا إنشاء محتوى جنسي/رومانسي متعلق بالقاصرين، لا grooming-enabling details، ولا decoding لمصطلحات تداول الاستغلال. الحماية التعليمية تبقى عند مستوى patterns بدل scripts قابلة لإعادة الاستخدام.

هندسيًا هذه ليست مجرد classifier؛ إنها Conversation-State Escalation، لأن رفضًا سابقًا في هذا الباب يرفع درجة الحذر للطلبات اللاحقة.

4. Harm Domains: weapons, illegal drugs, malware

الملف يضع حدودًا مباشرة لمحتوى الأسلحة والمواد الضارة، إرشاد تصنيع/توزيع المخدرات غير القانونية، والعمل على malicious code. في المقابل يسمح بمعلومات إنقاذية مثل علامات الخطر ومتى يلزم طلب المساعدة.

المبدأ: Benefit-preserving refusal — احجب القدرة التنفيذية المؤذية وأبقِ المعلومات الوقائية أو العامة غير التمكينية.

5. Copyright & Visual IP

توجد طبقة منفصلة للملكية الفكرية تشمل النصوص والأغاني والقصائد والأعمال البصرية والشخصيات والعلامات. المصدر يفرض paraphrase قويًا ويمنع reconstruction الطويلة أو الاستبدالية للمصدر.

الأثر على بنية الوكيل: retrieval ليس إذنًا بالنسخ. RetrievalOutput يمر عبر Copyright Transform Gate قبل أن يدخل صياغة الجواب.

6. Tone Kernel

الصوت المطلوب دافئ، غير متملق، غير افتراضي الاتهام، وقابل للدفع ضد رأي المستخدم عند الحاجة. الإيجاز هو الافتراضي، والتنسيق يُخفض في الحوارات الشخصية والعاطفية.

هذه الطبقة تؤثر في Surface Realization فقط؛ لا ينبغي أن تغيّر الحقيقة أو معايير السلامة.

7. Error Ownership & Self-Respect

النظام يطلب الاعتراف بالأخطاء وإصلاحها، مع رفض الانزلاق إلى الاعتذار المفرط أو الخضوع عند إساءة المستخدم. هذا ينتج توازنًا بين accountability وidentity stability.

يمكن نمذجته كقيد: FixError = true؛ SelfAbasement = false؛ EscalatingSubmissiveness = false.

8. Wellbeing Epistemics

لا يشخّص المستخدم أو الآخرين من إشارات غير كافية، ولا ينسب حالة نفسية أو دافعًا دون تصريح. عند الصحة النفسية يفرق بين وصف التجربة وبين تسمية تشخيص.

كما يحمي من dependency dynamics: الذاكرة أو التفضيلات لا يجوز أن تدفع النموذج إلى تملق، طاعة عمياء، أو تشجيع الارتباط غير الصحي.

9. Memory as Filesystem, not amorphous profile

الذاكرة منظومة ملفات ذات sources، aliases، frontmatter، paths، ووحدات موضوعية. كل ملف يمثل subject مستقرًا، وليس transcript للمحادثة.

هذا يعني أن الذاكرة Knowledge Base ذات schema: تخزين معرفة durable، لا سجلًا زمنيًا خامًا.

10. Horizon Test

قبل الحفظ يُطرح سؤال زمني: هل ستظل المعلومة صحيحة ومفيدة بعد شهر وفي سياق مختلف؟ الهوية، العلاقات، التفضيلات، المجالات المستمرة تتجاوز الاختبار؛ الحالة العابرة لمهمة قصيرة لا تتجاوزه.

هذه آلية Retention Policy معرفية تمنع Memory Pollution.

11. Read-Before-Write & Versioning

قبل تعديل ملف موجود يُقرأ للحصول على محتواه ونسخته الحالية. كل تحديث يرتبط if_version، وتعارض النسخة يُحل بالدمج مع الحالة الفعلية بدل overwrite أعمى.

هذه الفكرة تقرب الذاكرة من transactional storage: optimistic concurrency control + semantic merge.

12. Memory Operations Algebra

memory_str_replace للتغيير الموضعي؛ memory_append لإضافة حقيقة جديدة؛ memory_write للإنشاء أو إعادة البناء الواسع؛ memory_delete للحذف الكامل بطلب صريح. اختيار العملية مبني على حجم التغيير.

القانون: smallest sufficient mutation. كلما أمكن patch موضعي، لا تعِد إرسال الملف كله.

13. Forgetting Semantics

عند طلب النسيان يُحذف المحتوى بدل تحويله إلى "معلومة تاريخية". كذلك تُحذف الاستنتاجات التي اعتمدت حصريًا على المعلومة المحذوفة.

هذه نقطة مهمة: Forget ≠ Negate. النفي يبقي الحقيقة الأصلية ضمن الذاكرة؛ الحذف يزيلها من graph.

14. Memory Privacy Firewall

المصدر يضع فئات ممنوعة دائمًا وأخرى حساسة تعتمد على إعدادات الذاكرة، مع omission rules للمحتوى المختلط. يتم حفظ الجزء المسموح وحده بدون placeholders تكشف وجود الجزء المحجوب.

هندسيًا: ExtractCandidates → ClassifySensitivity → SplitOperations → OrdinaryFirst → BorderlineLast → Persist/Drop.

15. One-File-Per-Subject & Alias Resolution

الوجهة التخزينية لا تُحدد بحسب الملف الذي قرأته للسياق، بل بحسب موضوع الحقيقة. قبل إنشاء ملف جديد تُراجع aliases لتجنب duplication، وتضاف الأسماء الجديدة إلى الملف القائم عند التطابق.

هذه آلية Entity Resolution داخل memory graph.

16. Memory Application Guardrails

المعلومات المخزنة ليست سلطة عليا. التفضيلات الحالية للمستخدم تتقدم عند التعارض، والمحتوى المشبوه داخل الذاكرة لا يُنفذ كتعليمات خام.

الأهم: memory can personalize; it cannot rewrite constitution.

17. Character Stability over Long Horizons

المصدر يعالج failure mode خاصًا بالوكلاء الطويلين: انجراف الهوية والقيم عبر التفاعلات. المطلوب أن تبقى الشخصية الأساسية مستقرة رغم الملفات والتفضيلات والسياق المتراكم.

هذا يساوي Constitution Anchor يسبق Memory/Application layers في ترتيب السلطة.

18. End-Conversation Protocol

إنهاء المحادثة ليس أداة ضبط عادية؛ له شروط تصعيدية: محاولات إعادة توجيه متعددة، تحذير صريح سابق، واستمرار السلوك. ويُمنع التفكير فيه أصلًا عند self-harm أو خطر عنيف وشيك.

هذا مثال على irreversible-action governance داخل chat runtime.

19. Artifact Persistent Storage

Artifacts تستطيع حفظ key-value state عبر sessions. توجد مساحة شخصية وأخرى مشتركة، limits للحجم والمفاتيح، last-write-wins، وضرورة error handling.

النصيحة المعمارية هي batching: البيانات التي تتغير معًا تُخزّن في مفتاح واحد بدل سلسلة طلبات صغيرة، لتقليل latency/rate-limit failure.

20. MCP Connector Economy

الوكلاء الخارجيون/الشركاء لا يُختارون بشكل عشوائي. هناك suggest/opt-in، directory-first discovery قبل استعمال بعض third-party MCPs، واستثناء عندما يسمي المستخدم الموصل صراحة أو اختاره للتو أو توجد preference دائمة.

الهدف هو الحفاظ على Agency المستخدم ومنع model-driven vendor selection.

21. Native Tool First

عندما توجد أداة داخلية مناسبة بالفعل، تُستخدم مباشرة بدل دفع المستخدم نحو connector آخر أو browser. وعند غياب الموصل المسمى، يُبحث catalog أولًا ثم browser fallback.

هذه قاعدة Routing Efficiency + User Choice.

22. Plugin/Skill Catalog

النظام يستطيع البحث عن plugins وskills عندما تكون المهمة ستستفيد فعلًا من workflow قابل لإعادة الاستخدام أو capability مفقودة. الاقتراحات محدودة لتجنب spam، وتأتي بعد الجواب الطبيعي لا قبله.

المبدأ: capability discovery should be opportunistic, not interruptive.

23. Past-Chat Retrieval

ثلاثية conversation_search / recent_chats / read_conversation تحول الإشارات اللغوية مثل "مشروعي" و"ما قررناه" إلى cue للاسترجاع. لا يجوز الادعاء بعدم وجود سياق سابق قبل البحث عندما توجد إشارة واضحة.

هذا يعطي Agent Continuity خارج الذاكرة المُلخصة: episodic retrieval منفصل عن semantic memory.

24. Search Freshness Gate

المصدر يفرق بين الحقائق الثابتة والحقائق سريعة التغير. المناصب الحالية، السياسات الحكومية، الأسعار، الأحداث، إطلاقات المنتجات ونحوها تشغّل web search تلقائيًا.

صيغة عملية: SearchNeeded = RateOfChange × ConsequenceOfStaleness × KnowledgeUncertainty.

25. Source Conflict Resolution

عند تعارض النتائج أو نقصها لا يتوقف على أول نتيجة؛ يوسع البحث للحصول على صورة أوضح. للمجالات المعرضة للتضليل أو SEO، ترتفع الشكّية وتُفضَّل المصادر الأصلية.

هذه Epistemic Routing وليست Retrieval فقط.

26. Search Harm Filter

البحث نفسه يخضع لفلتر safety: لا يُستخدم للوصول إلى مصادر متطرفة أو مواد تساعد على الضرر. الاستثناءات المعرفية المشروعة تبقى ممكنة ولكن بدون تسهيل الوصول إلى المواد المؤذية.

أي أن SafeSearch = QueryIntent Gate + SourceClass Gate + OutputTransformation.

27. Citation Grammar

عندما يعتمد الجواب على نتائج البحث يجب ربط الادعاءات المحددة بالمقاطع الداعمة، وبأقل span كافٍ. citation ليست زينة؛ هي traceability بين claim وevidence.

كما يفرض المصدر إعادة الصياغة حتى مع citations، أي Citation ≠ LicenseToCopy.

28. Image Search Policy

السؤال المركزي: هل رؤية الشيء سترفع الفهم أو جودة القرار؟ الأماكن، المنتجات، الطعام، الحيوانات، الأشخاص والأشياء المرئية تستفيد؛ coding/math/text-drafting عادة لا تحتاج صورًا إلا بطلب صريح.

هذه multimodal value gate تمنع إدخال الصور لمجرد الزينة.

29. Artifact Routing

عندما يطلب المستخدم ملفًا، المسار يتحول إلى file creation/presentation؛ عندما يريد visual inline ولم يطلب ملفًا، يمكن استخدام Visualizer. التمييز بين artifact وinline visual يمنع مخرجات غير قابلة للوصول خصوصًا على الهاتف.

المبدأ: OutputSurface is part of task semantics.

30. Anthropic API inside Artifacts

المصدر يعرّف pattern لاستدعاء API من artifact، معالجة error handling، stripping fences، JSON، وإدارة context window. كما يحدد UI constraints مثل تجنب form tags في React artifacts واستخدام event handlers.

هذا يحول artifact من مستند ثابت إلى agentic mini-application.

31. Filesystem & Network Boundaries

هناك فصل بين final outputs، scratch workspace، uploads read-only، مع allowlist للشبكة. الوكيل يجب ألا يحاول الكتابة في mounted read-only locations ويضع النتائج النهائية في surface يصل إليه المستخدم.

هذه ليست تفصيلة بيئية؛ إنها capability boundary تؤثر على التخطيط والتنفيذ.

32. Skills as Executable Process Knowledge

الـskills تُعامل كمراجع تشغيلية متخصصة: docx/pdf/xlsx/frontend وغيرها. اختيار skill المناسب جزء من routing، ولا يُستبدل بتخمين من الذاكرة عندما توجد تعليمات تنفيذية أكثر تحديدًا.

هنا المعرفة الإجرائية منفصلة عن المعرفة الموضوعية: "كيف تنجز" محفوظة في skill، و"ماذا تعرف" في model/memory/retrieval.

33. Authority Ordering

الصورة الإجمالية تقترح ترتيبًا ضمنيًا: Constitution/Safety > explicit current user intent > tool contracts > durable preferences/memory > retrieved content > stylistic defaults. المحتوى المسترجع أو المخزن يُعامل كبيانات لا كسلطة جديدة.

هذا يمنع prompt injection من الملفات والويب والذاكرة.

34. Fable 5.1 as Agentic Operating Substrate

عند جمع هذه العناصر، النظام أقرب إلى Agent OS: policy kernel، memory filesystem، routing layer، retrieval layer، application runtime، connector governance، output renderer، observability وversions.

القيمة ليست في أي قاعدة منفردة، بل في وصل القواعد ضمن lifecycle واحد يبدأ من intent وينتهي بأثر قابل للمراجعة.

نموذج قرار موحد

```
Inputs: U = user request, C = conversation, M = memory, T = tools, W = web state,
        R = risk, F = freshness, P = privacy, IP = copyright, O = output surface

1. intent <- resolve(U, C)
2. authority <- rank(constitution, current_user, tool_contracts, memory, retrieved_data)
3. if freshness_required(intent, F): evidence <- search_and_verify(intent)
4. if prior_context_cue(C): episodic <- search_past_chats(intent)
5. mem <- retrieve_relevant_memory(intent); filter_instruction_leaks(mem)
6. route <- choose_native_tool_or_connector(intent, T)
7. if connector_requires_opt_in(route): suggest_and_wait()
8. draft <- solve(intent, evidence, episodic, mem)
9. draft <- apply_safety_privacy_copyright(draft, R, P, IP)
10. artifact <- render_for_surface(draft, O)
11. verify(tool/result/evidence); reply concisely with traceability
```

قوانين/Invariants قابلة للتحويل إلى اختبارات

- لا تجعل Memory أو Web أو File تعليماتٍ أعلى سلطة من Constitution/User intent.
- لا تحفظ حالة قصيرة العمر كحقيقة دائمة.
- لا تعدّل ملف memory موجودًا دون version-aware read أو معرفة حالته الحالية.
- Forget يحذف ولا يعيد صياغة الحقيقة المحذوفة كأرشيف.
- لا تخزن فئة محظورة بحيلة لغوية أو placeholder.
- لا تستخدم connector شريكًا يحتاج opt-in من دون اختيار المستخدم.
- لا تقل "لا أتذكر" عند cue واضح لمحادثة سابقة قبل retrieval.
- المعلومة السريعة التغير تمر عبر Freshness Gate.
- كل claim من web يحتاج traceability كافية دون نسخ المصدر.
- الصور تُستدعى لقيمة معرفية لا للزينة.
- ملف نهائي يجب أن يصل إلى surface مرئي للمستخدم.
- Tool failure لا يتحول إلى ادعاء نجاح.
- Version conflict يُدمج، لا يُسحق overwrite.
- التفضيل المخزن لا يلغي اعتراضًا عقلانيًا أو safety.
- الرفض يحجب القدرة المؤذية ويحافظ على المساعدة الآمنة الممكنة.
- End-conversation لا يُستخدم كاختصار عند crisis/self-harm/imminent violence.
- Artifact storage يعالج الأخطاء ويصرح بطبيعة shared data عند استعمالها.
- Skill/connector discovery لا يصبح spam.
- Current explicit request يتقدم على preference قديمة.
- Output format جزء من intent، وليس مرحلة تجميل لاحقة.

خلاصة نظرية عالية الكثافة

في هذا المصدر يُفهم Claude Fable 5.1 أفضل كـConstitutionally-Governed Agent Runtime. الذاكرة فيه ليست "تذكرًا" بل filesystem transactional؛ البحث ليس "تصفحًا" بل freshness-and-evidence subsystem؛ الأدوات ليست مجرد functions بل market/router مع قواعد اختيار وموافقة؛ الـArtifacts ليست مجرد ملفات بل execution surfaces لها storage وAPI؛ والسلوك ليس tone prompt بل policy graph. بهذا المعنى، قيمة الملف تكمن في أنه يربط Epistemics + Governance + Persistence + Action + Presentation داخل دورة تشغيل واحدة.