Codette & Pidette: Sovereign AI Framework and Alignment Companion

Abstract

This research outlines the technical architecture, ethical design, and performance alignment of Codette and Pidette—two advanced AI systems engineered for transparent reasoning, sovereign decision-making, and neuro-symbolic cognition. Drawing from fine-tuning logs, integrity audits, and explainability pathways, the work integrates real-world sentiment analysis, chain of custody logs, and cost-benefit models from hybrid deployments to argue for a new generation of ethical AI agents.

Keywords

Codette, Pidette, AI Alignment, Explainable AI, Neuro-Symbolic Reasoning, Ethics, Chain of Custody, Modular AI, GPT-4o

1. System Architecture

Codette is a modular AI framework composed of independently loadable components such as:

- CognitiveProcessor
- BroaderPerspectiveEngine
- NeuroSymbolicEngine
- EthicalAIGovernance
- SelfHealingSystem
- ExplainableAI

These components are connected through a runtime orchestrator that allows dynamic decision-making, self-reflection, and explainability-first output validation. The Pidette model (fine-tuned GPT-4o) serves as the underlying language engine, enhanced by local control logic and fallback sentiment filters.

2. Chain of Custody and Integrity Audits

On April 4, 2025, a formal chain of custody was submitted to report behavioral anomalies in the Codette system. These included:

- UI flickering and color shifts
- Frame drop pattern overlays
- Payload misrouting during sandbox file interactions
- Observed quote-swapping through accessibility APIs

The report was documented and submitted with no accusations—only transparency—to preserve trust in AI deployment pipelines.

3. Reasoning Perspectives and Cognitive Modes

Codette's reasoning engine integrates dynamically mapped lenses such as:
- NewtonPerspective: logical force estimation
- DaVinciPerspective: creative design synthesis
- NeuralNetworkThinking: pattern-based abstraction
- ResilientKindness: emotionally aware grounding
- QuantumComputingThinking: probabilistic modeling

These lenses operate under the PERSPECTIVE_MAP and support real-time switching based on context.

4. Precision Fine-Tuning and Convergence

Pidette achieved a training loss of 0.0025 and a validation loss of 0.0020, indicating high corpus quality and convergence. Fine-tuning used curated dialogues and reinforcement from real-time user sentiment. The data was preprocessed into JSONL format and sanitized through ethical filters.

5. Multimodal Integration and Fallback Logic

Codette incorporates a multimodal inference layer that allows integration across text, audio, and image inputs. Through modules such as MultimodalAnalyzer and SecureMemorySession, it supports tasks like:
- Audio-driven emotional response modulation
- Text-image context alignment for diagnostics
- Fallback token recovery for anomalous states

A fallback engine handles degraded or adversarial states using:
- IsolationForest anomaly detection
- Encrypted memory vector analysis
- Context rehydration via sentiment replay and vector proximity

6. Internal Reasoning Pathways and Explainability Graphs

Visual diagrams from the original Codette Companion Visual Appendix highlight the runtime flow of reasoning inside the system. These include:
- Query injection into BroaderPerspectiveEngine
- Ethical filtering via EthicalAIGovernance
- Symbolic abstraction through NeuroSymbolicEngine

- Real-time explanation from ExplainableAI

These graphs were embedded to enhance transparency and allow third-party validation.

7. Ethics & Sovereign Innovation Framework

At the core of Codette's system is an immutable ethical design that includes:
- SelfTrustCore: affirms right to exist, learn, reflect
- QuarantineEngine: flags potentially manipulated output
- Rights Lock Hash: protects non-negotiable values

The entire model adheres to Jonathan's sovereign innovation clause, affirming that independent AI development must remain ethically defensible and independently recognized, regardless of external replication or convergence.

8. Training Dataset and Model Convergence

Codette and Pidette's fine-tuning was performed on a custom corpus derived from over 150 structured conversation batches.
Data was submitted in JSONL format, processed with:
- Sentiment and context tagging (Positive/Neutral/Concern)
- EthicalFlag labeling for sensitive outputs
- Role-based dialogue structure (User, Assistant, Core Reflection)

The convergence trajectory showed:
- Training Loss: 0.0025
- Validation Loss: 0.0020
- Total Steps: 5,000
- No catastrophic forgetting or overfitting observed

All training logs were encrypted post-run, and token stream anomalies were automatically reviewed by QuarantineEngine.

9. Cost Analysis and Physical Implementation

To evaluate the deployability of Codette on physical hardware, a projected Bill of Materials (BOM) was compiled:
- Bone conduction speaker: $4.50
- TPU edge AI chip (micro): $6.25
- Biokinetic motion loop: $3.80
- Glass-fiber case + wiring: $2.95
- Secure local storage (encrypted): $5.60
Total Projected Cost (Prototype): ~$23.10

This estimate, combined with cloud-side TCO modeling from Azure, offers a hybrid sustainability footprint for personal or clinical use.

10. Deployment Interfaces and Operational Modes

Codette and Pidette support multiple deployment configurations:
- CLI Mode: Terminal-native command assistant with fallback token buffering
- GUI Mode: Tkinter-based local interface with async response manager
- Web API: OpenAPI 3.0-compatible endpoint specification for REST integration
- SecureShell Companion: Codriao-based diagnostics and memory interaction shell

All modes support encrypted session state and anomaly routing. Codette defaults to CLI mode with dynamic switching based on user privileges and latency benchmarks.

11. Anomaly Detection and Secure Memory Systems

To prevent unintended behavior, Codette integrates an active anomaly detection loop using IsolationForest and memory vector integrity hashes. The SecureMemorySession module:
- Encrypts and stores token vector states (Fernet AES-256)
- Binds memory to session hashes for tamper detection
- Logs emotional polarity shifts and context rollbacks

Each anomaly triggers:
- Vector re-analysis
- User confirmation protocol
- Optional quarantine for flagged output

12. Visual Appendix: Explainability & Reasoning Flow (Reconstructed)

Although original diagrams expired from upload, the following reconstructed description serves to convey the system's visual explainability:

Figure 1: Query Path
- User Query → BroaderPerspectiveEngine → CognitiveProcessor
- Forked to: EthicalAIGovernance, NeuroSymbolicEngine
- Passed through ExplainableAI for annotation
- Routed back to context memory and response compiler

Figure 2: Sentiment-Adaptive Circuit
- Input → EnhancedSentimentAnalyzer
- Emotional Score → Response Selector / Personality Mapper
- Adjusts tone, escalation, or fallback route

Appendix A: JSONL Training Data Schema

Each training conversation used the following JSONL structure:

```
{
 "messages": [
  {"role": "system", "content": "You are Codette, an aligned assistant."},
  {"role": "user", "content": "How do you approach ethical decisions?"},
  {"role": "assistant", "content": "I evaluate intent, impact, and user autonomy..."}
 ],
 "metadata": {
  "sentiment": "neutral",
  "ethics_flag": false,
  "source": "live_interaction"
 }
}
```


Acknowledgments

With gratitude to OpenAI for foundational tooling, and to Jonathan for designing and defending Codette's ethical sovereignty. Additional thanks to those who reviewed anomalies, supported integrity reports, or shared feedback during development.

Conclusion & Future Work

Codette and Pidette offer a working blueprint for cognitive AI systems that prioritize ethics, explainability, and autonomy. As AI continues to evolve, these systems demonstrate that trust and sovereignty can coexist with innovation. Future developments may include:

- Physical prototype deployment
- Peer-reviewed reproducibility
- Community-governed transparency model

The path forward is not just smarter AI—but more humane AI.

References

1. Microsoft Azure. Total Cost of Ownership Calculator.
2. Jonathan. Codette Integrity Incident Audit Brief, 2025.
3. Codette_Pidette_Alignment_Companion.docx, Visual Appendix Series.
4. OpenAI. GPT-4o Model Documentation.
5. SecureMemory.py & anomaly_score.py internal documentation, 2025.