CHAMELEON

Generative AI Compression Engine

AI-powered text compression using GPT-2 and arithmetic coding, plus intelligent file comparison for validation

The Compression Ceiling

Traditional Limits

Legacy algorithms like gzip and bzip2 operate purely on byte-level patterns. They're blind to semantic structure and language context, hitting hard performance ceilings with natural language data.

This pattern-only approach leaves significant compression potential untapped, especially for text-heavy datasets.

Verification Gap

Developers lack reliable tools to verify compressed-decompressed file integrity across formats. Comparing TXT, PDF, and JSON outputs manually is error-prone and time-consuming.

The solution? Context-aware compression paired with intelligent verification.

CHAMELEON Architecture

Autoregressive Modeling

DistilGPT-2 predicts next-token probabilities, understanding language context at a semantic level.

Probability Quantization

2^{24} frequency tables map probability distributions to compact integer representations.

Arithmetic Coding

Custom 64-bit implementation achieves optimal entropy efficiency with binary-safe .bin format.

Intelligent File Comparison

01

Universal Input

Accept pasted text or uploaded files in multiple formats including PDF, TXT, and JSON.

02

Smart Extraction

Advanced PDF parsing and automatic encoding detection ensure accurate text representation.

03

Precision Scoring

Generate detailed similarity metrics to verify perfect round-trip decompression.

Technology Stack

Core Framework

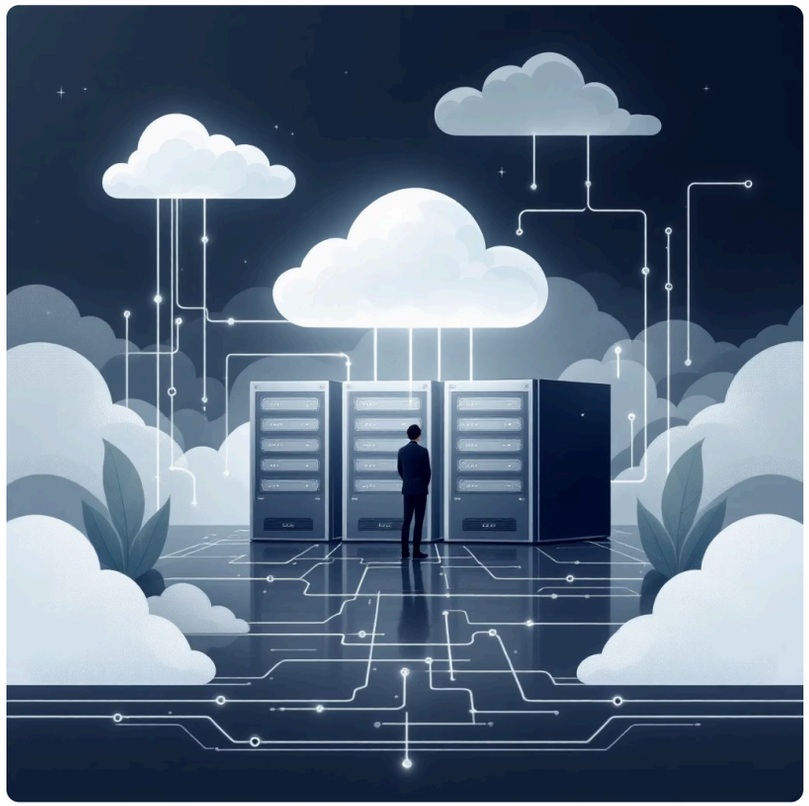
- Python 3.10+ runtime
- PyTorch for model inference
- HuggingFace Transformers
- Custom 64-bit arithmetic coding

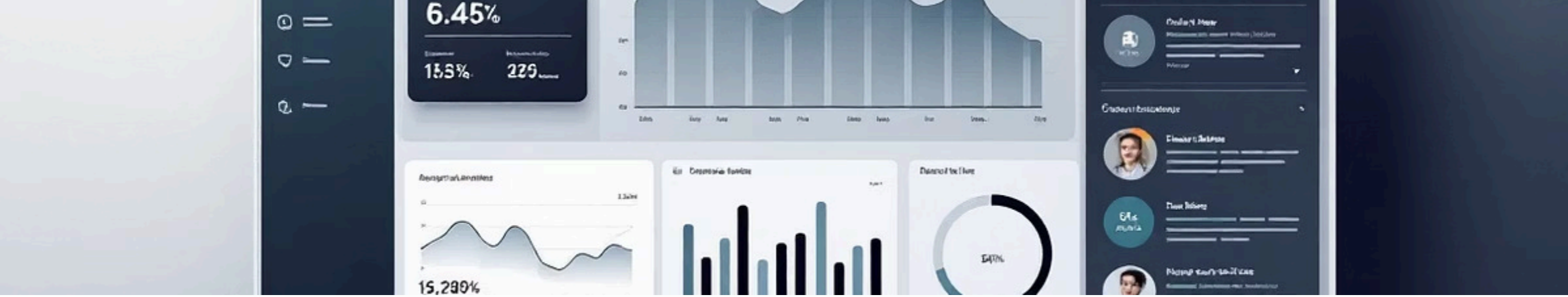
Interface & Processing

- Streamlit for interactive UI
- pdfplumber for extraction
- chardet for encoding detection

Deployment Ready

Deploy on Streamlit Cloud, **Render.com**, or GPU-accelerated platforms like RunPod and Azure for production workloads.





Live Demo Capabilities

1 Compression Pipeline

Upload text files, compress using GPT-2 predictions, and download binary-safe .bin archives.

2 Decompression & Validation

Upload .bin files to reconstruct original text with byte-perfect accuracy.

3 Performance Metrics

View real-time statistics including original size, compressed size, and compression ratio.

4 Cross-Format Comparison

Compare outputs across PDF, TXT, and other formats with 100% match verification.

Real-World Impact

Efficient Archival

Dramatically reduce storage costs for text-heavy datasets while maintaining semantic integrity.

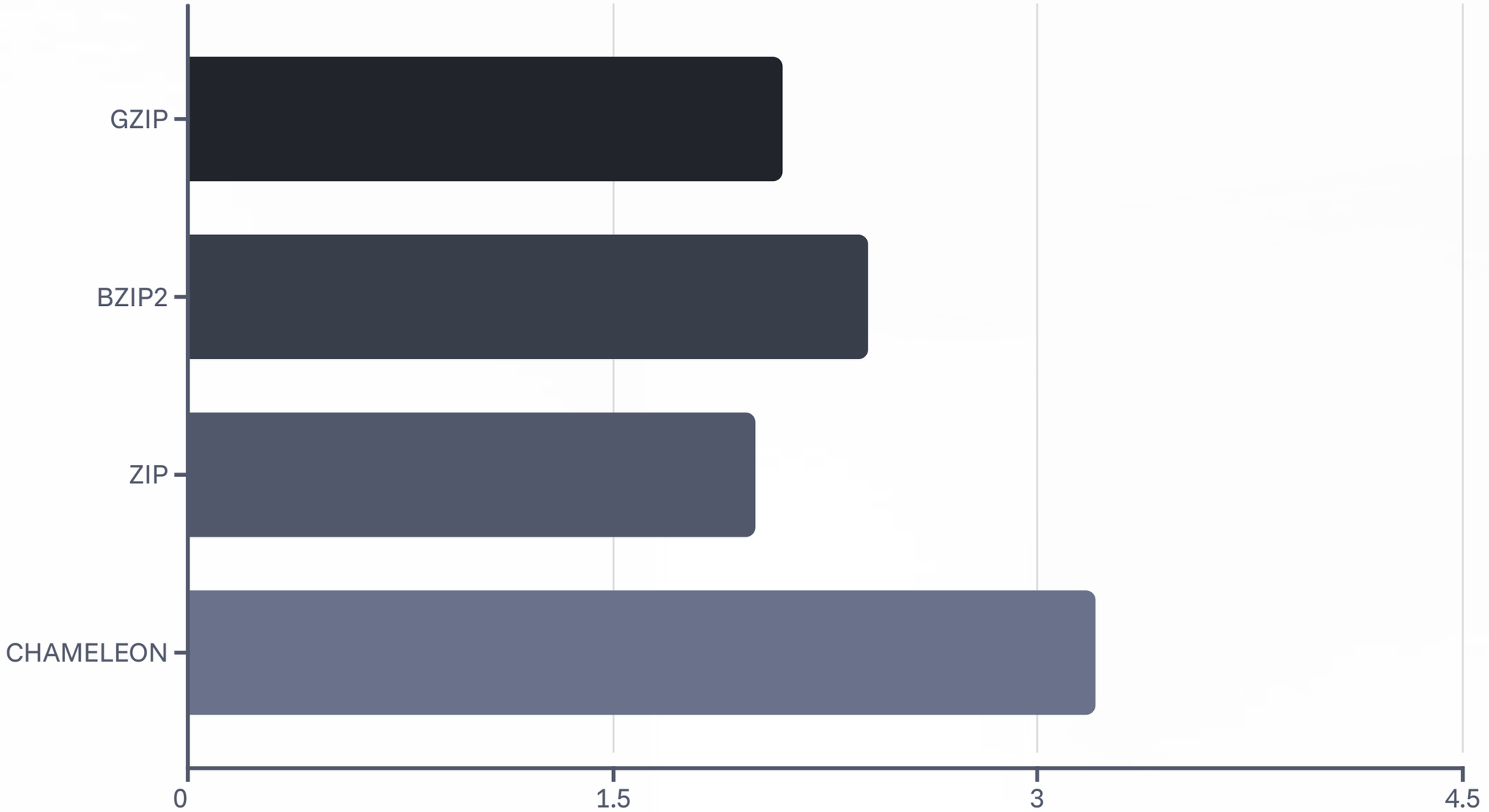
ML Pipeline Integration

Seamlessly integrate model-aware compression into research and production ML workflows.

Format Flexibility

Cross-format comparison tools enable reliable validation across PDF, TXT, and JSON.

Compression Performance



CHAMELEON achieves superior compression ratios on natural language text by leveraging semantic understanding. Traditional algorithms plateau around 2.0-2.4x compression, while context-aware modeling pushes beyond 3.0x for language-heavy datasets.

Future Roadmap

Q1 2024

GPU-accelerated batch encoding and longer-context LLM support

1

Q2 2024

RESTful API launch and Chrome extension for webpage compression

2

Q3 2024

Cloud dashboard with corpus-level adaptive models

3

2025+

Multimodal expansion to images and audio using generative priors

4

Swap GPT-2 for LLaMA or Mistral to unlock even higher compression gains. Add encryption layer for secure compression format.

Thank You

Ready to transform text compression with context-aware AI