DrugTrial

AI-Powered Clinical Trial Automation Platform

From Protocol Upload to Patient Recruitment
Built-In Drug Safety Modeling • Local LLMs • Zero Data Leakage

Confidential • February 2026

Table of Contents

1. Executive Summary

DrugTrial is a multi-agent AI platform that automates the end-to-end clinical trial lifecycle for US healthcare — from FDA document processing and patient eligibility screening to in-silico drug analysis and principal investigator workflows.

It replaces months of manual work with **13 intelligent, autonomous agents** that handle regulatory compliance, patient matching, literature research, and safety evaluation — all running on local infrastructure with **zero data leaving your network**.

13	$0	4
AI Agents	API Costs	Intelligence Subsystems

Target Market: US Pharmaceutical companies, Contract Research Organizations (CROs), Biotech startups, Academic Medical Centers.

2. The Problem

US Clinical Trials Are Broken

$2.6B
Avg Cost to Market Per Drug

80%
Trials Miss Enrollment Targets

$8M
Cost Per Day of Trial Delay

Pain Point	Impact
80% of trials fail to meet enrollment deadlines	$8M+ per day in delays per Phase III trial
Manual FDA form processing takes weeks	High error rates; ~30% FDA form rejection rate
Patient screening is slow & biased	Coordinators spend 15-20 hrs/week reviewing charts
No centralized intelligence	Literature, safety, eligibility are siloed processes
HIPAA compliance is a guessing game	PII exposure risks during data sharing across teams
Drug-Drug Interaction checks are reactive	Safety signals caught late, post-enrollment

The root cause: Clinical trials are drowning in manual, disconnected, error-prone processes that AI can automate end-to-end.

Metric	Value
Average clinical trial duration	6-7 years
Screen failure rate	25-50% of enrolled patients
FDA form rejection rate	~30% due to incomplete/incorrect submissions
Drugs entering Phase I that receive approval	Only 12%
Active trials on ClinicalTrials.gov	48,000+

3. Our Solution

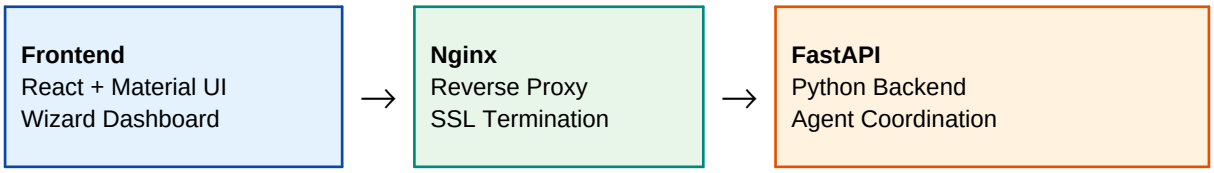
DrugTrial: An Autonomous Multi-Agent AI Platform

DrugTrial is a **full-stack intelligent system** that deploys specialized AI agents working in concert — like a virtual clinical operations team:

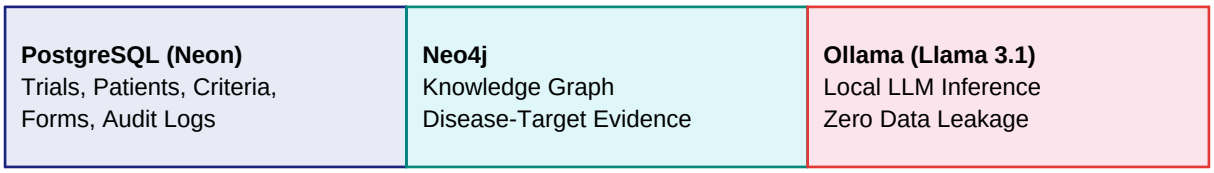
What We Automate	Traditional	DrugTrial
FDA Form Processing	Manual data entry (weeks)	AI extracts FDA 1571/1572 in minutes
Patient Eligibility	Chart review (hours/patient)	Rule-based + NLP matching in seconds
Protocol Analysis	Manual criteria extraction (days)	NLP parses criteria automatically
Literature Review	Reading dozens of papers (weeks)	PubMed agent + knowledge graph (minutes)
Drug Safety (In Silico)	Separate toxicology studies (months)	RDKit toxicity + PK/PD simulation
DDI Screening	Pharmacist review (hours)	Automated drug-drug interaction checking
HIPAA Compliance	Manual de-identification	AI-powered PII detection + pseudonymization
PI Coordination	Email chains & paper (weeks)	Digital submission, review, approval

4. System Architecture

High-Level Architecture



Data Layer



Technology Stack

Layer	Technology	Purpose
Frontend	React 18, Material UI 5, Recharts	Interactive clinical trial dashboard
AI Engine	Python, FastAPI, LangChain	Multi-agent orchestration & NLP
NLP/ML	SciSpaCy, UMLS Linker, NegSpaCy	Biomedical entity extraction
LLM	Ollama + Llama 3.1 (Local)	Zero-hallucination protocol analysis
Chemistry	RDKit, PubChem	Molecular descriptors, toxicity prediction
Databases	PostgreSQL (Neon), Neo4j	Relational data + knowledge graph
Bio APIs	PubMed, HGNC, UniProt	Literature mining & validation
Deployment	Docker Compose + NVIDIA GPU	Multi-container orchestration
Security	SHA-256 audit chain, DeID Agent	HIPAA-aligned, tamper-evident

5. The Agentic Architecture

DrugTrial is not a monolithic application — it is a **network of autonomous AI agents**, each with specific domain expertise, that collaborate through an event-driven orchestration layer.

What Makes DrugTrial "Agentic"

- **Event-Driven Orchestration:** When a trial is created, a TRIAL_CREATED event fires. The Orchestrator subscribes and autonomously decides what to do — no human intervention.
- **Autonomous Planning:** The Orchestrator inspects trial metadata (disease, drug, phase) and generates an execution plan using rule-based logic.
- **Parallel Execution:** LTAA and InSilico pipelines execute concurrently via asyncio.gather, each involving multiple sub-agents working in sequence.
- **Specialist Agents:** Each agent has dedicated data sources — FDAProcessor uses pdfplumber + SciSpaCy + LLM; LTAAAgent uses PubMed + HGNC + UniProt + Neo4j.
- **Cryptographic Accountability:** Every agent action is logged to a hash-chained audit trail. Each entry contains SHA-256 of previous entry, creating a tamper-evident chain.

Four Intelligence Subsystems

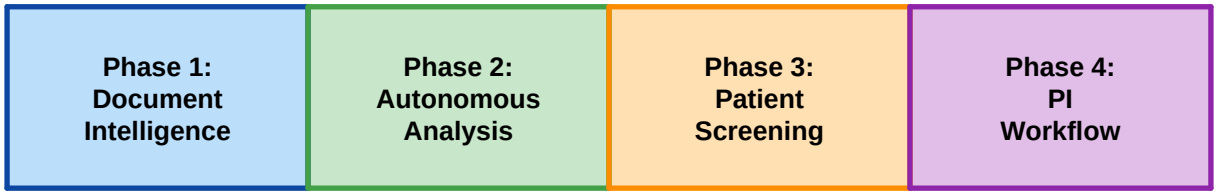
Document Intelligence FDAProcessor ProtocolRuleAgent	**Research Intelligence** LTAAAgent (PubMed, HGNC, UniProt, Neo4j)
InSilico Modeling DrugExtraction, ChemicalResolver, Toxicity, DDI, MolTarget, PK/PD	**Patient Intelligence** EligibilityMatcher DeIDAgent, MedicalNLPAgent

6. Agent Profiles

Agent	Role	Key Capability
Trial Orchestrator	Autonomous planner	Subscribes to TRIAL_CREATED, plans which agents to invoke, executes in parallel via asyncio.gather
FDA Processor	Regulatory document parser	PDF -> Table Extraction (pdfplumber) -> NER (SciSpaCy+UMLS) -> LLM Refinement -> Validation
Protocol Rule Agent	Criteria extraction	PDF OCR -> Section Detection -> Criterion Splitting -> Entity Extraction -> LLM Normalization
Eligibility Matcher	Patient screening	15+ category-specific rule engines with weighted confidence scoring and compound logic
LTAA Agent	Literature & target discovery	PubMed Search -> NER -> Bio-Validation (UniProt/HGNC) -> Neo4j Knowledge Graph -> Ranked Targets
De-ID Agent	HIPAA de-identification	Pseudonymization (SHA-256), age generalization, string masking, PII vault separation
Drug Extraction Agent	Drug identification	LLM-guided chunking to extract drug names, dosages, prohibited medications
Toxicity Agent	Safety assessment	RDKit Lipinski Rule of 5: MW, LogP, H-bond donors/acceptors, TPSA, QED
DDI Agent	Interaction detection	Rule-based pharmacological database for drug-drug interaction screening
Molecular Target Agent	Target identification	SciSpaCy + UMLS semantic types for protein/gene/pathway extraction
PK/PD Simulator	Pharmacokinetic modeling	1-compartment oral model: Cmax, half-life, steady-state, concentration curves
Chat Agent	Natural language Q&A;	LLM with trial context injection for interactive trial queries
Medical NLP Agent	Text understanding	SciSpaCy + UMLS for diseases, drugs, procedures, anatomy extraction

7. Core Workflow & Data Flow

End-to-End Trial Processing — 4 Phases



Phase 1 — Document Intelligence

Upload protocol PDF. FDA Processor extracts Form 1571/1572. Protocol Rule Agent parses eligibility criteria. LTAA + InSilico analyses begin in background.

Phase 2 — Autonomous Analysis (Event-Driven)

EventBus fires TRIAL_CREATED. Orchestrator plans analysis. LTAA queries PubMed and builds Neo4j knowledge graph. InSilico suite runs in parallel (drug extraction, chemical resolution, toxicity, DDI, molecular targets, PK/PD).

Phase 3 — Patient Screening

User triggers batch eligibility check. Eligibility Matcher evaluates all patients against extracted criteria with confidence scores and per-criterion reasoning.

Phase 4 — PI Workflow

Organization sends trial + eligible patients to Principal Investigator. PI reviews, approves/rejects individual patients, signs documents.

8. Security & HIPAA Compliance

Control	Implementation
Data De-Identification	DeID Agent: pseudonymization (SHA-256), age generalization, PII masking
PII Vault	Separated encrypted PII storage (patient_vault table)
Audit Trail	Every action logged: timestamp, agent, target, status, details, hash chain
Privacy Scanning	Dedicated privacy router scans for data exposure
Access Control	JWT + Role-Based Access Control (Admin, User, PI, System Admin)
Transport Security	Helmet.js headers, CORS whitelisting, rate limiting
Document Integrity	SHA-256 hashing of all uploaded documents
On-Premise LLM	All inference via local Ollama — no cloud API calls for sensitive data

Role-Based Access Control

Role	Permissions
System Admin	Full system access — manage all organizations, users, configuration
Organization Admin	Manage org trials, users, documents; trigger analysis pipelines
Organization User	View trials, run screening, access results
Principal Investigator	Review submissions, approve/reject patients, sign documents

9. Database Design

Dual-Database Architecture

PostgreSQL (Neon) — Primary Relational Storage

Tables	Purpose
patients, patient_vault	De-identified demographics + separated PII vault
conditions, medications, observations	Clinical data linked via pseudonymized IDs
allergies, immunizations	Patient allergy and vaccination records
clinical_trials, eligibility_criteria	Trial definitions and structured criteria
patient_eligibility	Screening results per patient per trial
fda_documents, fda_form_1571, fda_form_1572	Uploaded protocols and extracted regulatory forms
audit_trail	Hash-chained, tamper-evident action log

Neo4j — Biomedical Knowledge Graph

Disease nodes linked to Target nodes via HAS_EVIDENCE relationships. Each edge carries source citations, page numbers, weights, and timestamps. Targets are ranked using aggregated confidence scores with entity-type weighting (Gene=5x, Protein=4x, Chemical=3x).

10. Key Platform Features

Intelligent Dashboard

Organization-level trial overview with status tracking, real-time analysis progress, patient screening results with confidence scores.

FDA Document Intelligence

PDF upload with automatic text + table extraction, AI-powered form population for FDA 1571/1572, validation scoring, digital signature support.

Protocol Criteria Engine

Visual rule builder with interactive criteria cards, compound logic editor (AND/OR), real-time status with manual override.

Patient Screening Console

Batch screening across hundreds of patients in seconds, per-criterion breakdown, weighted confidence scoring.

In-Silico Drug Analysis Dashboard

Molecular descriptor visualization, Lipinski toxicity analysis, PK/PD concentration curves, drug-drug interaction matrix.

Literature & Target Intelligence

PubMed-powered research, knowledge graph visualization, ranked therapeutic targets with citations, scientific report generation.

Principal Investigator Workflow

Submission management pipeline (draft -> submitted -> review -> approved), patient-level approval/rejection.

Comprehensive Audit Trail

Every action logged with agent attribution, tamper-evident hash chain, exportable compliance reports.

11. Target Market

Primary: US Clinical Research & Pharma

Segment	Market Size	Pain Fit
Contract Research Organizations	$80B+ market	Need automation to reduce trial timelines
Pharmaceutical Companies	4,000+ in the US	Spend $2.6B avg per drug; need efficiency
Academic Medical Centers	400+ conducting trials	Limited staff for growing trial volumes
Biotech Startups	5,000+ in the US	Can't afford large clinical ops teams

Secondary Markets

- FDA-registered clinical trial sponsors
- Hospital-based research programs
- Government-funded clinical research networks (NIH, PCORI)

Market Sizing

Segment	Market Size (2024)	Growth
Global Clinical Trials	$81.5B	6.5% CAGR -> $120B+ by 2030
Clinical Trial Management Systems	$2.2B	12% CAGR
AI in Drug Discovery	$1.5B	30%+ CAGR -> $8B+ by 2030
Patient Recruitment Technology	$2.1B	15% CAGR

12. Business Model

Revenue Stream	Target	Pricing
SaaS Platform	Biotech / CROs	$50K-200K/year per organization (tiered by trial volume)
Enterprise License	Top 20 Pharma	$500K+/year — on-premise deployment with EHR integration
Per-Trial Processing	CROs (pay-as-you-go)	$10K-50K per trial
API Access	CTMS/EDC vendors	Usage-based pricing for integration
Academic Tier	Universities / Research hospitals	Free — community support

Unit Economics Advantages

- **Zero marginal cost for AI inference** — local LLM eliminates per-query charges
- **High switching costs** — once integrated with EHR and trained on institutional protocols, migration is expensive
- **Network effects** — more trials processed = better criteria templates, richer knowledge graph

13. Competitive Advantages

Capability	DrugTrial	Medidata	Veeva	Deep6 AI
Local LLM (no data leakage)	Yes	No	No	No
InSilico drug safety modeling	Yes	No	No	No
Literature intelligence (PubMed)	Yes	No	No	Partial
FDA form auto-extraction	Yes	Manual	Manual	No
AI patient matching	Yes	Yes	Partial	Yes
Knowledge graph	Yes	No	No	No
Cryptographic audit trail	Yes	Yes	Yes	Partial
Zero API costs	Yes	No	No	No
PK/PD simulation	Yes	No	No	No
Open-source LLM	Yes	No	No	No

- **Multi-Agent Architecture**: Each agent is specialized — not a one-size-fits-all LLM wrapper
- **Event-Driven Autonomy**: Agents self-trigger analysis without human prompting
- **Anti-Hallucination Guardrails**: Source text validation on every LLM extraction
- **Dual-Database Intelligence**: Relational data for ops + knowledge graph for discovery
- **HIPAA-Native Design**: De-identification built into the pipeline, not bolted on
- **Full Audit Trail**: Every agent action logged — critical for regulatory submission

14. Key Metrics & Impact

Metric	Before DrugTrial	With DrugTrial
FDA Form Processing	2-3 weeks	Minutes
Patient Screening	15-20 hrs/week	Seconds (batch)
Protocol Criteria Extraction	2-3 days	Under 5 minutes
Literature Review	2-4 weeks	Minutes
Drug Safety Pre-Screening	Months (separate studies)	Instant (in-silico)
Audit Compliance	Manual logging (error-prone)	Automatic (100% coverage)

Year 1 Targets

50+

Trials Processed

8-10

Paying Customers

$1M+

Annual Recurring Revenue

15. Product Roadmap

Phase	Timeline	Key Deliverables
Phase 1 (Complete)	MVP — Now	Multi-agent platform, FDA processing, eligibility screening, in-silico analysis, audit trails, HIPAA de-identification
Phase 2	Q2 2026	EHR/FHIR integration, medical ontology mapping (ICD-10, LOINC, RxNorm), semantic matching, encrypted PII vault
Phase 3	Q3 2026	21 CFR Part 11 compliance, digital signatures, GCP workflow enforcement, expanded DDI database, ML toxicity models
Phase 4	Q4 2026	ClinicalTrials.gov integration, full-text paper access, multi-site support, enhanced PK/PD models
Phase 5	2027	eCTD/FDA gateway submission, real-time EHR streaming, Kubernetes deployment, predictive enrollment analytics

16. Gap Analysis & Future Work

P0 — Critical (Production Blockers)

#	Gap	Effort
15	Fix Birthdate De-ID — year-only (HIPAA Safe Harbor)	0.5 days
14	Encrypted PII Vault — Fernet symmetric encryption, KMS-ready	1-2 days
1	Medical Ontology Mapping — ICD-10, SNOMED-CT, LOINC, RxNorm	3-4 days
6	21 CFR Part 11 — digital signatures, trusted timestamps, RBAC	5-7 days
2	EHR/FHIR Integration — FHIR R4 client, live EHR sync	5-7 days

P1 — High Value (Accuracy & Compliance)

#	Gap	Effort
3	Integrate MedicalNLPAgent into eligibility matching	2 days
4	Semantic similarity for medical term matching (UMLS synonyms)	1-2 days
5	Temporal reasoning for conditions/medications	1-2 days
9	ML-based toxicity models (QED + Tox21 deep learning)	2-5 days
10	Expand DDI database (500+ pairs + OpenFDA API)	2-3 days
8	GCP workflow state machine + deviation tracking	3-4 days
16	Re-identification risk analysis (K-anonymity, L-diversity)	2 days

P2 — Nice-to-Have (Enhanced Capabilities)

#	Gap	Effort
12	ClinicalTrials.gov integration	2 days
13	Full-text paper access (PMC open access)	2 days
11	Enhanced PK/PD modeling (2-compartment, drug-specific params)	2-3 days
7	eCTD/FDA gateway electronic submission	10-15 days

16
Total Gaps

~50
Days to Close All

5
P0 Critical Gaps

17. Why Now?

AI Maturity

SciSpaCy + local LLMs (Llama 3.1) make biomedical NLP production-ready. Tasks that required GPT-4 just 18 months ago now run on a single GPU on-premise.

Regulatory Pressure

FDA's 2024 guidance on AI/ML in drug development signals increasing acceptance. Companies adopting AI early gain a competitive advantage in regulatory interactions.

Cost Crisis

Drug development costs have tripled in 20 years. With $2.6B per drug and 80% enrollment failures, the industry is desperate for automation. $8M/day delay costs create massive willingness to pay.

COVID Catalyst

The pandemic exposed how slow traditional trial infrastructure is. Decentralized and AI-assisted trials are now expected, not aspirational.

Data Availability

EHR adoption at 96% means patient data is digitized and screenable. The infrastructure for automated matching finally exists.

18. Team & Ask

Team Expertise Required

Role	Domain Expertise
CEO / Product	Clinical trial operations, pharma industry relationships
CTO	AI/ML systems, healthcare data, distributed systems
Chief Medical Officer	Clinical research, regulatory affairs, FDA interactions
Head of Engineering	Full-stack development, DevOps, security
Head of Data Science	NLP, knowledge graphs, cheminformatics

Funding: Seed Round

Use of Funds	Allocation
Engineering (EHR integration, compliance, scaling)	50%
Clinical validation studies	20%
Sales & marketing (conferences, partnerships)	15%
Regulatory consulting (21 CFR Part 11 certification)	10%
Operations	5%

Year 1 Key Performance Indicators

Metric	Target
Trials processed	50+
Paying customers	8-10
Annual Recurring Revenue	$1M+
Patient matching accuracy vs manual	>90% concordance
Time from protocol to screening	<30 minutes (vs 2-4 weeks manual)

Thank You

DrugTrial — AI-Powered Clinical Trial Automation

13 Specialized AI Agents • Zero Data Leakage • HIPAA-Aligned

Confidential — February 2026