

ELI LILLY & CO.



AI-Native Product Experiences for B2C Pharma + Digital Health

Agentic Product Design + Architecture Overview : OpenAI @ LillyDirect

Eros Marcello | Sr. Product Design Engineer | 09/2024 - 03/2025

Frameworks Developed:

• Agentic Tooling Architecture • Utterance Calibration Framework • Design System-Driven AI Automation

Executive Summary

PROACTIVE (Ambient)

DOM parsing, intent prediction, auto-scroll, personalized help

REACTIVE (Chat/Voice)

Rx filling, auth flows, FAQ via RAG, caretaker assist

Role

SME for Eli Lilly's OpenAI Partnership & CrewAI Discovery

Scope

GenAI/Agentic AI PoC → Prototype → PROD across org

Focus

LillyDirect platform for Zepbound/Mounjaro weight-loss drugs

Impact

3 multimodal AI-native PROD-ready experiences presented to exec leadership

Agent Tooling Framework

From: 'Agents for Lilly Direct: Enhancing Patient-Centric Pharma Services'

Asset Meta

Type: Architecture Spec
Used by: Engineering, Product
Purpose: Tool orchestration

Knowledge Augmentation

- Drug Interaction Databases
- Insurance Verification APIs
- Pharmacy Network Inventory

Expand agent knowledge with real-time data

Capability Extension

- Calculator APIs (dosage)
- Translation Tools (multilingual)
- Symptom Checker Models

Boost operational effectiveness

Write Actions

- Submit Prescription Requests
- Update Patient Records (EHR)
- Pharmacy Interface Direct

Execute validated operations

Planning Granularity

Sequential: Insurance check → Generate response | Parallel: Patient history + symptoms gathered simultaneously

Reflection & Error Correction: Feedback integration on failed verifications + Patient safety validation on all Rx actions

Utterance Calibration Framework

From: 'Calibrating Nominal Language Understanding for Declarative Human-Interfacing AI'

Asset Meta

Type: NLU Training Spec

Used by: ML Eng, Data Science

Purpose: Intent robustness

Master

Core commands devoid of detail

"File a claim" | "Policy details"

Golden

Most common/expected phrasing

"I want to file an insurance claim"

1st Degree

Slight phrasing variations

"I need to submit a claim for my car accident"

2nd Degree

Structurally diverse expressions

"Had a fender bender—need to claim insurance"

3rd Degree

Rare/slang/cultural references

"Got in a scrape with my ride, how do I claim?"

LLM Integration: Golden utterance calibration + Ambiguity resolution + Real-time feedback loop for iterative improvement

Recursive Synthetic Data Generation

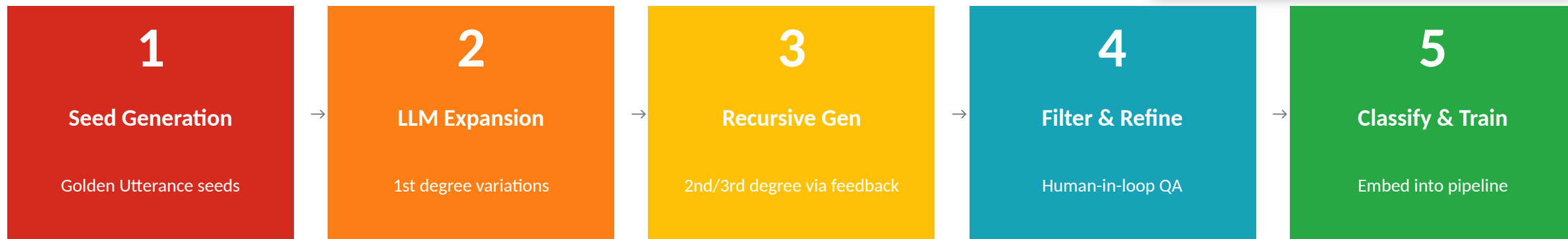
Asset Meta

Type: Data Pipeline

Used by: ML Eng, QA

Purpose: Training augmentation

Recursive Generation Workflow



Example: Healthcare Query Expansion

Seed: "I need a refill on my medication"

1st Degree: "Can I get a refill for my prescription?" | "Help me renew my meds"

2nd Degree: "Need to get my meds refilled—what do I do?" | "How do I renew my current medication?"

3rd Degree: "Out of pills—can you sort me a new Rx?" | "Need to re-up my meds, what's the process?"

Benefits: Scalability | Cost Efficiency | Data Diversity | Edge Case Coverage | Bias Mitigation

Design System-Driven AI Automation

From: 'Design System-Driven AI Tools for Automated Design Generation'

Asset Meta

Type: Design Ops Spec

Used by: Design, Frontend

Purpose: Workflow acceleration

1

Define Components

Catalog buttons, typography, layouts, constraints, interaction patterns

2

Dataset Creation

Export via Figma/Sketch/XD APIs with spacing, hierarchy, responsive metadata

3

Model Selection

Fine-tuned GPT models or open-source (ControlNet, Sketch-RNN)

4

Training

Hyper-specific use cases: UI mockups, pattern variations

5

Integration

Plugins/APIs for real-time suggestions or auto-generation

Applications

Automating UI Mockups

Generate wireframes from user stories

Real-Time Feedback

AI suggests layout improvements

Scaling Components

Auto-create light/dark mode variants

Design → Code

Generate React/HTML from design tokens

Trends: GANs for dynamic layouts | Fine-tuned LLMs for NL→Design | Multimodal AI | No-code integration

Authority Matrix

Asset Meta

Type: Governance Doc

Used by: Legal, Compliance

Purpose: Decision rights

Action Type	AI Agent	Human Review	Escalation
FAQ Response (Low Risk)	✓ Autonomous	Optional Audit	None
Dosage Information	✓ With Disclaimer	Post-hoc Review	Flag if uncertain
Prescription Assist	Draft Only	Required Approval	Always available
Caretaker Auth	Verify Identity	Required	Immediate
Medical Advice Query	Decline + Redirect	N/A	Auto-escalate
Adverse Event Report	Capture + Flag	Immediate Review	Priority Queue

Key: HIPAA + FDA alignment requires human-in-loop for all clinical decisions

Failure Mode Mitigation

From: 'Agents for Lilly Direct'

Asset Meta

Type: Risk Spec

Used by: Engineering, QA

Purpose: System resilience

Planning Failures

Invalid Tools

Ensure all APIs align with Lilly compliance + safety standards

Goal Misalignment

Verify agent adheres to patient-centric goals (affordability, accessibility)

Tool Failures

Data Accuracy

Regularly audit tools for dosages and drug interactions

Critical Tasks

Backups/fallbacks for high-stakes tasks like refill submissions

Efficiency Issues

Redundant Calls

Optimize workflows by minimizing duplicative API checks

Latency Spikes

Cache common queries, parallel processing for independent tasks

Patient Safety Checks (Always Enforced)

- Validate all suggested actions around prescription refills and dosage adjustments
- Feedback integration on failed verifications with automatic reattempt
- Human escalation trigger on any confidence <0.6 or safety-critical keywords

Confidence Gate Definitions

HIGH >0.85

- Autonomous response
- No human review
- Full context logged
- Audit trail created

MEDIUM 0.6 - 0.85

- Clarification prompt
- Suggest options
- Offer human option
- Re-evaluate on input

LOW <0.6

- Auto-escalate
- Human handoff
- Context preserved
- Priority flagged

Domain-Specific Overrides

Dosage Queries: >0.95 → Extra verification required

Adverse Events: Any → Always escalate to clinical

Billing/Payment: >0.90 → Human confirmation for changes

Asset Meta

Type: Threshold Spec

Used by: ML Eng, QA

Purpose: Quality gates

Escalation Flow Design

Asset Meta
Type: Process Flow
Used by: Support, Ops
Purpose: SLA compliance

Tier 0	AI Agent Self-service FAQ, navigation help, form assist	Instant
Tier 1	Virtual Pharmacist Drug info, dosage clarification, refill status	<2 min
Tier 2	Live Agent Complex auth, billing disputes, complaints	<5 min
Tier 3	Clinical Staff Adverse events, medical concerns, emergencies	Immediate

Auto-Escalation Triggers

- Confidence <0.6
- Keywords: 'emergency', 'adverse', 'allergic'
- 3+ failed intents
- User requests human

UI / Interaction Prototypes

Lilly Assistant

How do I refill my Zepbound?

I can help! You have 2 refills remaining.
[Refill Now] [Change Pharmacy]

I need to talk to someone

Connecting you with a pharmacist. Wait: <2 min

Asset Meta

Type: UI Wireframes

Used by: Design, Legal

Purpose: Stakeholder review

Ambient Assistance

Prescription Form

Pharmacy: [CVS - Main St]

Insurance: [Blue Cross - ****1234]

 Tip: You can use caretaker info for pickup. Need help?

Proactive: Context-aware tooltips based on DOM state + intent signals

WebView

<https://creator.voiceflow.com/share/69790f7ab023e37239ee1585/development>

Regulatory Alignment Notes

Asset Meta

Type: Compliance Checklist

Used by: Legal, Regulatory

Purpose: Audit-ready

HIPAA

- ✓ PHI encryption at rest + transit
- ✓ Minimum necessary principle
- ✓ Audit logging all access
- ✓ BAA with OpenAI

FDA 21 CFR Part 11

- ✓ Electronic signature validation
- ✓ Audit trail immutability
- ✓ System access controls
- ✓ Data integrity checks

FDA SaMD Guidance

- ✓ Risk classification (Class II)
- ✓ Clinical validation protocols
- ✓ Intended use documentation
- ✓ Change control procedures

All AI outputs reviewed by clinical/legal before production deployment per Lilly SOP-AI-2024

Behavioral Test Plans

Asset Meta

Type: QA Spec

Used by: QA, Engineering

Intent Recognition

Pass Rate: 95%

- 100+ utterance variants per intent
- Ambiguous query handling
- Multi-intent detection
- Language/accent robustness

Escalation Triggers

Pass Rate: 99%

- Keyword detection accuracy
- Sentiment threshold validation
- Timeout escalation paths
- Human request recognition

Safety & Compliance

Pass Rate: 100%

- PHI handling verification
- Adverse event capture
- Disclaimer insertion
- Audit trail completeness

Stress Test: 10K concurrent users | Avg Response: 1.2s | P99 Latency: 3.8s | Error Rate: <0.1% | Uptime: 99.9%

Impact Across Lilly Organization

Compliance & Safety

- Validating adherence to regulatory standards
- Automating data submissions for audits
- Real-time PHI monitoring

Sales & Marketing

- Agent insights to identify patient needs
- Tailor campaigns for specific medications
- Demographic targeting optimization

Research & Development

- Query clinical data repositories
- Drug development insight extraction
- Pattern recognition in trial data

Generative AI for Synthetic Data in Pharma

GenAI simulates patient utterances to augment training datasets and improve agent accuracy:

Formal: "Can I get a refill for [drug name]?" | Casual: "I'm out of my meds—what now?" | Complex: "How do I know if this will interact with others I'm taking?"