

Contents

1 Pi Monorepo Architecture	1
1.1 Overview	1
1.2 System Architecture	1
1.3 Azure OpenAI Managed Identity Architecture	1
1.4 Data Flow Architecture	1
1.5 Package Dependency Graph	1
1.6 Provider Architecture (pi-ai)	1
1.7 File System Architecture (Coding Agent)	1
1.8 Key Features by Package	1
1.8.1 1. @mariozechner/pi-ai	1
1.8.2 2. @mariozechner/pi-agent-core	1
1.8.3 3. @mariozechner/pi-coding-agent	5
1.8.4 4. @mariozechner/pi-mom	5
1.8.5 5. @mariozechner/pi-tui	5
1.8.6 6. @mariozechner/pi-web-ui	5
1.8.7 7. @mariozechner/pi-pods	5
1.9 Security Features	6
1.9.1 Azure OpenAI Managed Identity (Implemented)	6
1.9.2 Other Security Features	6
1.10 Performance Optimizations	6
1.11 Development Workflow	6
1.12 Technology Stack	6
1.13 Version	6

1 Pi Monorepo Architecture

1.1 Overview

Pi is a monorepo containing 7 interconnected packages for building AI agents and managing LLM deployments. The architecture follows a layered approach with clear separation of concerns.

1.2 System Architecture

1.3 Azure OpenAI Managed Identity Architecture

1.4 Data Flow Architecture

1.5 Package Dependency Graph

1.6 Provider Architecture (pi-ai)

1.7 File System Architecture (Coding Agent)

1.8 Key Features by Package

1.8.1 1. @mariozechner/pi-ai

- **Multi-Provider Support:** 700+ models across 17 providers
- **Streaming:** First-class streaming support with SSE
- **Tool Calling:** Unified tool/function calling interface
- **Model Discovery:** Auto-generated model catalog
- **Cost Tracking:** Token usage and cost estimation
- **Cache Management:** Prompt caching support
- **Azure Integration:** Managed Identity authentication (NEW)

1.8.2 2. @mariozechner/pi-agent-core

- **Transport Abstraction:** Pluggable I/O layer

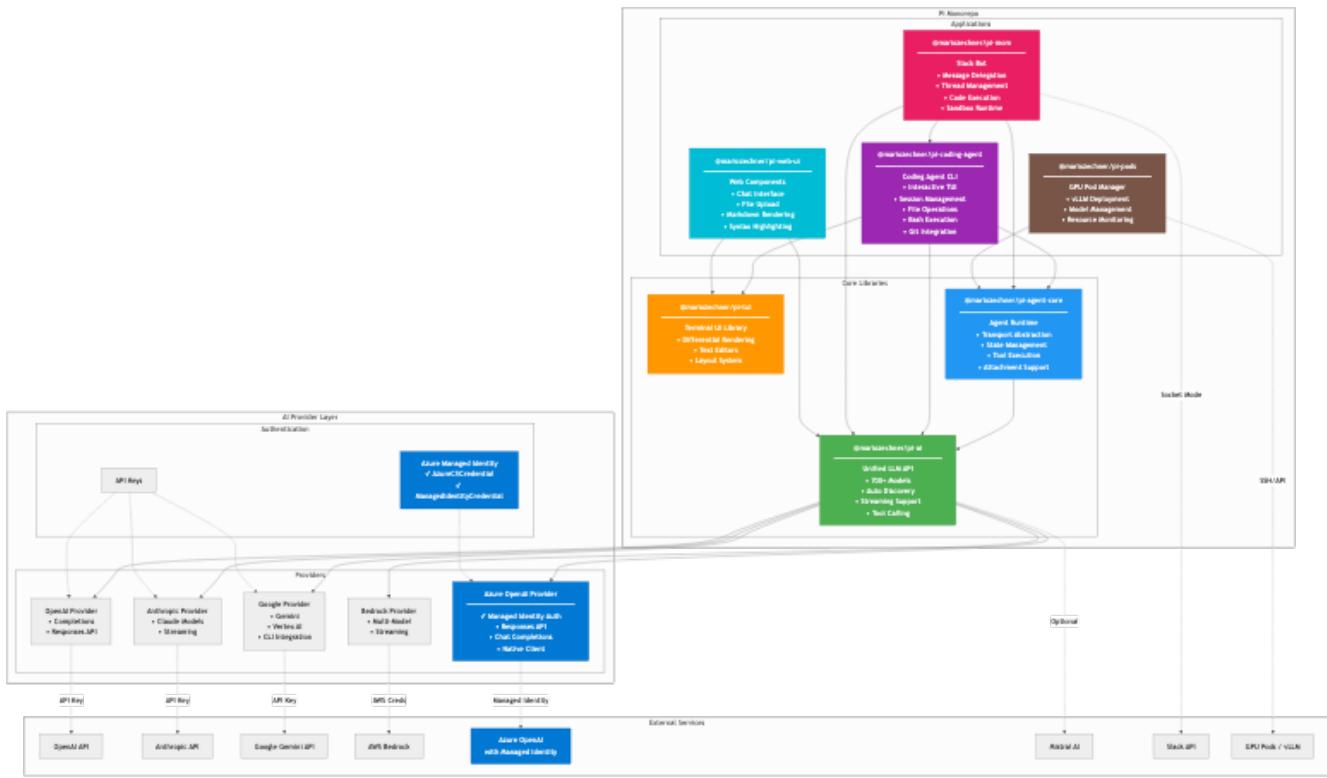


Figure 1: Diagram 1



Figure 2: Diagram 2

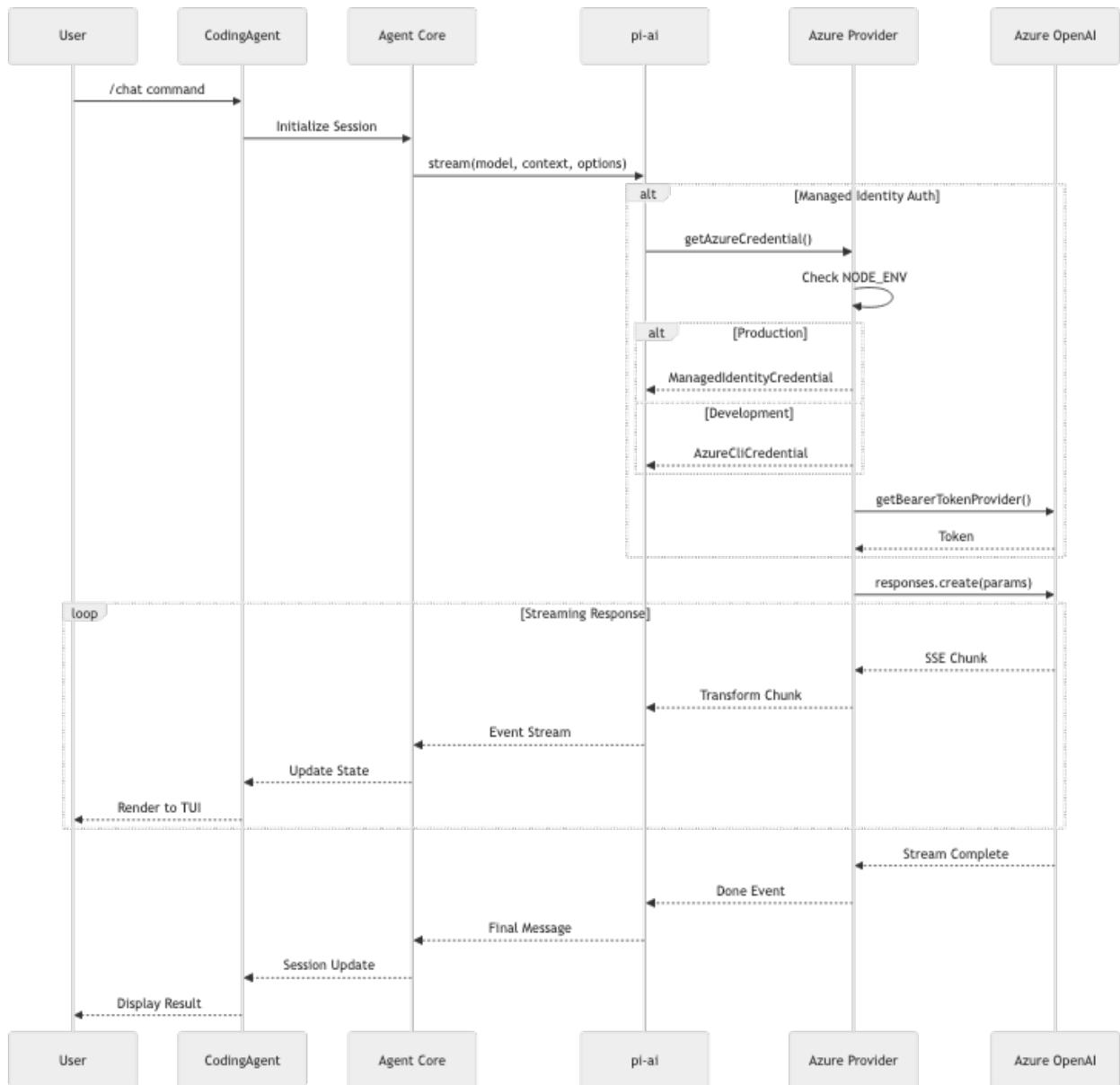


Figure 3: Diagram 3

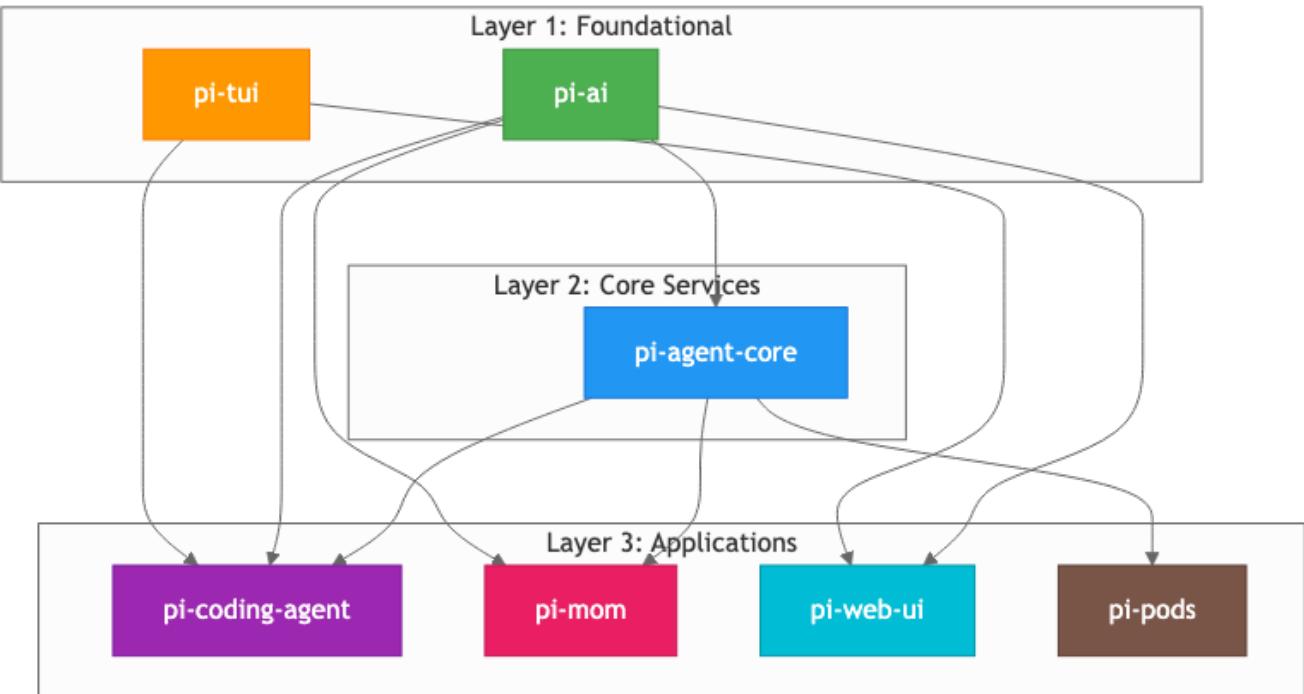


Figure 4: Diagram 4

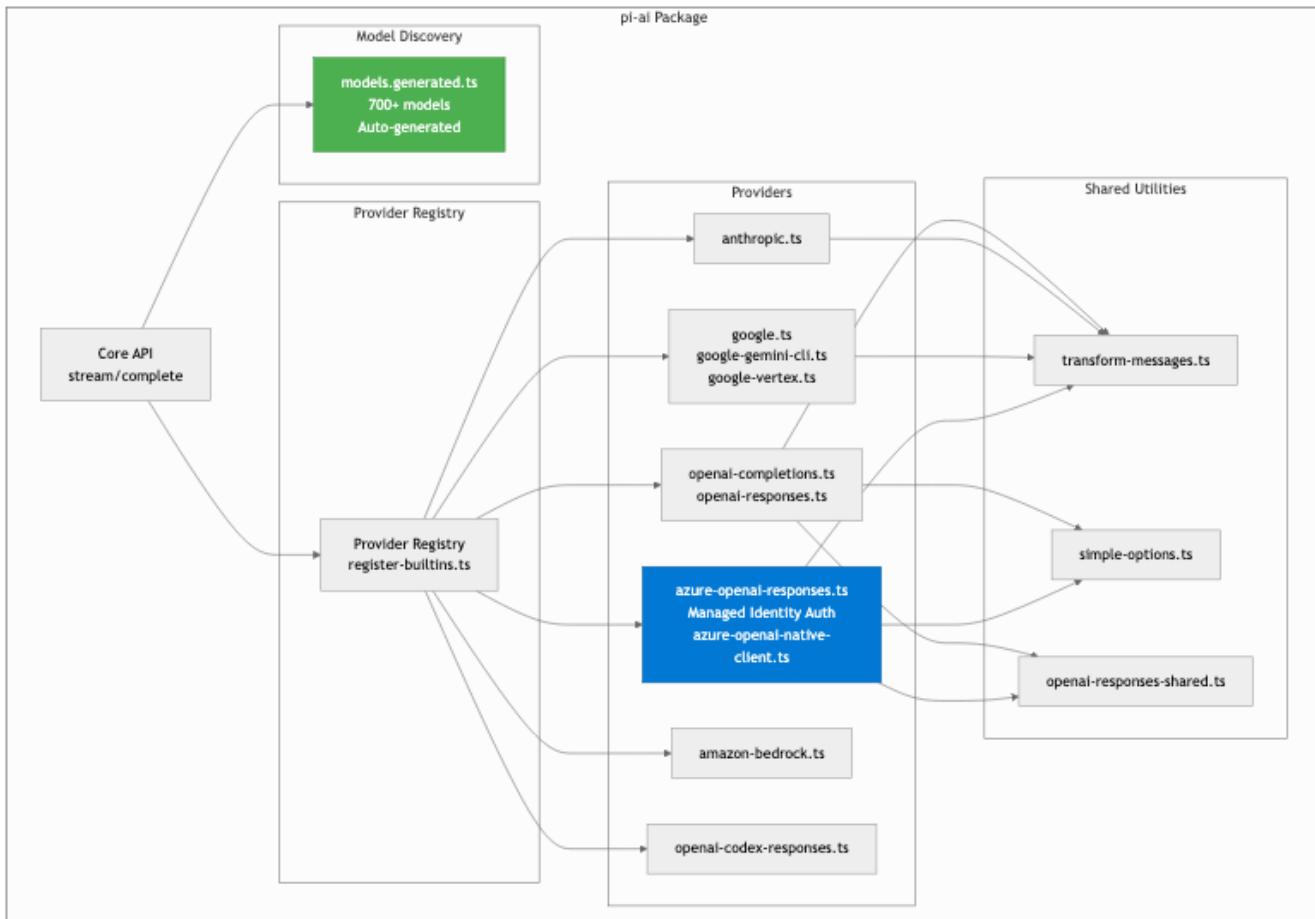


Figure 5: Diagram 5

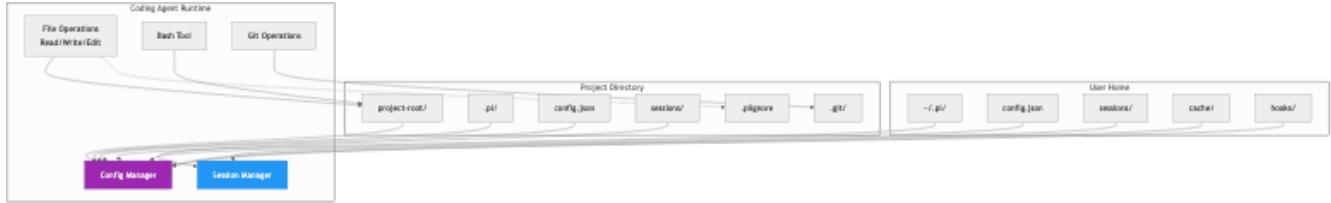


Figure 6: Diagram 6

- **State Management:** Conversation and tool state
- **Attachment Support:** File and image handling
- **Event System:** Typed event emission
- **Tool Execution:** Async tool call handling

1.8.3 3. @mariozechner/pi-coding-agent

- **Interactive TUI:** Full-featured terminal interface
- **Session Management:** Save/restore conversations
- **File Operations:** Read, Write, Edit with diff support
- **Bash Execution:** Safe command execution with sandboxing
- **Git Integration:** Commit creation, PR generation
- **Context Management:** Smart file inclusion
- **Extension System:** Custom tools and providers

1.8.4 4. @mariozechner/pi-mom

- **Slack Integration:** Socket mode for real-time messages
- **Thread Management:** Conversation threading
- **Sandbox Execution:** Safe code execution with E2B
- **Scheduled Tasks:** Cron-based job execution
- **Multi-Channel:** Support multiple Slack workspaces

1.8.5 5. @mariozechner/pi-tui

- **Differential Rendering:** Efficient screen updates
- **Text Editors:** Full-featured text input components
- **Layout System:** Flexbox-like terminal layouts
- **Markdown Rendering:** Terminal-based markdown display
- **Syntax Highlighting:** Code syntax coloring

1.8.6 6. @mariozechner/pi-web-ui

- **Web Components:** Lit-based reusable components
- **Chat Interface:** Full-featured chat UI
- **File Upload:** Drag-and-drop file support
- **Document Preview:** PDF, DOCX, XLSX rendering
- **Syntax Highlighting:** Code block rendering
- **Dark/Light Theme:** Theme support

1.8.7 7. @mariozechner/pi-pods

- **vLLM Deployment:** GPU pod orchestration
- **Model Management:** Download and configure models
- **Resource Monitoring:** GPU utilization tracking
- **SSH Integration:** Remote pod management

1.9 Security Features

1.9.1 Azure OpenAI Managed Identity (Implemented)

```
No API keys in code  
Azure AD authentication  
Automatic token refresh  
Development/Production modes  
Least privilege access  
Audit trail
```

1.9.2 Other Security Features

- **API Key Management:** Environment-based configuration
- **Sandboxing:** Safe code execution in Mom
- **.piignore:** Prevent sensitive file access
- **Git Safety:** No force operations without explicit consent
- **Bash Safety:** No destructive commands by default

1.10 Performance Optimizations

1. **Streaming:** All providers support streaming responses
2. **Caching:** Prompt caching for repeated context
3. **Differential Rendering:** Only redraw changed terminal regions
4. **Lazy Loading:** Models loaded on-demand
5. **Connection Pooling:** Reuse HTTP connections
6. **Batch Operations:** Group file operations

1.11 Development Workflow

```
npm install          # Install dependencies  
npm run build       # Build all packages  
npm run dev         # Watch mode for all packages  
npm run check        # Lint, format, type check  
npm test            # Run test suites
```

1.12 Technology Stack

- **Language:** TypeScript 5.7+
- **Runtime:** Node.js 20+
- **Build:** Custom tsgo compiler
- **Testing:** Vitest
- **Linting:** Biome
- **UI:** Terminal (blessed-like), Web (Lit)
- **AI SDKs:** Official provider SDKs
- **Authentication:** Azure Identity SDK

1.13 Version

Current Version: **0.52.9**

Last Updated: February 11, 2026