

PHUSE US CONNECT 2026 · AI AND AUTOMATION TRACK

# Leveraging AI for Clinical Data Transparency

*Automating Complex Disclosure Workflows*

Woo Song, CEO · Xogene

Tuesday, 24 March 2026 · Nueces Room

# AI is a commodity.

## Domain expertise is the moat.



### Same Models

Everyone has access to Claude, GPT, Gemini — the LLMs are not the differentiator



### The Hard Part

Engineering design patterns + deep domain knowledge  
separate demos from production systems



### Without Both

AI produces confident nonsense — plausible outputs with dangerous errors

## WHAT RESULTS DISCLOSURE ACTUALLY REQUIRES

# Partial capability ≠ production-ready disclosure.

### Regulatory Knowledge

FDAAA requirements, PRS field definitions, mandatory vs optional

### Document Navigation

Knowing AE summary is in CSR §12.2, not 'somewhere in the document'

### Source Data Selection

Choosing the right sections for the right fields

### PRS Field Mapping

Exact format, data types, and constraints for each structured field

### Cross-Referencing

Participant counts must match across Arms, Flow, Baseline, and AEs

### Sponsor-Specific Rules

Every client has different conventions, thresholds, and preferences

## A SPECTRUM OF AI APPROACHES

# Different architectures produce very different results

### AD-HOC / CHATBOT

"Upload a CSR section and ask"

- ✗ No document preprocessing
- ✗ Poor registry field knowledge
- ✗ No quality checks
- ✗ Single-shot output

**~30% numerical error rate**

in our testing on clinical data

*Looks plausible but numbers are often wrong. Good for exploration, not production.*

### GENERAL-PURPOSE AI

Enterprise copilots & assistants

- ✓ Summaries, meeting notes
- ✗ No clinical document structure
- ✗ Poor PRS schema awareness
- ✗ No multi-document reasoning

**Useful, but wrong tool**

for regulatory compliance tasks

*Designed for productivity, not regulatory precision. No domain knowledge layer.*

### BASIC RAG PIPELINE

Document retrieval + prompting

- ✓ Has document access
- ~ Generic chunking (by token count)
- ~ Prompts may lack domain nuance
- ✗ Limited quality assurance

**15-20% error rate**

better, but misses nuance

*Right direction. Gap is in domain-aware preprocessing and systematic QC.*

### DOMAIN-SPECIFIC AGENTS

Purpose-built agentic system

- ✓ Multi-pass preprocessing
- ✓ TOC-driven intelligent routing
- ✓ Domain knowledge - every layer
- ✓ Systematic QC + reflection
- ✓ Sponsor-specific customization

**<2% error rate with QC**

production-grade output

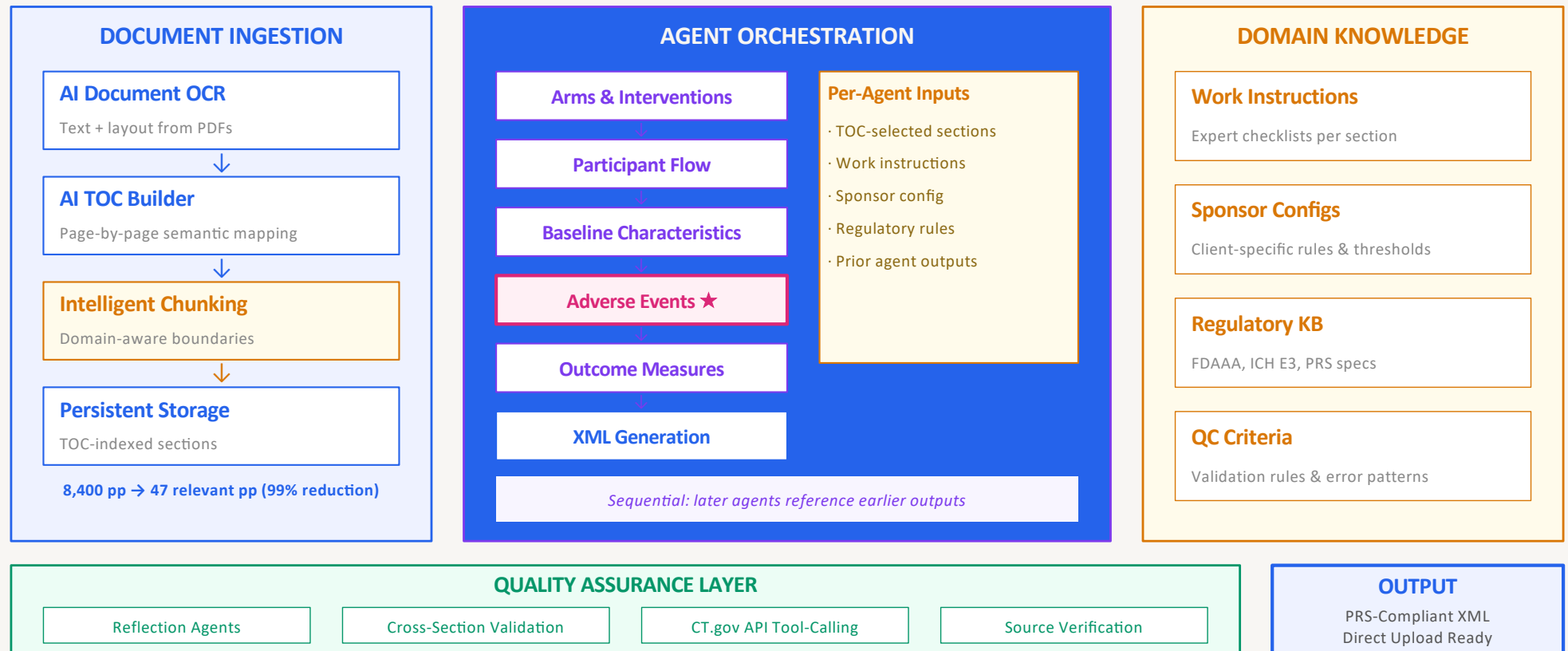
*Expert-reviewed drafts, PRS-compliant XML, ready for submission.*

Increasing domain specificity and production readiness →

## SYSTEM ARCHITECTURE

# Multi-agent orchestration built on domain expertise

■ AI / LLM   ■ Domain Knowledge   ■ Quality Assurance



## Multi-pass preprocessing: messy PDFs to structured intelligence

### Pass 1: AI OCR

- Extracts text + layout structure
- Preserves table formatting
- Handles scanned pages
- Processes images & figures

*Specialized OCR models with document layout understanding*

### Pass 2: AI TOC Builder

- Reads every page + images
- Builds semantic TOC
- Tags content types (table, narrative, figure)
- Maps CSR section structure

*Domain knowledge: knows CSR section numbers and types*

### Pass 3: Smart Chunking

- Split by content boundaries
- Tables stay intact
- Sections stay coherent
- Tagged with TOC metadata

*Not 500-token splits — logical content units*

Typical approach: Split every 500 tokens → tables broken, sections mixed, context lost

Domain-informed: Split by content boundaries → tables intact, sections coherent, precision extraction

# Garbage In Garbage Out

“

That's where 'Garbage in Garbage-out' (GIGO) enters the vocabulary. Mellin explains "If the problem has been sloppily programmed the answer will be just as incorrect.

Army Specialist William D. Mellin, 1957

## Work With New Electronic 'Brains' Opens Field for Army Math Experts

WASHINGTON — BIZMAC, UNIVAC, GARBAGE IN-GARBAGE OUT—all new terms in the Army.

They haven't replaced GOVERNMENT ISSUE or BY THE NUMBERS and probably won't, but in a small, select circle they're heard more frequently than the old expressions used by millions of veterans of World War II and Korea.

For these colorful expressions are part of the workday vocabulary of the military mathematicians who man the Army's electronic computers, the "brains" that fed the right figures to the men who know how to come up with the right answers in a matter of split seconds.

Specialist Third Class William D. Mellin of Gardner Mass. is a master of one of these machines, some of which are as big as fruit basket ball courts.

MELLIN, a man of unusual occupation has an unusual title. He is an electronic computer programmer, one of a group performing as highly specialized soldiers in this unique military field.

In his Signal Corps work with the computer Mellin might be given a problem to solve in connection with radio frequency data. Done the old fashioned way it might have taken 50 girls with hand calculating machines a week to come up with the answer.

One man however in this case Mellin "programs" the problem a pre-planning task that may take a few hours or a couple of days of figuring out the step by step process to be taken by the electronic "brain."

When this is completed the information is fed often as mathematical formulae—into the maze of the machine and depending on the magnitude of the problem the answer is obtained in seconds or minutes.



**A 'BRAIN'S' BRAIN** Mechanical brains are useless without first-class human mind to figure what information to feed them. Specialist Third Class William D. Mellin of Gardner Mass., does this work for the Army, which uses several types of electronic computers in supporting headquarters and combat unit activities.

need it. This means a smooth flow of weapons and equipment from manufacturer to combat soldier. Today that flow is made possible without miles of files.

Today one programmer "brain" can keep track of more than 100 million facts and then file all this information in a couple of reels of magnetic tape.

**NOW is the**

**NEXT STEP** is debugging or checking the accuracy of the work. Did the results meet the original requirements or should the problem be re-run with corrections? That's where Garbage in Garbage-out (GIGO) enters the vocabulary. Mellin explains "If the problem has been sloppily programmed the answer will be just as incorrect. If the programmer made mistakes the machine will make mistakes. It can't correct them because it can't do one thing: It can't think for itself.

Men like Mellin are not picked out of basic training units and taught their Army jobs by the numbers. But the Army taught him plenty just the same. A 1952 Duke University graduate in economics and mechanical engineering, he became interested in computers as he watched the television reports of that year's national elections.

Fascinated by the possibilities of a new field in the relatively new field he went to work for one of the "brain" manufacturers his starting job consisting of eight hours of school daily followed by six months apprenticeship on the machine.

**WHEN MELLIN** entered the Army he was quickly returned to several schools operated by manufacturers of various models that the Army uses. The idea, of course, was to help him become familiar with the different machines used by the Army and especially by the Signal Corps.

All the "brains" incidentally aren't found in the Pentagon or in the Army's research and development laboratories. They're also become almost indispensable in what old soldiers still fondly call the "field." The employment of new weapons systems requires the use of lightning calculations which has created a demand for trained programmers.

The complexity of the modern Army means equally complex supply and accounting systems to insure that field units have what they need when and where they need it.

**Observe Anniversary** SALT LAKE CITY, Utah (AP)—The famed Salt Lake Tabernacle choir is observing an anniversary—its 25th successive year of Sunday morning broadcasts over the Columbia Broadcasting System. President Eisenhower and a score of other government officials have sent congratulations.



**... to start your savings account!** When it comes to opening a savings account, there will never be a time as good as NOW. Nothing is easier than putting it off. But nothing is better (for your financial future) than getting started at once upon a program of systematic saving. Don't waste precious time. Open a savings account at HEGEWISCH FEDERAL... you'll always be glad you did.

Current Rate Of **3 1/2%** DIVIDEND ON SAVINGS **3 1/2%**

**HEGEWISCH FEDERAL SAVINGS**

13303 Baltimore Ave.  
Hegewisch, Ill.  
MI 6-1000  
Wm. R. Anderson, Mgr.

AND LOAN ASSN. OF CHICAGO

## Intelligent routing: send only what's relevant to each agent

### Sample TOC Output (anonymized)

#### §12.2 Adverse Events

pp 340-385 | 3 summary tables, 2 figures

##### §12.2.1 Treatment-Emergent AEs

pp 340-355 | SOC × PT frequency table

##### §12.2.2 Serious Adverse Events

pp 356-370 | SAE narratives

##### §12.2.3 Deaths

pp 371-375 | Death listing table

#### §16.1.9 MedDRA Coding

Appendix | MedDRA v25.1 preferred terms

### AE Agent queries TOC →

Gets ONLY: §12.2 + §16.1.9

**47 pages, not 8,400**

*99% context reduction*

### Why this matters

Standard retrieval searches by keyword or cosine similarity — returns 200+ chunks with lots of data noise.

TOC-driven routing understands document structure. It knows AE data lives in specific CSR sections. This is knowledge no vector database captures.

## LAYER 2: SECTION-BY-SECTION AGENTS

# Specialized agents mirror how experts actually work

### Arms & Interventions

Protocol, CSR arms

### Participant Flow

Disposition + Arms output

### Baseline Characteristics

Demographics, vitals

### Adverse Events ★

AE tables, SAE, MedDRA

### Outcome Measures

Results tables, SAP

### Sequential dependency: later agents reference earlier outputs

The AE agent knows the arm definitions because the Arms agent established them. A human expert handles this intuitively — you'd never fill out AE tables without knowing the study arms. We encoded that same logic.

Why multi-agent? A single pass tries to hold the entire posting in one context — losing cross-references, missing nuance, unable to customize per section. Multi-agent decomposition mirrors the complexity of the task.

## DEEP DIVE: ADVERSE EVENTS AGENT

# Where complexity meets precision

### 1 TOC Query

Pulls only §12.2 AE tables, SAE narratives, MedDRA appendix. ~47 pages out of thousands.

### 2 Domain Knowledge

Expert checklists: verify MedDRA version, check arm consistency, apply frequency thresholds.

### 3 Sponsor Config

Sponsor A:  $\geq 5\%$  threshold.  
Sponsor B: All AEs reported.  
Sponsor C: Custom AE grouping terms.

### 4 QC Reflection

Verify counts sum correctly, arm labels match prior agents, no numbers absent from source.

#### Extraction Example

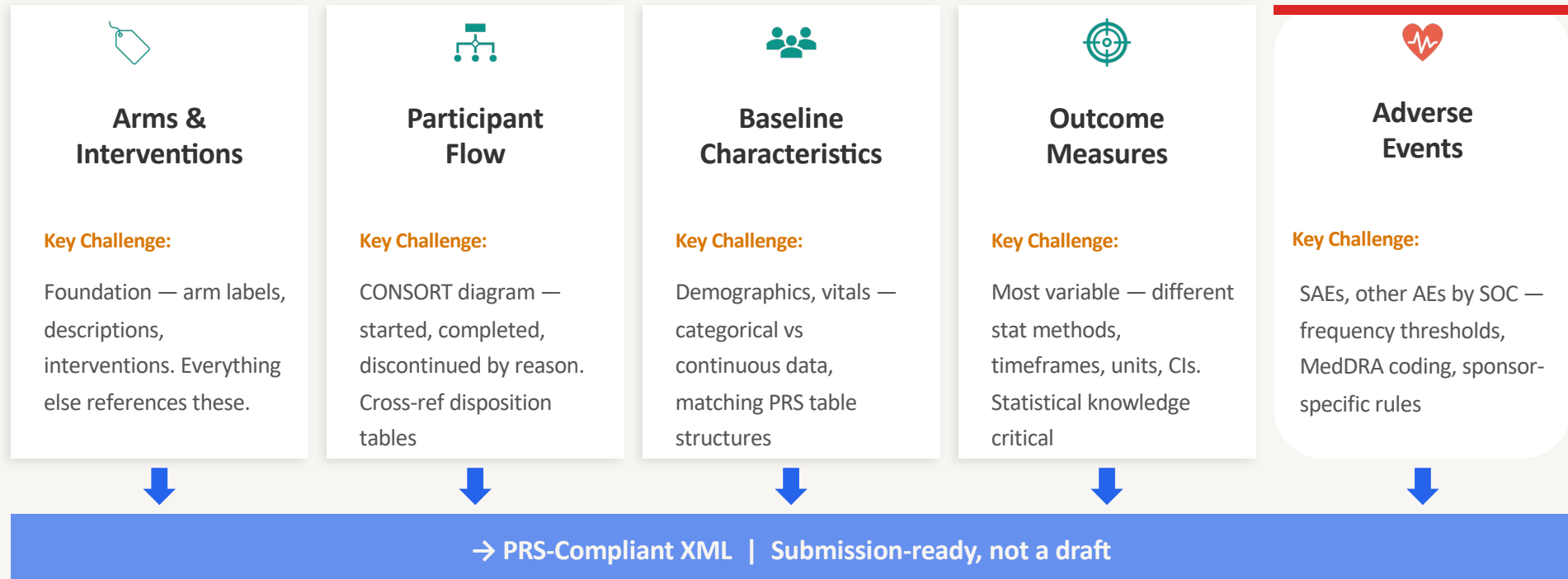
Source table → structured JSON with exact counts, percentages, arm labels. Precision matters: '243 participants' not '~245 participants.'

#### Error Caught by QC

Agent output: '245 participants'  
Source table: '243 participants'  
→ Flagged, corrected before human review

## THE FULL WORKFLOW

### Each section: same pattern, different domain knowledge



*Each section has its own agent, work instructions, QC checks, and sponsor rules.*

## LAYER 3: TRUST BUT VERIFY

### Three quality assurance mechanisms

#### Reflection Agents

Every section output is reviewed by a QC agent against original source chunks. Do these numbers appear? Are arm labels consistent? Does it pass our QC checklist?

~12% error rate → <2% after QC

#### Tool-Calling

Agents query ClinicalTrials.gov API directly — verify NCT numbers, pull existing registration data, compare with posted results.

Verification against external truth

#### Cross-Section Validation

Arms ↔ Flow (counts), Arms ↔ Baseline (denominators), Arms ↔ AEs (labels). Internal consistency checked systematically.

Catches cascading errors

*The goal: give human reviewers clean drafts to evaluate, not error-filled outputs to fix.*

## DOMAIN KNOWLEDGE AS INFRASTRUCTURE

**The model is above the waterline. Everything else is below.**

~15% of what matters

What everyone talks about

### THE AI MODEL

Claude · GPT · Gemini

### WHAT ACTUALLY MAKES IT WORK

~85% of what matters

Domain knowledge isn't  
a feature — it's the  
foundation.

#### Work Instructions

Expert checklists for each section

#### Sponsor Configurations

Client-specific rules & thresholds

#### Regulatory Knowledge Base

FDAAA, ICH E3, PRS field specs

#### QC & Validation Rules

Cross-section checks, error patterns

#### Preprocessing Pipeline

OCR, TOC building, smart chunking

#### Agent Architecture

Routing logic, section dependencies

## DESIGN CONSIDERATIONS

# Seven things that matter when building AI for disclosure

### How does the system handle documents?

- 1** Strong: Multi-pass preprocessing with domain-aware chunking  
Weak: Arbitrary token-count splits that break tables

### Are you trying to do too much at once?

- 2** Strong: Section-specific processing with PRS field mapping  
Weak: Single pass treats the entire posting as one task

### Where does domain knowledge live?

- 3** Strong: Maintained external knowledge base with versioned work instructions  
Weak: Embedded in a prompt, or not present at all

### How does it handle sponsor-specific rules?

- 4** Strong: Parameterized config — same system, different behavior  
Weak: Requires re-engineering for each client

### What does the QA architecture look like?

- 5** Strong: Automated reflection + cross-section validation  
Weak: Human reviewer is the only QC mechanism

### Can it produce machine-ready output?

- 6** Strong: PRS-compliant XML for direct upload  
Weak: A draft document requiring manual data entry

### How is accuracy measured and improved?

- 7** Strong: Systematic error tracking with feedback loops  
Weak: Ad-hoc feedback with no structured improvement

*These considerations apply whether you're building, buying, or evaluating AI for disclosure.*

## DESIGN PRINCIPLES THAT GENERALIZE

# Applicable to PLS, CTIS, anonymization, and beyond

### 1 Preprocess ruthlessly

Emphasize document preprocessing over other components. If your AI gets bad input, no amount of prompt engineering saves you.

### 2 Let domain experts design the routing

TOC categories and agent assignments should be designed by disclosure experts — not just engineers or data scientists.

### 3 Inject expertise at every layer

Not just the final prompt — in preprocessing, chunking, agent design, and QC criteria. Domain knowledge should be pervasive.

### 4 Build QC as a first-class citizen

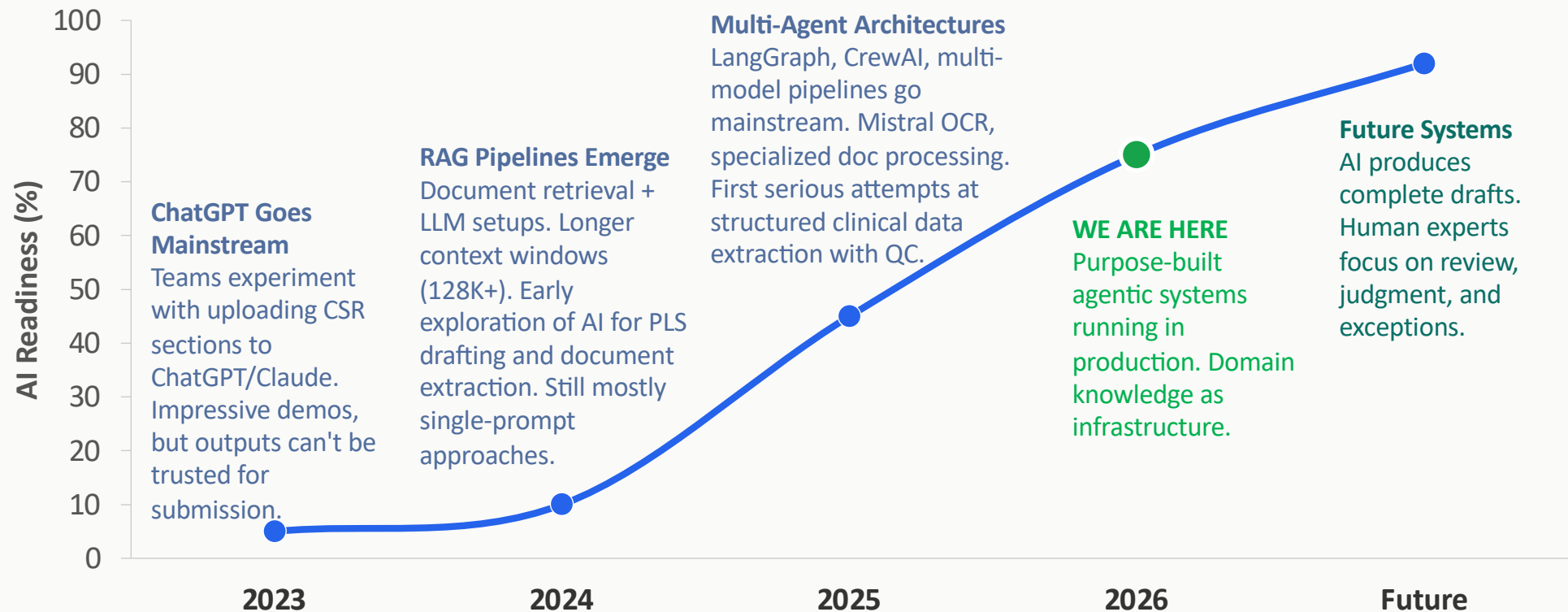
Not optional, not 'we'll add it later.' Reflection and cross-validation are core architecture, not afterthoughts.

### 5 The model is the least important decision

Claude, Mistral, and Nova, Nano Banana in different parts. Architecture and domain knowledge matter more than which LLM you pick.

## Where we are and where we're headed

### AI Readiness for Disclosure



## ANTHROPIC

### Labor market impacts of AI: A new measure and early evidence

Mar 5, 2026

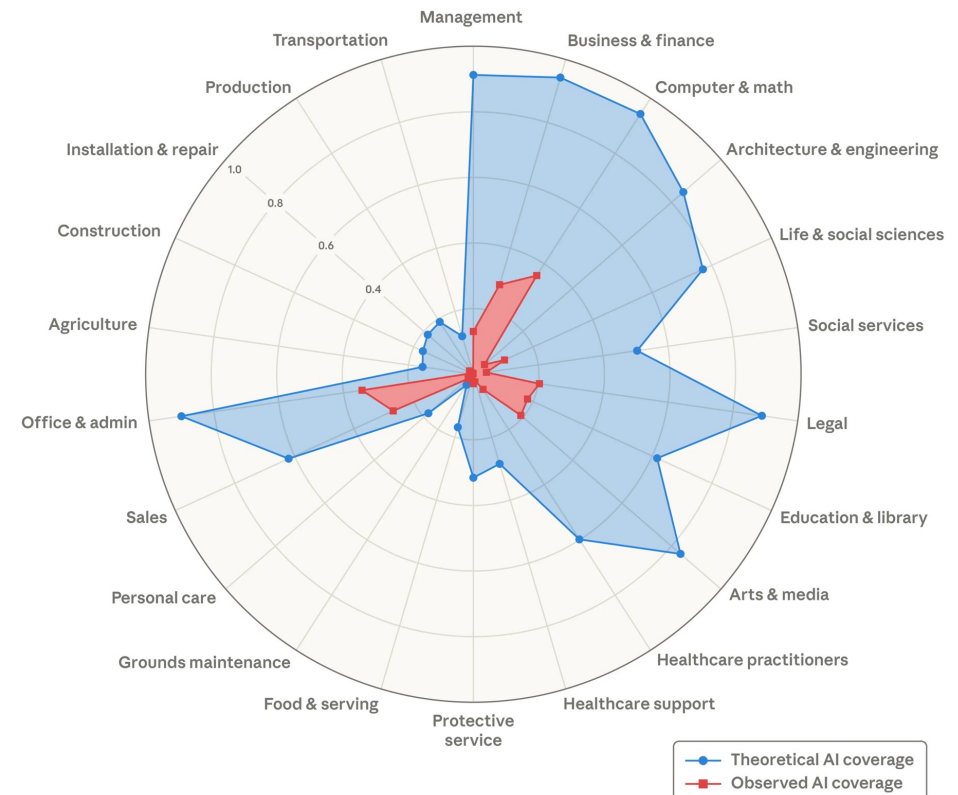
**Figure 2: Theoretical capability and observed exposure by occupational category**

Share of job tasks that LLMs could theoretically perform (blue area) and our own job coverage measure derived from usage data (red area).

<https://www.anthropic.com/research/labor-market-impacts>

Xogene · PHUSE US Connect 2026

**Theoretical capability and observed usage by occupational category**



# ANTHROPIC

## What 81,000 People want from AI

Mar 18, 2026

The largest qualitative AI study ever conducted — open-ended interviews with 80,508 Claude users across 159 countries and 70 languages.

81K

Interviews

159

Countries

70

Languages

<https://www.anthropic.com/features/81k-interviews>

AI



## What 81,000 people want from AI

Last December, tens of thousands of Claude users around the world had a conversation with our [AI interviewer](#) to share how they use AI, what they dream it could make possible, and what they fear it might do.



Jump to story

Each dot represents 4 respondents

## WHAT KEEPS PEOPLE UP AT NIGHT

### Voices from the Anthropic 81K Survey

#### UNRELIABILITY

*"I got caught in what I now recognize as a large, slow hallucination — answers that were internally consistent, confident, and wrong in subtle but compounding ways."*

— Researcher, United States

#### VERIFICATION BURDEN

*"An assistant that sounds sure but is often wrong forces you to treat everything as suspect. Instead of freeing attention, it creates a permanent 'fact-check tax.'"*

— United States

#### COGNITIVE ATROPHY

*"I don't think as much as I used to. I struggle to put the ideas I do have into words."*

— Heavy AI user, United States

## WHERE AI IS MAKING A DIFFERENCE

### Voices from the Anthropic 81K Survey

#### DOCUMENTATION RELIEF

*"I receive 100–150 text messages per day from doctors and nurses. So much of my cognitive labor was spent on documentation... Since implementing AI, the pressure of documentation has been lifted. I have more patience with nurses, more time to explain things to family members."*

— Healthcare worker, United States

#### TIME RECLAIMED

*"I used AI to cut a 173-day process down to 3 days. But the most meaningful part is the freedom to grow my career without sacrificing time with loved ones."*

— Software Engineer, United States

#### QUALITY + TIME FREEDOM

*"For the first time, I felt AI had surpassed human quality in a business task. That day I left work on time and picked up my daughter from daycare."*

— Software Engineer, Japan

## KEY TAKEAWAYS

1

AI is the engine.  
Domain expertise is the fuel.

2

Architecture > Model Selection  
How you design the system matters more than which LLM you pick.

3

Your regulatory expertise is the most valuable asset in the AI era.