

ML28



From Demo to Deployment: Bridging the AI Implementation Gap

March 25, 2026

Mel Hullings

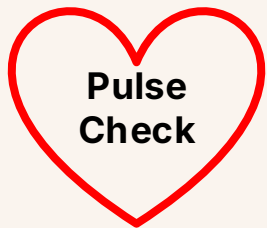
Emily Yates

Agenda

AI Readiness Journey

AI EDC Spec Builder Case Study

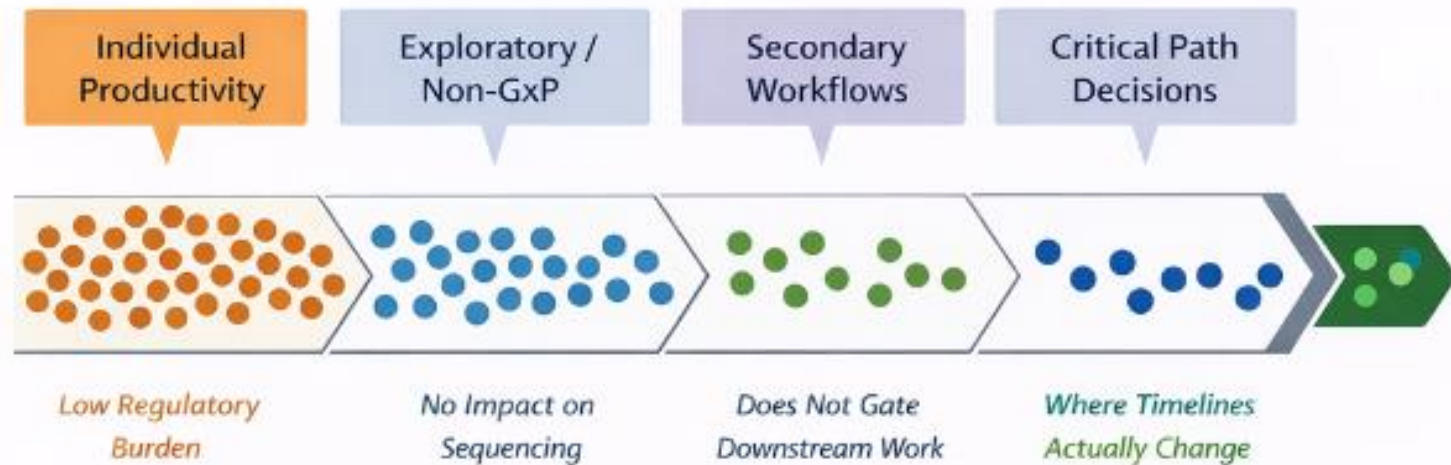
Learnings & Recommendations



Where does AI readiness sit today in your organization?



How can we reduce meaningful bottlenecks with AI?



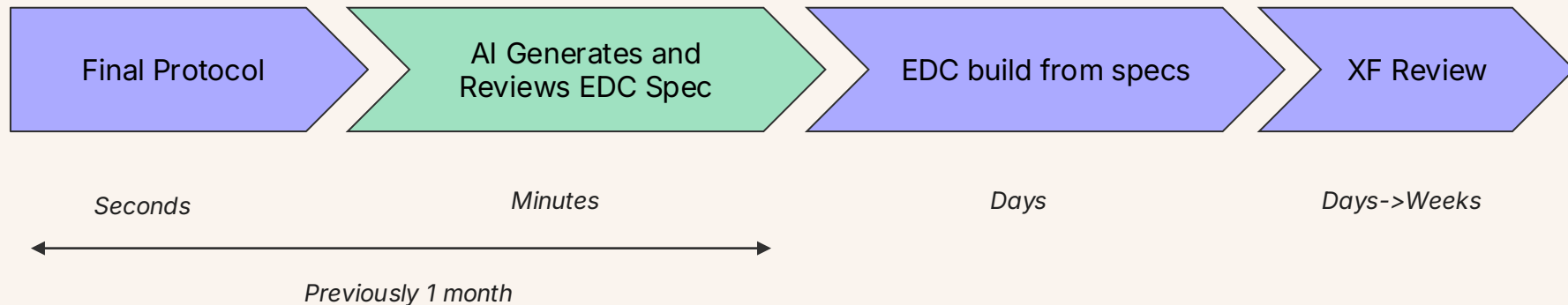
Implementing AI on the critical path is what drives impact to study speed, cost, and quality

Identifying High-Impact Use Case: EDC Build Case Study

	Today		Target
 Primary Focus: Time to Completion	12 week industry standard		50% improvement over industry
Scalability	1 Analyst <> 1 study		1 Analyst <> multiple studies
	Committee of human reviewers doing multiple rounds of review and testing		Committee of agent + human reviewers are more efficient without compromising on quality
Cost	\$100-200k Veeva professional services fees per study		\$0 per study *
Quality	Dependent on human reviewers and quality is limited by timeline		Automation + human reviewers lead to more thorough, standardized, and efficient reviews

Hypothesis: Initial speed savings from automated EDC spec creation

- ✓ We are able to automatically generate an EDC spec in minutes (previously ~1 month initial EDC build time)



However, the length of the end-to-end process depends on the quality of the AI generated output

AI output core challenges: legibility, trust, speed

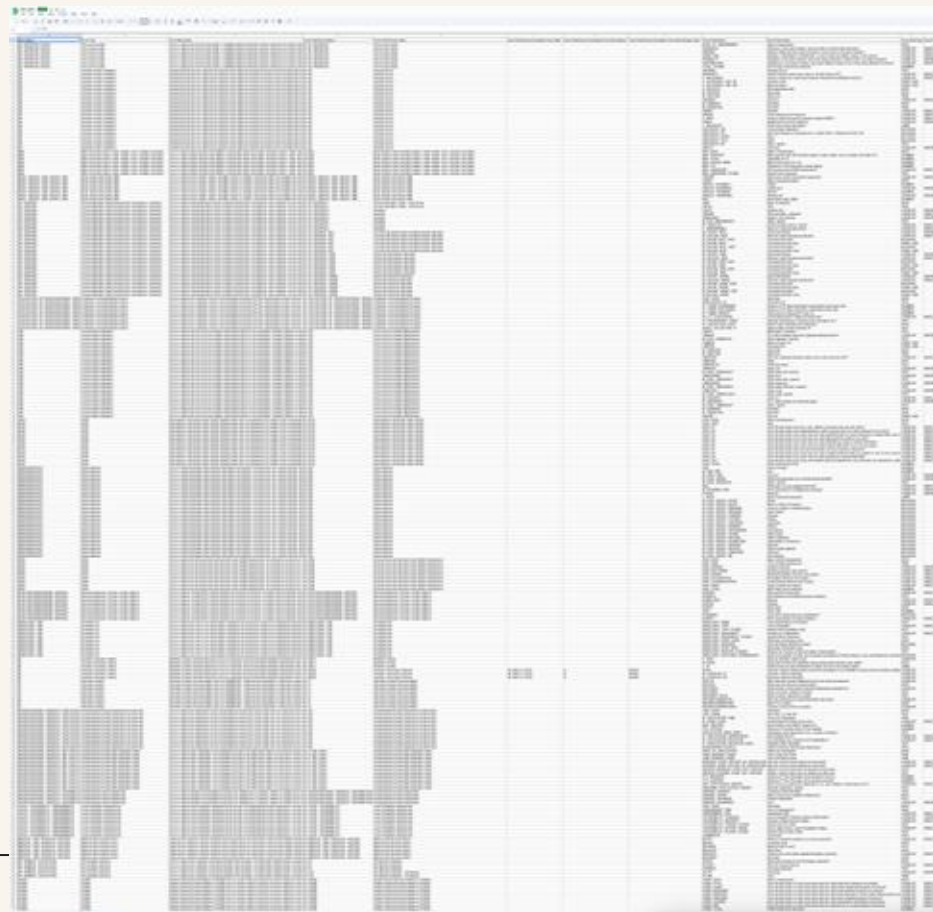
EDC Spec file is too difficult for a human to review

Some parts of the spec are more important to get right early

(i.e. if this one thing is wrong, 100 other things are wrong)

Lack of trust in output leads to difficulty to claim speed savings

(i.e. fear of having to start over manually because there are too mistakes in AI generated EDC spec)



Shift in Focus: legibility > trust in quality > speed

Break down the full EDC spec into components that are easier for human to review

Sequence the build so humans review the most pivotal decisions first, before the rest of the output is created

Humans review AI intermediate outputs, not just the final output

Expected Result: AI output, while not 100% accurate, is usually a trusted first draft

What are components?

Components = a definable unit of the EDC spec that can be generated, reviewed, and decided on independently

- Prioritized into low & high impact on the EDC spec
- Human can reason about an entire component
- Requirements, product build prioritization, and reviews all organized at the component level
- Having a defined unit let us make decisions consistently and build patterns across components

EDC Spec

High impact components

Events

Visit schedule and data collection timeline

Forms

Case Report Forms (CRFs)

Protocol Citations

Relevant sections of a protocol

Fields

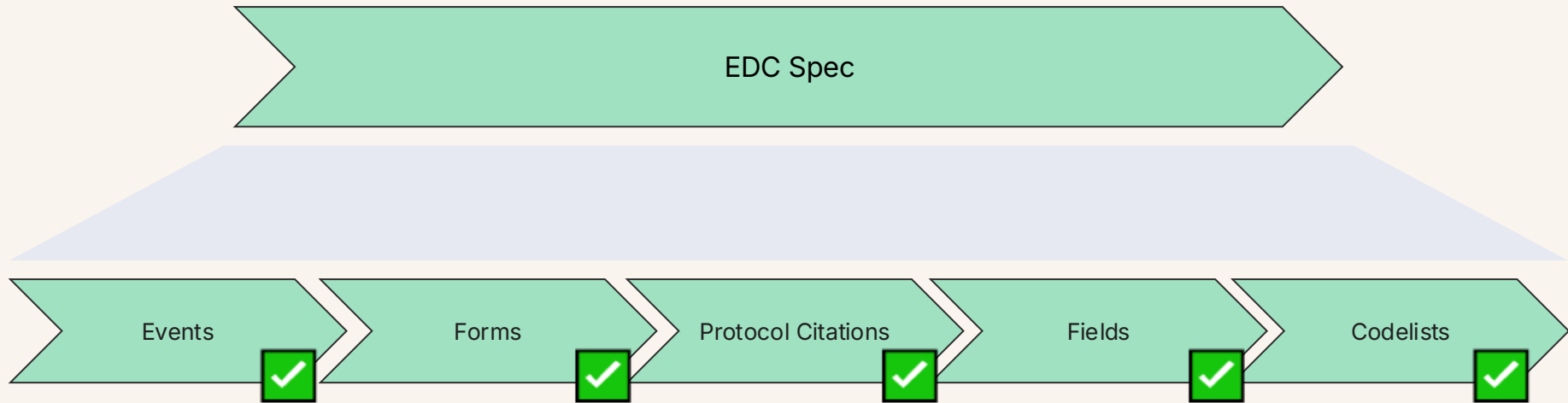
Each data point entered on a form

Codelists

Predefined allowable values for certain fields

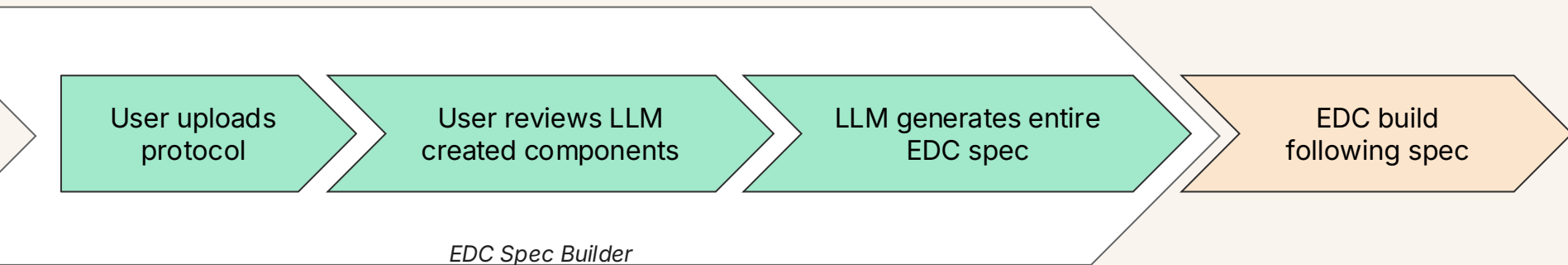
... and other low-impact components

High impact components are inputs to downstream components



LLM generates each component & then human reviews, ensuring an accurate list before moving to the next component

Result: EDC Spec Builder iteratively outputs EDC Spec document through human-in-the-loop process

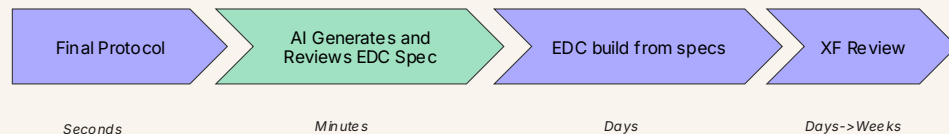


The system has a defined input & output allowing us to iterate within a defined problem space

Unexpected result: Automated EDC spec unlocks a new, upstream workflow with even greater impact

- ✓ We are able to use auto-generated EDC specs to have detailed discussions during protocol development

Typically, EDC design discussions did not begin until final protocol was available and study startup kicked off



With AI, discussions that typically cause delays during UAT can begin during protocol development and impact overall trial quality

Live protocol development testing results

EDC Spec Builder Goal	Live Protocol Demonstration
Reduce build time from 12 → 6 weeks	Draft spec generated in minutes, not weeks
Shift review earlier when changes are cheaper	40+ Issues caught before protocol finalization, while changes can still be incorporated Provided a way to deeply engage with the protocol & think through potential data issues

CDM: "Engaged with the output in a real sense. Made me really think about and engage with the protocol."

Medical Writing:

"Without the tool this wouldn't be happening so early."

"Found these things now instead of study team finding them later."

"Findings likely to have caused some amendments or at least a clarification letter."

AI Critical Path Implementation Framework

Decisions

Workflows

Quality

1

What operational metrics define success?

(What actually changes if this works?)

2

What decision is the AI intended to influence?

(Who decides differently, and when?)

3

What does the new workflow look like?

(How does work move before vs after AI?)

4

Where are the upstream and downstream impacts?

(What new conversations, risks, or dependencies appear earlier?)

5

How will interpretability and review be supported?

(Can experts reason and give feedback on output?)

6

How will this be tested and measured over time?

(How to ensure risk-based validation?)

Alignment with evolving regulatory AI guidance

Decisions

Clear role and scope for how AI is being used

Workflows

Clear, accessible, and contextually relevant output

Quality

Risk-based expert oversight and performance testing

FDA Guiding Principles

1. Human-centric by design
2. Risk-based approach
3. Adherence to standards
4. Clear context of use
5. Multidisciplinary expertise
6. Data governance and documentation
7. Model design and development practices
8. Risk-based performance assessment
9. Life cycle management
10. Clear, essential information



Questions?

Mel Hullings: melanie@formation.bio

Emily Yates eyates@formation.bio