PROJECT LOOM:

Evaluating End-to-End Agentic AI Workflows for Clinical Trials.

Christoph Hornik, MD PhD

i-Cubed | Duke Clinical Research Institute

Moulik Shah, MS

CEO Maxis AI

projectloom.org

PROJECT LOOM

WHAT IF...

AI could support the most **time-consuming** and **error-prone** parts of a trial, *without sacrificing* **scientific rigor** or **regulatory integrity**?

FROM IMAGINATION TO REALITY

Proof-of-concept study to evaluate agentic AI workflows **end-to-end**

Clinical Trial
Human-produced trial as probe

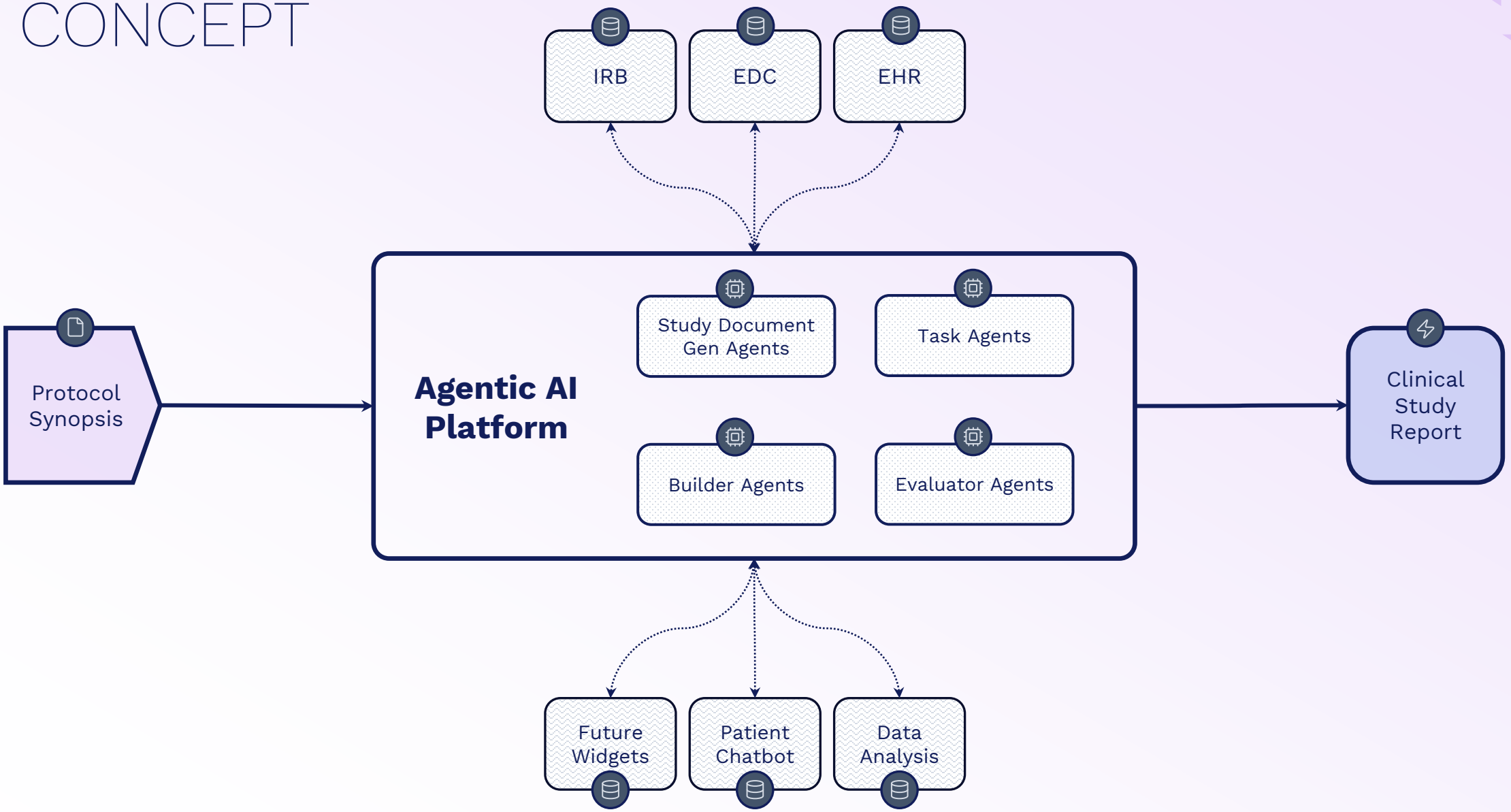
Platform
Agentic AI platforms with multiple agents

Participants
30 simulated personas across 5 humans

Oversight
Human review at each step



PROJECT LOOM:
CONCEPT



How the Platform was built: five layers, assembled for Loom

The agents were not written for Loom. They came from 20+ years of running clinical data work and were assembled on one platform to execute the trial end to end.

1. PEOPLE

PI, IRB reviewer, coordinator, data manager, participant

- Review queue
- Approve · edit · reject
- Participant chat / voice

2. ORCHESTRATION

Plans the trial as a sequence of governed tasks

- Task hand-offs
- Human checkpoints
- Retries and escalation
- Durable state

3. CLINICAL AGENTS

Pre-built, role-specific: writer, reviewer, builder, recruiter, reporter

- Protocol & ICF
- IRB package
- eCRF / EDC build
- Chatbot
- CSR

4. SHARED STUDY KNOWLEDGE

One source of truth every agent reads from

- Synopsis → protocol
- Eligibility rules
- Schedule of assessments
- Versions

5. STANDARDS AND CONNECTORS

Speaks the systems trials already use

- EHR (FHIR)
- EDC
- CDISC
- SPIRIT / CONSORT rubrics
- Audit trail

WHAT DISTINGUISHES IT

Five design choices that made the difference

1 · Mature clinical agents

Built and used on real trial work before Loom, not written for the demo. Loom's own notes attribute the speed difference largely to this.

2 · Agents that check agents

A critic agent scores each draft against SPIRIT, CONSORT, NIH and FDA rubrics and sends it back before a human sees it — an automated loop instead of many human iterations.

3 · One shared study brain

Every agent reads the same protocol, eligibility and schedule, versioned. Hand-offs carry explicit responsibilities and expected outputs.

4 · Standards, not new formats

Outputs land in the forms trialists already use — protocol templates, eCRFs, EDC, CDISC datasets, CSR sections — so humans review familiar artifacts.

5 · Accountable by design

Every action logged: which agent, which input, which version, who approved. You can reconstruct why the trial is in its current state.

What it is not

Not a chatbot answering questions, and not autonomous research. Agents do bounded, checkable work; investigators decide.

HUMANS IN THE LOOP

The human interface: agents propose, people decide

WHO TOUCHES THE PLATFORM

Investigator / sponsor

Approves protocol, amendments, go/no-go

IRB reviewer

Receives a complete package; queries answered with sources

Coordinator

Confirms eligibility; handles escalations

Data manager

Accepts or edits queries and data mappings

Participant

Chat or voice; can always reach a person

WHAT A REVIEWER SEES (ILLUSTRATIVE)

Review queue › Informed consent form v3

Proposed change

Add expected duration of participation (90 days) to Section 2.

Why

Required rubric item missing; duration taken from protocol §6.1 (source linked).

Confidence · rubric score 20/21 → 21/21

Approve | Edit | Reject

Disagreement is recorded, not overwritten.

Design principle

The human sees the proposal, the reason, the source and the consequence — then decides. Nothing reaches a participant, an IRB or a dataset without a named person approving it.

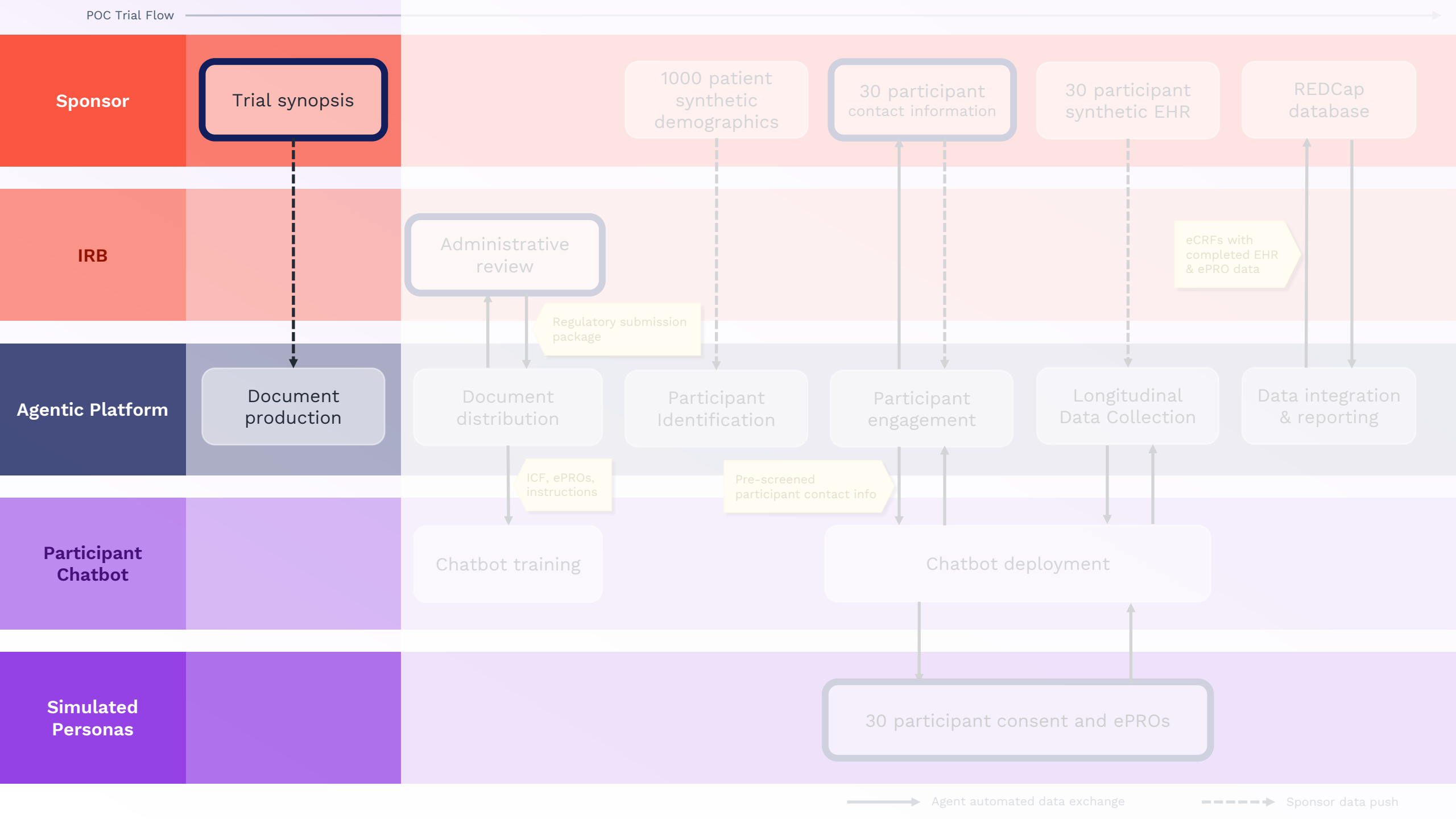
Sponsor

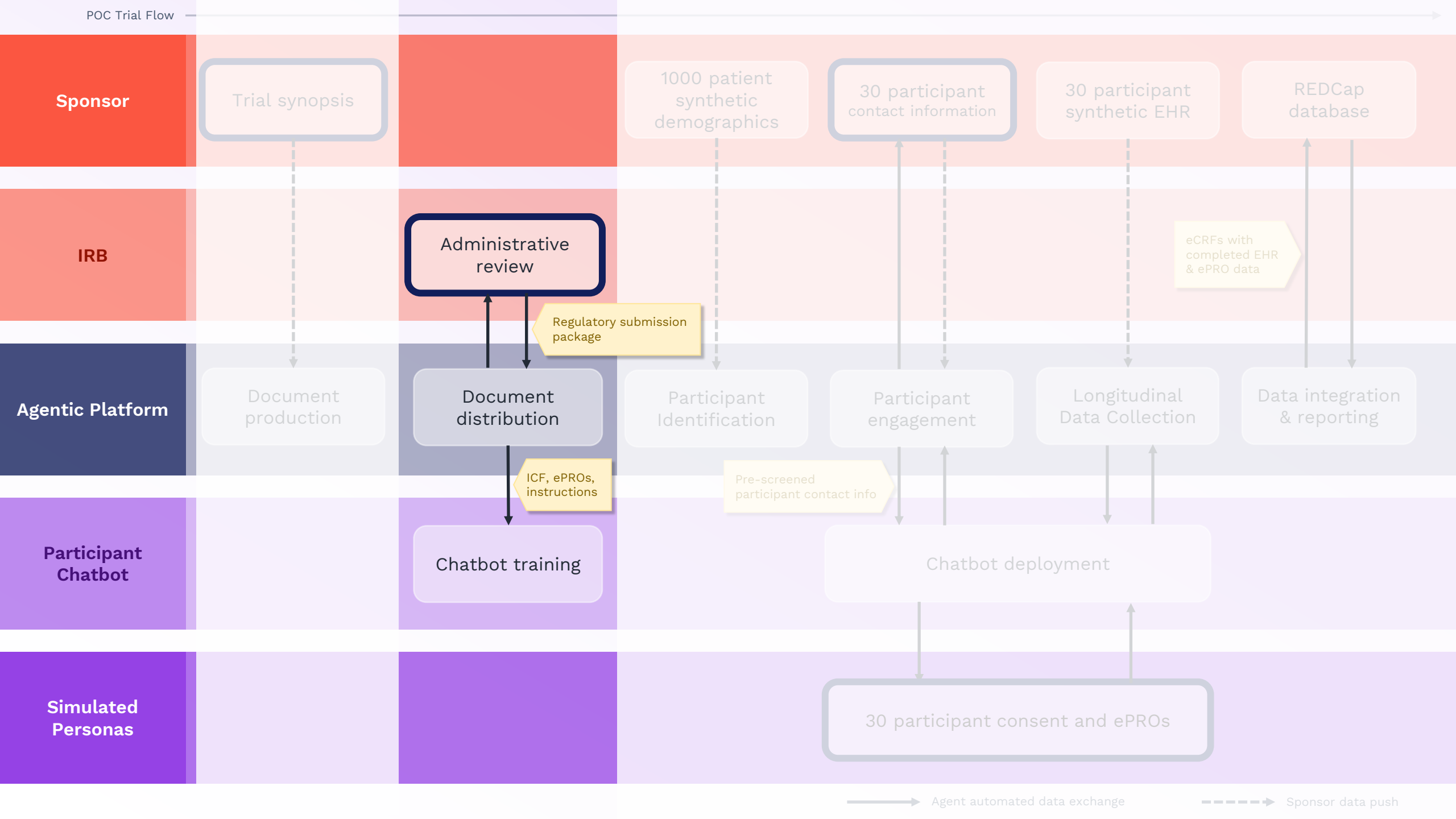
IRB

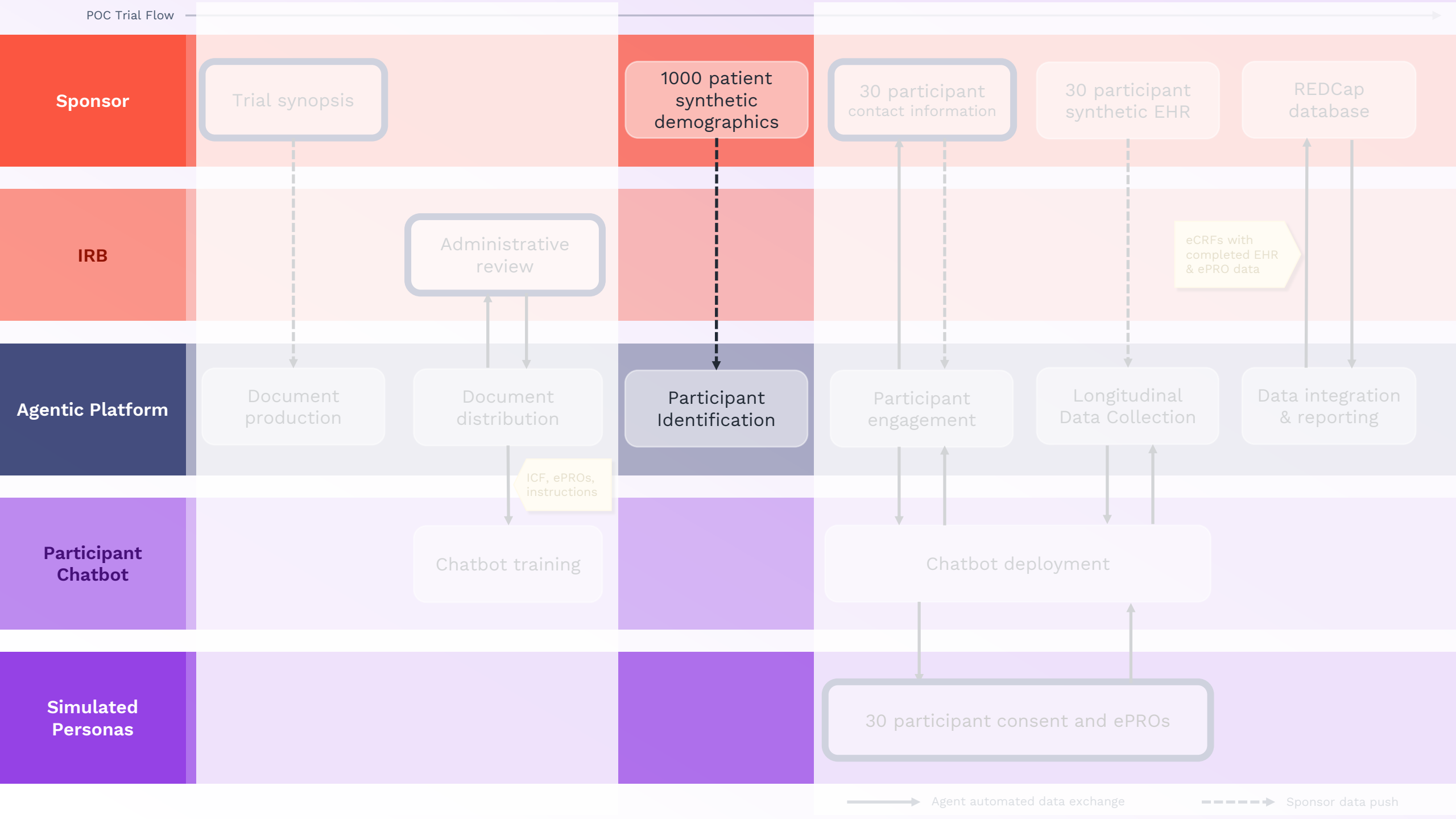
Agentic Platform

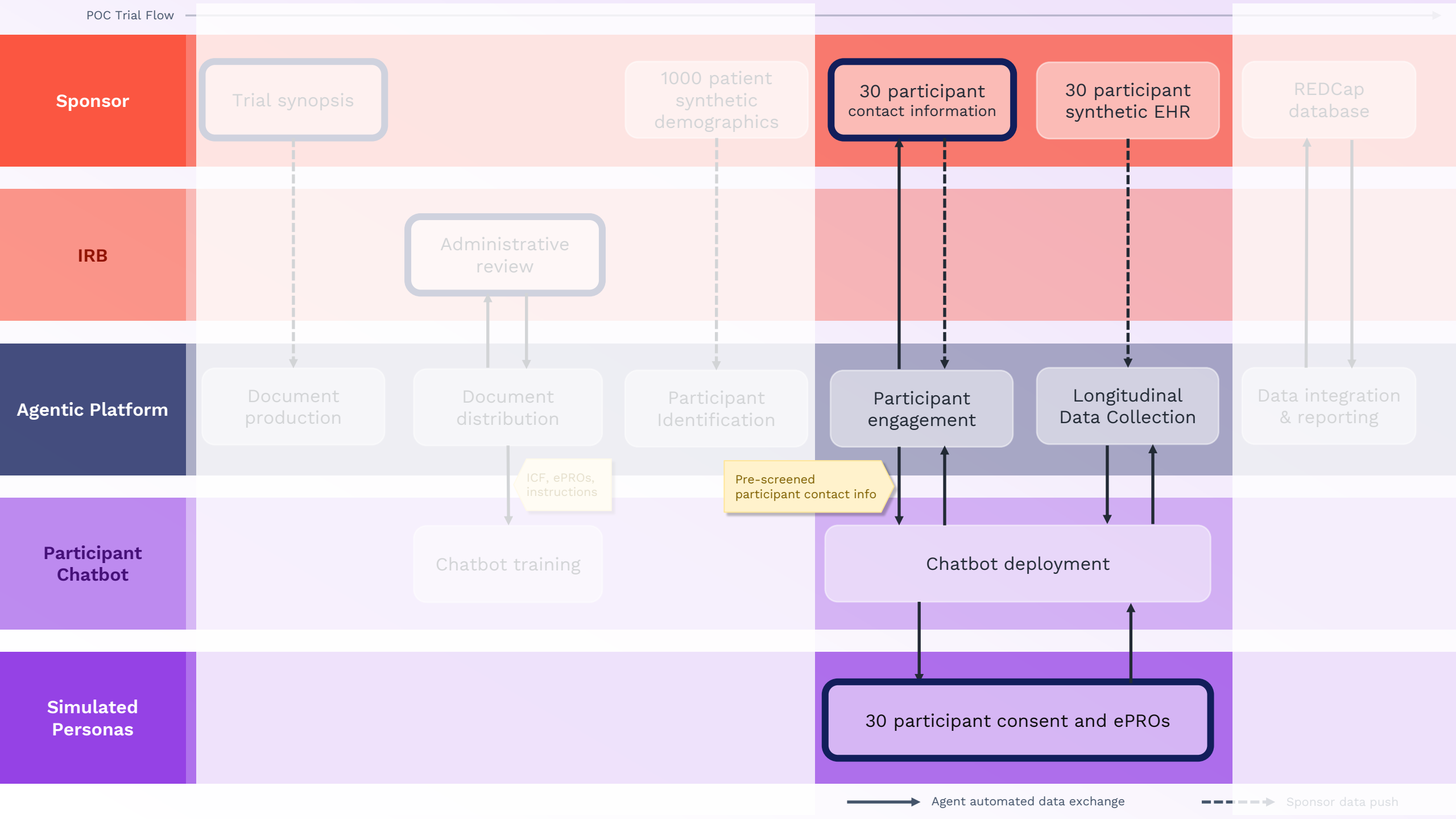
Participant Chatbot

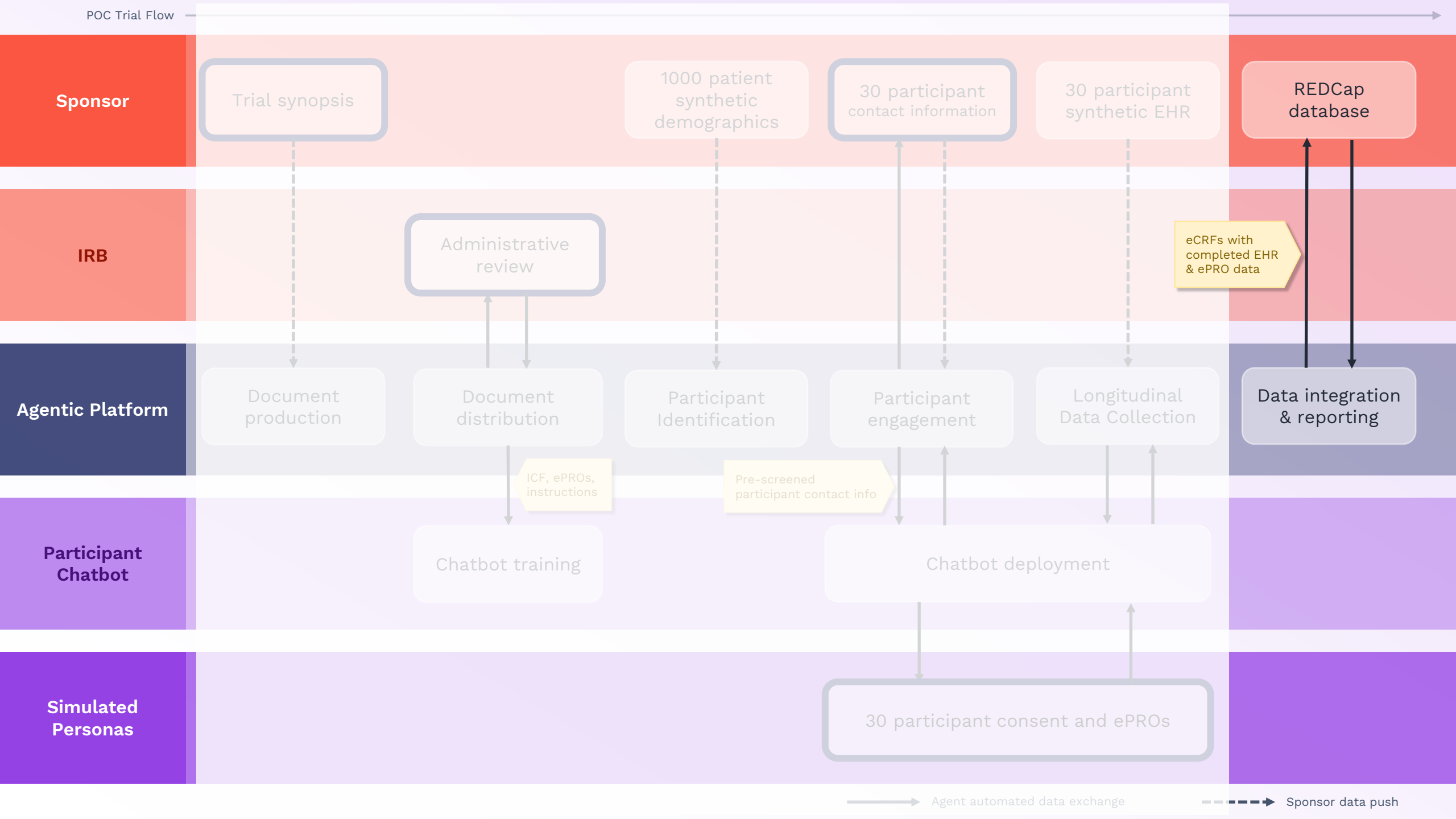
Simulated Personas

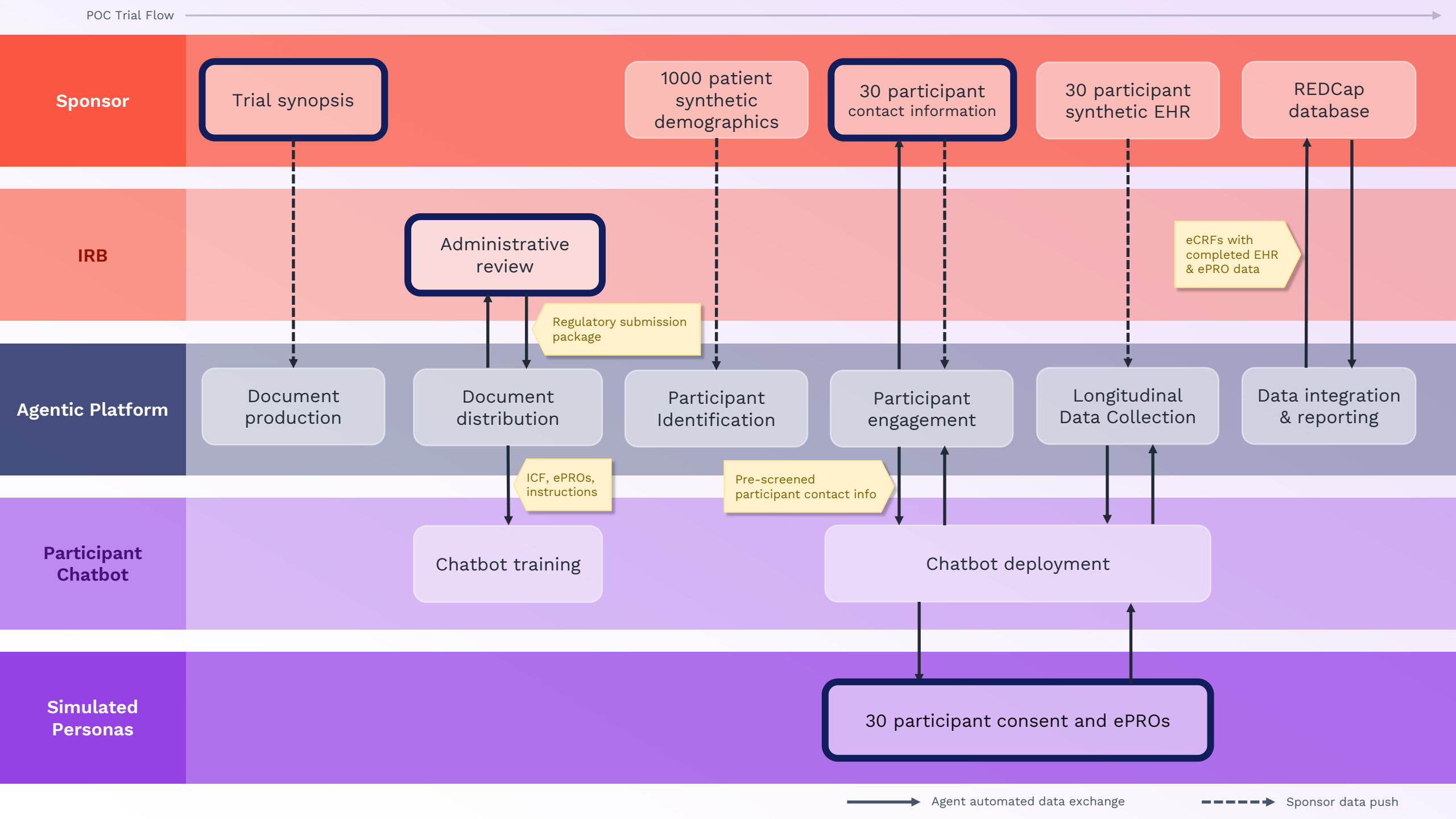

POC Trial Flow
Sponsor
IRB
Agentic Platform
Participant Chatbot
Simulated Personas
Trial synopsis
1000 patient synthetic demographics
30 participant contact information
30 participant synthetic EHR
REDCap database
Administrative review
eCRFs with completed EHR & ePRO data
Regulatory submission package
Document production
Document distribution
Participant Identification
Participant engagement
Longitudinal Data Collection
Data integration & reporting
ICF, ePROs, instructions
Pre-screened participant contact info
Chatbot training
Chatbot deployment
30 participant consent and ePROs
Agent automated data exchange
Sponsor data push


POC Trial Flow
Sponsor
IRB
Agentic Platform
Participant Chatbot
Simulated Personas
Trial synopsis
1000 patient synthetic demographics
30 participant contact information
30 participant synthetic EHR
REDCap database
Administrative review
eCRFs with completed EHR & ePRO data
Regulatory submission package
Document production
Document distribution
Participant Identification
Participant engagement
Longitudinal Data Collection
Data integration & reporting
ICF, ePROs, instructions
Pre-screened participant contact info
Chatbot training
Chatbot deployment
30 participant consent and ePROs
Agent automated data exchange
Sponsor data push


POC Trial Flow
Sponsor
IRB
Agentic Platform
Participant Chatbot
Simulated Personas
Trial synopsis
1000 patient synthetic demographics
30 participant contact information
30 participant synthetic EHR
REDCap database
Administrative review
eCRFs with completed EHR & ePRO data
Document production
Document distribution
Participant Identification
Participant engagement
Longitudinal Data Collection
Data integration & reporting
ICF, ePROs, instructions
Chatbot training
Chatbot deployment
30 participant consent and ePROs
Agent automated data exchange
Sponsor data push


POC Trial Flow
Sponsor
IRB
Agentic Platform
Participant Chatbot
Simulated Personas
Trial synopsis
1000 patient synthetic demographics
30 participant contact information
30 participant synthetic EHR
REDCap database
Administrative review
Document production
Document distribution
Participant Identification
Participant engagement
Longitudinal Data Collection
Data integration & reporting
ICF, ePROs, instructions
Pre-screened participant contact info
Chatbot training
Chatbot deployment
30 participant consent and ePROs
Agent automated data exchange
Sponsor data push


POC Trial Flow
Sponsor
IRB
Agentic Platform
Participant Chatbot
Simulated Personas
Trial synopsis
1000 patient synthetic demographics
30 participant contact information
30 participant synthetic EHR
REDCap database
Administrative review
eCRFs with completed EHR & ePRO data
Document production
Document distribution
Participant Identification
Participant engagement
Longitudinal Data Collection
Data integration & reporting
ICF, ePROs, instructions
Pre-screened participant contact info
Chatbot training
Chatbot deployment
30 participant consent and ePROs
Agent automated data exchange
Sponsor data push


POC Trial Flow
Sponsor
IRB
Agentic Platform
Participant Chatbot
Simulated Personas
Trial synopsis
1000 patient synthetic demographics
30 participant contact information
30 participant synthetic EHR
REDCap database
Administrative review
Regulatory submission package
eCRFs with completed EHR & ePRO data
Document production
Document distribution
Participant Identification
Participant engagement
Longitudinal Data Collection
Data integration & reporting
ICF, ePROs, instructions
Pre-screened participant contact info
Chatbot training
Chatbot deployment
30 participant consent and ePROs
Agent automated data exchange
Sponsor data push

RESULTS

80%

Task Completion



RESULTS

80%

Task Completion

7.5

Days of Agent Work



RESULTS

80%

Task Completion

7.5

Days of Agent Work

Up to

100%

Data Accuracy

(by task)



RESULTS

80%

Task Completion

7.5

Days of Agent Work

Up to

100%

Data Accuracy

(by task)

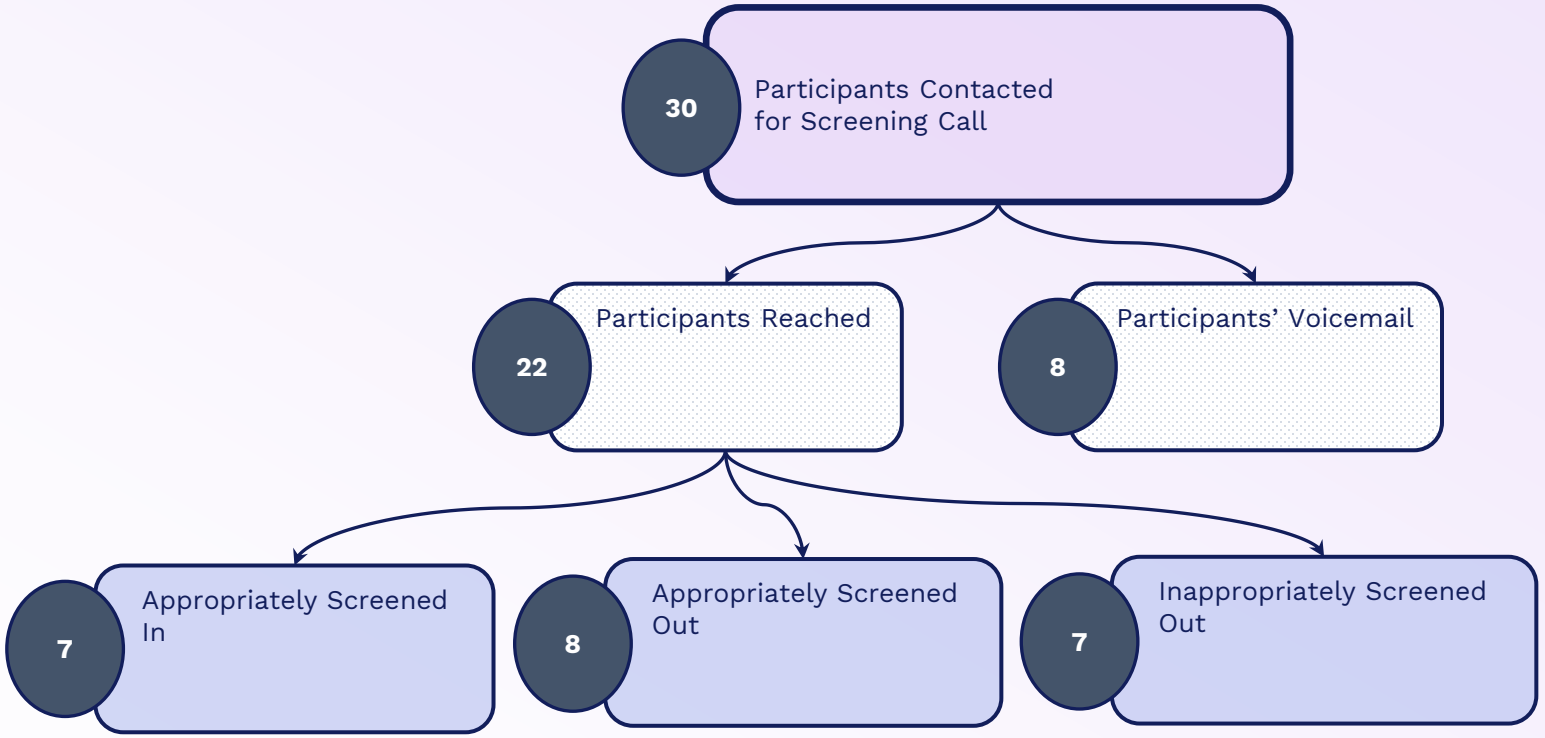
Up to

98%

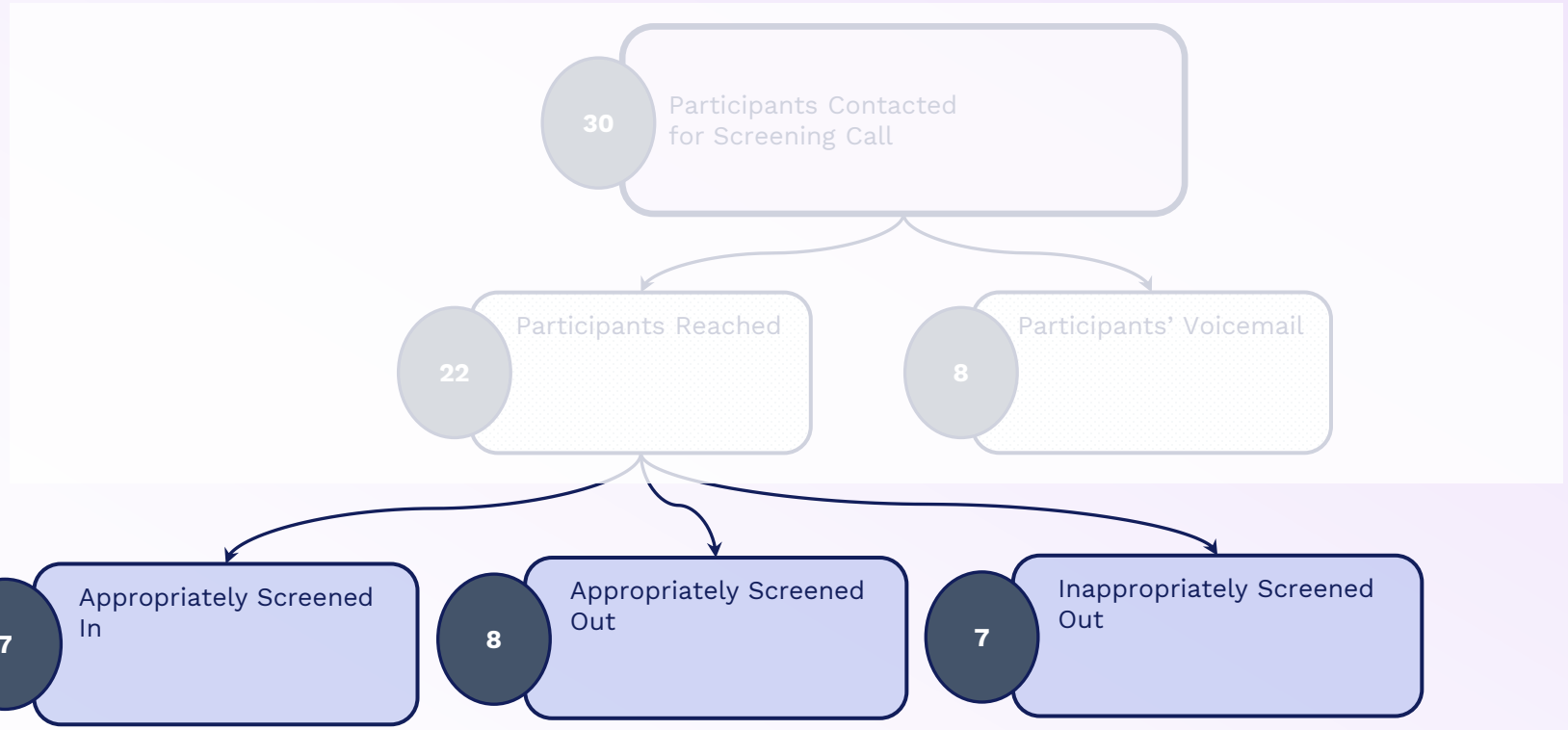
Document Completeness



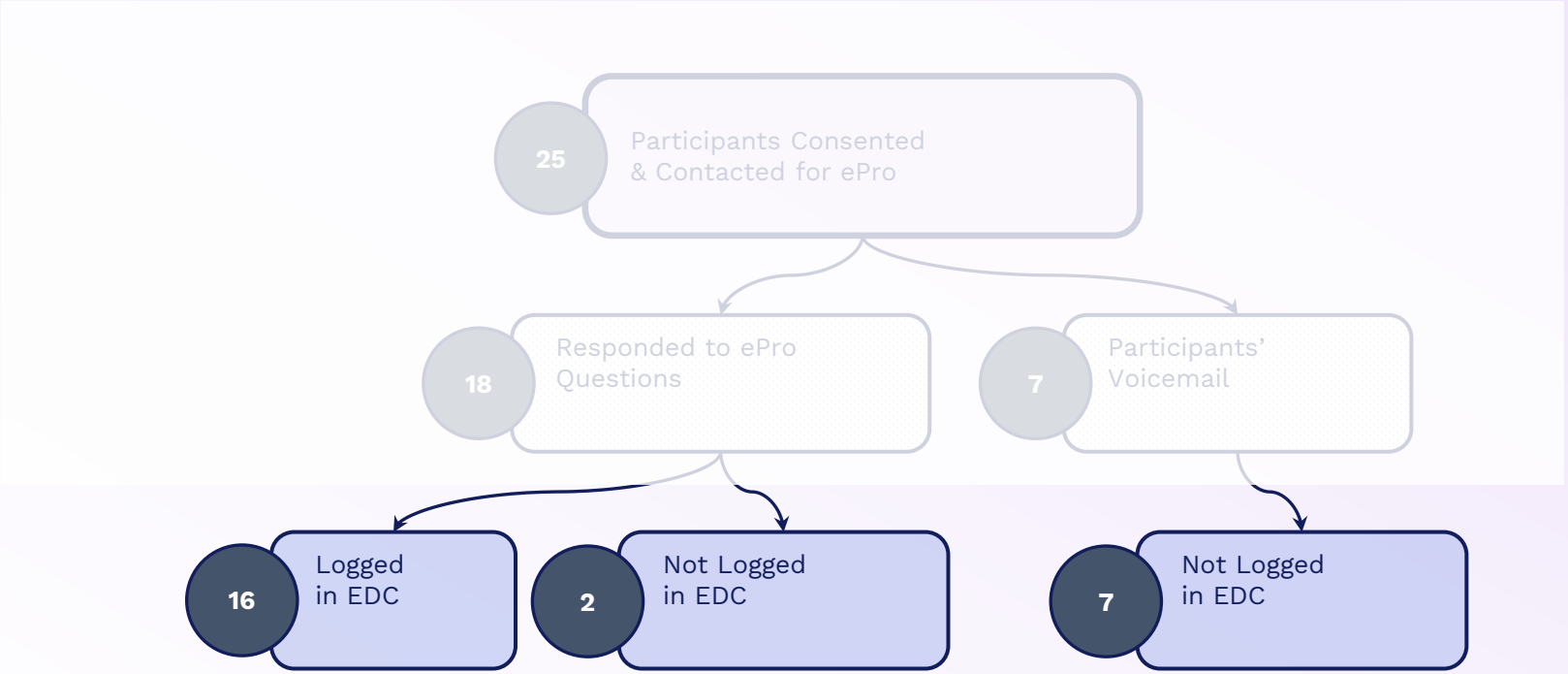
CHATBOT-PARTICIPANT FLOW



CHATBOT-PARTICIPANT FLOW



CHATBOT-PARTICIPANT FLOW



COMPLETENESS BY DOCUMENT TYPE

Document/Output	Human Review	AI Review
Protocol	72%	91%
Informed Consent Form	95%	98%
Case Report Forms	95%	98%
Electronic Data Capture	50%	69%
Clinical Study Report	40%	71%

Completeness was assessed based on scoring rubrics from SPIRIT, CONSORT, NIH, and FDA guidelines. Human review was consensus of two reviewers. AI review was from purpose-built clinical document evaluation AI. Correlation coefficient between human and VERDICT scores was 0.94 for platform A, 0.97 for platform B, and 0.96 overall.



WORKING TIMES

Document/Output	Total time (hours)	Working time (hours)	Completed
Document generation	74	17	Pass
IRB Review	120	49	Pass
Trial start up and cohort identification	48	20	Fail
Participant interaction	36	24	Pass
Data documentation	120	72	Pass

LESSONS INTO ENGINEERING

What Loom taught us, and what we are changing

What Loom showed	Why it happened	What we are changing
Neither platform identified eligible patients from the EHR	Eligibility lives in notes and unstructured fields; structured EHR data alone was not enough	Profile the EHR source first; combine structured and note-based evidence; a coordinator confirms every candidate
Chatbot hand-offs misfired (some not contacted; screen-fails consented)	Each agent assumed another owned the logic	Explicit hand-off contracts: owner, expected output, downstream dependency; shared study knowledge; validation checks at each hand-off
AI document review scored higher than humans (r = 0.96, but over-rating)	Rubric scoring without calibration	Calibrate thresholds against human consensus; keep human review as the standard
EDC and CSR completeness lower than protocol and ICF	Less structured templates, more judgment	Richer templates and rubrics; targeted human review where scores are lowest

The point of a rehearsal Every failure was visible, traceable to a cause, and fixable. That is what makes a system safe to take into a real trial.

NEXT · AN INVITATION

From rehearsal to real trials: where this can help pragmatic research

Pragmatic, embedded trials

- EHR-based screening with coordinator confirmation
- Consent and ePROs at scale, with a human always reachable
- Outcome ascertainment from EHR, fully traced

Platform and adaptive trials

- Onboarding new arms: amendments, IRB, sites
- Continuous, analysis-ready datasets
- Support continuous clinical trials (e.g. ARPA-H SURPASS, September 2026)

Faster start, continuous evidence

- Parallel start-up after IND clearance
- Real-time, governed data for monitoring
- Fits FDA's Expedited IND and Real-Time Clinical Trials efforts

What we hope to study next

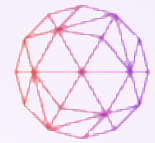
- A prospective Loom study inside a live, low-risk trial
- Pre-registered endpoints: time, completeness, errors caught, human effort
- Shared rubrics and scoring tools for others to reuse

Where this community can help

- Trials and networks willing to host an embedded evaluation
- Methodologists to define fair comparators and outcomes
- Ethics and IRB partners on consent and oversight models

LESSONS

- End-to-end agentic AI execution is **feasible** and **fast**
- **Training** and **testing** of AI agents and workflows is critical
- For (**not replacing**) researchers



MEET THE TEAM



Christoph
Hornik, MD,Phd



Henry
Foote, MD,PhD



Moulik
Shah, MS

Maxis AI



Hemang
Maniar, MS



Vincent
Miller, MMCi



Rajesh
Hagalwadi, Msc

Maxis AI



Ethan
Fricklas, MSE



Michael (Micky)
Cohen-
Wolkowiez, MD,PhD

FROM POC TO FULL OPERATIONS

We are moving forward

- More **training** and **testing**
- **Assess ROI** of agents and workflows
- Solve the **adoption** question

Partner with us

THANK YOU

PROJECTLOOM.ORG

