USER RESEARCH PLAN

AI Model Interaction & Visualization Pilot | Department of Veterans Affairs (VA)

2025 UX Research

1. Background

This research supports the discovery, validation, and design of an AI model interaction interface within the Department of Veterans Affairs (VA). As AI systems increasingly support analytics, oversight, and operational decision-making, the interface must ensure trust, explainability, accessibility, and compliance with design.va.gov and USWDS standards.

The VA's AI modernization context has evolved significantly. In 2025, AI interfaces are no longer purely analytical dashboards — they increasingly serve as decision-support surfaces where the stakes of poor UX include incorrect benefit determinations, missed oversight, and eroded staff trust. This research must account for that elevated-stakes environment.

> ▶ **Addendum:** Agentic AI Interfaces: Research plans in 2025 increasingly scope for agentic AI systems — where the AI takes sequences of actions, not just surfaces information. This plan should capture user expectations around AI agency boundaries, override mechanisms, and accountability trails.
>
> ▶ **Addendum II:** Human-AI Teaming Models: 2025 UX practice distinguishes between AI-as-tool and AI-as-collaborator interaction paradigms. Surfacing which model VA users expect is critical to interface trust and adoption.
>
> **Justification:** *Early-stage research reduces implementation risk, prevents costly redesign, and ensures alignment with federal AI governance expectations.*

2. Objectives & Goals

GOAL 1:

Understand user mental models, workflows, and trust thresholds.

GOAL 2:

Identify high-value UI components and visualization requirements.

GOAL 3:

Define transparency, explainability, and auditability expectations.

GOAL 4:

Translate findings into prioritized, compliant design requirements.

GOAL 5

Evaluate user comprehension and mental models around AI confidence, uncertainty, and failure — not just outputs.

GOAL 6

Assess the emotional and cognitive safety of the interface — including fatigue, over-reliance risk, and trust calibration under uncertainty.

▶ **Addendum:** Outcome-Framed Goals: 2025 best practice rewrites goals as decision-linked outcomes rather than activity descriptions. Each goal should be accompanied by a stated decision it unlocks (e.g., 'GOAL 2 informs which visualization patterns advance to prototype').

▶ **Addendum II:** Continuous Discovery Readiness: Goals should specify whether they apply to a one-time discovery sprint or a rolling discovery cadence. In 2025, ongoing research programs attach goals to product milestones rather than fixed timelines.

Justification: *Clear goals ensure findings directly inform design decisions and backlog prioritization.*

3. Stakeholders & Team

- Product Owner / Program Lead: Decision authority and prioritization.
- UX Strategist / Research Lead: Research integrity and synthesis oversight.
- UX Designer: Design translation and prototype validation.
- Data Science Lead: Technical feasibility and evaluation validation.
- Accessibility / 508 Specialist: WCAG 2.1 AA compliance oversight.

Justification: *Cross-functional participation ensures research findings are actionable and implementable.*

4. Existing Research to Reference

- Prior VA dashboards
- AI pilot evaluations
- design.va.gov standards
- Federal AI ethical guidelines

Justification: *Leveraging prior research avoids duplication and accelerates insight validation.*

5. Audience & Participant Recruiting

- Business Users (6–10): Consume AI outputs for operational decisions.
- Data Scientists (6–10): Evaluate, test, and tune AI models.

Justification: *Distinct personas ensure tailored UI requirements and avoid over-generalization.*

2025 ENHANCEMENTS — RECRUITING

Recruiting criteria should be expanded to capture 2025-relevant participant characteristics:

6. Metrics

Goal	Metric(s)
Clarify interaction needs	Ranked feature priorities per persona

Reduce trust risk	Documented trust signals and transparency requirements
Design readiness	Validated UI direction and backlog inputs
Accessibility alignment	508-compliant component mapping

Justification: *Metrics define what success looks like and ensure research produces measurable value..*

7. Research Focus & Scope

In Scope:

- Goals, workflows, trust signals, visualizations, compliance needs
- **In Scope (additions):** AI failure mode comprehension testing; override and human-review UI flows; notification and alerting UX for model-generated recommendations; multi-modal output formats (text, chart, table, narrative).
- **Explicitly Deferred (not out of scope, but Phase 2):** Mobile/tablet optimization; integration with external data pipelines; longitudinal behavioral tracking dashboar

Out of Scope:

Algorithm training optimization and backend engineering.

Justification: *Clear scope prevents research drift and protects delivery timelines.*

8. Research Questions (Sample)

Business Users:

Insight clarity, confidence, interpretation, workflow integration.

Data Scientists:

Evaluation controls, model comparison, and auditability.

Cross-Cutting:

Explainability, comprehension, and governance alignment.

Justification: *Questions are designed to directly inform UI structure, transparency layers, and control mechanisms.*

Business Users — Expanded

- When an AI recommendation contradicts your prior judgment, what do you need to see before you change your decision?
- What does 'I trust this output' look like to you — what specific signals are present vs. absent?
- How do you currently document or explain decisions influenced by AI outputs to your supervisor or auditors?
- Describe the last time an AI tool surprised you. What happened next?

Data Scientists — Expanded

- Walk me through how you currently compare two model candidates. Where does the current tooling fall short?

- What metadata about a model run is non-negotiable for you to have visible during evaluation?
- How do you communicate model limitations to non-technical stakeholders — what format works?
- If a model's performance degrades post-deployment, how should the UI surface that to you vs. to a business user?
- If the AI made a decision that turned out to be wrong, how would you want that surfaced in the interface after the fact?
- What is the right level of explanation detail — too little leaves you unsure, too much overwhelms. Where's your line?
- Are there contexts where you would want to turn off AI assistance entirely? What would trigger that?

▶ **Addendum:** Question Mapping: Plans map each research question to a specific design decision and deliverable (e.g., 'Q4 → Trust & Transparency Requirements List → UI Pattern: Confidence Indicator'). This prevents 'interesting but unactionable' findings.--Themetaic coding

9. Budget

- Participant incentives: $1,125–$2,000
- Research tools: Existing licenses

Justification: *Incentives ensure participation quality; existing tools reduce overhead.*

10. Timeline (6 Weeks)

- Week 1: Alignment and artifact review.
- Week 2: Recruiting and guide finalization.
- Week 3: Business User interviews.
- Week 4: Data Scientist interviews.
- Week 5: Synthesis and prioritization.
- Week 6: Validation survey and executive readout.

Justification: *A phased timeline balances depth of insight with delivery speed.*\

▶ **Addendum:** Sprint Alignment: Research timelines are increasingly aligned to engineering sprint boundaries (typically 2-week cycles). Week 5 synthesis should ideally land at a sprint planning boundary so findings immediately feed backlog refinement.

▶ **Addendum II:** Parallel Accessibility Track: Schedule 508 specialist review of stimuli/prototypes during Week 2–3, not as a final gate. This surfaces accessibility issues before fieldwork reveals them through participant struggle.

▶ **Addendum III:** Post-Readout Retro: Build a 1-hour research retrospective into Week 6 — what worked, what to change for the next phase. This is now standard in 2025 continuous discovery programs.

11. Methods

Core Methods:

Interviews, contextual inquiry, artifact review, thematic synthesis.

Advanced Additional Methods:

- **Wizard-of-Oz AI Testing:** Simulate AI responses manually behind the interface to test user reactions to different output formats, confidence levels, and failure states — without needing a live model.
- **Paired Cognitive Walkthrough (Business User + Data Scientist):** Have one participant from each persona walk through the same AI output together. Surface where their interpretation diverges — a uniquely revealing method for multi-stakeholder AI tools.
- **Expectation Mapping Protocol:** Before showing any interface, ask participants to sketch or describe what they expect the AI output screen to look like. Compare expectation maps to actual prototypes to identify the largest mental model gaps.
- **Failure Mode Scenario Testing:** Present controlled AI error scenarios (wrong confidence level displayed, model drift not surfaced, misleading chart) and measure whether users catch them, how long it takes, and what in the UI helped or hindered detection.
- **Card Sort for Transparency Hierarchy:** Determine which AI metadata fields (model version, training date, confidence %, data sources) users consider mandatory vs. optional vs. distracting — results feed directly into information architecture decisions.

▶ **Addendum:** Multimodal Session Capture: 2025 research increasingly captures screen recording, facial expression (with consent), and think-aloud audio in parallel — tools like Lookback, Maze, or UserZoom support this natively. Richer capture accelerates synthesis.

▶ **Addendum II:** AI-Assisted Synthesis with Human Validation: Aurelius, Dovetail, and similar tools now support LLM-powered theme generation. 2025 practice treats AI synthesis as a first-pass hypothesis generator, not a final output — always requiring researcher review and participant voice preservation.

12. Tools

Tool	Purpose	2025 Consideration
Zoom / Teams	Remote session facilitation	Enable live closed captions; confirm recording consent flows are compliant with VA policy
Miro / FigJam	Synthesis boards, journey maps	Use Miro AI for first-pass sticky clustering; validate all clusters manually
Aurelius / Dovetail	Research repository + AI synthesis	Confirm data residency is VA-compliant; avoid storing PII in cloud synthesis tools
Figma	Prototype and stimuli creation	Use FigJam for co-design workshop facilitation alongside Figma prototypes

Confluence / SharePoint	Documentation and stakeholder sharing	Establish a single source of truth for findings — avoid parallel tracking in email threads
Lookback / Maze (new)	Unmoderated task testing	Enables async prototype checks between moderated sessions; 2025 standard for hybrid research programs

Justification: *Tool selection supports efficiency, collaboration, and compliance.*

Appendix A: design.va.gov Component Mapping

User Need	UI Component	VA Reference
Insight summary	KPI Cards	Layout & Grid
Trend analysis	Charts	Data Visualization
Model comparison	Side-by-side tables	Tables
Confidence signaling	Alerts	Alerts
Accessibility	Semantic tables / ARIA labels	Accessibility Standards

Appendix B: Enhancements

A. Decision Outcomes and How Findings Will Be Used

2024 research plans increasingly specify the decisions research will unlock. This study will explicitly support: (1) selection of core visualization patterns, (2) definition of transparency layers and controls by persona, (3) prioritization of MVP features and backlog, and (4) governance-ready documentation of trust and accessibility.

Planned decision artifacts:

- Feature Priority Matrix (Must/Should/Could) by persona
- UI Pattern Recommendations with rationale and examples
- Trust & Transparency Requirements List (signals, thresholds, failure states)
- Backlog-ready Epics/Stories with acceptance criteria (508 and auditability included)

B. Research Types and Options

To improve speed and confidence, 2024 best practices emphasize triangulation — combining depth methods (interviews/contextual inquiry) with rapid validation (concept testing, unmoderated checks) and stakeholder alignment (workshops). The following options can be used based on access and timeline.

Recommended additions (select as appropriate):

- Rapid Concept Testing (low-fi prototypes) to validate layout, terminology, and interaction expectations
- Unmoderated Task-Based Prototype Checks for quick signal on comprehension and findability
- Co-Design / Design Studio Workshop to align Business Users, Data Scientists, and stakeholders on UI trade-offs
- Diary / Longitudinal Light Touch (1–2 weeks) for trust calibration and workflow integration learnings
- Expert Review (heuristics + accessibility-first) of early patterns, especially data visualization and form controls
- Journey/Workflow Mapping Session using observed artifacts and system touchpoints
- Quant-lite Validation Survey to rank feature priorities and trust signals across a larger group

C. Analysis Plan and Evidence Confidence

Commonly include an explicit analysis plan to reduce ambiguity and improve repeatability. Analysis will use thematic analysis supported by affinity mapping, with findings tagged by persona, workflow stage, and UI implication. Each key recommendation will include an evidence confidence label (High/Medium/Low) based on recurrence, severity, and triangulation across methods.

Analysis steps:

- Session debriefs within 24 hours (top insights, surprises, open questions)
- Affinity mapping into themes (needs, pain points, trust signals, controls, visualizations)
- Persona-by-persona synthesis and comparison (where needs converge/diverge)
- Translate themes into requirements and UI pattern recommendations

- Prioritize with stakeholders using an impact/effort and risk lens
- Document evidence, confidence, and unresolved questions for next-phase testing

D. Responsible AI, Ethics, and Data Management

Typically include privacy, ethics, and data-handling details — especially for AI systems. This study will ensure participant consent, minimize sensitive data exposure, and document AI-specific risk considerations.

Additions to include in execution:

- Consent language covering recording, data retention, and de-identification
- Data handling plan: storage location, access controls, retention window, and deletion process
- Protocol for handling exposure to sensitive/PII content during contextual inquiry
- Bias/impact prompts where applicable (who is helped/harmed; failure modes)
- Auditability requirements captured as design constraints (logs, provenance, versioning visibility)

E. Expanded Timeline Detail

Below is a more detailed breakdown that preserves the 6-week structure while clarifying day-to-day activities and outputs.

Week 1 (Alignment & Setup)

- Kickoff with stakeholders: confirm decisions needed, constraints, and success measures
- Review existing artifacts (dashboards, model docs, metrics, governance inputs)
- Finalize participant criteria and recruiting plan; confirm incentive approach
- Draft moderator guides (Business Users + Data Scientists) and prototype/concept stimuli plan
- Set up repository structure (notes, recordings, synthesis board, findings doc)

Week 2 (Recruiting & Pilot)

- Recruit and schedule participants; run the screener and confirmations
- Pilot 1–2 sessions internally; refine questions and stimuli
- Finalize session templates and observation roles; confirm accessibility review approach
- Prepare rapid concept artifacts (low-fi) for Week 3–4 testing

Week 3 (Business Users Fieldwork)

- Conduct 3–5 Business User interviews + contextual walkthroughs
- Daily debriefs; log trust signals, visualization preferences, workflow actions
- Optional: quick unmoderated check of comprehension (if prototypes exist)

Week 4 (Data Scientists Fieldwork)

- Conduct 3–5 Data Scientist interviews + evaluation workflow walkthroughs
- Capture requirements for controls, comparisons, metrics, and auditability
- Inventory artifacts (dashboards, notebooks, eval reports) for pattern extraction

Week 5 (Synthesis & Prioritization)

- Affinity mapping and persona-by-persona synthesis
- Create feature priority matrix + UI pattern recommendations
- Map findings to design.va.gov components and accessibility requirements
- Hold stakeholder prioritization workshop (impact/effort + risk)

Week 6 (Validation & Readout)

- Optional: short validation survey to rank features/trust signals
- Finalize readout deck and written report; include evidence confidence labels
- Deliver backlog-ready epics/stories and next-phase testing plan

F. Deliverables

- Executive Summary (1 page) + Readout Deck
- Persona-specific Workflow Maps and Key Jobs-To-Be-Done
- Feature Priority Matrix (Must/Should/Could) with examples
- UI Pattern Recommendations mapped to design.va.gov components
- Trust, Transparency, and Auditability Requirements List
- Backlog-ready Epics/Stories + Acceptance Criteria
- Open Questions and Next-Phase Research Plan (usability + longitudinal)

Document Legend

Marker	Meaning
Justification: ...	Original rationale text preserved from source document
2025 ENHANCEMENTS label	New section added under each original section — no original content removed
▶ **Addendum: ...**	Specific 2025 UX research trend or emerging practice — clearly distinguished from original content