



# EP Gen AI Hub: PoC Phase 1 User Feedback Report

---

Pilot Assessment, Performance Metrics, and Strategic Roadmap

Internal Assessment Document | December 2025  
Internal Assessment Document | December 2025

# Executive Summary: A Proof of Necessity, Not Yet of Reliability

The Pilot Phase confirms that EP staff are eager to adopt AI tools to increase efficiency. However, trust is fragile due to "hallucinations" in the on-premise models.



## Adoption Driver

The "Safe Harbor" promise. Users prioritize data privacy above all else to process sensitive EP documents.



## Performance Gap

A stark contrast exists between high-performing Cloud models (GPT-4) and struggling On-Premise models.

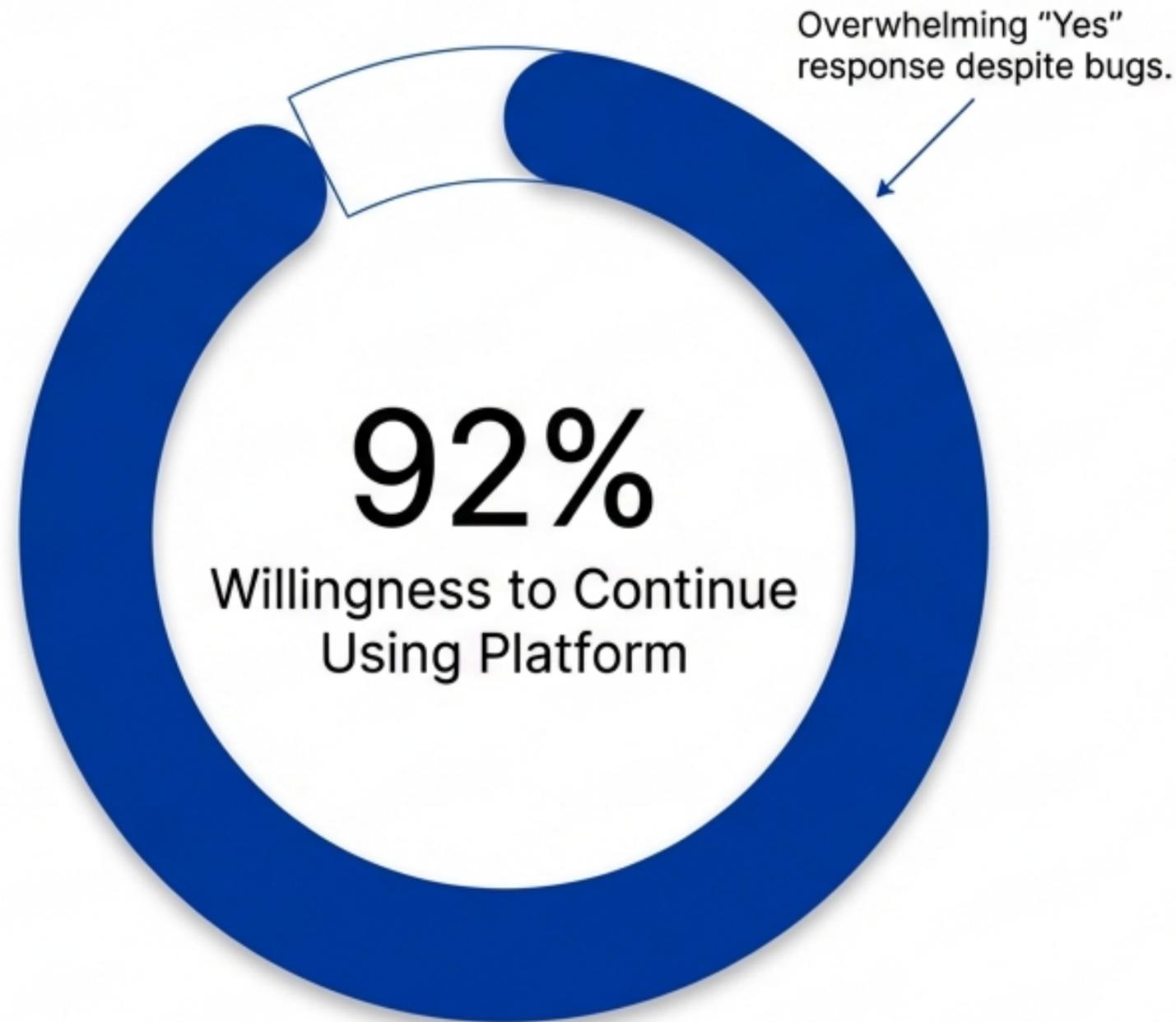


## Critical Risk

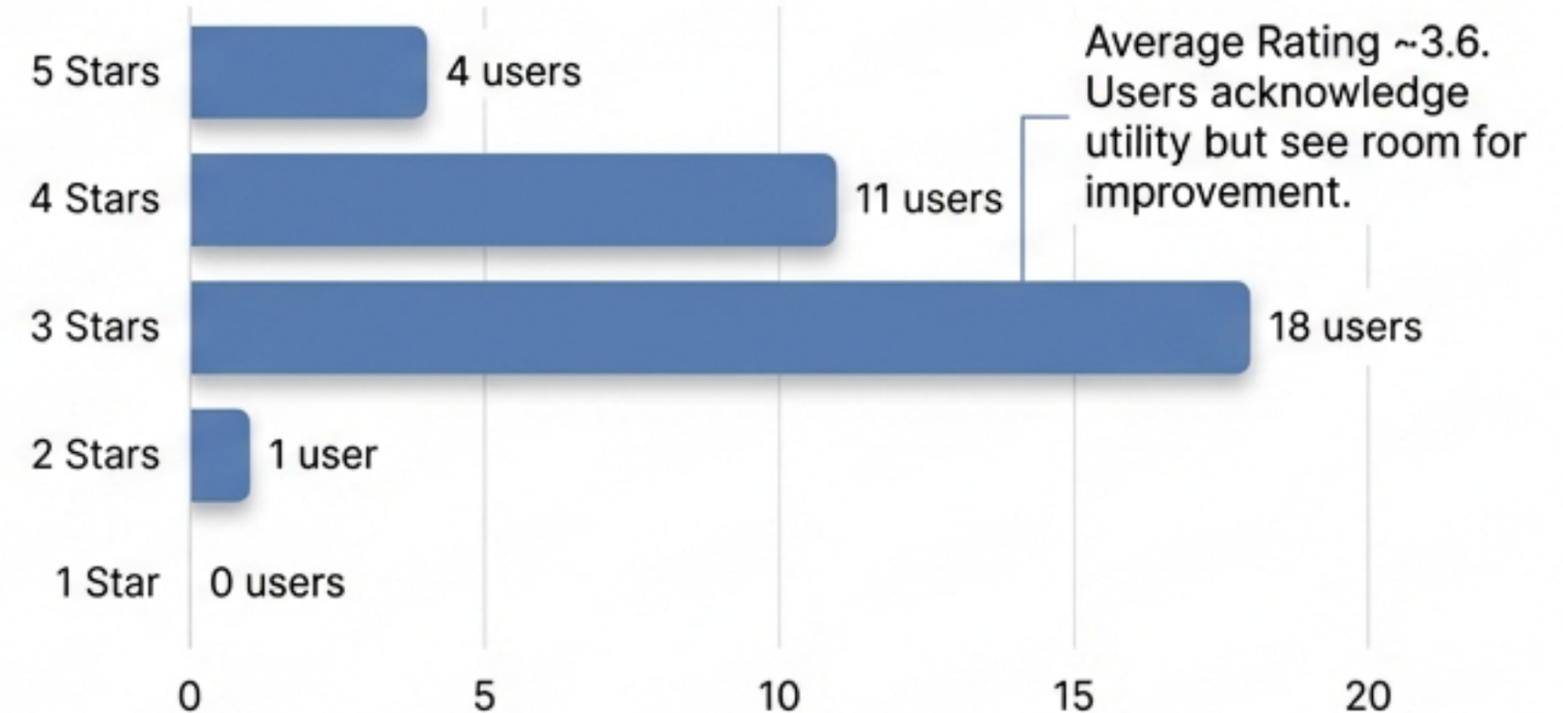
"Hallucinations" regarding Staff Regulations and legal texts pose a significant liability if unmanaged.

**The Verdict:** The PoC was successful in validating the use case, but the technology requires requires a "Human-in-the-Loop" strategy before full deployment.

# Users Encountered Friction but Remain Committed to Adoption



## User Experience Ratings (1-5)



“The willingness to adopt is high despite moderate performance ratings—users want this to work.”

# The Core Value Proposition: A ‘Safe Harbor’ for Sensitive Work

“It represents the only safe environment for certain type of work (e.g., draft briefings/studies that are not public). — User 9

“I occasionally use AI for work-related reasons, and the thing that worries me about the commercial ones is that they are not private.”  
— User 19



“The fact that I can share practically anything without being concerned about data privacy.”  
— User 15

**Differentiation:** In a landscape of commercial AI, the EP Gen AI Hub's “Killer Feature” is data privacy.

# The Performance Gap: Cloud Intelligence vs. On-Premise Security

High Performance

## Cloud Models (GPT-4 / Claude / Mistral)

- Praised for structure, reasoning, and following complex instructions.
- Public models were obviously more powerful.

“GPT provided the best replies... Llama provided decent results but required significant effort.” — User 25

The Trade-off:  
Security vs. Capability

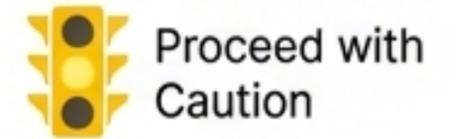
Privacy Secure / Reliability Issues

## On-Premise Models (Gemma / Llama)

- Struggled with complex instructions.
- Prone to stalling on workflows and formatting errors.

“I fear the internal models are still hallucinating a lot... it is difficult to believe anything coming out of the internal models.” — User 15

# Feature Deep Dive: Document Analysis (RAG)



## The Good

- ✓ Efficient at summarizing simple texts.
- ✓ Effective for drafting emails from bullet points.
- ✓ Good for “diplomatic rewriting”.

## The Bad: Friction Points

### Broken References

““The RAG feature always showed the same highlighted parts of the document independently of the reference selected.” — User 25

### Format Blindness

- ✗ Failed to recognize “crossed out text” in amendments, leading to confusion on deletions vs. additions.

### Upload Bugs

- ⚠ Persistent “10MB limit” errors and “Contact Administrator” alerts.

# Specialized Assistants: Clear Writing Wins, Legal Summary Struggles

**EFFECTIVE**



## Clear Writing Assistant

Used for “diplomatic rewriting” and “proofreading in accordance to EU/EP style guides”.

“The clear language Assistant does indeed draft in a language that is more appropriate for communication purposes.” — User 18

**UNRELIABLE**

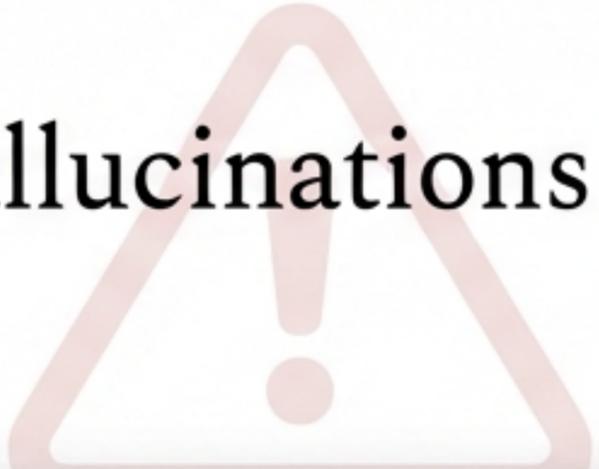


## Legal Summary Assistant

Used for analyzing Directives and Staff Regulations.

“The Legal Summary Assistant... failed to analyze the entire document” and often hallucinated links or regulations.” — User 25

# The 'Red Flag' Risk: Hallucinations in Legislative Contexts

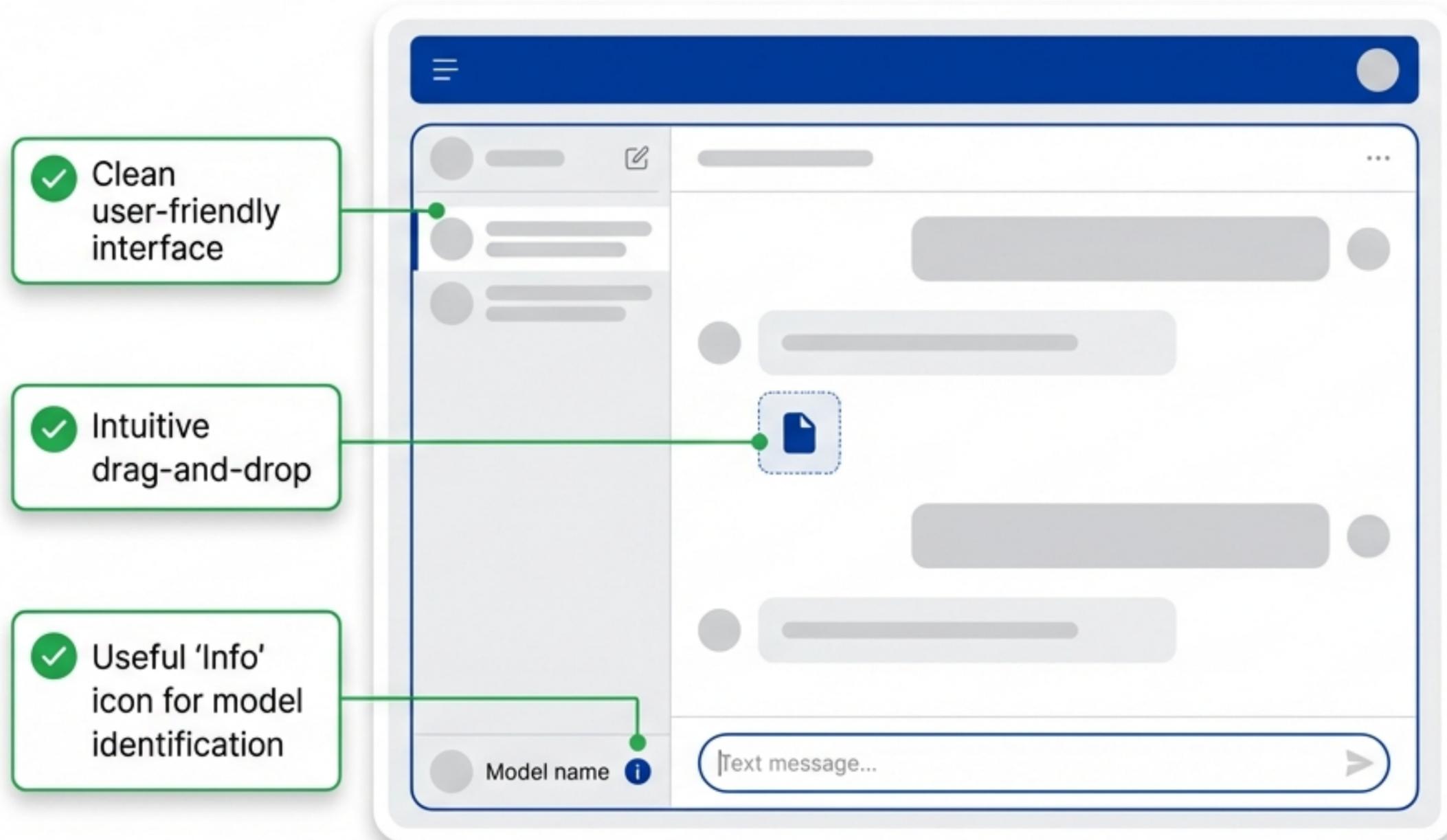


## Case Study: The Staff Regulations Disaster (User 22)

- 1. **The Prompt:** Summary of rights... Art 42a of EP Staff Regulations.
  - 2. **The Hallucination:** AI Claimed 18 weeks of leave.
  - 3. **The Reality:** Actual regulation grants 6 months.
- 
- 1. **The Prompt:** What is Art 60 about?
  - 2. **The Hallucination:** Discussed 'ethics'.
  - 3. **The Reality:** Actually covers 'absences'.

**Implication: The system cannot be used for legal/HR advice without mandatory verification protocols. "With full confidence the system is providing information which is not backed up."**

# User Experience: Clean Interface Marred by Stability Issues



## The Bug List

- 🐛 **Generic Errors:** "Something went wrong" messages requiring refreshes.
- 🐛 **Persona Confusion:** Collaborative chats mix up who is speaking (User 40).
- 🐛 **Disappearing History:** Chat tabs vanish or fail to load.
- 🐛 **Language Switching:** Prompting in English but receiving answers in Italian (User 8, 23).

# Current Utility: How the EP is Using AI Today



## Drafting

Diplomatic rewriting, emails, converting bullet points to prose.



## Summarization

Condensing long directives, reports, and background notes.



## Coding

VBA code for Excel and troubleshooting IT issues (High satisfaction). ✓



## Translation

Quick drafts (noted as "too literal" by some users). ⚠

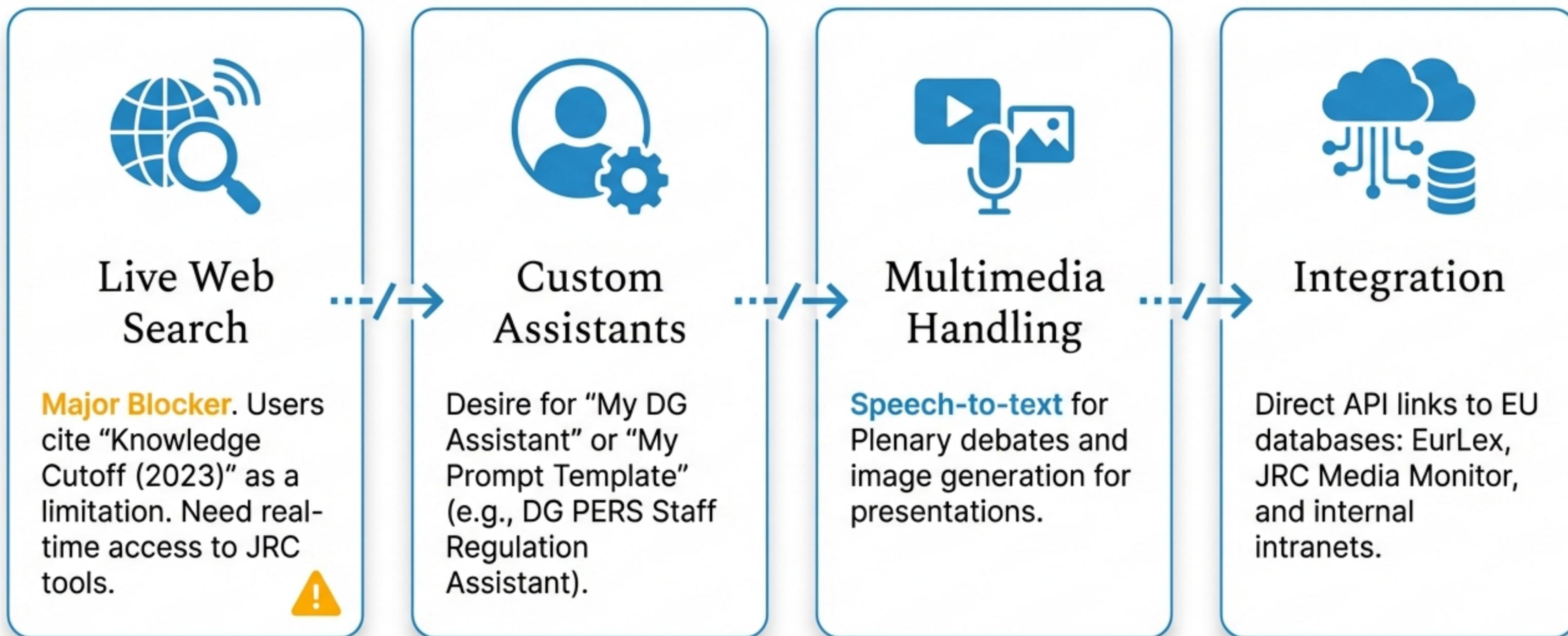


## Brainstorming

Generating structure for documents and speeches.

# User Wishlist: The Gap Between Prototype and Daily Driver

## Top 4 Requested Features



# Quick Technical Wins for Phase 2



## Fix Reference Behavior

Links must open in a New Tab, not replace the chat window (User 22 feedback).



## Clarify Model Capabilities

Add 'Best for...' tooltips to guide users on when to use Gemma vs. Llama vs. GPT.



## Error Messaging

Replace generic 'Contact Admin' errors with specific guidance (e.g., 'File too large' for 10MB limit).



## PDF Parsing

Fix OCR for 'crossed-out' text in legislative amendments.

# Strategic Recommendation: The "Human-in-the-Loop" Workflow



## Spectral bold

**Positioning:** Market the tool as a "Drafting Partner", not an "Oracle".

## Spectral bold

**Training:** Users must be trained to verify output, specifically for legal citations.

## Spectral bold

**Disclaimers:** Add context-specific warnings on legal/HR queries.

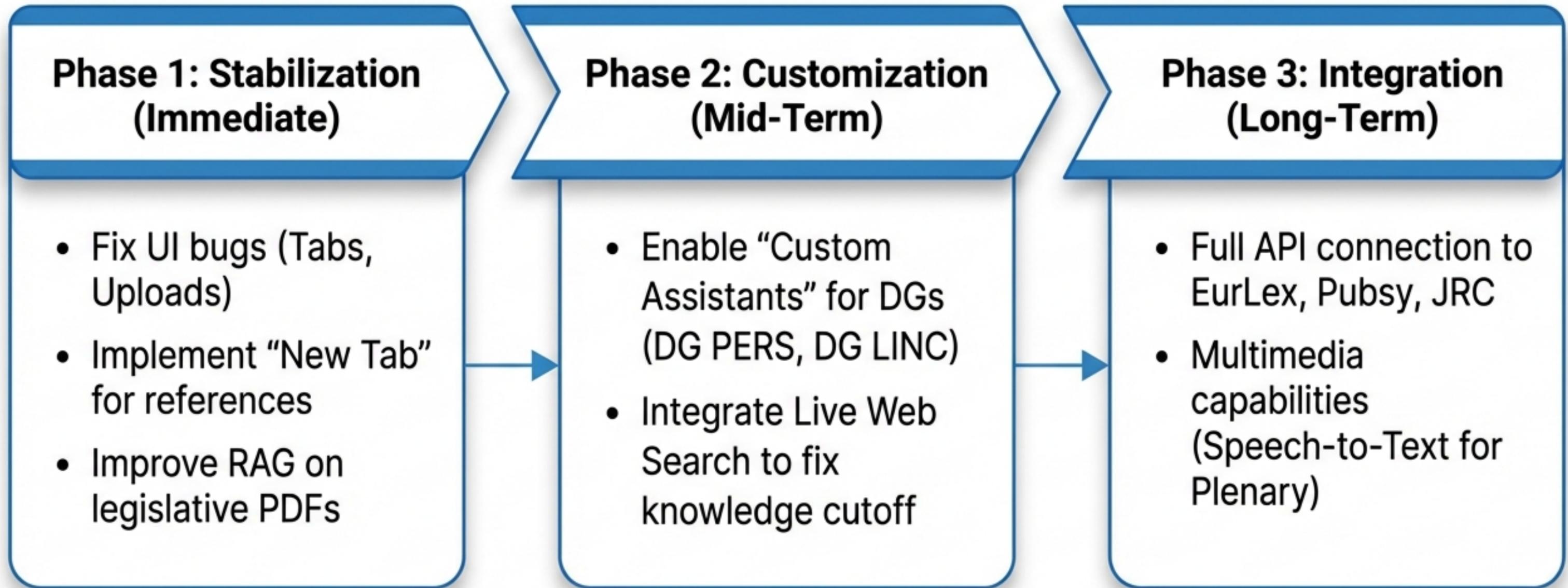
# Verdict: Strong Readiness for Adoption, Conditional on Trust

“With improvements, this is the tool that would simplify a lot the work in the EP.” — User 15

## Condition of Success

Trust is fragile. If the secure 'On-Prem' models don't improve, users will revert to insecure public tools on personal devices. Phase 2 must prove reliability.

# Strategic Roadmap: Path to Maturity



**The PoC proved the hunger. Phase 2 must prove the reliability.**