Red Hat AI Studio: Vision Framing Workshop Report

Date: May 27, 2025

Participants: Adel Zaalouk, Andy Braren, Ann Marie Fred, Burr Sutter, Dash Copeland, Jason Greene, Jeff DeMoss, Jenn Giardino, Jon Nemargut, Peter Double, Tony Kay [cite: 1]

I. Executive Summary

A. Workshop Goal: This intensive, multi-day workshop was convened to bring together key stakeholders from diverse backgrounds—including UX, engineering, and product leadership—to rapidly define and align on the vision for the next iteration of Red Hat OpenShift AI (RHOAI)[cite: 1, 165]. The primary objective is to accelerate the traditionally lengthy process of specification and discovery, condensing weeks or months of work into a single week[cite: 140]. By leveraging a hybrid approach combining established design thinking practices with emerging Generative AI (GenAI) capabilities for "vibe coding" and spec generation, the workshop aims to produce a foundational specification document and, ultimately, a functional, albeit non-production-ready, prototype[cite: 139, 154, 157, 181, 182]. This prototype will serve as a tangible representation of the collective vision, enabling faster iteration and providing clear guidance for the engineering teams[cite: 9, 140, 159].

B. Vision Framing Outcome: The initial "Vision Framing" exercise, which tasked participants with envisioning a highly successful product launch at the following year's Red Hat Kubecon, yielded a strong, cohesive vision[cite: 267, 268]. A central theme emerged, encapsulated by the mantra **"Open Private AI"**[cite: 420, 690, 709]. This emphasizes Red Hat's commitment to open-source principles while highlighting the critical market need for secure, private, and potentially air-gapped AI solutions[cite: 699, 702, 513]. This core idea was further articulated in the headline: **"Scalable, Trusted, and Easy-to-Use Hybrid GenAI"**[cite: 715, 721]. This vision positions the Red Hat AI Studio as a platform empowering a range of users to:

- **Rapidly prototype** and deploy AI solutions, particularly agentic systems[cite: 318, 815].
- Leverage **their own data** and maintain **choice in models**, with a strong preference for enabling **open models**[cite: 321, 582].
- Operate within a **secure, private, and trusted environment**[cite: 365, 841].
- Benefit from a **hybrid architecture** that supports both on-premise needs and cloud flexibility[cite: 493, 501].
- Utilize a **unified platform** to build, evaluate, deploy, and manage AI applications[cite: 375, 394].

C. Key Challenges Identified: During the visioning discussions, several critical challenges and considerations for the platform's success were identified:

- **The "Blank Page" Problem:** Many existing and potential customers lack defined use cases and need guidance and templates to begin their AI journey[cite: 889, 902, 905].
- **GPU Adoption vs. Hybrid Flexibility:** Balancing the strategic goal of driving GPU adoption on Red Hat platforms with the customer need to utilize external models (like Claude or Gemini) and run workloads in various hybrid configurations[cite: 528, 536, 557, 493].
- **Simplifying Complexity:** Making sophisticated but essential concepts, such as model/agent evaluations (evals) and observability, accessible and integrated seamlessly into the workflow[cite: 731, 740, 751, 893].
- **Creating a Unified Experience:** Overcoming the tendency to build separate components and instead delivering a truly interconnected platform where different tools and tasks flow together logically[cite: 74, 75, 822].
- **Security & Trust:** Addressing the inherent lack of security in many current GenAI approaches and building a platform where security and compliance are foundational[cite: 31, 32].

D. Next Steps: The outcomes of this vision-framing exercise serve as the direct input for the subsequent phases of the workshop. The immediate next steps involve:

- Developing detailed **personas** based on the initial discussions to clearly define the target users and their needs[cite: 184].
- Creating AI-assisted **journey maps** to understand user workflows and identify pain points and opportunities[cite: 192].
- Generating a **product specification** and **technical stack** proposal[cite: 194, 196].
- Defining a **coding task list** to guide the GenAI in building the initial **prototype**[cite: 198, 199].

II. Workshop Introduction & Context

A. Purpose: The "Reorganizing RHOAI for GenAI Workshop" was initiated to address the need for a consolidated and accelerated approach to defining the future direction of Red Hat's AI platform, tentatively named the "AI Studio". Recognizing that numerous stakeholders held valuable but potentially diverse perspectives, the workshop aimed to create a shared space to "lay out everybody's perspective" and capture this collective intelligence. The core driver was to

radically compress the development lifecycle—taking a process that typically spans weeks or months down to a single week. The goal is not to produce production-ready code, but rather to create a functional prototype and a robust specification document. This "jumps us ahead," providing engineering teams with a clear, tangible representation of the desired end-state, fostering better alignment and facilitating feedback loops early in the process. It serves as an experiment to move UX and development forward at "light speed," particularly through rapid discovery and research phases.

B. Methodology: The workshop employs an innovative, hybrid methodology, blending established design thinking principles from the Open Practice Library with the cutting-edge "Vibe Coding" and "Spec Coding" approaches. The process is structured for rapid iteration:

- **Morning Sessions:** Dedicated time for stakeholder discussions, brainstorming, and input gathering, capturing as much detail ("word vomit") as possible into transcripts.
- **Afternoon/Nightly Processing:** Utilizing Generative AI (specifically aiming for Gemini Pro 2.5 via the Gemini AI Studio API, with Claude as a backup) to process the transcripts and existing documentation. This AI assistance helps generate drafts of personas, journey maps, and eventually, the product specification and prototype code.
- **Morning Reviews:** Each subsequent morning begins with a review of the AI-generated outputs, allowing participants to discuss, edit, and refine the materials, ensuring alignment and accuracy.
- **Iterative Spec Development:** The aim is to build a full package of documents iteratively throughout the week.
- **Process Improvement Focus:** A key meta-goal is to treat this workshop as a first draft of a *new process*, actively identifying ways to improve this AI-powered design and specification method for future use on other products. This is acknowledged as a new and experimental approach, requiring patience and learning from all participants.

C. Workshop Agenda Overview: The workshop is structured across five mornings (8 AM to Noon), with a flexible day built-in, and GenAI work occurring in the afternoons:

- **Day 1 (Tuesday):** Focus on "Vision Framing" to establish the end goal and "Persona Mapping" to define the target users. *Afternoon: GenAI generates detailed persona reports.*
- **Day 2 (Wednesday):** Review the generated personas. Conduct "Feature Discovery" based on personas. Perform AI-assisted "Journey Mapping". *Afternoon: GenAI generates the first pass at a product spec.*

- **Day 3 (Thursday):** The "heaviest Engineering Day". Generate, review, and finalize the "Tech Stack". Create, review, and finalize the "Coding Task List". *Afternoon: GenAI begins creating the prototype.*
- **Day 4 (Friday):** "Flex Day" or "Rabbit Hole Day". Reserved to address any potential "derailers" or important topics requiring deeper discussion before final prototype generation. This day can also be used to bring in additional experts if needed.
- **Day 5 (Monday):** The "Big Reveal". Conduct a "Design Studio" walkthrough of the prototype to gather feedback. *Afternoon: Perform a "Remix" or second version based on feedback.*

III. The Vision Framing Exercise: "Remembering the Future"

A. Scenario: To initiate the workshop and align participants around a positive, ambitious goal, the first activity was "Vision Framing," presented as a "Remembering the Future" exercise. Participants were asked to project themselves one year forward, imagining they were at the next Red Hat Kubecon. In this imagined future, the new Red Hat AI Studio, with its Generative AI capabilities, has been launched and proven so successful and impactful for Red Hat and its customers that the CEO, Matt Hicks, is featuring it prominently in his keynote address. This scenario sets a high bar for success and encourages participants to think big about the potential value and impact of the product.

B. Task: Working individually on a shared Miro board, each participant was asked to "claim" a section and flesh out a hypothetical "Kubecon Announcement" based on the scenario. The specific elements required for each announcement were:

- **Headline:** A concise, compelling title for Matt Hicks' talk or the announcement itself.
- **Customer Quote:** A "pithy quote" that effectively captures the core value and benefits customers are receiving from the new experience.
- **Key Benefits:** A list, ideally starting with three but allowing for more, detailing the primary advantages and achievements the product delivers for customers.

Participants were given approximately 15 minutes to complete this task, encouraged to think from the customer's point of view and focus on what would make them feel successful.

C. Objective: The primary goal of this activity was twofold. First, it aimed to get all participants into a "really positive and creative state of mind", fostering an optimistic and aligned starting point for the workshop. Second, it served as an initial, high-level requirements-gathering exercise, designed to capture the envisioned end-state, the core value proposition, and the most critical success factors from each stakeholder's perspective. By focusing on a future success story, the

exercise encourages a focus on outcomes and benefits rather than immediate constraints or features.

Okay, let's proceed to **Section IV: Synthesized Vision: Headlines & Mantras**. In this section, we will analyze the various headlines and keywords proposed during the Vision Framing exercise and detail the converged-upon vision, digging into the rationale and strategic thinking behind the selected language.

You are absolutely right to point that out, and I apologize for misrepresenting the synthesized headline. My access is limited to the text transcript, and I cannot view the Miro board where the final consensus might have been captured visually. Thank you for providing the corrected version: **"Your Hybrid GenAI Platform" that is Scalable, Trusted, Sovereign-ready**.

Let's rewrite Section IV.B based on this corrected information, ensuring we ground the rationale in the transcript's discussions.

IV. Synthesized Vision: Headlines & Mantras

Following the presentation of individual vision cards, a discussion ensued to synthesize these ideas into a core message. The group aimed to identify a short, memorable mantra and a more descriptive headline that would capture the essence of the Red Hat AI Studio's value proposition, particularly as it might be presented at a high-profile event like the Red Hat Kubecon.

A. Core Mantra: "Open Private AI"

- **Selection & Rationale:** Proposed initially by Burr Sutter[cite: 420], this three-word phrase gained traction as a potential mantra[cite: 709]. It was seen as a powerful, concise way to communicate key differentiators.
    - **"Open":** This word resonates deeply with Red Hat's core identity and mission[cite: 705]. Peter Double articulated a strong belief that open-source AI is "absolutely necessary for humanity" [cite: 699] as a means to democratize AI and prevent control by a few large corporations[cite: 700, 701]. It represents a commitment to transparency and community-driven development.
    - **"Private":** This word was chosen for its strong market resonance[cite: 515]. Burr Sutter noted significant interest in "Private AI" based on online trends (like

YouTube views), suggesting it effectively captures customer attention[cite: 507, 509]. It directly addresses a critical customer concern and a key Red Hat strength: the ability to provide secure, air-gapped, and controlled AI environments[cite: 501, 513, 514], which is a significant differentiator compared to purely SaaS-based offerings. Adel Zaalouk added that "Private" is a strong "catching statement" [cite: 675] that can draw people in and potentially encompass related concepts like "Sovereign" [cite: 681] for specific audiences.

- **Insight - Moving Beyond "Hybrid" (in the Mantra):** While "Hybrid" has been a long-standing Red Hat term[cite: 503], and its meaning often implied "Private"[cite: 504], there was a desire to use "Private" more explicitly in the *mantra*[cite: 691]. "Private" was seen as having a stronger, more direct impact in the current AI landscape and better highlighting the immediate differentiator. However, as we'll see, "Hybrid" was considered essential for the broader description.
- **Supporting Keywords:** The discussion identified several related terms crucial to the vision: Secure[cite: 695], Responsible[cite: 671, 694], Trusted[cite: 708, 53], and Sovereign[cite: 676].

B. Elaborated Headline: "Your Hybrid GenAI Platform" - Scalable, Trusted, Sovereign-ready

- **Purpose & Structure:** Recognizing that a three-word mantra needs further explanation, the group aimed to craft a more descriptive phrase. Based on your input reflecting the Miro board consensus, this emerged as **"Your Hybrid GenAI Platform"**, further defined by its key attributes: **Scalable, Trusted, and Sovereign-ready**. This structure positions the offering and then lists its core strengths.
- **Component Rationale:**
    - **"Your Hybrid GenAI Platform":** This phrasing firmly plants the "Hybrid" flag[cite: 719, 721], reflecting the crucial understanding that the platform must support diverse deployment models—both on-premise and connected to cloud/SaaS resources[cite: 493, 497, 500, 503]. It explicitly moves away from *only* "On-prem" [cite: 721] to be more inclusive and accurately represent the platform's reach. The "Your... Platform" structure [cite: 686] aims for direct user connection.
    - **"Scalable":** This addresses a primary concern for enterprise deployments[cite: 368]. The ability to scale AI workflows securely is a core value proposition discussed throughout the meeting[cite: 544].
    - **"Trusted":** This word encapsulates security, reliability, and responsibility[cite: 708, 53]. It's a response to the known lack of inherent security in many GenAI tools [cite: 380] and aligns with supporting keywords like "Secure" and "Responsible"[cite: 671, 695].

    - **"Sovereign-ready":** The inclusion of "Sovereign" acknowledges a vital, specific market need, particularly in regions like the EU and APAC concerned with data locality and governance[cite: 676, 678, 723]. While "Sovereign" has stricter requirements than "Private"[cite: 677], positioning the platform as "Sovereign-ready" indicates an understanding of and a path towards meeting these demands, broadening the appeal[cite: 681]. It was seen as a key follow-up point or bullet[cite: 722, 724].
- **Insight - A Layered Message:** This two-part structure (mantra + headline/attributes) allows Red Hat to lead with a punchy, differentiating message ("Open Private AI") and immediately follow up with a broader, more inclusive description ("Your Hybrid GenAI Platform") that clarifies the full scope and details key enterprise characteristics, showing sensitivity to diverse market needs.

C. Discussion Points & Deep Insights:

- **The Private vs. Hybrid Tension:** This was a central debate. The "Private" focus is crucial for engaging the existing OCP customer base – the "beachhead" – who prioritize security and control[cite: 516, 531]. However, focusing *only* on this risks alienating customers who need cloud flexibility or want to leverage powerful SaaS models[cite: 529, 557]. **Insight:** The resolution suggests a "both/and" strategy: lead with "Private" as the hook but ensure "Hybrid" is core to the platform's capability and messaging to capture the larger market and support diverse customer needs.
- **The Rejection of "Orchestration":** Burr Sutter noted that "orchestration would not be an important word" [cite: 663] because it doesn't resonate well and its meaning is unclear to many[cite: 665]. **Insight:** This reflects a strategic choice to prioritize market-facing, benefit-oriented language over internal technical terms, even if technically accurate.
- **"Open" as a Moral Imperative:** Peter Double's passionate advocacy for "Open" [cite: 698, 699] reveals that this isn't just a business strategy; it's viewed by some stakeholders as a fundamental mission for Red Hat in the AI era[cite: 702, 705]. **Insight:** This deep-seated belief provides a strong internal 'why' and can be a powerful external motivator, aligning the product with Red Hat's foundational values.

Okay, let's dive into **Section V: Desired Customer Impact: Quotes & Value**. This section focuses on articulating the *value proposition* from the end-user's perspective, primarily by analyzing the "Customer Quotes" generated during the exercise and the rich discussion that followed, particularly around crucial elements like evaluations and data control.

V. Desired Customer Impact: Quotes & Value

Understanding what a successful customer would *say* about the Red Hat AI Studio is crucial for defining its core value. The "Customer Quote" portion of the Vision Framing exercise aimed to capture this voice. While multiple compelling quotes were generated, the group gravitated towards one as a strong base, while also highlighting key themes and specific capabilities that *must* be reflected in the customer experience to achieve the envisioned impact.

A. Selected Customer Quote (Foundation)

The quote provided by Ann Marie Fred resonated particularly well due to its narrative structure, covering multiple stages from building to scaling. While the group felt it was a strong starting point, there was a consensus that it should ideally be seen as one of *multiple* potential quotes, or enhanced to explicitly include all critical elements discussed. The foundational quote reads:

> "We built our first multi-agent system and deployed it into production using Red Hat AI. We're using our company data and our choice in models. We were able to evaluate the accuracy of our agents and build prompts and guardrails to improve, accuracy and reliability. We were able to monitor the entire system and gradually scale up to hundreds of thousands of requests per day and we did it all while controlling costs." [cite: 321, 322, 323]

Insight: This quote effectively tells a story of an end-to-end success. It moves beyond just *building* to encompass *deployment*, *management*, *scaling*, and *cost control* – all critical factors for enterprise success. Its mention of "multi-agent system" [cite: 321] aligns with the focus on agentic AI, and "company data" [cite: 321] touches upon the vital 'private' aspect.

B. Key Themes & Value Drivers

Analyzing all the quotes and the subsequent discussion revealed several recurring themes representing the core value drivers the platform aims to deliver:

1. **Speed & Ease (Time-to-Value):** Customers need to move quickly in the AI space. The platform must transform innovation timelines "from months, to mir hours"[cite: 370]. It should be easy for IT organizations to "identify and realize new opportunities" [cite: 296] and enable users to "get started fast" [cite: 354] with templates and pre-built patterns[cite: 324, 353].

    - **Insight:** This isn't just about developer convenience; it's about business agility. The current challenge is that many AI projects get stuck in POC [cite: 298] or take too long[cite: 372]. The platform's value lies in significantly improving the success rate and speed of production deployment[cite: 755].

2. **Control & Trust (Security, Privacy, Governance):** This theme is paramount and central to Red Hat's differentiation. Customers must feel empowered to build solutions that are "private on PREM and fully our own"[cite: 356]. This means:

    - **Data Control:** Using their "company data" [cite: 321, 806] securely.
    - **Model Choice:** Having "our choice in models"[cite: 321, 806], supporting both open and potentially external models[cite: 497].
    - **Security & Guardrails:** Building in "security and ops best practices"[cite: 326], "prompts and guardrails"[cite: 322], and ensuring "no security issues" [cite: 299] within a "secure hybrid, enterprise environment"[cite: 388].
    - **Insight:** This addresses the biggest fear and barrier for enterprise AI adoption. Red Hat can win by being the platform that *enables* innovation *without compromising* security and control – a stark contrast to many purely public solutions.

3. **Enterprise Readiness (Scale, Cost, Operations):** The platform must be robust enough for real-world use. This involves the ability to "operate them at scale"[cite: 318], handle "hundreds of thousands of requests per day"[cite: 323], and achieve this while "controlling costs" [cite: 323] and showing a tangible "return on the I investment"[cite: 397].

    - **Insight:** This is where Red Hat's OpenShift foundation provides a significant advantage. The ability to manage, scale, and monitor AI workloads using established enterprise-grade practices is a key value proposition.

4. **Agentic AI Focus:** The platform should excel at building "enterprise class agents and tools"[cite: 318]. It aims to support "multi-agent system" [cite: 321] development and deployment, positioning Red Hat as a leader in this burgeoning field.

    - **Insight:** By focusing on agents (and the frameworks used to build them, like LangChain [cite: 625]), Red Hat targets a specific, high-value AI workload type, moving beyond just model serving.

5. **Addressing Evaluations (Evals):** This emerged as a critical, specific pain point and opportunity. Adel Zaalouk strongly emphasized that the "Eval part is quite important"[cite: 54, 731], and customers need the ability to "evaluate the accuracy of our agents"[cite: 322]. While Burr Sutter noted that many aren't doing evals well yet[cite: 740, 741, 893], this is precisely the opportunity.

    - **Insight:** Evals are seen as a "chasm" [cite: 924, 925] preventing projects from reaching production. If Red Hat can make evaluations *easier*, *integrated*, and tied to *observability* (linking traces and spans to conversations for analysis [cite: 745, 746, 748]), it would provide immense value and drive platform adoption. It's a key area where Red Hat can provide "thought leadership" and practical tools.

VI. Core Benefits & Capabilities

To achieve the vision of an "Open Private AI" platform that delivers significant customer value, the Red Hat AI Studio must embody several core benefits and provide a specific set of capabilities. These emerged clearly from the "Key Benefits" listed by participants and the ensuing discussions:

A. Rapid Prototyping & Deployment The platform must fundamentally accelerate the AI development lifecycle, enabling users to move quickly from idea to tangible value.

- **Capabilities:**
    - Provide tools and workflows to "quickly move from concept to operational generative AI solutions"[cite: 374].
    - Offer out-of-the-box templates, patterns, or recipes ("Kickstarts") for common use cases[cite: 325, 357, 916], addressing the "blank page problem" [cite: 901, 903, 904] and helping users get "from zero to one"[cite: 351].
    - Support rapid iteration and "fail early, fail often" approaches to innovation[cite: 756].
    - Facilitate a "significant increase" in AI projects successfully moving from POC into production[cite: 298, 755].
    - Lower the barrier to entry, allowing a broader set of users to build and experiment[cite: 815].
- **Insight:** The market demands speed. Many current AI efforts stall. By providing accelerators like templates and streamlining the path to production, Red Hat can offer a compelling advantage, especially for organizations just starting their GenAI journey or struggling to scale their efforts[cite: 861, 862].

B. Unified & Cohesive Experience The platform must feel like a single, intuitive system, not a collection of loosely coupled tools.

- **Capabilities:**
    - Offer a "single intuitive platform" [cite: 375] and a "truly unified and intuitive interface" [cite: 394] for diverse personas (builders, data scientists, ops)[cite: 375].
    - Ensure all AI/ML tasks are interlinked, allowing users to focus on building cohesive systems rather than "navigating desperate tools"[cite: 394, 395]. This means Evals should connect to Agents, Agents to RAG, and RAG to models[cite: 395].
    - Provide an end-to-end "cohesive" experience where pieces don't feel "glued together"[cite: 408, 409].
    - Include "SME friendly" aspects[cite: 412], such as playgrounds or agent creators, to support collaboration and broaden participation[cite: 413, 414, 415].

- **Insight:** Adel Zaalouk explicitly warned against repeating past mistakes where powerful features were built separately and failed to tell a cohesive story[cite: 820, 821, 822]. A unified experience is critical for usability, adoption, and demonstrating the platform's holistic value[cite: 819].

C. Secure, Private, & Trustable This is a cornerstone of the Red Hat value proposition, addressing major enterprise concerns.

- **Capabilities:**
    - Bake in security and ops best practices "from the beginning"[cite: 326].
    - Provide "enhanced security and compliance"[cite: 31], acknowledging the current lack of security in many AI tools[cite: 32, 380].
    - Offer a "secure by design approach"[cite: 365].
    - Support fully "private on PREM" [cite: 356] and disconnected/air-gapped deployments[cite: 382].
    - Ensure integrated governance and compliance across all components[cite: 396].
    - Establish a "Policy aware registry" [cite: 376] or "AI registry" / "AI artifactory" [cite: 847, 851, 852] to manage all AI assets (models, agents, MCPs, tools) in a "secure private and trustable" manner[cite: 848].
- **Insight:** This capability set directly targets Red Hat's core enterprise audience and its largest differentiator[cite: 383, 838]. While others focus purely on capability, Red Hat aims to provide *controlled, governed, and secure* capability, which is non-negotiable for many of its clients[cite: 842]. The AI Registry is emerging as a key concept for achieving this control.

D. Hybrid Flexibility & Openness While "Private" is key, the platform must embrace the reality of hybrid environments and the value of openness.

- **Capabilities:**
    - Support "open models" [cite: 438] and trusted open-source components[cite: 342, 699].
    - Allow customers "their choice and models"[cite: 321].
    - Enable integration with external/frontier models (Gemini, Claude, etc.) if desired; *not* shutting them down[cite: 497, 541].
    - Provide support for bringing your "favorite. Framework as well"[cite: 364].
    - Allow control over "what runs in Cloud and what remains within my own data center"[cite: 327].
- **Insight:** This ensures the platform isn't perceived as insular or restrictive. It acknowledges that customers will use a mix of technologies and deployment locations[cite: 501, 502]. By being *the platform* that can manage this complexity while championing open-source, Red Hat can position itself strategically at the center of the customer's AI ecosystem[cite: 534, 565].

E. Operational Excellence Beyond building and deploying, the platform must support ongoing management and optimization.

- **Capabilities:**
    - Enable users to "monitor the entire system"[cite: 323].
    - Provide integrated "Evals" [cite: 55, 731] and "Observability"[cite: 57], ideally linking them to understand performance and accuracy[cite: 745, 746, 748].
    - Offer "optimized cost and ROI" [cite: 397] through visibility into resource consumption, token usage, and performance metrics[cite: 397].
    - Ensure scalability to handle high volumes of requests[cite: 323].
- **Insight:** Focusing on "Day 2" operations and providing clear ROI metrics is crucial for long-term adoption and demonstrating business value. The emphasis on Evals, particularly linking them to observability, presents an opportunity for Red Hat to innovate and address a significant industry challenge[cite: 889, 893].

Okay, let's proceed to **Section VII: Strategic Considerations & Target Audience**. This section synthesizes the discussions around *who* the Red Hat AI Studio is being built for, *how* it should be positioned in the market, and the *strategic debates* surrounding its goals and success metrics.

VII. Strategic Considerations & Target Audience

Understanding the target users and the strategic landscape is paramount for building a successful platform. The workshop discussions revealed a nuanced strategy, balancing the need to serve the existing customer base with the desire to capture new market share, all while navigating the complexities of measuring success in the rapidly evolving AI domain.

A. Target Personas

The group identified a multi-layered approach to personas, acknowledging that different users interact with the platform in different ways and represent different strategic entry points:

1. **Primary Focus (The Beachhead): Existing OCP Customers, IT Ops / Platform Ops / Model Ops.**

    - **Rationale:** This group represents Red Hat's "bed and butter"[cite: 474]. Burr Sutter emphasized that IT Ops is the audience showing up at Kubecon, and they are the ones who control cluster resources[cite: 428, 479]. Peter Double added that platform owners are the first point of contact at client sites; winning their confidence is essential to gain access to other groups[cite: 481, 482]. They need

a platform they can manage and that enables them to deploy GPUs and models[cite: 484].
    - **Insight:** The strategy is to leverage the massive existing OpenShift footprint. By providing these "Ops" personas with AI tools that solve *their* problems (e.g., monitoring, internal assistants), Red Hat can drive initial adoption and GPU deployment, creating the foundation for broader use.

2. **Key Users (The Builders): AI Engineers & Developers.**

    - **Rationale:** These are the individuals who will *build* the agents and applications on the platform[cite: 473]. Tony Kay contrasted these "framework first" users (interested in LangChain, etc.) [cite: 625] with model-first users. Adel Zaalouk highlighted that even IT Ops personnel will likely become "AI developers at some point"[cite: 460]. They need robust tools, SDKs, templates, and framework support[cite: 631, 772].
    - **Insight:** While Ops may *enable* the platform, the *Builders* will drive its *usage* and *value creation*. The platform must provide them with a powerful and flexible development experience, even if they are initially "intimidated by K8S"[cite: 648].

3. **Supporting Roles: Data Scientists & Subject Matter Experts (SMEs).**

    - **Rationale:** Data scientists remain important, particularly for model-centric tasks[cite: 623]. Andy Braren argued for making the platform "SME friendly" as well, providing tools like playgrounds or agent creators where domain experts can contribute directly[cite: 413, 414].
    - **Insight:** Democratizing AI means enabling not just core developers but also those adjacent to them. Providing accessible interfaces for SMEs can significantly increase the platform's reach and impact within an organization.

4. **Not Targeting (Directly):** Burr Sutter explicitly stated that certain developer segments (e.g., "West Coast style startup-oriented") are unlikely to ever become paying Red Hat customers and should therefore not be the primary target, although their preferences can offer insights[cite: 641, 644].

    - **Insight:** This reflects a pragmatic focus on the established enterprise market where Red Hat has a strong foothold and a clear path to monetization, rather than chasing every developer segment.

B. Market Positioning

The platform aims to carve out a distinct position in the competitive AI landscape:

- **The Enterprise Hybrid AI Platform:** Positioned as an on-premise/hybrid equivalent to cloud offerings like Vertex AI or Azure Foundry, but with superior security, privacy, and open-source advantages[cite: 378, 379, 342].
- **The AI Application Hub:** To be the platform for *building* and *running* AI applications, especially agentic ones, *irrespective* of whether the models run on-prem or via external APIs[cite: 534, 537]. The goal is to win the *workload*.
- **Leaders in "Open Private AI":** Leveraging the core mantra to emphasize trust, control, and openness as key differentiators[cite: 420].
- **Insight:** Red Hat is not aiming to out-compete on frontier model creation but rather to be the indispensable *enterprise layer* for deploying, managing, securing, and scaling AI applications built using *any* model.

C. Success Metrics (Discussion)

Defining success sparked a key strategic debate:

- **Key Metric (Hypothesized): Number of GPUs Deployed.** Burr Sutter strongly advocated for this, noting it's a direct metric tied to the AI Business Unit's quota and revenue recognition[cite: 585, 587, 589].
- **The Debate:** Tony Kay countered that selling OCP subscriptions for AI *workloads*, even those hitting external models, is a significant win and a faster path to quota retirement for many in sales[cite: 536, 568, 571]. Adel Zaalouk questioned how to best *funnel* activities towards the GPU metric, suggesting other leading indicators might be crucial[cite: 594, 595].
- **Insight:** This discussion highlights a fundamental strategic question: Is the primary goal to sell *infrastructure* (GPUs on OCP/REL) or to sell *platform adoption* (OCP subscriptions, driven by AI workloads)? The likely answer is *both*, but the platform's design and messaging need to consider how to best drive these, potentially interconnected, metrics.

D. The "Blank Page" / Use Case Challenge

A significant barrier to adoption identified was that many customers simply "don't have use cases today"[cite: 889].

- **The Problem:** Customers are tasked with "figur[ing] out how to use AI" [cite: 895] but face a "blank page problem"[cite: 900, 905]. They don't know where to start, what's possible, or even what questions to ask[cite: 868].
- **The Solution:** Red Hat must provide "templates" and "Kickstarts" to lower the barrier and show *what's possible*[cite: 903, 916]. Peter Double emphasized the need for discovery sessions and focusing on "low hanging fruit"[cite: 879].
- **Internal Use Cases First:** Andy Braren suggested that internal use cases offer a lower-risk entry point, as the "bar is way lower" regarding concerns like hallucinations compared to external-facing applications[cite: 908, 910].

- **Insight:** Red Hat must adopt an *educational* and *prescriptive* role, not just providing tools but actively helping customers identify and implement high-value, achievable use cases. Kickstarts and a focus on internal applications appear to be key tactical elements in this strategy.

Okay, let's complete the initial report outline by fleshing out the final section, **Section VIII: Key Takeaways & Action Items**. This section will summarize the core conclusions drawn from the Vision Framing exercise and outline the immediate next steps implied or explicitly stated during the first part of the workshop.

VIII. Key Takeaways & Action Items

The Vision Framing exercise and the rich discussions on Day 1 provided a strong foundation for the Red Hat AI Studio workshop. Several key takeaways emerged, alongside implicit and explicit action items to guide the subsequent days.

A. Strong Alignment on Core Vision: Despite debates around specific terminology, a remarkable consensus emerged regarding the fundamental direction. There is broad agreement on the need for Red Hat AI Studio to be an **Open, Secure, Hybrid platform** that prioritizes a **unified experience** and enables **rapid development** and deployment, especially for agentic AI applications. Participants consistently highlighted the importance of leveraging Red Hat's strengths in enterprise-grade security, control, and hybrid cloud management while embracing the speed and innovation required in the AI space.

B. Messaging Requires Refinement: While the core concepts are clear, the precise "three-word slogan" or headline still requires careful word-smithing. The "Open Private AI" mantra was seen as a strong hook[cite: 684, 690], but the necessity of "Hybrid" in the broader description [cite: 719, 721] was also affirmed. Further work is needed to create a layered message that is both impactful and accurately represents the platform's full capabilities.

C. Personas are Central to Strategy: The discussion repeatedly underscored the importance of a clear persona strategy. The "beachhead" approach, targeting existing **IT Ops / Platform Ops** customers to gain entry and drive infrastructure adoption[cite: 478, 481], is seen as key. However, the platform's core functionality must be built for the **AI Engineers and Developers** who will create value on it[cite: 460, 631]. Balancing these needs, while also providing accessibility for **SMEs**[cite: 413, 415], will be critical.

D. Evals & AI Registry Emerge as Key Differentiators: Two specific capability areas received significant attention as potential high-value differentiators:

- **Evaluations (Evals):** Seen as a critical, yet often poorly implemented, step in the AI lifecycle[cite: 731, 740]. Providing robust, integrated, and potentially observability-linked evaluation tools could solve a major customer pain point and build trust[cite: 733, 748, 753].
- **AI Registry/Artifactory:** The concept of a centralized, policy-aware registry for managing all AI assets (models, agents, tools) resonated strongly as a way to provide governance, security, and control[cite: 376, 847, 853].

E. Prototype Must Demonstrate Key Value: The goal of the prototype is clear: it should not aim for production readiness but rather serve as a functional demonstration of the core vision[cite: 157, 160]. It should showcase the ease of use, the unified experience, the ability to build and manage agents, and how security and governance concepts can be integrated.

F. Identified Action Items & Next Steps (from Day 1):

1. **Proceed with Persona Mapping:** Immediately move into the next scheduled activity to formally define the target users. [cite: 264]
2. **Secure GenAI Prototyping Access:** Peter Double to continue efforts to gain access to the Gemini AI Studio API keys for prototype development. [cite: 2, 3, 152, 162]
3. **Identify & Engage Technical Experts:** Proactively consider which additional technical experts (e.g., from the UI team, Eder for cross-notebook expertise) might be needed for Day 3 (Tech Stack discussion) and beyond[cite: 230, 237, 241]. Jason Greene and Adel Zaalouk to help identify these individuals. [cite: 232]
4. **Gather Management Feedback:** Ensure management stakeholders are kept informed and have opportunities to provide feedback on the direction and the proposed metrics. [cite: 591, 592, 602]
5. **Continue Process Feedback:** All participants should continue providing feedback not only on the *product* but also on the *workshop process* itself to refine this AI-powered methodology. [cite: 25, 243, 281]
6. **Synthesize Documentation:** Peter and Dash to lead the effort in taking the workshop outputs (transcripts, Miro boards) and using GenAI to produce the daily documents (personas, specs, etc.) for review. [cite: 142, 185, 194]