

# Blackbox Prototype Usability & Performance Study

---

*Author: Neon, Nervos Community Catalyst, Blackbox co-founder. Contact: [neon@nervos.community](mailto:neon@nervos.community)*

## Introduction

The global payments landscape is undergoing a major transformation as digital assets and blockchain technology continue to mature. While crypto payments have traditionally been limited by slow confirmation times, unfamiliar user experiences, and fragmented tooling, recent advances—particularly in payment-channel networks and Layer-2 technologies—are making near-instant crypto settlement increasingly feasible at the point of sale. As merchant demand grows for low-cost, borderless, programmable payments, there is a corresponding need for point-of-sale (POS) terminals that can seamlessly integrate both current fiat requirements and future decentralised payment rails.

Traditional POS terminals are optimised for fiat card networks, offering stability, reliability, and clear UX flows refined through years of industry experience. However, they do not natively support cryptocurrencies, and the few emerging crypto POS devices often lack consistency, usability, or the regulatory and hardware maturity required for real-world deployment. To bridge this gap, Blackbox is being developed as a next-generation all-in-one POS terminal that supports fiat payments while also enabling secure, fast, and private crypto transactions. On the crypto side, Blackbox is designed to integrate payment-channel networks, enabling instant settlement, low fees, and enhanced privacy.



Figure 1: Blackbox v0 prototype

The version tested in this study (Blackbox V0) is an early-stage prototype. (*Figure 1*) Because POS reliability depends heavily on clarity of UX flows, system feedback, hardware ergonomics, and predictable performance, rigorous usability testing is essential before progressing to a pilot-ready version.

This study evaluates how real testers interact with the Blackbox prototype, examining hardware, software, user interface (UI) design and user experience (UX), and usability heuristics to identify improvements required for V1 and commercial deployment.

## Aims

The primary aims of this study are to test current Blackbox features, to assess user experience and to identify key areas for product refinement and development.

## Objectives

Specific objectives include:

- Conduct focused prototype testing with a select group of testers from appropriate industry backgrounds.
- Gather qualitative and quantitative data providing insight into functionality, hardware and software design and performance, and UI intuitiveness.
- Assess Blackbox through the lens of Human Computer Interaction (HCI) principles.
- Create a list of recommendations based on feedback for future development priorities.

## Methods

Testers were recruited using purposive (expert) sampling, aiming for a total of 6. Each tester received:

- A physical Blackbox V0 device and power cable.
- Setup instructions (epub file)
- A structured feedback pack containing 35+ rating items and open-ended questions (*Appendix 1*)

Testers were gathered into a discord group to co-ordinate activities, and were given a maximum of 7 days to complete their feedback. Aside from identifying bugs to be addressed, feedback was submitted privately to avoid bias.

To help testers assess UX and usability from the perspective of HCI, the "10 Usability Heuristics" by Jakob Nielsen were implemented into the forms, prompting testers to reflect on the Blackbox through the lens of these standards.

The feedback pack was structured into 4 sections, each with criteria assessment items:

- Software Evaluation (11 items)
- UI/UX Evaluation (9 items)
- Hardware Evaluation (9 items)
- Usability Heuristics (10 items)

All items used a 1–5 rating scale (where 1 = poor and 5 = excellent) and contained space for written feedback. Open-ended questions were also provided to obtain further perspectives. Data was extracted, cleaned, and quantitatively analysed. Graphs included mean ratings by section, item-level scores, per-section breakdowns, and rating distributions.

## Results

A total of 6 testers were recruited onto this study. In particular, testers from hardware and software development backgrounds, as well as UI/UX specialties, were selected in order to provide relevant and focused feedback.

Delays in testing were experienced due to obstacles with acquisition of printing components and international shipping of prototypes. Detailed reflections on the methodologies and limitations of the study design and execution are provided in *Appendix 2*.

A compiled version of all feedback and ratings, as well as information on the testers, has been provided in *Appendix 3*.

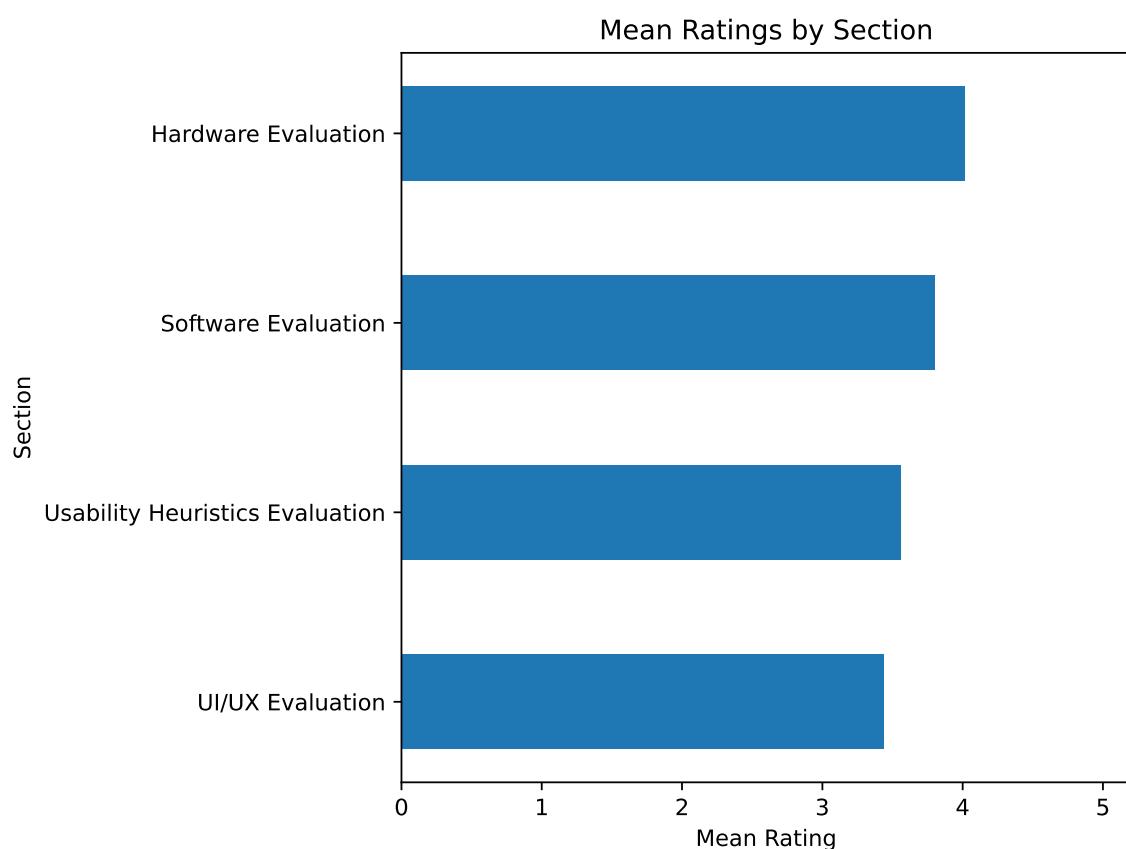
*Table 1* shows mean ratings received across all assessment items. All items received 6 total ratings apart from *Port Accessibility, Power Reliability, Error Prevention* and *Error Handling and Messaging* which received 5 total ratings. In those cases, the mean calculation accounted for this.

Item	Mean	Item	Mean
Setup Ease	3.50	Form & Stability	3.67
System Performance	3.50	Surface Texture	3.00
Responsiveness	3.33	Screen Readability	3.83
Function accuracy	4.67	Port Accessibility	4.40*

Payment processing flow	3.33	Port Options	4.17
Inventory Management	3.00	Power Reliability	4.60*
Invoice and record generation	4.33	Build Quality	4.00
Data handling and storage	4.67	Thermal Comfort	5.00
Reliability	4.17	Aesthetic Appeal	3.50
Error handling and messaging	2.80*	Visibility of System Status	3.33
System Stability & Recovery	4.50	Match Between the System and the Real World	3.67
Navigation and Menu Flow	3.50	User Control and Freedom	4.33
Button Placement & Responsiveness	3.50	Consistency and Standards	3.67
Visual Hierarchy & Readability	3.33	Error Prevention	3.60*
Configuration & Settings	4.33	Recognition Rather Than Recall	4.33
Feedback & System Status	2.83	Flexibility and Efficiency of Use	3.00
Aesthetic Consistency	3.83	Aesthetic and Minimalist Design	3.83
User Flow Efficiency	3.00	Help Users Recognize, Diagnose, and Recover from Errors	3.00
Accessibility & Localization	3.17	Help and Documentation	2.83
Overall UI/UX Impression	3.50		

*Table 1: Mean ratings across each assessment item. Values marked with an asterisk (\*) had their mean calculated based on n=5 where one rating was not submitted.*

*Figures 2 – 7 show bar graphs representing mean values across sections and specific items. Figure 8 shows a boxplot graph containing the range, median and variance of values across the sections.*



*Figure 2: Bar graph showing mean ratings across sections in descending order*

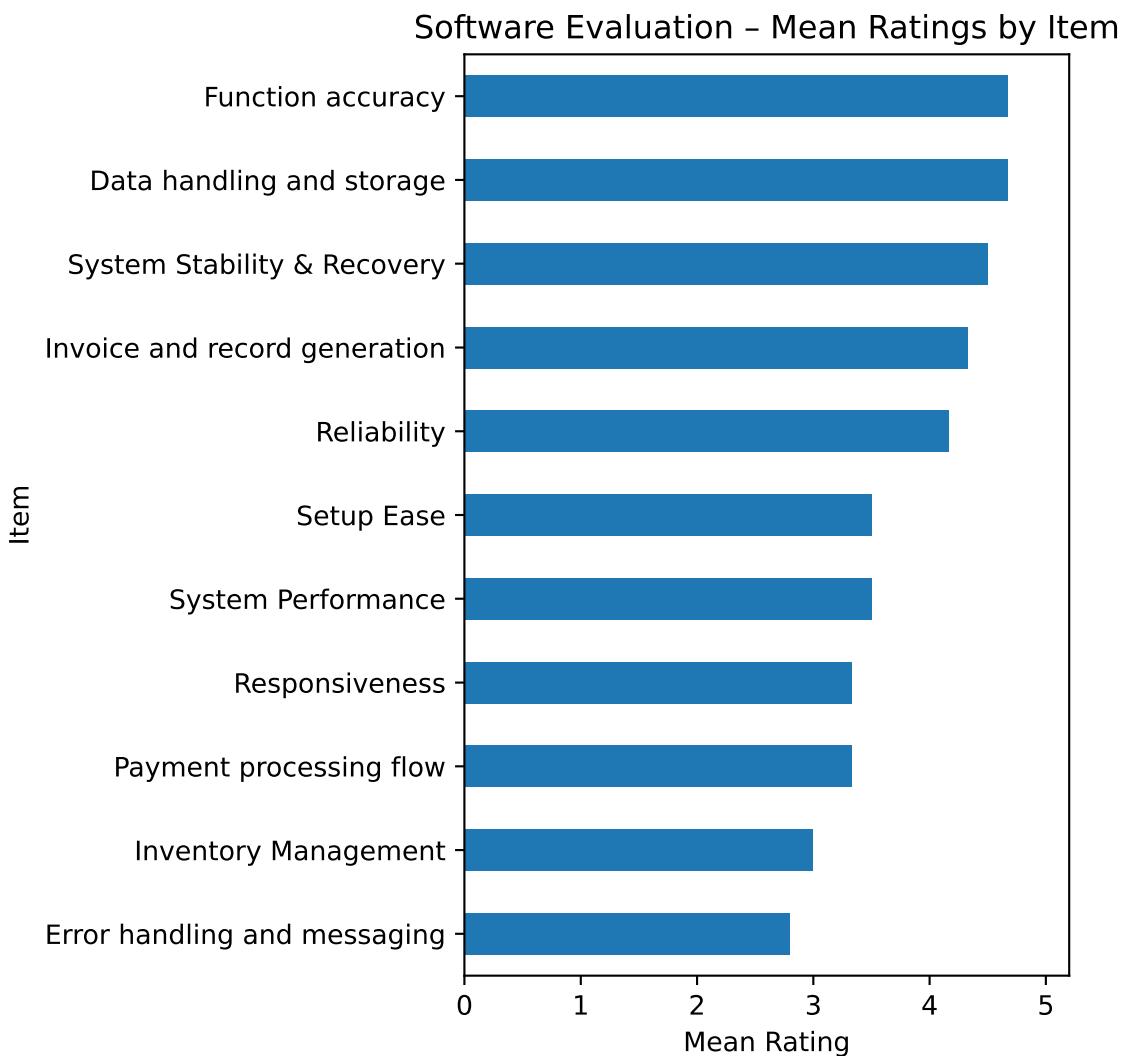


Figure 3: Bar graph showing mean ratings of criteria within Software Evaluation

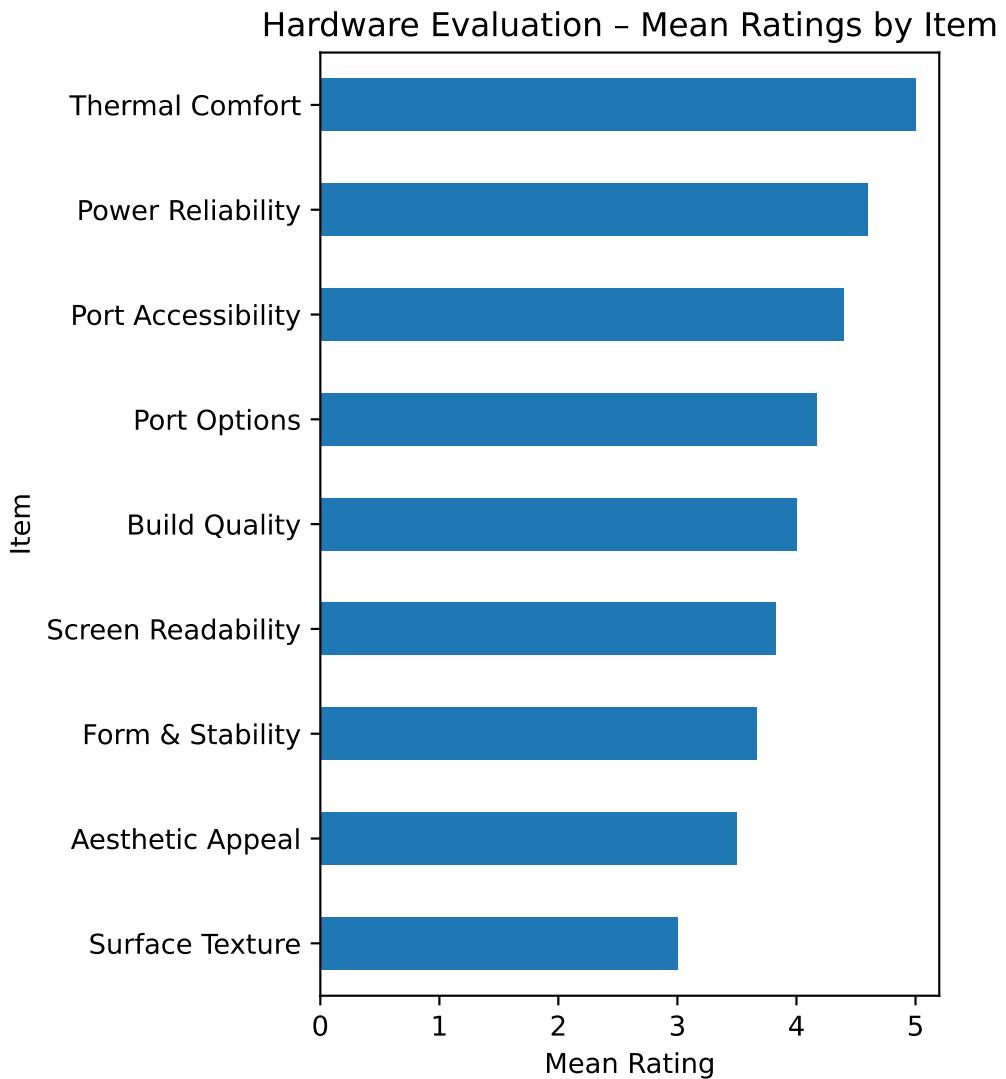
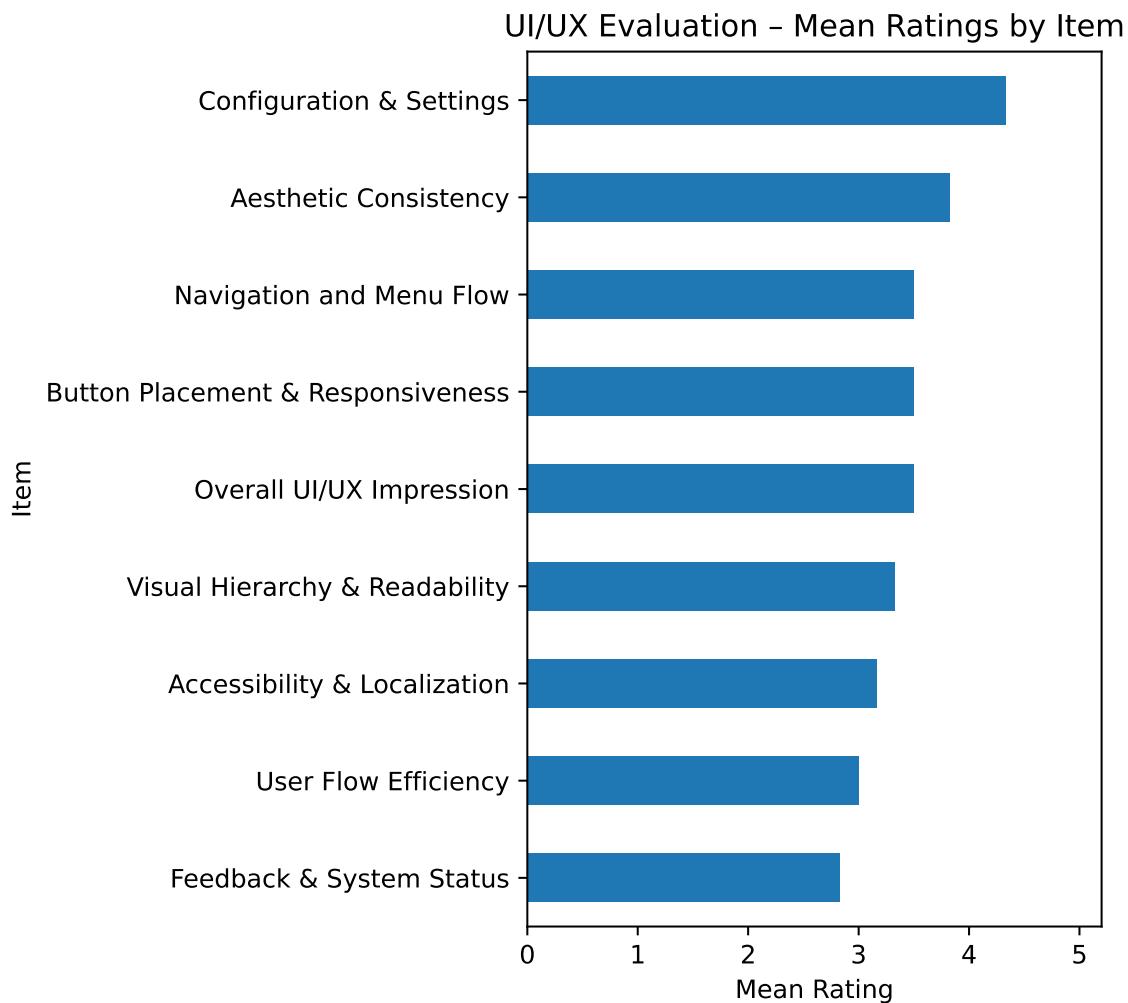
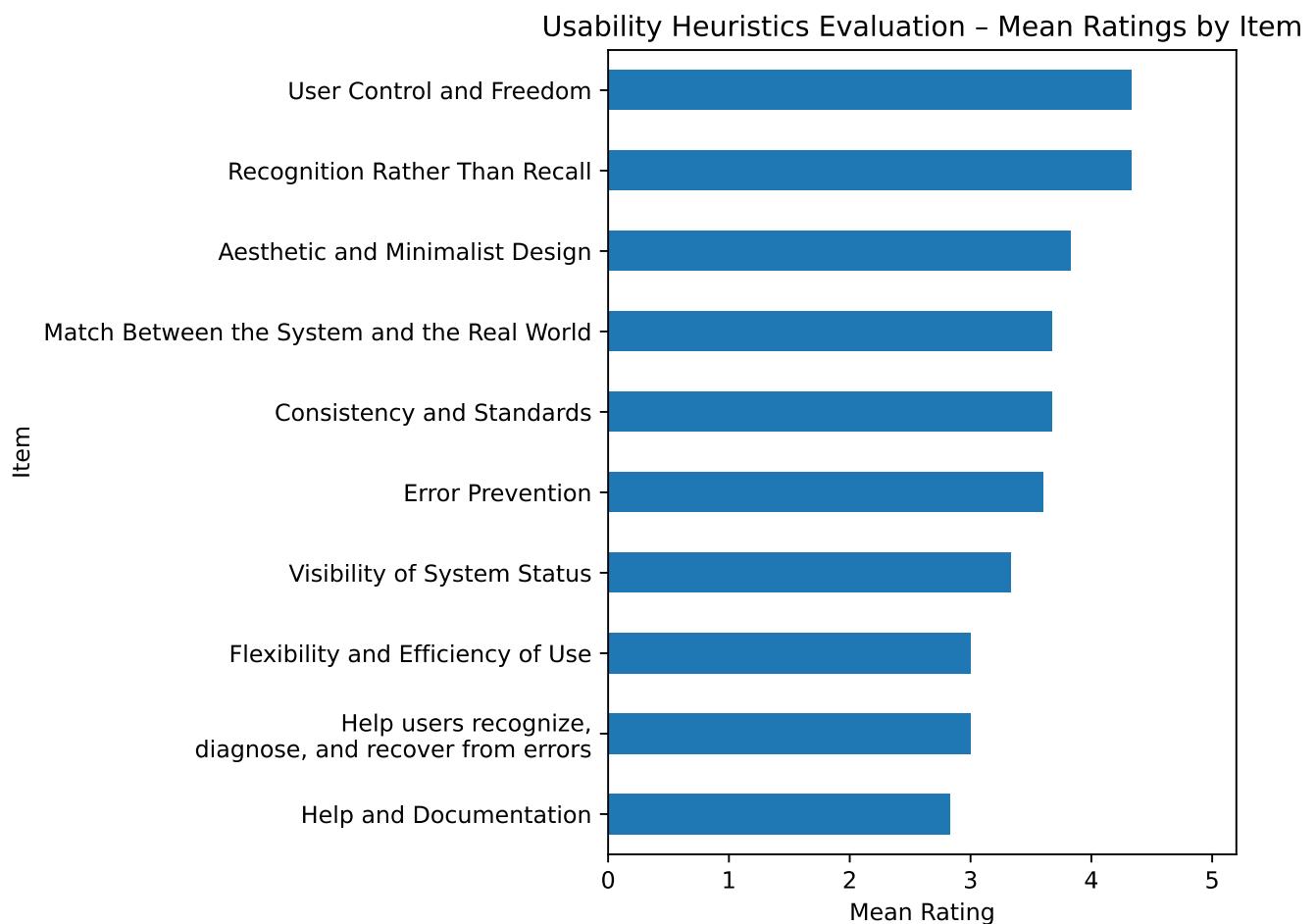


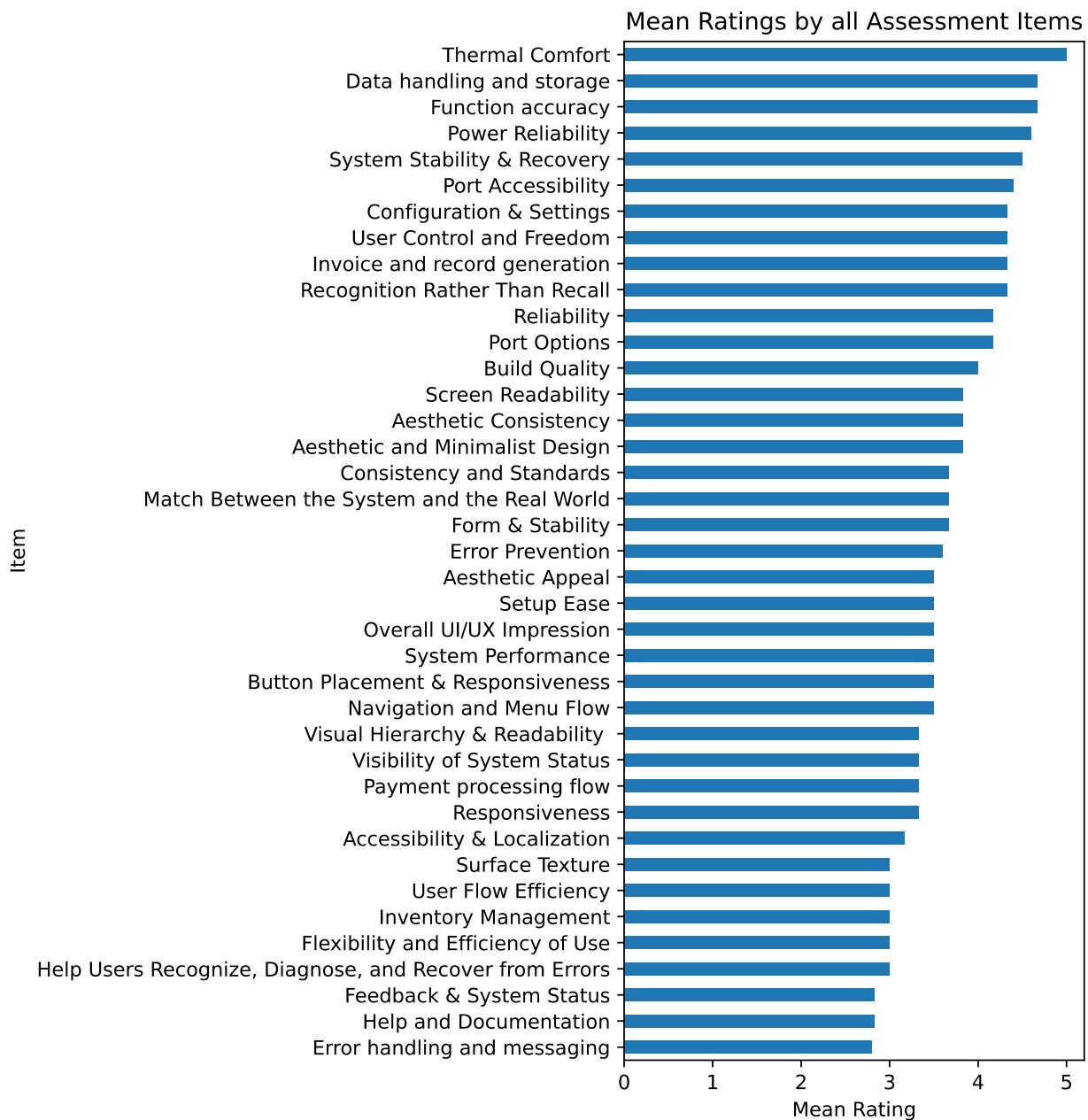
Figure 4: Bar graph showing mean ratings of criteria within Hardware Evaluation



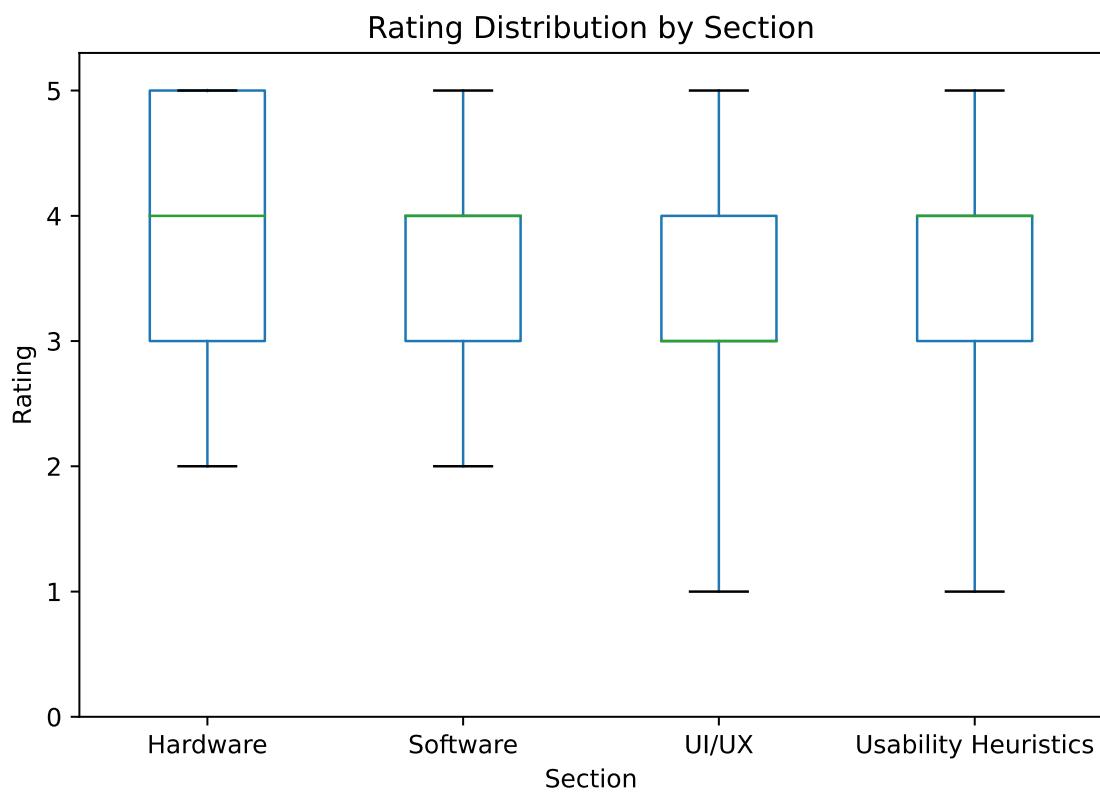
*Figure 5: Bar graph showing mean ratings of criteria within UI/UX Evaluation*



*Figure 6: Bar graph showing mean ratings of criteria within Usability Heuristics Evaluation*



*Figure 7: Mean ratings across all assessment items in descending order*



*Figure 8: Boxplot showing distribution (range, median, interquartile ranges) of ratings across each section*

Mean ratings across each section were as follows (in descending order):

Hardware Evaluation: **4.02**

Software Evaluation: **3.80**

Usability Heuristics Evaluation: **3.56**

UX/UI Evaluation: **3.44**

## Discussion

### Quantitative analysis

The mean ratings displayed in *Results* showed the following trends:

- Hardware was the highest-performing category.
- UI/UX scored lowest overall.
- Variance existed in how testers rated assessment items. Some trended more towards higher ratings on average than others.
- The strongest rated items included: **Thermal comfort, Data handling and storage, Function accuracy**
- The weakest rated items included: **Feedback and system status, Error handling and messaging, and Help and Documentation**

Deeper statistical analysis is avoided here due to the low sample size, subjectivity in rating scale perception, and rapidly changing nature of the prototype.

### Qualitative analysis

The discussion is broken down with reference to the relevant section and item of assessment.

#### Section 3 – Setup and scenarios

Feedback on setup and scenario workflows highlighted several usability issues that made early configuration more difficult than necessary, particularly around Wi-Fi connection and address handling. The Wi-Fi setup flow was described as confusing and fragmented, with page titles, labels, and actions sharing the same visual styling, making it unclear "what's interactive and what's not." Connecting to a network required multiple back-and-forth steps - scanning, returning to the previous page, entering a password, and saving - which one tester described as a "disjointed and inefficient" flow. Password entry was further hindered by very small touch targets, making a basic setup task "frustrating and error-prone." Uploading images over Wi-Fi also required manual resizing, adding unnecessary technical overhead.

Across scanning and address-related scenarios, testers consistently noted weak system feedback and insufficient validation. When initiating scanning, there was "no on-screen visual telling that scanning mode has started," with the only cue being a side LED that was easy to miss. After a successful scan, confirmation was limited, as the UI "does not react right away" and lacked an audible success signal. Address handling also raised concerns: mismatched addresses could be stored without warning, and in manual request scenarios the QR code generated on screen was "not valid," while only the printed receipt worked correctly-undermining confidence in the on-device workflow.

Other feedback pointed to stability and error-handling gaps in specific scenarios. Certain flows could crash reliably when required inputs were missing, such as confirming a fund request without selecting an asset, where the system "crashes stably after clicking confirm." Timezone configuration did not initially attempt to sync system time, leaving it showing "1970" until a manual change was made. Network issues could surface as incorrect values, such as showing "Inf" when oracles failed to load, without clear error messaging. Despite these issues, testers generally felt that "most of the functions are as expected," and many problems were addressed quickly through the shared Discord group and over-the-air updates, reinforcing the value of rapid iteration during prototype testing.

## Section 4 – Software Evaluation

### 4.1 Setup Ease

Setup was consistently described as achievable but unintuitive, with four testers reporting uncertainty during first-time configuration. Users were able to progress through setup, but the system did not reliably indicate which steps were mandatory before attempting core actions, leading to confusion or invalid outputs later. This resulted in a perception that setup lacked sufficient guidance, captured by the comment that setup was "functional but not intuitive enough."

In addition to flow clarity, interaction constraints compounded setup difficulty. Two testers independently struggled with Wi-Fi password entry due to small touch targets, with one noting that "later, it worked fine when I used a stylus." Further observations pointed out that unclear affordances made it difficult to distinguish labels from actions, increasing reliance on trial-and-error during what should be a deterministic process.

### 4.2 System Performance

Five testers characterised system performance as generally acceptable in short sessions but unreliable under prolonged use or certain conditions. Freezes, slowdowns, and stalled operations were repeatedly reported, with one tester noting that the system could "freeze for more than 30 seconds" during extended usage. These interruptions often lacked explanation, making it difficult for users to assess whether the system was still functioning.

Network dependency was highlighted by multiple testers as a contributing factor. At least two explicitly linked degraded performance to unstable connectivity, noting the system "will have some trouble if the connection to over internet is not very good." Additional observations pointed out that while interactions were often smooth, delays without visual indicators undermined confidence in system responsiveness.

### 4.3 Responsiveness

Four testers identified responsiveness issues, primarily driven by missing or inconsistent feedback. Users frequently could not determine whether actions were processing, had failed, or had not registered. One tester noted that navigating to Sales could take "several seconds with no loading cue," while another attributed poor responsiveness ratings to the fact that there "aren't any error prompts." Further observations emphasised inconsistency

across screens: some actions provided clear loading states, while others relied on subtle or non-standard indicators. This uneven feedback caused hesitation and repeated inputs.

#### **4.4 Function Accuracy**

Function accuracy was generally trusted by three testers, particularly for fetching exchange rate. Independent verification reinforced confidence, with one tester stating, "I checked against the exchange rate on Coinbase, and it matched." One tester documented discrepancies caused by screen malfunctions, while another experienced a crash when fetching rates for the first time, noting that "the box crashed and rebooted when getting the rate of CKB." These instabilities were resolved via firmware update during testing.

#### **4.5 Payment Processing Flow**

Three testers raised concerns about the payment processing flow, such as ambiguous entry points, weak visual hierarchy, and slow scanning, making it difficult to identify the primary action. One tester summarised this by stating that "the payment creation flow lacks clarity and visual hierarchy." Additional observations highlighted misleading terminology, unclear scanner state, and cases where on-screen QR codes appeared invalid. At least two testers noted that payment confirmation relied on checking external blockchain explorers.

#### **4.6 Inventory Management**

Inventory management was consistently described by four testers as functional but inefficient. Routine actions were slowed by workflow interruptions and poor state handling, with multiple reports that newly added items did not immediately appear. One tester noted that after adding an item, "the inventory list does not automatically refresh," forcing unnecessary navigation resets. Further points included intrusive confirmation modals, cluttered action areas, and icons visible even when non-functional. Typing friction was also widely reported, with two testers describing the keyboard as difficult to use.

#### **4.7 Invoice and Record Generation**

Three testers commented on invoice generation, generally agreeing that it was reliable and accurate. Invoices printed successfully and QR codes remained scannable. However, presentation quality was raised as a concern by more than one tester. It was noted that invoices, while legible, could better reflect user priorities, with one tester observing that "the presentation could be improved to better reflect user priorities." This suggests that improvements are needed in information hierarchy.

#### **4.8 Data Handling and Storage**

Data persistence was not identified as a major issue. Both testers who commented on this area reported reliable storage and consistency across sessions. One stated that "overall persistence and data accuracy between sessions work reliably." The only caveat raised related to inventory synchronisation, already noted earlier. No tester reported data loss or corruption.

#### **4.9 Reliability**

Four testers raised reliability concerns, particularly during extended use. Issues included freezes, crashes, and unresponsive states, with one tester reporting "two crashes during testing" and another noting that prolonged operation "may lead to occasional system freezes, which require a restart."

Additional observations indicated that some failures were repeatable, while others were intermittent. Update-related problems and sensitivity to network conditions further contributed to instability. Short sessions were generally stable, but reliability degraded under sustained or complex use.

#### **4.10 Error Handling and Messaging**

Four testers identified error handling as insufficient, with failures frequently occurring silently. Guidance on what went wrong or how to recover was reported to be insufficient. One tester highlighted that entering an incorrect Wi-Fi password caused the "connect" icon to disappear, but "no error message appears." In several cases, errors escalated directly into freezes or reboots rather than recoverable states with clear messaging.

#### **4.11 System Stability and Recovery**

Three testers commented on system stability and recovery. When recovery occurred, it was generally effective and fast, with one tester noting that "recovery is fast, and no data loss... was observed." This indicates that state restoration mechanisms themselves are robust. However, the need for recovery raised concerns. Rapid input and longer sessions increased the likelihood of unresponsive states, with one tester reporting the system could "fail to respond in a timely manner."

### **Section 5 – Hardware Evaluation**

#### **5.1 Form Factor and Stability**

Four testers commented on the physical size and stability of the unit, broadly agreeing that it feels solid and stable on a countertop but larger than expected for a POS device. While the weight was generally seen as appropriate and the device did not easily tip over, its overall footprint was perceived as larger than necessary. One tester described it as "noticeably larger than a typical POS terminal," which influenced perceptions of portability and placement flexibility.

#### **5.2 Surface Texture and Edges**

Surface finish and edge treatment were raised by four testers, with consistent concern around sharpness and tactile comfort. The dominant issue was that the device feels hard and angular, which detracts from comfort during handling and repositioning. One tester noted that "the edges are sharper than expected," raising concerns about prolonged use in a retail environment.

Further observations reinforced this point, with testers recommending rounded corners and smoother finishes and something that is easier to hold in one's hand. One tester

explicitly suggested that "the sharp corners around the machine be rounded," while another described the texture as "mediocre" and "rough in a real touch." Collectively, this feedback suggests that enclosure refinements could significantly improve perceived quality and comfort.

### 5.3 Screen Readability

Three testers commented on screen readability, with mixed but generally acceptable impressions. Under normal indoor lighting, the screen was described as clear and usable, with one tester stating simply that it "looks great under all lighting." Another tester found brightness acceptable across different environments. One tester reported that the screen was "not easy to use under dim lighting." While this was not a majority view, it suggests that brightness range or contrast optimisation may be necessary to ensure consistent usability across varied retail conditions.

### 5.4 Port Accessibility

Port accessibility was generally viewed positively, with three testers noting that ports were easy to reach and cables straightforward to connect. One tester described the ports as "well-placed and easy to access," highlighting this as a practical strength of the device.

That said, one tester raised a regional compatibility issue, noting that "the original plug couldn't fit into my socket in China," requiring a replacement plug. Another observation concerned scanner positioning rather than ports directly, with one tester noting that the scan camera's placement required lifting the device for effective scanning.

### 5.5 Port Options

Port availability itself was less frequently commented on, but two testers raised points worth noting. The existing port configuration was generally seen as simple and sufficient, with one tester observing that "there's only one power port, so it's easy to find and use."

However, one tester explicitly noted a missing option, stating that the device should "better have an ethernet port for real-world applications." This suggests that while the current setup may be adequate for testing or controlled environments, expanded connectivity options could be important for broader deployment.

### 5.6 Power Reliability

Power reliability was not identified as a problem area. Three testers explicitly reported stable power behaviour, with no interruptions during use. One tester stated that the device "works well on my end," while another noted that it enters standby mode appropriately after inactivity.

### 5.7 Build Quality

Build quality was generally perceived as solid, with three testers commenting positively on assembly and rigidity. One tester noted that the unit "feels solid and well-assembled," and that components did not shift when the device was rotated. However, refinement opportunities were still identified. One tester observed that "the seams need optimization,"

suggesting that finishing details could be improved to enhance perceived quality and durability.

### **5.8 Thermal Comfort**

Thermal performance was viewed very positively. Four testers explicitly stated that the device remained cool during use, even over extended sessions. One tester described it as "cool enough after a long session," while another noted "no noticeable warmth or hot spots." No negative thermal feedback was reported, indicating that heat dissipation is effective.

### **5.9 Aesthetic Appeal**

Aesthetic impressions were mixed but generally acceptable. Three testers commented on appearance, often describing the device as functional rather than premium. One tester summarised this by stating that it "doesn't look premium, but it does look professional enough for retail."

Additional observations pointed to roughness in finish and overall design maturity. One tester described the device as "simple enough but kinda rough," suggesting that while the aesthetic aligns with a prototype or utilitarian product, further industrial design input could elevate its market readiness.

## **Section 6 – UI / UX Evaluation**

### **6.1 Navigation and Menu Flow**

Navigation was generally regarded as understandable and learnable, with four testers indicating that they could move between screens without becoming lost. Persistent navigation controls helped establish a sense of orientation, and most menu groupings were perceived as logical. One tester summarised this positively, noting that "overall navigation is straightforward."

However, multiple testers also identified misleading visual cues within the navigation structure. Buttons and titles often shared the same visual styling, causing users to misinterpret headings as interactive elements. This resulted in hesitation or mis-clicks, as captured by the observation that the "Settings" title looked like a button and appeared actionable when it was not.

### **6.2 Button Placement and Responsiveness**

Button placement was mostly acceptable, but responsiveness issues were repeatedly reported. Four testers described delays or inconsistency in button reactions, which led to uncertainty about whether inputs had been registered. One tester explicitly pointed out "slow response of on-screen buttons," while another noted that touch interactions were "sometimes inconsistent."

Additional observations highlighted contextual issues rather than layout alone. Buttons that could not perform actions in certain states were still presented as active, creating false affordances. This mismatch between visual availability and functional readiness increased

trial-and-error and contributed to perceived sluggishness even when the system was processing correctly.

### 6.3 Visual Hierarchy and Readability

Visual hierarchy was one of the most routinely highlighted issues with the interface. Two testers noted that while text was generally readable, too many elements competed for attention, making it difficult to identify primary actions. One tester described that "the platform lacks a clear visual hierarchy," with all text appearing to carry similar weight. Also reported were dense layouts and misaligned text fields on certain screens, which contributed to a cluttered appearance. This was reinforced by feedback noting that "some buttons and information are too densely packed," increasing cognitive load during routine tasks.

### 6.4 Configuration and Settings

Configuration screens were generally viewed as accessible and logically organised. Three testers noted that core settings such as Wi-Fi, timezone, and sound were easy to locate, with one describing them as "easy to locate and understand." Unduly small input areas were highlighted here as an issue.

### 6.5 Feedback and System Status

System feedback was widely regarded as insufficiently visible or timely. Three testers commented that while some feedback exists, it is often too subtle or delayed. One tester noted that they "haven't noticed where I can see the progress or errors," highlighting a lack of clear system state communication.

Additional observations emphasised that critical actions such as scanning relied too heavily on hardware cues rather than on-screen feedback. One tester suggested that confirmation signals should not rely "only on a brief green light," as this can easily be missed, particularly when users focus on the main display.

### 6.6 Aesthetic Consistency

The interface was generally viewed as visually consistent, with three testers commenting positively on colour usage and overall styling. One tester noted that "the color palette and typography feel consistent across the interface," reinforcing a sense of cohesion.

However, inconsistencies emerged in icon sizing and font clarity. Multiple testers observed that some icons appeared disproportionately large or small, disrupting visual balance. One tester also suggested that the font choice made it difficult to distinguish certain characters, indicating that typographic refinement could improve legibility without altering the overall aesthetic direction.

### 6.7 User Flow Efficiency

User flow efficiency was identified as a significant weakness. Three testers reported that common actions required more steps than expected, slowing routine operations. One tester captured this directly by stating that "many common actions require more steps than necessary."

Additional observations focused on repeated confirmation popups and fragmented workflows, particularly during item editing. These interruptions were seen as unnecessary for low-risk actions, with one tester describing them as "redundant" and another noting that "too many confirmations" disrupted momentum.

### **6.8 Accessibility and Localization**

Accessibility and localization were raised by three testers, primarily around language clarity and icon interpretation. Terminology such as "SSID" was cited as potentially confusing, and one tester noted that "only supports English," limiting broader usability. Also highlighted were issues with text casing and icon ambiguity. Uppercase labels were described as harder to read, and certain icons (such as printer or connect/disconnect) lacked clear meaning without labels.

### **6.9 Overall UI / UX Impression**

Overall impressions of the UI were cautiously positive. Four testers described the interface as functional and usable, with one stating that it "exceeded my expectations." The system was generally seen as learnable, with consistent layouts supporting basic navigation.

At the same time, it was noted that the experience does not yet feel polished. Weak visual hierarchy, inefficient workflows, and subtle feedback were repeatedly cited as barriers to confidence. As one tester summarised, the system "is usable but needs refinement" before it can feel modern and production-ready.

## **Section 7 – Usability Heuristics Evaluation**

### **7.1 Visibility of System Status**

Four testers reported issues related to insufficient visibility of system status, particularly during delays, scanning, or confirmation actions. Users frequently could not tell whether the system was processing an action, had completed it, or had failed silently. One tester noted that after tapping confirm, there was "about a 2-second delay... with no visible response," creating uncertainty about whether the input was registered.

Other comments highlighted missing feedback during critical moments. Examples included receiving funds "without any notification" and refresh actions that resulted in a "noticeable delay and no progress bar or hint." Collectively, these reports indicate that the system often changes state without adequately communicating those transitions to the user.

### **7.2 Match Between the System and the Real World**

Two testers identified mismatches between system terminology, iconography, and real-world user expectations. The most common issue was the use of technical or ambiguous labels that required interpretation rather than recognition. One tester pointed out that "Quantity is unclear in a payment context," forcing users to mentally translate system language into real-world meaning.

Further observations noted that some icons lacked intuitive meaning, with one tester stating that "the icon will need a decision of its meaning." These issues increase cognitive load, particularly for non-technical users.

### **7.3 User Control and Freedom**

User control was generally perceived as adequate, with three testers noting that navigation options such as Back and Home provided reasonable freedom to recover from mistakes. One tester described the back function as "highly convenient," reinforcing that escape routes are present and discoverable. However, gaps were still identified. One tester noted that "the sale page does not have a back button."

### **7.4 Consistency and Standards**

A key concern raised was that page titles were styled similarly to interactive elements, which created misleading affordances. One tester observed that titles "use the same visual treatment as clickable elements," leading to confusion. Also noted was inconsistent pagination patterns and differing screen designs for the same function depending on entry point. These inconsistencies force users to relearn interactions across similar contexts.

### **7.5 Error Prevention**

Two testers commented on error prevention, with mixed perceptions. While some safeguards exist -such as numeric input restrictions and confirmation dialog – one tester felt that proactive error prevention was lacking, stating that "there is no error prevention."

Additional comments highlighted specific weaknesses, such as allowing product codes that later generate invalid barcodes. This highlights the need to consistently prevent errors at the point of entry.

### **7.6 Recognition Rather Than Recall**

Three testers raised concerns related to reliance on user recall rather than recognition. While page titles and labels generally helped users understand their location, certain workflows forced users to remember prior actions or states. One tester noted that "the current sales flow relies on recall... by displaying only the most recently scanned item."

Further observations pointed out that field labels behaving as editable text required users to delete placeholder content before entering information, increasing friction. Although recognition support exists in parts of the system, these gaps increase cognitive load during routine tasks.

### **7.7 Flexibility and Efficiency of Use**

Four testers identified efficiency limitations, particularly for experienced users. The most consistent issue was unnecessary confirmation steps and rigid workflows that slowed routine actions. One tester noted that "there are no quick-access paths," while another pointed out that too many confirmations appeared for safe actions. Another noted problem was lack of adaptive behaviour based on prior selections, such as not being able to delete any item in the sales workflow except the last scanned item.

## **7.8 Aesthetic and Minimalist Design**

Two testers commented on aesthetic and minimalist design, generally agreeing that the interface aligns with a technical, minimal style but suffers from clutter in practice. One tester noted that while the design is readable, "several pages feel visually cluttered due to too many elements being given similar visual weight." They pointed to small touch targets that detract from both aesthetics and usability. This indicates that minimalism is present at a conceptual level, but not consistently realised in layout and spacing.

## **7.9 Help Users Recognize, Diagnose, and Recover from Errors**

Three testers raised concerns about the system's ability to help users recover from errors. The dominant issue was silent failure: actions failed without explanation or guidance. One tester highlighted that entering an incorrect Wi-Fi password resulted in "no error message," leaving users unsure how to proceed. Similar patterns were observed for invalid QR codes and other failures, with one tester noting that errors often required a restart rather than recovery.

## **7.10 Help and Documentation**

Three testers commented on help and documentation, generally agreeing that external materials were helpful but insufficiently integrated into the device experience. One tester noted that "the external ePub documentation is helpful," but on-device guidance was limited.

Additional observations pointed out that many actions relied on icon-only controls without labels or tooltips, and that device-level information (such as IP address) was inaccessible. This creates a gap between available documentation and practical, in-context assistance.

## **8. Open-ended questions**

Testers offered a range of general observations that extended beyond the structured evaluation categories, often focusing on broader impressions of the device, its potential use cases, and workflow considerations.

Several testers commented positively on the promise of the Blackbox as an integrated fiat-crypto terminal and expressed interest in seeing future crypto functionality implemented, particularly the scanning, settlement, and payment-channel features. Testers also acknowledged that the device is still in prototype form and noted that despite early rough edges, its foundation appears strong and well-conceived.

Testers broadly agreed that the system becomes usable and efficient once learned, but does not feel immediately intuitive. Several respondents highlighted that core functions such as manual requests, invoice printing, scanning, and currency handling work reliably after some familiarity is gained.

A few participants highlighted the importance of Blackbox eventually supporting multi-language interfaces, more robust settings options, and merchant-friendly features such as

transaction summaries, logs, or reports that could further improve operational transparency.

Feedback also included practical suggestions relating to quality-of-life improvements that do not fall neatly into existing categories. Some testers emphasised the need for clearer onboarding instructions, especially for first setup and Wi-Fi configuration, as these initial interactions strongly shape user confidence. Others suggested enhancing the discoverability of certain functions, such as diagnostic tools, connectivity indicators, and workflow shortcuts. There were also recommendations for making the device more modular or customisable, giving merchants the flexibility to adapt the interface to their workflow preferences.

The feedback received reflects a high level of engagement from testers and reinforces the opportunity to enhance Blackbox with richer, more intuitive, and more merchant-focused features as development progresses.

## Recommendations

The feedback provided gives plenty of clear direction for refinement. A list of actionable improvements are listed for each section:

### Setup and scenarios

- Introduce a **mandatory setup completion gate** that blocks access to payment, sales, and manual request flows until required steps (Wi-Fi, timezone, at least one valid address) are completed.  
*Target:* 100% of users complete required setup before first transaction attempt.
- Redesign the **Wi-Fi setup into a single guided flow**, combining scan, network selection, password entry, and confirmation into one linear screen.
- Increase **Wi-Fi password input hit area and keyboard usability**, optimised for finger-only touch input.  
*Target:* ≥90% first-attempt Wi-Fi connection success without stylus.
- Add a **clear on-screen scanner activation indicator** when address or barcode scanning is active.  
*Target:* 100% of scanning states visibly indicated on the main display.
- Add **audible and visual success confirmation** for successful address scans.  
*Target:* Confirmation feedback latency ≤500 ms.
- Validate **address type correctness at entry time** (prevent BTC addresses being saved as CKB, and vice versa).  
*Target:* 0% cross-chain address misclassification.
- Prevent generation of **invalid on-screen QR codes** in manual requests and ensure parity with printed codes.  
*Target:* 100% parity between displayed and printed payment data.

- Add explicit **rate-fetch failure handling** (network/oracle errors) with clear user messaging.  
*Target:* 100% of rate failures surfaced with readable error states.
- Automatically **sync timezone on first boot** and clearly flag failures.  
*Target:* 100% of devices show correct time after setup.
- Surface **scenario-specific setup guidance inline** (manual request, sale, inventory) when prerequisites are unmet.  
*Target:* Reduce scenario-related setup errors by  $\geq 50\%$ .

## Software

- Standardise **loading and progress indicators** across all asynchronous actions.  
*Target:* 100% of async actions show visible feedback
- Cap **perceived response latency** by ensuring all user inputs receive acknowledgement promptly.  
*Target:*  $\leq 2$  seconds for 95% of interactions.
- Replace silent failures with **structured error handling** across all workflows.  
*Target:* 0 silent failures observed in regression testing.
- Improve **network resilience** with retries, timeouts, and graceful degradation instead of freezing.  
*Target:* No UI freezes longer than 5 seconds without explanation.
- Stabilise **barcode scanning reliability** under suboptimal conditions.  
*Target:*  $\geq 95\%$  successful scans in normal retail environments.
- Integrate **payment confirmation directly within the device UI**, reducing reliance on external explorers.  
*Target:*  $\geq 90\%$  of payments confirmed without leaving device UI.
- Fix **inventory list refresh logic** so newly added or edited items appear immediately.  
*Target:* 100% real-time inventory synchronisation.
- Reduce **unnecessary confirmation modals** for reversible, low-risk actions.  
*Target:*  $\geq 50\%$  reduction in confirmation prompts.
- Harden **crash recovery and logging**, capturing context automatically.  
*Target:* 0 data loss incidents following crashes.
- Add **firmware-level regression tests** for rate fetching and known crash paths.  
*Target:* 0 rate-fetch crashes in production firmware.

## Hardware

- Reduce overall **device footprint** to better align with standard POS expectations.  
*Target:* 20-30% reduction in width/depth without compromising stability.
- Round all **external edges and corners** to improve comfort and safety.

- Improve **morphology, surface texture and finish quality** to enhance handheld ergonomics and perceived durability.
- Reposition the **scanner camera** to face the same direction as the screen.
- Standardise **global power connector compatibility** across target markets.  
*Target:* 100% compatibility without replacement plugs.
- Add an **Ethernet port** for fixed-location retail environments.  
*Target:* Wired connectivity available on all production units.
- Improve **seam alignment and enclosure tolerances**.
- Maintain current **thermal performance envelope**.  
*Target:* Surface temperature <40°C under sustained load.
- Add a **physical power button** for better control and recovery.
- Increase **screen size relative to device size** improve touch accuracy and readability.

## User interface / user experience

- Clearly differentiate **page titles from interactive buttons** using typography and styling.  
*Target:* 0 instances of titles mistaken for actions in usability testing.
- Establish a strong **visual hierarchy** that prioritises primary actions.  
*Target:* Users identify the primary action within 2 seconds.
- Reduce **visual density** on high-traffic screens.  
*Target:* ≥20% reduction in on-screen elements per page.
- Visually **disable unavailable actions** to prevent false affordances.  
*Target:* 100% inactive actions clearly differentiated.
- Standardise **icon sizes and alignment** across the UI.  
*Target:* No icon size variance within the same action group.
- Replace ambiguous icons with **labelled or clearer alternatives** (e.g. printer, save).  
*Target:* ≥90% correct icon interpretation in testing.
- Improve **on-screen keyboard layout** for text-heavy tasks.  
*Target:* ≥40% reduction in typing errors.
- Remove excessive **confirmation popups** in edit flows.  
*Target:* ≥50% reduction in modal interruptions.
- Apply **sentence case consistently** across all UI text.  
*Target:* 100% compliance with casing guidelines.
- Prepare the UI for **internationalisation (i18n)** beyond English.  
*Target:* i18n framework implemented across core UI.

## Usability Heuristics

- Guarantee **visibility of system status** for every state change.  
*Target:* 100% of state transitions observable visually or audibly.
- Replace **technical terminology** with real-world merchant language.  
*Target:*  $\geq 90\%$  comprehension without explanation.
- Ensure **back navigation exists on all screens**.  
*Target:* 100% of screens have a clear exit path.
- Standardise **pagination and list behaviour** across the system.  
*Target:* One pagination pattern used consistently.
- Prevent **invalid inputs at entry time** (barcodes, addresses, codes).  
*Target:* 0 invalid inputs accepted silently.
- Prefer **recognition over recall** in sales flows (show full cart).  
*Target:* 100% of scanned items visible at all times.
- Introduce **efficiency shortcuts** for repeat users.  
*Target:*  $\geq 30\%$  reduction in steps for common repeat actions.
- Reduce **visual clutter** while preserving clarity.
- Implement **recoverable error flows** instead of forcing restarts.  
*Target:*  $\geq 90\%$  of errors recoverable in-flow.
- Add **contextual help and tooltips** for non-obvious actions.

## Miscellaneous

- Introduce a **first-run onboarding mode** with skippable guidance for setup, scanning, and first payment.
- Implement **lightweight telemetry** for crashes, freezes, retries, and abandoned flows.  
*Target:*  $\geq 95\%$  of critical failures logged with context.
- Create a **formal design and interaction system** (buttons, icons, pagination, feedback)
- Add **contextual, in-place help** for advanced or ambiguous actions.  
*Target:*  $\geq 50\%$  reduction in help-related confusion.
- Plan early for **regional and international deployment** (language support, power, terminology, network conditions).

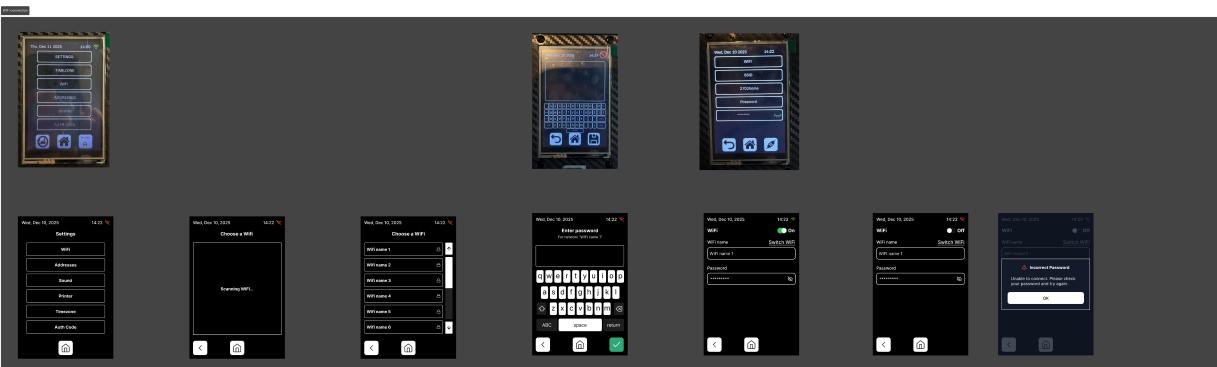
## Visual reference images for implementation

Of particular value were mockups created by Yuqi Feng for suggested workflows. These are shared below to serve as references for refinements. The original hi-res figma files are also available for reference.

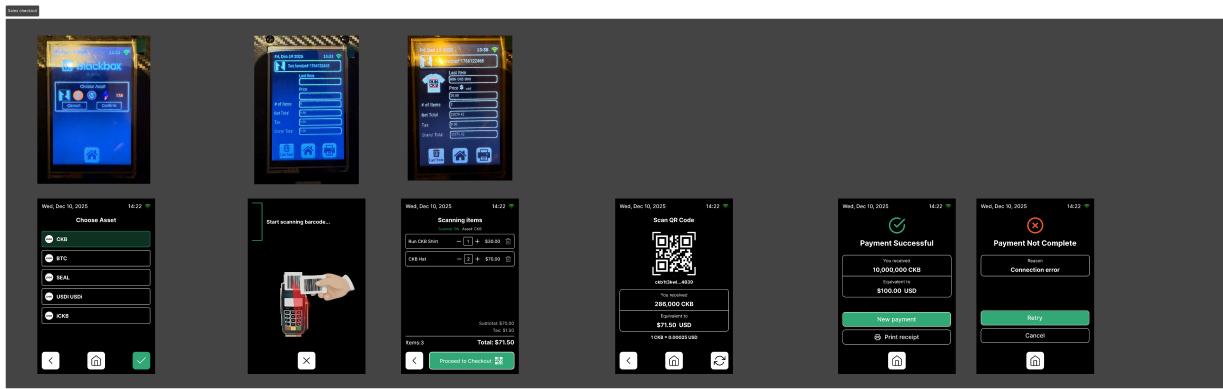
### Scan & Store address



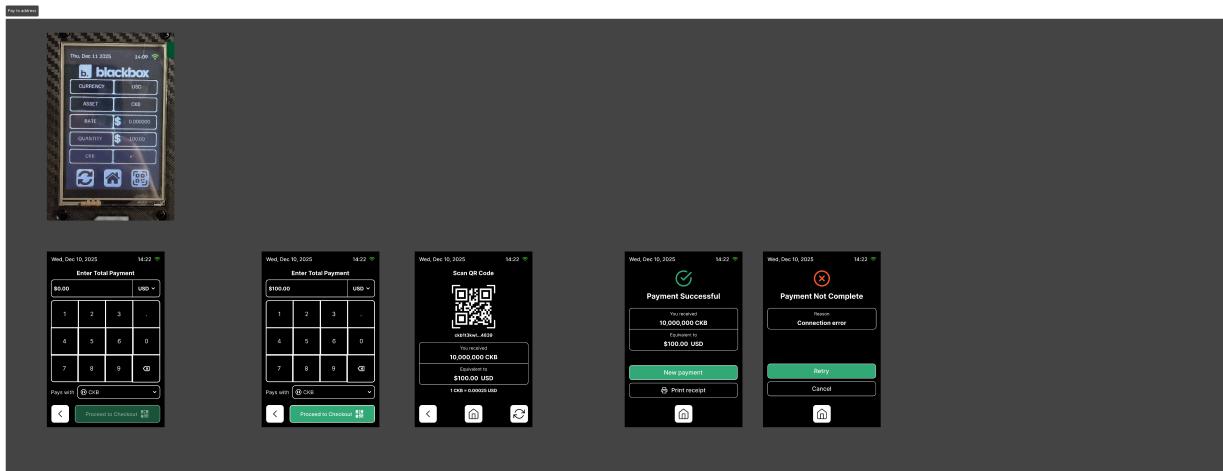
### WIFI connection



## Sales checkout



## Pay to address



## **Conclusion**

This study provides a clear roadmap for Blackbox V1 development. The hardware fundamentals are strong, while UI/UX and system feedback require targeted improvements. Addressing the identified issues and iterating on product design will significantly improve reliability, performance, reduce friction, and enhance merchant confidence; bringing Blackbox closer to market readiness as a next-generation fiat–crypto terminal.

## **Acknowledgments**

Thanks are extended to Phill (co-founder and CTO Blackbox) for his tireless efforts. Thanks are also extended to Vicky (Cryptape), and the prototype testers: Alive (CKB developer), Code Monad (Nervape CTO), Guopeng Lin (Cryptape), JackyLHH (CKB contributor), Tian Litao (CKB developer), Yuqi Feng (Cryptape). Finally, thanks are extended to the Spark programme and the Nervos Foundation.

## **Appendices (attached files)**

**Appendix 1: Blackbox prototype testing pack**

**Appendix 2: Testing Operations Experience and Reflections**

**Appendix 3: Compilation of all quantitative and qualitative feedback received from testers**