Quest Love: Revision Log

Comments: Conference Chair

1) Clearly document the data extraction process (Step 1), even if the raw data itself cannot be released.

We added a new section entitled "Dataset and Limits on Reproducibility" and in it, we documented how the quest platform backend worked, collected data, and how we accessed it.

2) Explain why Step 1 is non-reproducible (e.g., business confidentiality).

In the new section "Dataset & Limtations: Reproducibility," we describe the choice of the data provider, why we find ourselves in this situation, clearly indicates how this limits reproducibility, but also states one main advantage (we had input into quest design which allowed us to "experiment" with engagement factors). Additionally we changed the language we used for the dataset in a few place where we referred to it as public (this was an oversight, we did not expect the company to withdraw their name and did not fully update the paper in all places)

3) Perhaps, provide an example dataset or synthetic version if feasible.

We actually provide the real data, we just anonymized details that would easily identify it (such as abbreviating addresses and hashes to a few characters). We add a link in the paper to GitHub with the data (https://github.com/MadibaGroup/2025-Quests)

4) I agree with your proposal for step 2. Ensure Step 2 is sufficiently detailed so that others can apply the methodology to their own data.

We refer to GitHub but will add documentation (after submitting the paper to you, if ok, it will take some time) to the dataset (markdown files) that explain each CSV file and how it is used to generate our results.

Comments: Shepherd

1. Transparency and Reproducibility

- The paper claims that all measurements are reproducible from public blockchain data, but no details are provided. Perhaps, disclose the contracts addresses and block number range so that others can gathered the same data as you.
- Please specify the dataset source and provide clear instructions on how to access it.
- Indicate whether the dataset will be made publicly available or explain why it cannot be shared.
- Identify the specific blockchain project(s) and DApps covered by the dataset.

We added a new section entitled "Dataset and Limits on Reproducibility" that deals with the limitations related to these concerns. We could not do all of these suggestions (e.g., disclosing specific addresses and project names) however we did upload our dataset after anonymizing identifying details.

2. Research Questions

- Reviewers found the number of research questions excessive. Consider consolidating RQ1-RQ5 into subquestions under a primary RQ1.

We made this change.

3. Methodological Clarity

- Quest Completion Timing (Section 2.2): Currently, it appears that completion times are based on a single person's assessment. A more robust approach would involve multiple participants and using an average/median measure.

This is a valid criticism. We decided to disclaim it as such. Hiring new research associates to code difficulty cannot be done in a short timeframe and we do not suspect it would change much, as the criteria was pretty clearcut.

- Quest Cost Calculation: Clarify how quest costs in USD were computed. Did you use exchange rates at the time of transaction, or a uniform later exchange rate? You can use data feed from Yahoo Finance (off-chain prices) or Chainlink Labs (on-chain prices).

We added "Quest costs in USD were computed using token exchange rates at the time of each transaction, based on price data provided by CoinGecko, aligning precisely with the prices displayed to users on the quest platform UI."

- Sybil Attacks & Bots: Address how you handle bot identification and Sybil attacks, as these impact the validity of the findings. Please, also discuss the limitations of your approach.

We do not have ground truth on bots, sybils, and genuine users. We feel the submitted version is already clear in this regard: "From our data, we are not able to directly classify users between humans and bots, however we do offer some insights and discussion points about the influence of automated task-completion…" and "We currently lack data revealing how many task

completions come from unique individuals versus repeat participants using multiple addresses (farming), whether by manual sybils or automated bots"

- Descriptive vs. Causal Analysis: Some reviewers noted that your findings imply causality but are based on descriptive statistics. Be cautious in making causal claims or explicitly acknowledge the limitations.

We proofread the document thoroughly and agree that the previous language implied causation through words like **"impact"** and **"to drive completion."** To address this, we replaced these terms with more appropriate alternatives such as **"correlates"** and **"associated."**
These changes were made in Sections 2.1, 2.2, 2.3, and 2.4, including both paragraphs and titles.
Here's a link to the commit: [Updated wording to ensure descriptive analysis without causal claims](#).

4. Others

- Ensure all figures are vectorized (e.g., PDF format) to avoid pixelation.
- Increase font sizes in figures for better readability.

Figures were replaced with PDF format and font increased for better readability.

- Clarify Table 2 by providing details on how stakeholder data was collected (e.g., surveys, sample size, methodology).

Stakeholder analysis is not based on any data, it is based on our evaluation as researchers as to the preferences of different groups. This is how it was done in the other FC paper we cite. We clarified this.

- Clarify the broader relevance of your study beyond the analyzed dataset and how your findings apply to other blockchain ecosystems.

We tried to address this in a new "representativeness" section

---

REVIEW 1:

Copied TEXT:

The paper is quite an easy and pleasant read, and comes across as occasionally insightful, although it is not at all rigorous, being based on one blockchain project's experience
1. Carried out without any notion of ground truth with respect to whether participants are humans or bots.

This is correct and a limitation of our study and it is also an opportunity for future work. We do disclaim this fact (e.g., "From our data, we are not able to directly classify users between humans and bots, however we do offer some insights and discussion points about the influence of automated task-completion") and we added this to future work.

2. The attempt to introduce Fogg theory is laudable, but it doesn't seem to lead anywhere.

We leveraged the Fogg theory to develop our heuristics around bot detection. We clarified this a bit in the relevant section.

3. It is rather frustrating that it is claimed "All of our measurements are replicable from the public website and blockchain data" but then it is not reported what is the public website, what is the blockchain project and where the blockchain data can be found, rendering the paper in reality entirely irreproducible.

We agree fully and we replaced this statement (and any similar) with a forward pointer to a new section called "datasets and reproducibility." In this section, we explain the data, why we choose to work with a company, the advantages and disadvantages for that decision, and the consequences/limitations to reproducibility. We also uploaded our data in anonymized fashion to GitHub and put the link in the paper so that some of our results can be reproducible.

4. It is also frustrating that the question inherent in the title is left unanswered.

We agree fully and we changed the title of the paper to better reflect the preliminary nature of the study ("Quest Love: A First Look at Blockchain Loyalty Programs"), while also highlighting it as a novel and unstudied area of blockchain. We also tried to answer the question of the original paper's title better in the discussion section, in particular regarding the concept of loyalty.

Tense is inconsistent and noun numbers (singular/plural) are all over the show in the paper, which needs a good proofread.

Example typos/issues:
1. "factor into who receive airdropped" -> "factor into who will receive airdropped" (or receives...)
   Corrected to "factor into who will receive airdropped"
2. "the analysis of quest completion data revealed" -> you might want to keep tense consistent throughout (reveals...)
   Corrected to "reveals"
3. "12x" -> "12\times"
   Fixed
   "However this did" -> "However, this did"
   Fixed
4. "By contrasts, bot" -> "By contrast, bot"
   Fixed

5. "Heuristic 1: If a quest goes from gradually becomes profitable over time" -> does not parse
   Replaced by : If a quest gradually becomes profitable over time, bots will detect the exact moment it turns marginally profitable and act.
6. "however they want" -> "however, they want"
   Fixed

REVIEW 2:

TEXT:
The paper addresses an interesting problem by exploring whether, how, and why quests are effective. I believe the topic could spark engaging discussions during the workshop.

However, there are some issues that need to be addressed before a camera-ready version can be considered. I think that the authors can fix these concerns. My comments are as follows:

1. Citation and Literature Review:
   The paper would benefit from additional citations to support several arguments made in the introduction, dataset overview, discussions on bots and farming, etc. With only six citations included so far, there is opportunity to incorporate more relevant works that can justify the claims made—either through the authors' analyses or by referencing previous work.

We put a concerted effort into citing more literature and now have many more relevant citations

2. Data Source: The source of the dataset is not disclosed, nor is there a characterization of this dataset. For acceptance, the paper should:
   a. Clearly indicate the source of the dataset.
   b. Discuss how representative the data is.
   c. State whether the dataset will be made publicly available to promote reproducibility of their results.
   d. Specify which blockchain projects and DApps are covered by the dataset.

As described above, we address these issues to the extent that we can (with the company being unidentified) in the new "Dataset and Limits on Reproducibility" section. We also added a section after called "Representativeness"

3. Figures and Plots:
   Please ensure that all figures and plots are generated in a vectorized format (e.g., PDF

when using Matplotlib) to avoid pixelation in the final PDF. Additionally, increasing the font size within these figures will improve readability, especially when the paper is printed.

We have addressed this by regenerating all figures in PDF formats and increased font size. You can find the github commit here

4. Section 2.2 – Quest Completion Timing: The current approach appears to be based on the completion time recorded by a single person. It would be more robust to involve multiple participants (e.g., at least 10) and use their average or median completion times to obtain a more reliable measure.

This is a valid criticism. We decided to disclaim it as such. Hiring new research associates to code difficulty cannot be done in a short timeframe and we do not suspect it would change much, as the criteria was pretty clearcut.

5. Quest Cost Calculation: Clarification is needed on how the quest costs in USD were computed. Did the authors use the exchange rate at the time each transaction was issued, or was a later exchange rate applied uniformly? This detail is important for assessing the accuracy of the cost analysis.

We added "Quest costs in USD were computed using token exchange rates at the time of each transaction, based on price data provided by CoinGecko, aligning precisely with the prices displayed to users on the quest platform UI."

6. Table 2 – Stakeholder Data: Additional details are required regarding the data presented in Table 2. For example:
   a. Was this information obtained from stakeholder surveys? If so, how many stakeholders were surveyed, and what was the sampling method?
   b. What was the overall process for data collection?

Stakeholder analysis is not based on any data, it is based on our evaluation as researchers as to the preferences of different groups. This is how it was done in the other FC paper we cite. We clarified this.

7. Typos and Minor Corrections:
   a. In the introduction, the first line should be revised to: "Blockchain projects understand that even the best technical ideas and code cannot be achieved without user adoption of the platform."
      We changed this sentence not to the exact suggestion but removed the awkward grammar.

b. In Section 2.3, please confirm whether the reference should be to Figure 4 instead of Figure 3.
The reference should be Figure 4, that has been fixed.

c. Add more citations to support some of the arguments and affirmations the paper raises on the text!
Agree and see above

REVIEW 3:

TEXT:
The paper empirically analyzed a deployed quest system on the blockchain. Amount of reward, monetary value of reward, difficulty, and cost were identified as factors that impact task completion. The authors also discussed how farming and bots can impact the quest system.

The research is timely, the methodology is rigorous, and the paper is well-written. I only have some minor concerns, based on which the authors can revise the paper for the camera-ready version.

1. First, there are currently too many research questions, which is a bit distracting for the readers. I suggest the authors put RQ1-RQ5 as sub-RQs under RQ1.
Agree and changed

2. Second, the paper is not structured as an academic paper. I suggest the authors structure the sections as 1. Introduction 1.1. RQs 2. Related Work 3. Methodology 3.1 Dataset 3.2 Analysis (this is currently missing in the paper) 4. Results 4.1. RQ1 4.2. RQ2 (this part is more like speculation instead of empirical results currently) 5. Discussion (the stakeholder analysis is a bit surprising in the current structure, as no RQ is proposed relating to this analysis) 6. Conclusion.
We think this is a matter of taste and there is no "standard" structure (at least for us in CS). We took some of the advice (added analysis and discussion) but also left some of the structure alone.

3. Third, I encourage the authors to provide more details about the statistical analysis, which will help readers better contextualize the results.
We will add this to document the dataset on GitHub

4. Fourth, the current section 2.6 lacks empirical results, and the current section 3 is not linked to any RQs. The authors might want to put them in a discussion section instead of a results section.

We renamed the research questions as "discussion questions" and placed them in a discussion section as suggested.

REVIEW 4

TEXT:

General Remarks:

1. The authors should provide more details on the blockchain itself. «a new blockchain offered by a company» is a very vague statement. At the very least, I want to know some architectural characteristics of that blockchain (UTXO or account-based, rich statefulness or simple scripting, consensus, pseudonym-system, potential access restrictions…) and get some metrics why this blockchain is relevant for a broader context.
   We cannot say too much without de-identifying the company but we added: "offering a new EVM blockchain based on proof of stake"

2. The statement that data is only available from epoch 5 without any further explanation why they decided not to collect and include 1-4 is somewhat puzzling.
   Clarified

3. RQ2-RQ4 (task difficulty / time, transaction fees, capital cost) all are different representations of economic costs. I recommend to start with a simple theoretical model that shows the expected relationship between the costs and reward.
   We will add this to future work, we are not economists or game theorists and the chances we blow (after it is has been reviewed so it won't be checked again by the reviewers) this is non-trivial

4. I recommend professional proofreading. The text was incredibly hard to read, due to some awkward sentences and grammatical errors.
   We did the best we could 🙂

5. The paper would benefit from a more thorough discussion — ideally in its own section.
   Added this section and moved some discussion points into it

Methodology:

1. The authors should more carefully outline how they treat and control for sybil attacks. Otherwise, any metrics and visualization based on completion and addresses have to be interpreted with great care.

We do not control but will add this as a limitation and make it clearer that indeed the reader needs to execute "great care"

2. The paper only uses descriptive statistics and reports correlations. Yet, the research questions and the way the results are portrayed, implies causality. I strongly recommend choosing a more careful interpretation or use a methodology that is better suited to support the claims.
We agree that the previous language implied causation through words like **"impact"** and **"to drive completion."** To address this, we replaced these terms with more appropriate alternatives such as **"correlates"** and **"associated."** These changes were made in Sections 2.1, 2.2, 2.3, and 2.4, including both paragraphs and titles. Here's a link to the commit: Updated wording to ensure descriptive analysis without causal claims.

3. The authors should put more effort into the identification of bots. I agree that this is a challenging task, but many of the findings are severely impaired by the lack of this data. In addition to the «number of distinct completions», variables like task difficulty can be completely off, if for example the external dependency is in fact just a coordination between 2 bot instances of the same entity.
This is very understandable that it would have been great if we could have arrived at identifying bots. While we didn't completely answer this we did propose 3 heuristics that can help identify bots and an experiment that was done to identify them, however being 100% sure requires much more dedicated research for this. We hope that as a first look this paper will inspire future investigations.