International Journal of Graphic Design (IJGD)
Volume 3 No. 1 May 2025
E-ISSN: 2987-9434; P-ISSN : 2988-0343, Page 196-210
DOI: https://doi.org/10.51903/ijgd.v3i1.3715
Available online at: https://journal.stekom.ac.id/index.php/ijgd

Financial Risk Dashboard Design for Institutional RWA Investors: Visual Hierarchy, Chart Comprehension, and Explainability in FinChart-Bench

Zeyi Li [1], Sihan Zhou [*2], Zoe Zhou [3]
[1]Industrial Engineering, New York University, NY, USA
[2]Enterprise Risk Management, Columbia University, NY, USA
[3] Human-Computer Interaction, Georgia Tech, GA, USA

Email Address: irisnycli@gmail.com

Abstract. *Institutional investors who review real-world asset (RWA) credit pools must read dense financial charts while judging default pressure, dilution, liquidity, and liquidation-related exceptions. This paper develops a benchmark-based dashboard design analysis using FinChart-Bench financial chart images and question records. The parsed data release contains 1,202 unique base chart images and 7,019 question records: 2,384 true/false, 2,350 multiple-choice, and 2,285 numeric or short-answer QA records. Each question was matched to its base chart image, classified by chart-reading cue, and mapped conservatively to RWA dashboard panels. Image-level visual features were also extracted from the chart files to estimate the interface burden attributable to visual density, edge structure, colour variation, and mark coverage. The analysis shows that numeric arithmetic is the dominant chart-comprehension demand, covering 3,263 records (46.49%). Comparison, extreme/rank lookup, and trend/direction cues account for another 3,031 records (43.18%). The risk-panel mapping is intentionally cautious: 3,992 records remain general financial context, while 1,345 map to dilution risk, 964 to liquidity risk, 616 to probability of default, and 102 to liquidation anomaly. The findings support a dashboard design principle for institutional RWA review: chart assistance should preserve the original evidence, mark the relevant visual region, attach a concise risk-specific explanation, and keep caveats visible. The contribution is a data-driven interface requirement profile for future VLM-connected dashboards and analyst user testing, rather than a claim of observed analyst performance.*

Keywords: *Chart Comprehension, Design Mapping, Explainable AI, Financial Dashboard, Visual Hierarchy.*

INTRODUCTION

Financial risk dashboards are design systems for decision making, not merely repositories of charts. Institutional RWA investors inspect collateral pools, receivables flows, credit transitions, liquidity buffers, and liquidation outcomes under time pressure. The design problem is therefore a graphic design and UI/UX problem: a chart must be legible, but it must also be ranked by risk relevance. Classical work on graphical perception shows that viewers read position and length more accurately than color hue or area, which explains why visual hierarchy matters for financial comparisons (Cleveland & McGill, 1984; Ware, 2019). Dashboard design literature likewise argues that metrics become useful only when they are selected, grouped, and sequenced for a specific managerial action rather than displayed as a dense collection of available signals (Few, 2006; Pauwels et al., 2009).

The recent rise of vision-language models (VLMs) changes the design space, but it does not remove the interface problem. A VLM may parse chart marks, recognize labels, and answer questions; an LLM may transform extracted observations into short explanations or caveats. Yet

Received: Januari 2025; Revised: February 2025; Accepted: March 2025; Published: May 2025
*Corresponding author, irisnycli@gmail.com

an RWA investor does not need a generic caption. The investor needs a reliable view of probability of default (PD), dilution risk, liquidity, and liquidation anomalies because these dimensions affect credit enhancement, reserve triggers, advance rates, and risk-weighted asset interpretation. Basel-style risk measurement emphasizes disciplined exposure classification and transparent capital logic (Basel Committee on Banking Supervision, 2017, 2023). The dashboard must therefore translate chart evidence into decision-relevant structure rather than simply attach generated text to a chart.

This paper analyzes how financial chart benchmarks can inform that structure. It uses FinChart-Bench as a design corpus and asks which chart-reading demands are most common, which questions contain risk-relevant financial cues, and which image features suggest higher visual complexity. The focus is not a model-performance leaderboard and not a laboratory user study. The focus is the interface layer that will eventually sit between model output and institutional review: explanation cards, threshold callouts, risk panels, and caveat lines.

The design problem is especially important in private-credit, receivables, infrastructure, and tokenized asset workflows, where institutional users must evaluate heterogeneous real-world asset pools. A dashboard may need to summarize borrower delinquency, seller concentration, reserve utilization, collateral liquidation, cash-flow timing, and servicer performance. A layout (Kuhn et al., 2024) that places many charts on one screen can create a false sense of analytical completeness. The more difficult task is to identify which chart features are decision-relevant and make them visible without concealing the underlying evidence.

The paper therefore contributes a revised, benchmark-grounded dashboard design workflow. It first parses the chart-question corpus and computes question-demand, risk-panel, and image-complexity evidence. It then uses that evidence to refine four interface treatments: Plain Chart, Explanation Card, Visual Annotation, and Hybrid Dashboard. The central claim is deliberately design-oriented: financial chart assistance should be auditable and spatially anchored. The original chart remains the primary evidence, while the assistance layer identifies the relevant region, states the risk implication, and preserves the caveat needed for institutional review.

LITERATURE REVIEW

Chart comprehension research has moved from synthetic chart answering toward realistic visual reasoning. DVQA demonstrated that chart question answering is difficult because answers may depend on chart-specific words and values rather than a fixed vocabulary (Kafle et al., 2018). PlotQA extended the challenge to scientific plots with real-valued answers and open-ended numerical reasoning (Methani et al., 2020). ChartOCR showed that chart understanding often

requires a hybrid pipeline: detecting marks and key points, applying chart-specific rules, and reconstructing data tables (Luo et al., 2021). ChartQA introduced human-written and generated questions involving visual and logical reasoning, demonstrating that complex chart questions require both perception and arithmetic (Masry et al., 2022). Chart-to-Text shifted the task from answering to summarizing, but it also reported factual errors and difficulty explaining complex trends, which is directly relevant to explanation-card design (Kantharaj et al., 2022). FinChart-Bench extends this line of work to real-world financial charts, a domain in which temporal labels, market terminology, and mixed chart encodings create additional reading demands (Shu et al., 2025).

VLMs provide the technical basis for future chart-aware dashboards. CLIP showed that image-text contrastive pretraining can produce transferable visual representations at scale (Radford et al., 2021). BLIP-2 connected frozen image encoders to frozen large language models through a lightweight querying transformer, reducing the training cost of multimodal generation (Li et al., 2023). LLaVA used visual instruction tuning to connect a vision encoder and an LLM (Chen & Chan, 2023) for general-purpose visual dialogue (Liu et al., 2023). Qwen-VL added text reading, localization, and grounding capabilities that are especially important for charts because financial slides often include small labels, legends, and embedded numeric annotations (Bai et al., 2023). These developments make VLM-connected dashboards feasible, but they do not automatically solve the design problem of how model outputs should be placed, prioritized, and verified.

Information visualization and human-computer interaction research explains why dashboard outputs must be structured. The visual information-seeking mantra emphasizes overview, zoom/filter, and details on demand (Shneiderman, 1996). Visualization analysis frameworks treat design as a mapping between user tasks, data abstractions, and visual encodings (Munzner, 2014). Cognitive load theory and multiple-resource theory predict that users slow down when an interface forces simultaneous search, comparison, reading, and arithmetic without grouping cues (Sweller, 1988; Wickens, 2008). Norman's design principles further suggest that system state and possible action should be perceivable without excessive inference (Norman, 2013). In a risk dashboard, this means that a chart annotation should show what changed, an explanation card should state why it matters, and the panel layout should indicate which risk dimension is affected.

Explainable AI literature adds a second requirement: explanations must be useful to the intended decision maker. Local surrogate explanations improved trust calibration by showing why a classifier produced a particular output (Ribeiro et al., 2016). Later work argued that

interpretability should be evaluated by task, user, and decision context rather than by explanation existence alone (Doshi-Velez & Kim, 2017). Human-centered explanation theory emphasizes contrastive and selective explanations: people want to know why this event, not every possible fact (Miller, 2019). In financial dashboards, therefore, an explanation card should not repeat every chart label. It should identify the salient comparison, state the risk consequence, and disclose the caveat or uncertainty that affects use in RWA decisions.

Financial risk visualization has additional constraints. RWA investors interpret assets through credit risk, exposure, collateral, and capital lenses, and securitized or receivables-backed exposures can be vulnerable to liquidity shocks and information opacity (Gorton & Metrick, 2012; Hull, 2018). Basel guidance formalizes the need to classify exposures and disclose risk-weighted outcomes, but the interface work of translating complex metrics into readable investor views remains underdeveloped. This paper addresses that gap by treating risk mapping as a dashboard design taxonomy rather than as a hidden ground-truth label in the benchmark data.

METHODS

The analysis used the FinChart-Bench data release as the design corpus. The dataset contains three task files, MC_data.json, QA_data.json, and TF_data.json, and three corresponding image folders. The JSON records identify q-suffixed image references, while the image folders store the base chart files. The analysis normalized each image reference by removing the q-suffix and matching the resulting base file name to the corresponding chart image. This produced 1,202 unique base chart images and 7,019 question records. Table 1 summarizes the parsed dataset structure, and Figure 1 shows the analysis workflow.

Table 1. Dataset structure after parsing

Dataset item	Count	Use in analysis
Unique base chart images	1,202	Base image names after removing q-suffixes from task references
TF question records	2,384	True/false chart-comprehension prompts
MC question records	2,350	Multiple-choice chart-comprehension prompts
QA question records	2,285	Numeric or short-answer chart-comprehension prompts
Total question records	7,019	Question-level records used in the benchmark analysis
Images with all three task types	1,144	Base chart images represented in TF, MC, and QA task files
Images with extracted visual features	1,202	Chart images used for image-complexity calculations

The question-demand analysis classified each record into one primary chart-reading cue. The cues were defined before counting the records and were applied with transparent keyword rules. Numeric arithmetic captures differences, totals, averages, ratios, and changes from one period to another. Extreme/rank lookup captures highest, lowest, largest, smallest, and related ranking prompts. Trend/direction captures increase, decrease, growth, decline, and overall movement. Comparison captures higher/lower, greater/less, versus, and between prompts when

the question does not primarily ask for arithmetic. Percentage/composition captures share, proportion, rate, margin, and mix prompts. Point lookup captures direct value or period retrieval. Remaining records are labeled Other chart reading. Table 3 reports the resulting distribution.

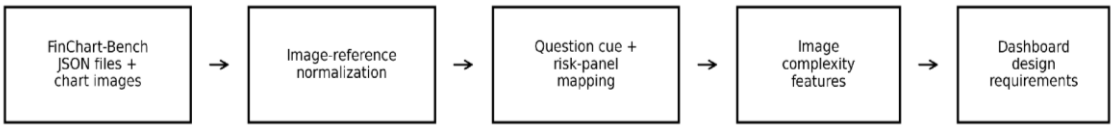

Figure 1. Benchmark-to-dashboard analysis workflow

The RWA panel mapping was designed as an interface taxonomy, not as a validated financial-risk label set. The benchmark questions do not provide ground-truth PD, dilution, liquidity, or liquidation-anomaly annotations. To avoid overstating the financial-risk status of the benchmark records, each record was mapped only when its question, choices, or QA reasoning contained explicit cue terms associated with one of the dashboard panels. Records without clear cue terms were retained as General financial context. Table 4 documents the mapping logic and the dashboard role of each panel.

Table 2. Base-image task coverage after reference normalization

Task coverage	Images	Percent of unique images
MC	1	0.08%
MC+QA	4	0.33%
MC+QA+TF	1,144	95.17%
MC+TF	45	3.74%
QA+TF	6	0.50%
TF	2	0.17%

Image complexity was computed from the actual chart files. Each base image was read from disk, resized only for feature extraction, and converted into visual-feature measures. Edge density approximates the density of lines, text strokes, axes, and mark boundaries. Color entropy captures the degree of color variation in the chart. Non-white pixel density approximates the portion of the image occupied by marks, text, shaded regions, and background elements. File size is included as a coarse proxy for image detail. Each feature was converted to a percentile rank across the 1,202 images, and the composite image-complexity index is the mean of the four percentile ranks scaled from 0 to 100. Table 6 reports the image-feature summary.

The four dashboard treatments are retained as design constructs rather than as user-performance conditions. Plain Chart preserves the benchmark image without assistance. Explanation Card adds a compact risk statement, trend description, and caveat. Visual Annotation adds mark-level callouts, threshold cues, or emphasized chart regions. Hybrid Dashboard combines the card and annotation layers inside a four-panel institutional risk layout. Figure 8 illustrates the treatments, and Table 8 defines the design constructs.

The statistical analysis is descriptive. It reports counts, percentages, cross-tabulations, medians, and image-complexity summaries. The results do not report observed analyst accuracy, observed task completion time, or live VLM inference accuracy. Those are appropriate next-stage validation studies. The present analysis establishes which chart-question and image features should guide the interface design of RWA dashboard modules.

RESULTS

The parsed benchmark has a balanced task profile across true/false, multiple-choice, and numeric or short-answer QA prompts. As shown in Figure 2, TF records account for 2,384 questions, MC records for 2,350, and QA records for 2,285. The task coverage is also broad at the image level: Table 2 shows that 1,144 of the 1,202 base chart images are represented in all three task files. This coverage supports using the corpus as a dashboard-design stress test because most chart images can be viewed through multiple comprehension tasks rather than a single prompt format.

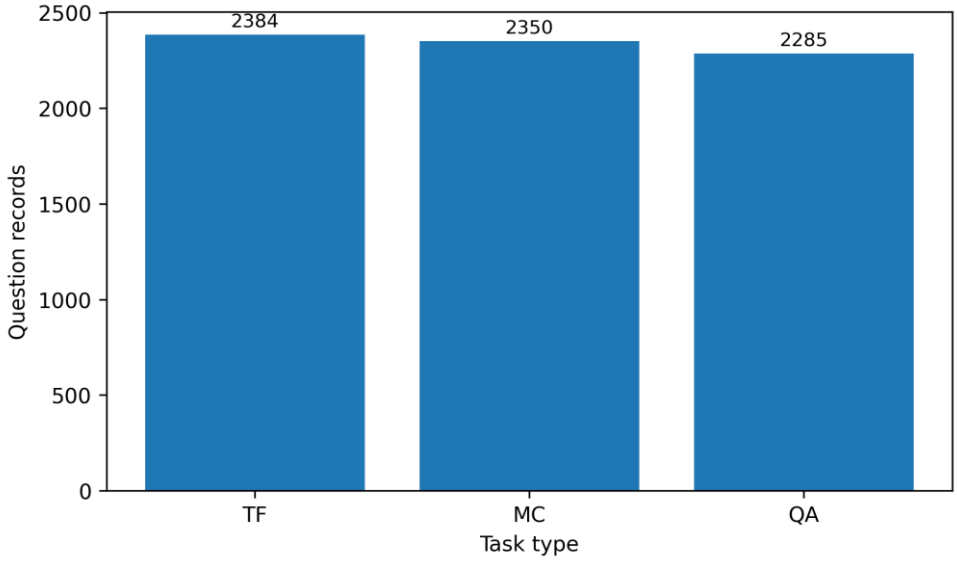

Figure 2. Distribution of question records by task type

The question-demand distribution reveals why a simple chart caption is not enough for institutional review. Table 3 and Figure 3 show that numeric arithmetic is the largest category, with 3,263 records, or 46.49% of all questions. Comparison, extreme/rank lookup, and trend/direction cues together account for 3,031 records, or 43.18%. These categories require users to locate the relevant marks and understand the relation among values. In dashboard terms, the evidence region and the explanation statement should be linked: a card that states a difference

without pointing to the chart evidence is incomplete, and an annotation that highlights a region without explaining the risk implication is also incomplete.

Table 3. Question cue distribution by task type

Question cue	TF	MC	QA	Total	Percent
Numeric arithmetic	481	805	1,977	3,263	46.49%
Comparison	990	44	4	1,038	14.79%
Extreme/rank lookup	202	820	15	1,037	14.77%
Trend/direction	381	303	272	956	13.62%
Other chart reading	251	42	6	299	4.26%
Point lookup	21	239	3	263	3.75%
Percentage/composition	58	97	8	163	2.32%

The RWA mapping confirms the need for caution. As Table 5 and Figure 4 show, 3,992 records, or 56.87%, remain General financial context because they do not contain explicit cue terms for the four RWA panels. Among mapped records, dilution risk is the largest panel with 1,345 records, followed by liquidity risk with 964, probability of default with 616, and liquidation anomaly with 102. Figure 5 shows the same pattern across task types. The result does not mean that the benchmark contains validated RWA risk labels. It means that a dashboard taxonomy can identify where risk-specific explanation is plausible and where the interface should instead preserve a general financial-chart reading mode.

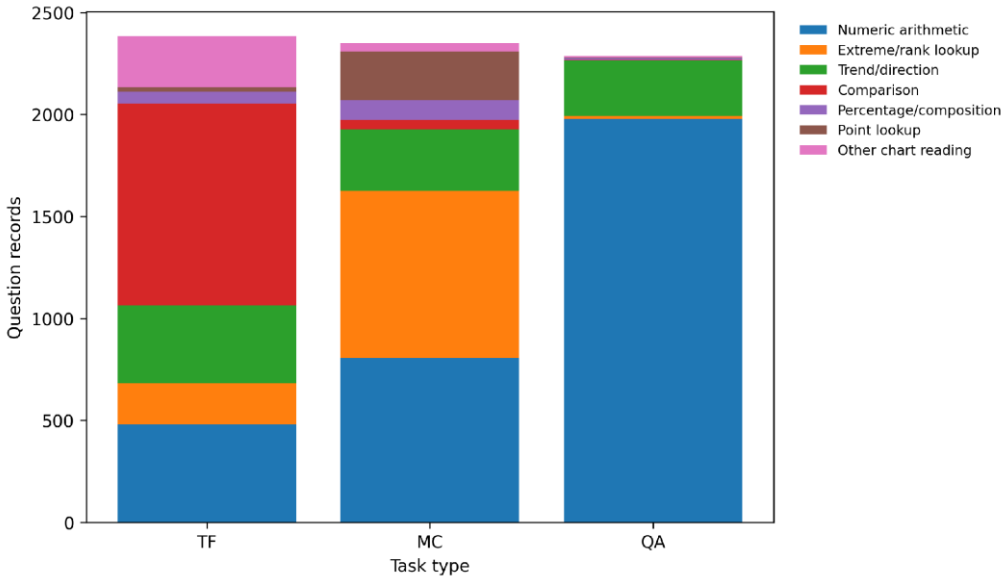

Figure 3. Question cue distribution within each task type

The image-feature analysis adds a visual-design perspective. Table 6 summarizes the 1,202 base chart images. The median image-complexity index is 50.10, and the interquartile range is 35.49 to 65.00. Because the index is based on percentile-ranked image features, it is best interpreted as a relative measure within the corpus rather than as a universal clutter score. Higher

scores indicate images with more edge structure, greater non-white coverage, richer color variation, or higher file-size detail. These conditions are exactly where dashboards need stronger scan order, callout discipline, and evidence-explanation coupling.

Table 4. RWA dashboard mapping rules used for design analysis

Panel	Cue examples	Dashboard role	Interpretive caution
Probability of Default	default, delinquency, charge-off, credit loss, leverage, coverage, provision	Surface borrower or issuer weakness; pair trend with credit-quality implication	Design mapping only; not a benchmark ground-truth PD label
Dilution Risk	receivable, revenue, sales, gross margin, discount, returns, allowance, refund	Flag revenue-quality or receivables-quality signals that may affect collateral value	Cue terms can also describe ordinary operating performance
Liquidity Risk	cash, cash flow, liquidity, funding, deposit, working capital, balance, maturity	Highlight timing, funding, and cash-conversion stress in the chart	Some balance and volume terms require context before risk interpretation
Liquidation Anomaly	liquidation, recovery, collateral, foreclosure, threshold, breach, spike, severe drop	Mark exceptions and threshold-like events that need review or caveat language	Sparse category; should be treated as anomaly-screening rather than validated event labels
General financial context	No clear cue term for the four RWA panels	Keep chart in general review queue or request analyst classification	Avoid forcing every financial chart into a risk panel

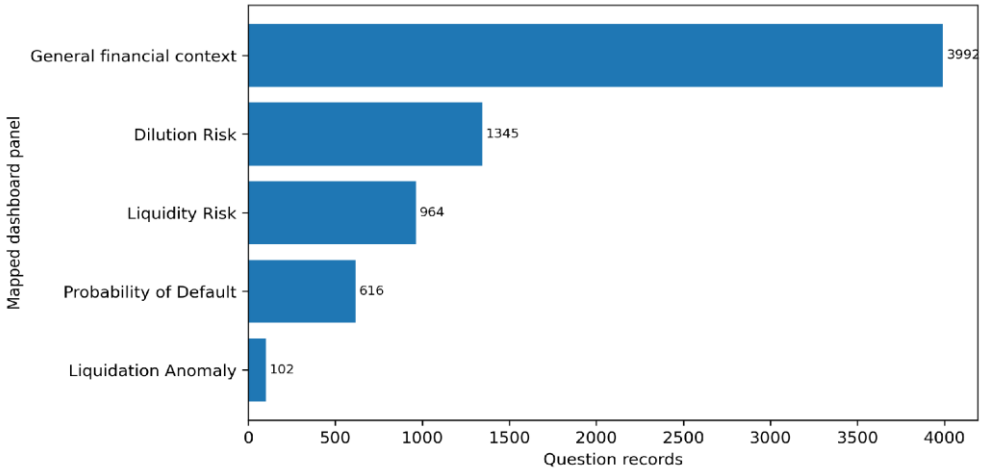

Figure 4. Total question records mapped to each dashboard panel

Complexity varies modestly across question cues and mapped panels. Figure 6 shows that numeric arithmetic and extreme/rank lookup records are concentrated around the middle-to-high portion of the complexity distribution. Figure 7 and Table 7 show a similar pattern across RWA panels: the mean complexity scores for mapped panels sit near the overall midpoint, while individual records still span a wide range. The practical implication is that risk-specific explanation should not be reserved only for visually complex images. Even moderately complex charts can require arithmetic, comparison, and temporal interpretation.

Table 5. RWA dashboard panel mapping by task type

Mapped panel	TF	MC	QA	Total	Percent
General financial context	1,379	1,303	1,310	3,992	56.87%
Dilution Risk	465	440	440	1,345	19.16%
Liquidity Risk	311	356	297	964	13.73%
Probability of Default	198	213	205	616	8.78%
Liquidation Anomaly	31	38	33	102	1.45%

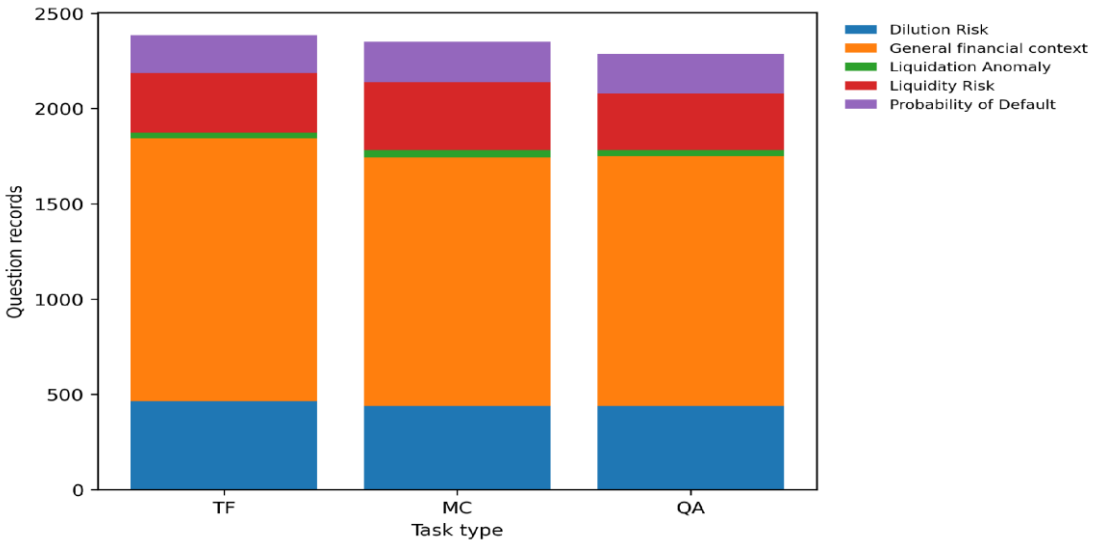

Figure 5. Dashboard panel mapping within each task type

The interface implication is summarized in Figure 8 and Table 8. Plain Chart is necessary as an evidence baseline but provides no risk translation. Explanation Card supports interpretation but can become detached from the chart evidence. Visual Annotation supports localization but does not fully explain why the highlighted region matters. Hybrid Dashboard is the most complete design pattern because it links the original chart, the visual evidence cue, the risk-specific explanation, and the caveat in one scan path. This is a design recommendation derived from benchmark demands, not a measured claim about analyst time or accuracy.

Table 6. Image-feature summary for base chart images

Metric	Mean	Median	IQR	Range
Image width (pixels)	1093.250	958.000	858.000-1221.750	389.000-3859.000
Image height (pixels)	748.755	658.000	549.000-924.000	329.000-2163.000
Edge density	0.210	0.205	0.162-0.251	0.059-0.501
Color entropy	0.185	0.181	0.142-0.224	0.042-0.418
Non-white pixel density	0.400	0.312	0.225-0.437	0.067-1.000
File size (KB)	67.443	57.918	43.792-80.591	17.411-278.152
Composite image complexity index	50.042	50.104	35.493-64.996	1.040-94.447

DISCUSSION

The results reposition the contribution of the study. The main finding is not that a dashboard condition has already improved analyst accuracy or completion time. The finding is that the FinChart-Bench corpus contains a high concentration of chart-reading tasks that require arithmetic, comparison, ranking, and trend interpretation. These tasks are exactly where financial dashboards need more than a chart caption. They need an anchored explanation that helps the reader verify the claim against the chart.

Table 7. Image complexity by mapped dashboard panel

Mapped panel	Records	Unique images	Mean complexity	Median complexity	Mean edge density	Mean color entropy
General financial context	3,992	813	50.82	51.34	0.217	0.184
Dilution Risk	1,345	285	49.18	49.35	0.193	0.188
Liquidity Risk	964	229	50.64	51.85	0.211	0.195
Probability of Default	616	136	49.60	46.55	0.209	0.183
Liquidation Anomaly	102	75	48.56	50.06	0.230	0.165

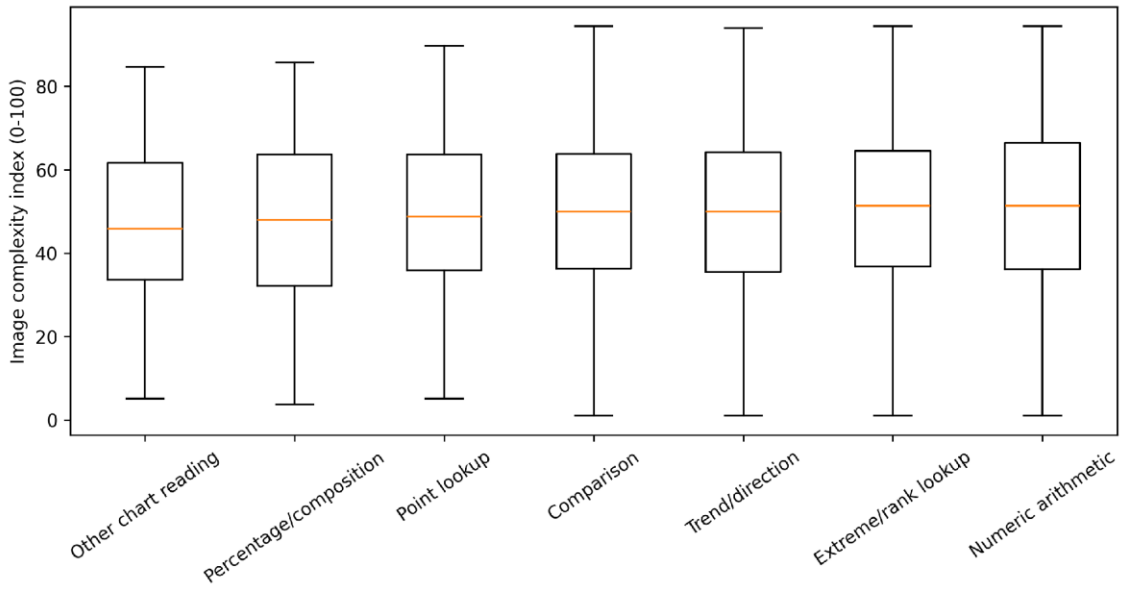

Figure 6. Image-complexity distribution by question cue

The RWA mapping also clarifies how risk semantics should be introduced. More than half of the records remain General financial context under the conservative mapping rules. This is an important design result. It shows that a dashboard should not force every financial chart into PD, dilution, liquidity, or liquidation-anomaly panels. When a chart lacks explicit risk cues, the interface should preserve a general review mode or ask for analyst classification. When a chart

does contain risk cues, the interface should make the category visible while still allowing the user to inspect the underlying chart evidence.

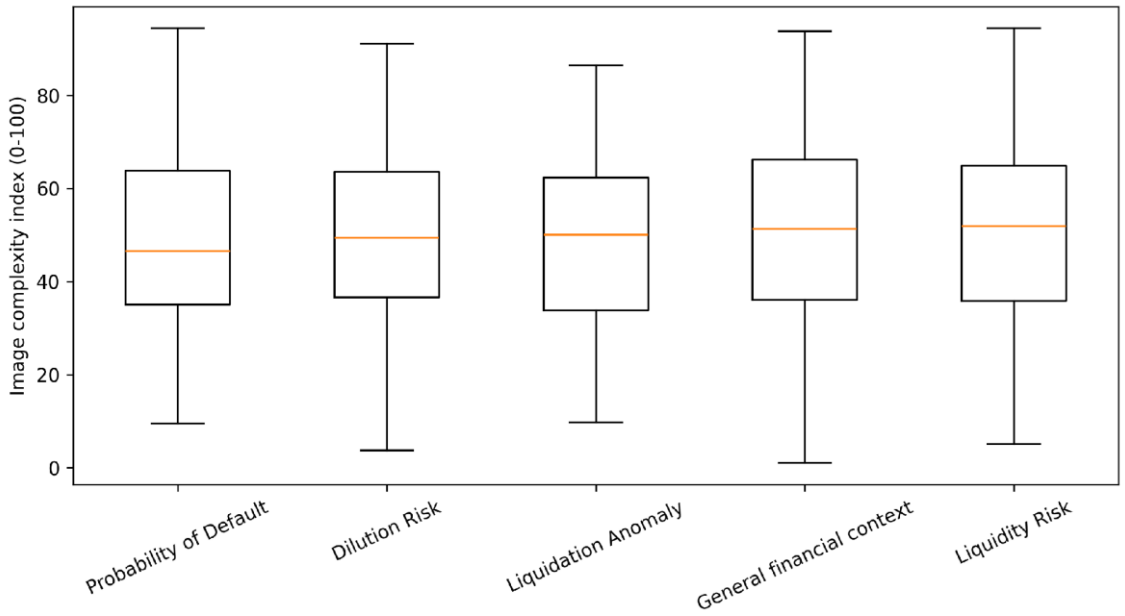

Figure 7. Image-complexity distribution by mapped dashboard panel

For RWA investors, the practical design rule is to couple one visual cue with one explanation cue. The visual cue identifies the relevant series, period, comparison, or threshold-like region. The explanation cue states the possible risk implication and caveat. This structure aligns with human-centered explanation theory because it is selective and contrastive: it explains the salient signal rather than every chart detail (Miller, 2019). It also aligns with visualization design because it reduces search costs and groups related marks into a purposeful scan path (Munzner, 2014; Shneiderman, 1996).

The image-complexity analysis suggests that interface support should not be triggered only by extreme visual clutter. Numeric arithmetic is the largest question-demand category, and arithmetic can be difficult even when the chart looks visually clean. Conversely, a visually dense chart may still be easy if it asks for a direct point lookup. This means a production dashboard should consider both visual features and task semantics. Edge density, color entropy, and mark coverage can help flag charts that need stronger visual hierarchy, but the question or decision context should determine what kind of explanation is needed.

The hybrid design avoids a common failure in AI dashboards: treating explanation as a detached sidebar. A sidebar can be ignored, and it can make the chart and text compete for attention. In the proposed layout, the explanation card is placed as a companion to the chart and uses the same risk label as the annotation. This coupling supports a verification loop: see the

highlighted evidence, read the risk interpretation, inspect the chart again, and decide whether the signal requires action.

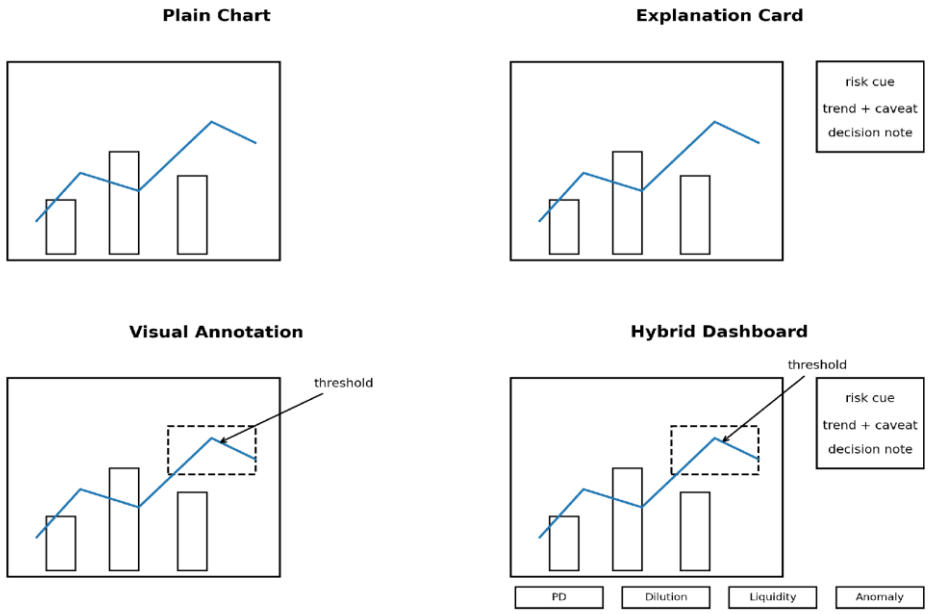

Figure 8. Dashboard treatment wireframes used to translate benchmark evidence into interface requirements

Table 8. Dashboard treatments and design constructs

Treatment	Interface treatment	Added elements	Design construct
Plain Chart	Original chart image only	No added interface layer	Evidence preservation and baseline review
Explanation Card	Chart plus compact text card	Metric summary, risk cue, caveat	Interpretive support and decision relevance
Visual Annotation	Chart plus mark-level cues	Leader lines, threshold box, highlighted region	Localization and visual hierarchy
Hybrid Dashboard	Chart, card, annotation, and RWA panels	Evidence cue, explanation cue, panel label, caveat	Auditable institutional dashboard module

The broader implication for graphic design research is that VLM performance should be evaluated at the interface level as well as at the answer level. A model may answer a chart question correctly, but the dashboard can still fail if the answer is visually detached from the chart. Conversely, a model explanation may be imperfect, but a well-designed interface can improve user verification by keeping the supporting visual evidence easy to inspect. The design target is therefore not automation alone. It is auditable augmentation.

Limitations

This study is a benchmark-based design analysis, not a live professional-investor user study. It does not report observed analyst accuracy, eye tracking, completion time, or subjective explanation-usefulness ratings. A subsequent validation study should recruit credit analysts and compare the same interface treatments under timed review conditions. The analysis also does not report live VLM inference results. The wording of the study has therefore been revised to avoid presenting the interface findings as model-performance evidence. A future VLM-connected version should run selected models directly on the chart images and report OCR errors, axis confusion, hallucinated values, instruction-following failures, and model-specific explanation quality.

The RWA panel mapping is a design taxonomy. FinChart-Bench does not directly label PD, dilution risk, liquidity risk, or liquidation anomalies. The mapping used here is transparent and conservative, but it should not be interpreted as a validated financial-risk annotation set. Real RWA portfolios may also require additional panels for concentration risk, servicer risk, collateral eligibility, legal maturity, hedging exposure, and covenant triggers. The image-complexity index is a relative corpus measure rather than a psychological measure of cognitive load. It is useful for identifying visually dense chart images within the corpus, but it does not replace user testing. The most important next step is to connect image features, task semantics, VLM outputs, and analyst behavior in one experimental workflow.

CONCLUSION

This paper revised a financial risk dashboard study into a benchmark-grounded design analysis using FinChart-Bench chart images and question records. The parsed corpus contains 1,202 unique base chart images and 7,019 chart-question records. Numeric arithmetic is the dominant question-demand category, and the conservative RWA mapping shows that many financial charts should remain in a general review mode unless explicit risk cues are present. The main design recommendation is operational. For institutional RWA dashboards, every AI-assisted chart module should contain four linked components: the original chart, a visually marked evidence region, a concise risk-specific explanation card, and a visible caveat or confidence statement. These components should be arranged so that the viewer can verify the explanation against the chart without changing screens. The contribution is a data-driven interface requirement profile for dashboard designers working with financial chart benchmarks and future VLM outputs. The revised study avoids treating generated measurements as observed analyst behavior and treats RWA categories as design mappings rather than validated financial-risk labels. This framing gives the paper a clearer and more defensible contribution for visual communication, dashboard design, and explainable financial AI.

REFERENCES

Bai, J., Bai, S., Yang, S., Wang, S., Tan, S., Wang, P., Lin, J., Zhou, C., & Zhou, J. (2023). Qwen-VL: A versatile vision-language model for understanding, localization, text reading, and beyond. arXiv. https://arxiv.org/abs/2308.12966

Basel Committee on Banking Supervision. (2017). Basel III: Finalising post-crisis reforms. Bank for International Settlements.

Basel Committee on Banking Supervision. (2023). Disclosure requirements: Pillar 3. Bank for International Settlements.

Card, S. K., Mackinlay, J. D., & Shneiderman, B. (1999). Readings in information visualization: Using vision to think. Morgan Kaufmann.

Cleveland, W. S., & McGill, R. (1984). Graphical perception: Theory, experimentation, and application to development of graphical methods. Journal of the American Statistical Association, 79(387), 531-554. https://doi.org/10.1080/01621459.1984.10478080

Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. arXiv. https://arxiv.org/abs/1702.08608

Few, S. (2006). Information dashboard design: The effective visual communication of data. O'Reilly Media.

Gorton, G., & Metrick, A. (2012). Securitized banking and the run on repo. Journal of Financial Economics, 104(3), 425-451. https://doi.org/10.1016/j.jfineco.2011.03.016

Hull, J. C. (2018). Risk management and financial institutions (5th ed.). Wiley.

Jason Kuhn, Yushan Chen, & Evelyn Chan. (2024). AI-Driven Mobile UI Pattern Recognition and Design Topic Mining on RICO: Semantic Clustering and Screenshot-Based Topic Classification. Journal of Advanced Computing Systems , 4(5), 67-83. https://doi.org/10.69987/JACS.2024.40506

Kafle, K., Price, B., Cohen, S., & Kanan, C. (2018). DVQA: Understanding data visualizations via question answering. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 5648-5656).

Kantharaj, S., Leong, R. T. K., Lin, X., Masry, A., Thakkar, M., Hoque, E., & Joty, S. (2022). Chart-to-text: A large-scale benchmark for chart summarization. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Vol. 1, pp. 4005-4023). https://doi.org/10.18653/v1/2022.acl-long.277

Li, J., Li, D., Savarese, S., & Hoi, S. (2023). BLIP-2: Bootstrapping language-image pre-training with frozen image encoders and large language models. In Proceedings of the 40th International Conference on Machine Learning (pp. 19730-19742). PMLR.

Liu, H., Li, C., Wu, Q., & Lee, Y. J. (2023). Visual instruction tuning. Advances in Neural Information Processing Systems, 36, 34892-34916.

Luo, J., Li, Z., Wang, J., & Lin, C.-Y. (2021). ChartOCR: Data extraction from chart images via a deep hybrid framework. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (pp. 1917-1925).

Masry, A., Long, D. X., Tan, J. Q., Joty, S., & Hoque, E. (2022). ChartQA: A benchmark for question answering about charts with visual and logical reasoning. In Findings of the

Association for Computational Linguistics: ACL 2022 (pp. 2263-2279). https://doi.org/10.18653/v1/2022.findings-acl.177

Methani, N., Ganguly, P., Khapra, M. M., & Kumar, P. (2020). PlotQA: Reasoning over scientific plots. In 2020 IEEE Winter Conference on Applications of Computer Vision (pp. 1517-1526).

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. Artificial Intelligence, 267, 1-38. https://doi.org/10.1016/j.artint.2018.07.007

Munzner, T. (2014). Visualization analysis and design. CRC Press.

Norman, D. A. (2013). The design of everyday things: Revised and expanded edition. Basic Books.

Pauwels, K., Ambler, T., Clark, B. H., LaPointe, P., Reibstein, D., Skiera, B., Wierenga, B., & Wiesel, T. (2009). Dashboards as a service: Why, what, how, and what research is needed? Journal of Service Research, 12(2), 175-189. https://doi.org/10.1177/1094670509344213

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, A., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021). Learning transferable visual models from natural language supervision. In Proceedings of the 38th International Conference on Machine Learning (pp. 8748-8763). PMLR.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). Why should I trust you? Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1135-1144). https://doi.org/10.1145/2939672.2939778

Shneiderman, B. (1996). The eyes have it: A task by data type taxonomy for information visualizations. In Proceedings of the IEEE Symposium on Visual Languages (pp. 336-343). https://doi.org/10.1109/VL.1996.545307

Shu, D., Yuan, H., Wang, Y., Liu, Y., Zhang, H., Zhao, H., & Du, M. (2025). FinChart-Bench: Benchmarking financial chart comprehension in vision-language models. arXiv. https://arxiv.org/abs/2507.14823

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. Cognitive Science, 12(2), 257-285. https://doi.org/10.1207/s15516709cog1202_4

Tufte, E. R. (2001). The visual display of quantitative information (2nd ed.). Graphics Press.

Wang, Z., Xia, M., He, L., Chen, H., Liu, Y., Zhu, R., Liang, K., Wu, X., Liu, H., Malladi, S., Chevalier, A., Arora, S., & Chen, D. (2024). CharXiv: Charting gaps in realistic chart understanding in multimodal LLMs. arXiv. https://arxiv.org/abs/2406.18521

Ware, C. (2019). Information visualization: Perception for design (4th ed.). Morgan Kaufmann.

Wickens, C. D. (2008). Multiple resources and mental workload. Human Factors, 50(3), 449-455. https://doi.org/10.1518/001872008X288394

Yushan Chen, & Evelyn Chan. (2023). Multimodal UI Representation Learning: Ablation of Screenshot, Wireframe, and View-Hierarchy Proxies on an Uploaded 168-Screen Dataset. Journal of Advanced Computing Systems , 3(1), 1-15. https://doi.org/10.69987/JACS.2023.30101