THE AUGMENTED BEHAVIOURAL CAPACITY MODEL: LARGE LANGUAGE MODEL ASSISTANCE AND CALIBRATED RETAIL-INVESTOR DECISION-MAKING

Dmitrii Gimmelberg
RISEBA University of Applied Sciences, Riga, Latvia
E-mail: 22-014729@riseba.lv

Abstract: This paper develops the Augmented Behavioural Capacity (ABC) Model as a domain-bounded theoretical extension of technology-acceptance and planned-behaviour frameworks for large language model (LLM) assistance in retail-investor decision-making. Using theoretical model development and existing construct-validation evidence for Perceived Cognitive Assistance (PCA), the paper argues that LLMs can enter the evaluative decision episode as conversational reasoning partners rather than merely adopted tools. ABC specifies an asymmetric Capacity-Calibration-Choice core: PCA captures felt decision-time cognitive capacity; Calibration distinguishes calibrated enhancement from miscalibrated confidence, deference, or offloading; and Choice captures proximal decision outputs. The model is motivated by a three-layer gap analysis and an eight-mechanism taxonomy of LLM-specific phenomena. Current evidence anchors PCA measurement and incremental prediction of complexity intentions, while Claim B remains a theoretically specified calibration-classification agenda requiring future matched-domain diagnostics.

Keywords: large language models, retail investors, behavioural finance, perceived cognitive assistance, calibration

INTRODUCTION

Retail-investor decision-making increasingly takes place in environments where digital tools do more than display information or execute predefined procedures. Large language models (LLMs) allow users to ask questions, structure ambiguous investment problems, compare scenarios, generate counterarguments, and translate complex financial material into user-specific language. This creates a theoretical problem for behavioural finance and information systems research. The central issue is not only whether retail investors adopt these systems, trust them, or find them useful. It is whether LLM engagement changes the evaluative stage of the decision itself: the point at which an investor defines the problem, judges uncertainty, compares alternatives, and forms a sense of whether action is cognitively manageable.

The established adoption and intention frameworks remain important but incomplete for that question. The Technology Acceptance Model explains system use through perceived usefulness and perceived ease of use; the Unified Theory of Acceptance and Use of Technology extends that logic through performance expectancy, effort expectancy, social influence, and facilitating conditions; and the Theory of Planned Behaviour explains intention through attitudes, subjective norms, and perceived behavioural control (Ajzen, 1991; Davis, 1989; Venkatesh et al., 2003, 2012). These frameworks remain necessary because LLMs are still technologies whose use depends on utility, effort, access, norms, and perceived control. Their limitation is that they mostly treat technology as the object being evaluated by the user, rather than as a participant in the evaluation of the decision problem itself.

This limitation is especially salient in retail investing. Retail investors face information overload, uneven financial literacy, attention-driven trading, and recurrent overconfidence (Agnew & Szykman, 2005; Barber & Odean, 2000, 2001; Lusardi & Mitchell, 2014). A system that can produce fluent explanations, compare strategies, and appear to validate an intended trade engages exactly the cognitive and metacognitive vulnerabilities documented in behavioural finance. Unlike a screener, dashboard, search engine, or robo-advisory interface, an LLM can sustain a multi-turn reasoning thread and co-construct the apparent structure of the choice. The relevant distinction is therefore not ordinary usefulness, but decision-time cognitive assistance.

This insufficiency is specified across three layers. Temporally, inherited frameworks model adoption and intention before or around use, not reasoning during the decision itself. Ontologically, they locate cognition in the individual user rather than in a user-model dyad where part of the reasoning is externalised to the system. Phenomenologically, they offer no clean measurement home for LLM-specific phenomena such as fluency-driven confidence, articulate deference, and cognitive offloading. This three-layer gap analysis, developed from the inherited frameworks and the post-2023 LLM literature, motivates a domain-bounded extension rather than another free-standing construct.

METHODOLOGY

The paper uses theoretical model development. It specifies the phenomenon, unit of analysis, construct registry, relationship logic, boundary conditions, rival explanations, and falsifiable propositions. This method is suited to an emerging phenomenon because a full causal model cannot be responsibly claimed before its constructs, levels of analysis, and refutation conditions are clarified (Bacharach, 1989; Whetten, 1989). Scale-development guidance makes the same point: construct conceptualisation, content validation, factor-analytic confirmation, discriminant testing, and early predictive evidence can establish a measurement backbone, but they do not by themselves validate a complete causal system (MacKenzie et al., 2011).

The unit of analysis is the evaluative decision episode: a bounded interaction in which a self-directed retail investor uses an LLM to construct, weigh, or revise a candidate trading or investing strategy at or near the moment of decision. ABC is therefore domain-bounded. It does not claim to explain all artificial intelligence adoption, all financial behaviour, professional investment decision-making, market efficiency, realised investment returns, or investor welfare. It addresses LLM-assisted retail-investor evaluative judgement under cognitively demanding conditions.

The model separates an inherited upstream layer, Adoption and Engagement, from the ABC-specific core and a downstream behavioural-footprint extension. The core contains three non-equivalent elements. Capacity is a construct, operationalised through Perceived Cognitive Assistance (PCA; Gimmelberg & Ludviga, 2025, 2026a). Calibration is a diagnostic and moderating apparatus, not a single reflective factor. Choice is a family of proximal behavioural outcomes, including complexity intentions, strategy evaluation, structural complexity of intended trades, and risk assertion. This registry prevents the architecture from being misread as a five-factor extension of existing adoption models.

The architecture is held to falsifiable discipline. Boundary conditions are stated explicitly, since a model that holds for all values of all variables is empirically empty (Bacharach, 1989; Whetten, 1989). The model is expressed as seven propositions (P1-P7) in a Lakatosian hard-core and protective-belt structure, each carrying a determinate refutation condition and required empirical operation (Lakatos, 1970; Popper, 1959). The central condition concerns Claim B: calibration must moderate the PCA-to-choice relationship. If PCA predicts the same proximal-choice pattern regardless of calibration diagnostics, the calibration overlay is not sustained.

FINDINGS

The first theoretical finding is that LLMs create a decision-time mechanism that inherited adoption frameworks do not name. Their dialogic, generative, explanatory, and context-retaining interface allows them to help users frame the problem, compare alternatives, simulate consequences, and ask whether a decision is manageable. This places the system inside the evaluative episode as a conversational cognitive artefact. The positive mechanism is cognitive scaffolding: the LLM may structure reasoning, reduce felt complexity, and support decomposition or comparison (Clark & Chalmers, 1998; Heersmink et al., 2024; Risko & Gilbert, 2016).

The second theoretical finding is the taxonomy that gives this mechanism its phenomenological content. The paper identifies eight LLM-specific mechanisms - hallucination, sycophancy, anthropomorphic dialogue, stochastic non-determinism, epistemic opacity under interactional transparency, asymmetric trust calibration, context-window adaptation, and co-construction of the option set. Each crosses the measurement boundary of an inherited adoption construct rather than fitting inside it: hallucination is not reducible to lower perceived usefulness, nor sycophancy to social influence. The contribution is the integrated taxonomy that groups these mechanisms by the construct boundaries they cross.

The third theoretical finding is that the same affordances create calibration risk. Fluent, coherent, human-like explanations can inflate perceived understanding and confidence even when underlying comprehension remains weak. Research on automation trust, overreliance, overconfidence, and the illusion of explanatory depth shows that people can become confident without becoming correspondingly accurate or knowledgeable (Lee & See, 2004; Moore & Healy, 2008; Rozenblit & Keil, 2002). ABC therefore does not treat higher perceived capacity as automatically beneficial. It makes calibration the condition under which the behavioural meaning of perceived capacity is interpreted.

The model advances two linked but distinct claims. Claim A, the behavioral-migration claim, proposes that LLM engagement may increase PCA and that elevated PCA may move proximal behavioral choice toward greater trade complexity, risk assertion, or decision-structure elaboration. This claim is value-neutral: movement toward complexity is not automatically good. Claim B, the calibration-contingent interpretation claim, proposes that the match between perceived capacity and its diagnostic referents conditions whether the same PCA-driven movement is best interpreted as calibrated enhancement or behavioural distortion.

Calibrated enhancement means that elevated PCA is matched by objective understanding, confidence accuracy, verification discipline, and retained independent reasoning. Behavioural distortion means that elevated PCA is paired with weak understanding, overprecision, uncritical deference, perceived understanding without objective comprehension, or unverified offloading. Enhancement therefore does not mean better realised results or even better ex ante decision quality. It means that the trader's felt increase in cognitive capacity is not merely a self-reported illusion.

Trading decision quality is positioned as a separate, open question rather than as part of Claim B. Whether choices made under calibrated enhancement are better structured than those made under behavioural distortion is an empirical question the model poses but does not answer, and on which it commits to no direction. No prediction is made that calibrated enhancement yields higher decision quality; a difference in either direction, or none, would be informative. The binding requirement is that ex ante decision quality be scored independently of the calibration diagnostics used for classification.

The current empirical anchoring is strongest for PCA. PCA was developed as a nine-item decision-time scale and tested with retail traders (Gimmelberg & Ludviga, 2026b). A confirmatory sample of 283 retail traders supported a one-factor structure, with comparative fit index = .966, Tucker-Lewis index = .955, standardised root mean square residual = .049, and reliability alpha = .881. Discriminant validity against proximal comparators was close but defensible, including a heterotrait-monotrait ratio of .845 against perceived usefulness. PCA also showed incremental prediction of complexity intentions over inherited proximal comparators, with beta = .130 and p = .034 in the full model, and beta = .164 and p = .009 after excluding straight-line respondents.

At the same time, evidence-status limits are reported openly. The cross-sectional indirect pathway from LLM use through PCA to choice was not confirmed. Claim A is therefore partially anchored at the component level, not established as confirmed end-to-end mediation. Claim B is theoretically specified but not operationally tested, because the matched-domain calibration battery needed to classify elevated PCA as calibrated enhancement or behavioural distortion does not yet exist as a coherent instrument package in retail-investor research.

CONCLUSIONS

The ABC Model contributes a domain-bounded architecture for understanding how LLMs may affect retail-investor evaluative decision-making. Its main theoretical contribution is to locate the missing mechanism between adoption and choice: tool-contingent perceived cognitive capacity, measured through PCA. Its second contribution is to make the interpretation of that capacity conditional on calibration. A retail investor who feels more capable after LLM interaction may indeed be better scaffolded; but the same feeling may also reflect fluency, deference, and unverified offloading. ABC makes that ambiguity the centre of the model.

The practical implication is that product designers, educators, regulators, and trading-platform operators should not evaluate LLM assistance only through usefulness, ease, adoption, or satisfaction. They should also ask whether users can explain the reasoning, verify outputs, retain independent judgement, and avoid confidence inflation. For behavioural finance, the model links LLM use to the established problem of retail-investor overconfidence without reducing the technology to a simple risk factor.

The model is intentionally limited. It does not claim improved returns, investor welfare, or full causal validation. PCA is empirically anchored as a Capacity-layer measure, while ABC remains a staged theoretical programme. The next empirical tasks are controlled decision-episode studies, stabilisation of a matched-domain calibration-diagnostic battery, direct tests of PCA-by-calibration moderation on proximal choice, and a later exploratory test, posed without directional prediction, of whether calibrated enhancement and behavioural distortion differ on independently scored ex ante decision quality.

REFERENCES

Agnew, J. R., & Szykman, L. R. (2005). Asset allocation and information overload: The influence of information display, asset choice, and investor experience. Journal of Behavioral Finance, 6(2), 57-70.

Ajzen, I. (1991). The theory of planned behavior. Organizational Behavior and Human Decision Processes, 50(2), 179-211.

Bacharach, S. B. (1989). Organizational theories: Some criteria for evaluation. Academy of Management Review, 14(4), 496-515.

Barber, B. M., & Odean, T. (2000). Trading is hazardous to your wealth: The common stock investment performance of individual investors. Journal of Finance, 55(2), 773-806.

Barber, B. M., & Odean, T. (2001). Boys will be boys: Gender, overconfidence, and common stock investment. Quarterly Journal of Economics, 116(1), 261-292.

Clark, A., & Chalmers, D. (1998). The extended mind. Analysis, 58(1), 7-19.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly, 13(3), 319-340.

Gimmelberg, D., & Ludviga, I. (2025). Strategic complexity and behavioral distortion: Retail investing under large language model augmentation. International Journal of Financial Studies, 13(4), Article 210.

Gimmelberg, D., & Ludviga, I. (2026a). Perceived cognitive assistance in LLM-augmented retail trading: Construct definition and content validation. International Journal of Financial Studies, 14(4), Article 83.

Gimmelberg, D., & Ludviga, I. (2026b). Psychometric validation of the Perceived Cognitive Assistance Scale: Exploratory and confirmatory evidence from LLM-augmented retail trading [Manuscript submitted for publication]. Social Sciences & Humanities Open.

Heersmink, R., de Rooij, B., Clavel Vazquez, M. J., & Colombo, M. (2024). A phenomenology and epistemology of large language models: Transparency, trust, and trustworthiness. Ethics and Information Technology, 26, Article 41.

Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In I. Lakatos & A. Musgrave (Eds.), Criticism and the growth of knowledge (pp. 91-196). Cambridge University Press.

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. Human Factors, 46(1), 50-80.

Lusardi, A., & Mitchell, O. S. (2014). The economic importance of financial literacy: Theory and evidence. Journal of Economic Literature, 52(1), 5-44.

MacKenzie, S. B., Podsakoff, P. M., & Podsakoff, N. P. (2011). Construct measurement and validation procedures in MIS and behavioral research: Integrating new and existing techniques. MIS Quarterly, 35(2), 293-334.

Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. Psychological Review, 115(2), 502-517.

Popper, K. R. (1959). The logic of scientific discovery. Hutchinson.

Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. Trends in Cognitive Sciences, 20(9), 676-688.

Rozenblit, L., & Keil, F. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. Cognitive Science, 26(5), 521-562.

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. MIS Quarterly, 27(3), 425-478.

Venkatesh, V., Thong, J. Y. L., & Xu, X. (2012). Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. MIS Quarterly, 36(1), 157-178.

Whetten, D. A. (1989). What constitutes a theoretical contribution? Academy of Management Review, 14(4), 490-495.