Financial advice behavior: humans versus AI

Ylva Baeckström, King's Business School, King's College London. Bush House, 30 Aldwych, London WC2B 4BG, UK. ylva.baeckstrom@kcl.ac.uk
Roman Matkovskyy, Finance and Accounting Department, Rennes School of Business, France, roman.matkovskyy@rennes-sb.com

November 25, 2025

Abstract

Financial advice can attenuate underinvestment but remains costly, biased, and skewed toward the wealthy. Artificial intelligence powered co-advisors have the potential to contribute to the development of a more scalable and affordable financial advice offering. To understand how, we use a vignette-based survey experiment to compare portfolio recommendations made by professional human advisors and GenAI (large language models, LLMs) under biased and unbiased prompts. Financial advice bias sees human advisors strongly projecting their own portfolios onto their clients, less with age and experience. AI advice projection is prompt and LLM dependent: ChatGPT is the least biased, while strong Gemini-Biased projection collapses when removing advisor demographics. LLMs are systematically more conservative than professional human advisors, recommending portfolios with approximately 1-7% lower 20-year terminal wealth for approximately 0.3-2.0 pp less volatility and growth.

Keywords: Financial advice; Lange Language Models; Artificial Intelligence; Portfolio Asset Allocation.

1. Introduction

Investment inertia and financial illiteracy contribute to the global retirement investment shortfall.[1] The potential antidote, financial advice, is plagued with problems. Mainly accessed by the wealthy, traditional advice is costly and biased.[2] Financial advisors educate their clients but misjudge client investment needs, they recommend underperforming, high-fee products and nudge clients toward mirroring the underperforming funds they themselves invest in (Choi and Robertson, 2020; Foerster et al., 2017; Gaudecker, 2015; Gennaioli, Shleifer and Vishny, 2015; Hoechle et al., 2017; Inderst and Ottaviani, 2012; Linnainmaa et al., 2021). Although research demonstrates that mediocre and biased advice contributes to advised portfolios underperforming investor self-managed portfolios, we have yet to appreciate the full extent of advisor biases and their economic consequences for clients. This is opportune given widespread client dissatisfaction with financial advisor performance and the rapid technological advancements that extend and challenge the current human financial advice model.[3]

The rapid diffusion of generative AI and large-language-model (LLM) systems has introduced algorithmic "co-advisors" whose artificial intelligence (AI) (Youyou et al., 2015) can be used to make investment decisions. Problematically, training data that has inherited statistical irregularities, can propagate harmful biases or exhibit brittle, non-transparent reasoning (Hagendorff et al., 2024; Lucy and Bamman, 2021). However, despite mounting commercial and academic interest in AI's potential to facilitate the creation of cheaper and more scalable financial advice, it is unknown whether AI advisors inherit human biases and if they make portfolio recommendations that are different from those made by professional human advisors.

[1] See for example: https://retirement.fidelityinternational.com/research-and-insights/global-retirement-survey/

[2] Only 10% of UK and 27% of US adults engage advisors with unaffordability and lack of trust deterring many See: https://www.fca.org.uk/publications/research/financial-advice-market-review-famr-baseline-report; https://business.yougov.com/content/50180-27-americans-use-financial-advisors-60-prioritizing-trust-as-the-top-factor

3 A recent survey shows that two thirds of wealthy clients are unhappy with the performance of their financial advisor with one third evaluating their advisors based on advanced technology use. See: https://www.forbes.com/sites/forbes-research/2025/07/07/the-wealthy-are-dissatisfied-with-their-financial-advisors/

Against this backdrop, we investigate how bias shapes the portfolio recommendations made by professional human and AI advisors and then compare the portfolio performance of their recommendations. We conduct our investigation through a custom-made survey-based experiment in which we ask human advisors and large language models to evaluate the investment needs and make portfolio recommendations to fictional investment clients. Our set-up allows us to address the central questions, including bias and portfolio performance, that surround the current and future state of human and AI financial advice. Two empirical gaps motivate our study. First, the underexplored magnitude and structure of advisor projection bias. With some exceptions e.g., Baeckström et al. (2021b), prior evidence is largely correlational, based on archival transaction data or single-scenario experiments, e.g., Foerster et al. (2017). Such designs cannot disentangle advisor-specific baseline leniency from scenario-specific sensitivity. Second, scholars have not yet pitted human professional and non-professional advisors against AI advisors using multiple LLM engines under identical information scenarios, using the exact same set of realistic client cases (vignettes) and prompts, that either include or withhold demographic context.

First, we collect responses from 190 experienced human financial advisors ('professional advisors') who, on average, are 39 years old, have 11.5 years of financial advice-giving experience and look after 39 clients, with US$623 million of assets under their management. Advisors report holding on average 41.1% of equities and 7.45% of fixed income in their personal 'self-owned' portfolios while their real-life clients hold 28.42% equities and 21.51% fixed income. Advisors' own portfolios therefore have higher exposure to the risky asset, equities, than the portfolios they manage on behalf of their clients. We then expose our advisors to the same set of ten richly scripted fictional client scenarios ('vignettes') that include client information such as age, gender, wealth, income, investment experience and lifestyle profiles. Advisors are asked to evaluate the investment attitudes of each vignette client and recommend one of seven investment portfolios with varied asset allocation for them to invest in. Then, to examine financial advice provided by Generative AI, we expose two LLMs (hereafter 'AI advisors') to the same survey questions and vignettes, as human participants. The models - OpenAI ChatGPT and Google Gemini - are accessed via their respective APIs under two conditions: (i) biased prompt, i.e., including human advisors' demographic and portfolio characteristics variables to to permit potential projection effects, and (ii) unbiased prompt, i.e., excluding human advisor variables to isolate cue-based inference from personal information leakage. This is to understand how closely client assessments and portfolio recommendations generated by state-of-the-art LLMs align with those of human advisors; and whether prompt design (demographic "biased" vs. stripped-down "unbiased") attenuates or exacerbates any human–AI divergence.

Our results, regarding how advisors invest for themselves, show that gender, age, confidence, risk tolerance and experience shape the equity and fixed income allocations in the personal portfolios of our advisors. Consistent with greater risk aversion in the gender literature, e.g., Charness and Gneezy (2012), female advisors invest less, both in equity and fixed income, than male advisors. Suggestive of life cycle investing (Parker et al., 2022), age raises equity and bond holdings. Self-rated risk tolerance strongly lifts equities but is near zero for bonds (Campbell, Chan and Viceira, 2003). Self-rated investment confidence exhibits a non-monotonic association with allocations: equity shares increase at moderate confidence but flatten or reverse at higher levels, with some evidence of renewed increases at the extreme upper tail. For fixed income, the pattern resembles a concave decline with a slight upturn at very high confidence (cf. Giannetti & Simonov, 2006). Greater investment experience is linked to higher equity allocations and dampens the influence of confidence on portfolio choices.

Second, assessing human 'advisor projection bias', we find that advisors project their personal equity and fixed income preferences onto their real-life clients. A +1 SD in self-owned equities translates to +0.3 SD in client equities, a slope that is similar across genders but falls with age and experience. Similarly, a +1 SD in self-FI is associated with +0.27 SD in client FI, on average; projection is about half for women (self-FI × Female is -0.139 SD) and again declines with age and experience. Third, turning to the portfolio recommendations advisors make to the fictional clients portrayed in the vignettes, we note small and inconsistent advisor covariates, reassuringly implying that advisors focus on the characteristics used to describe the fictional vignette clients rather than mirroring themselves when evaluating their clients' investment needs (Foerster et al., 2017). Advisor projection bias persists in relation to advisors' equity and their fictional-client projections. Higher self-allocated risk in advisors' personal portfolios increases the odds of suggesting the very high equity (risk) portfolios (tenure tempers this), an effect that is not present for lower risk portfolio recommendations.

Proceeding to our AI advisors, our models (ChatGPT vs. Gemini × biased vs. unbiased prompts) show that projection is not a generic property of LLM advisors: with the strongest effects for the highest-risk portfolio recommendations, it is strong for Gemini with demographics included but essentially absent for Gemini-unbiased and ChatGPT. Fifth, one-way ANOVAs with Tukey-HSD show systematic differences in portfolio recommendations across six respondent sources.[4] Relative to professional human advisors with a mean portfolio recommendation of 3.68 (of 7), AI recommendations are more conservative. Gemini-Unbiased is the most conservative with a relative difference of -1.15 and ChatGPT-Unbiased is the least conservative (-0.19). Benchmarking AI against human advisors with Deming regression, Lin's CCC and ICC, reveals systematic, directional disagreement that dominates random noise. Deming slopes for portfolio recommendations are below 1, indicating scale compression and systematically safer AI advisor over human advisor choices.

For robustness we estimate two mixed effects models using vignette-intercept (level-bias) and respondent-intercept (anchored on each advisor's baseline) models. Our results here show that, for the same client vignette, AI advisors recommend safer portfolios than human advisors and that AI deviations persist within-person (ICCs ≈ 0; vignette coefficients enlarge), implying that scenario cues, not advisor idiosyncrasies, drive the differences. AIs therefore do not track an advisor's

[4] I.e., ChatGPT vs. Gemini × biased vs. unbiased prompts, human professional advisors, human unprofessional advisors (in our student-based control sample).

personal risk level across cases. Prompt effects are asymmetric, i.e., removing demographics narrows ChatGPT's portfolio gap but intensifies Gemini's conservatism.

Finally, we analyze the economic cost of the differences in portfolio recommendations by human and AI advisors, measured by risk adjusted expected portfolio returns. Using the survey's seven risk tiers (portfolios) and capital-market assumptions, we show that a one-tier shift toward a more conservative portfolio lowers expected annual return by approximately 0.98 pp (under 1-yr assumption) and 1.07 pp annually (under 10-yr) with minimal 10-yr volatility change of approximately +0.35 pp.[5] Foregone return is thus not compensated by lower long-term risk. Applying these step effects, the results show that the robust pattern is systematically lower expected returns without commensurate risk relief. ChatGPT-Unbiased narrows this gap, while Gemini-Unbiased widens it. Thus, AI conservatism turns into sizable wealth gaps over time, i.e., Gemini-Unbiased loses about 17k (10y) and 49k (20y) per 100k invested, while ChatGPT-Unbiased losses are only approximately 1.9k (10y) and 7.9k (20y).

Economically, these safer recommendations deliver lower long-run Sharpe ratios with little volatility relief, risking chronic under-risking for growth-oriented clients. The upshot, regulatory-relevant and practical, is that replacing humans with AI does not deliver a first-best allocation. It introduces a new, measurable conservatism-compression distortion that extends current application of portfolio theory (Elton and Gruber, 1997; Markowitz, 1952) with advisory channels that can systematically shift investors below their efficient frontier, not through misestimation of inputs but through biased mapping from client characteristics to allocations.

We contribute to the literature on financial advice and delegated portfolio management (Bhattacharya and Pfleiderer, 1985; Stoughton, 1993). Although advisor projection bias is well documented (e.g., Bhattacharya et al., 2012) with client recommendations reflecting advisors' own personally owned mutual funds (Foerster et al., 2017; Linnainmaa et al., 2021), we measure how much an advisor's own equity-bond split "bleeds" into the portfolios they recommend. We also

[5] pp - percentage points; bp = basis points; 1 pp = 100 bp, 1 basis point = 0.01% = 0.0001 in decimal.

put a magnitude on the latent projection term that canonical agency models treat as exogenous. Second, our novel comparison of professional human and AI advisors strengthens evidence of the feasibility of AIs evaluating investment needs and making portfolio recommendations (Youyou et al., 2015). Our results support the view that state-of-the-art LLMs are replicating human decision-making patterns, simulating demographic-specific responses and mimicking human judgments on complex tasks (Argyle et al., 2022). As such LLMs can produce coherent survey answers comparable to humans (Horton et al., 2023) and emulate group-specific response styles when instructed (Aher et al., 2023). However, our results also underscore LLMs tendency to inherit stereotype-laden biases from their training data. Early GPT-3 exhibited gender, racial, and religious output biases (Abid et al., 2021) with explicit demographic cues triggering projection-like, biased outputs (Zhao et al., 2017), distortions that can be reduced with bias-aware prompts suppressing demographic information (Lucy and Bamman, 2021). Our novel evaluation of AI advisors' projection, portfolio recommendations, and risk-adjusted performance relative to professional human advisors can assist the practical challenges for researchers and retail investors seeking neutral AI advice. Third, we advance knowledge about AI judgement bias (Abid et al., 2021; Lucy and Bamman 2021; Zhao et al., 2017), extending it to financial decision making for others and we quantify its consequences. Our documented conservatism–compression pattern that persists when demographics are removed (Barocas and Selbst, 2016; Dwork et al., 2012), implies a structural (in our task design) rather than prompt-level artifact with prompt-level fixes insufficient for suitability tasks.

Fourth, the conservatism–compression issue reveals new technology-induced agency friction beyond the demographic and reputational distortions already catalogued in the literature. We extend portfolio theory by demonstrating that an AI advisory channel can systematically displace investors from the efficient frontier. Not by misestimating returns or risks but through biased mappings from client characteristics to allocations (functionally equivalent to misestimating client utility parameters), leading to suboptimal portfolio recommendations (Calcagno and

Monticone, 2015). This demonstrates the potential of a technological source of herding that goes beyond the benchmark-tracking incentives emphasized in traditional delegated portfolio theory. By showing a distinct mechanism when the "advisor" is a general-purpose LLM rather than a Markowitz optimizer, we complement and extend evidence that robo-advisors can improve diversification, lower volatility, and temper investor biases (D'Acunto et al., 2019). Modern, implemented robo-advisors, e.g., Vanguard PAS, (with human touch), improve diversification and Sharpe ratios while homogenizing portfolios (Rossi and Utkus, 2024). However, our study gives rise to a new frontier of general LLMs as advisors by (i) measuring human and AI projection, (ii) documenting conservatism–compression as a structural LLM property, and (iii) translating LLM–human gaps into explicit performance costs at standard risk tiers.

Finally, from a portfolio-theory perspective, current LLM advisors behave as if they overestimate investor risk aversion. Refining the Gennaioli et al. (2015) view of advisory intermediation, LLM co-advisors can curb human noise, but we propose hybrid, AI draft and human override designs to preserve the cue-responsiveness observed in practice. Current AIs behave more like "cautious counsellors" than "money doctors" Gennaioli et al. (2015). The AIs do not charge fees or manage to raise risky asset allocations to optimal levels. Instead, AIs appear to systematically under-risk clients, shrinking expected long-run returns, unless carefully prompted and governed. Therefore, the Gennaioli et al. (2015) welfare result may not apply to AI, and we identify a new, controllable effect - prompted projection, that is different from fee-driven human pandering.

Practically, our results are relevant to financial market regulators, the financial advice profession and LLM developers who cannot assume that omitting demographics or applying fine-tuning on "ethics" alone will make LLMs suitable for financial advice to retail clients. Without additional calibration layers, the models leave growth-oriented clients chronically under-risked and undermine the promise of genuinely personalized AI-advice.

2. Experimental design

2.1. Vignettes based survey experiment

We create a vignette-based experiment that we include in a tailor-made survey following the methodology by Baeckström et al. (2021b). Each of the ten vignettes aims to describe a typical wealthy investment client that advisors meet in their professional capacity. The vignettes contain variables that are often included in questionnaires used by financial advisors to establish the investment risk tolerance and investment needs of their clients. These variables are listed in Table 1 with a sample vignette in Appendix A1. While enabling advisors to make a rational evaluation about the investment needs of the fictional clients, the vignettes also contain contextual information about clients to represent early client interactions as closely as possible. Vignettes have been applied to measure judgement bias in a variety of domains, including management and economics (Aguinis and Bradley, 2014; Ambuehl and Ockenfels, 2017; Atzmüller and Steiner, 2010). In our case it is useful for assessing how advisors evaluate the client information included for arriving at their portfolio recommendations.

Akin to their real-life work method, advisors are asked to recommend one of seven portfolios with varied asset allocation and risk and return profiles to each vignette. Portfolio 1 has the lowest risk and return profile, increasing almost linearly to Portfolio 7 with the highest risk and return profile, see Table A2.[6] Following previous finance research, e.g., Bhattacharya et al. (2012), this method allows us to measure portfolio recommendations made by several (N=190) advisors to the same set of ten investors.

2.2. Professional human advisors

Our human financial advisor ('professional advisors') respondents are experienced financial advisors working for the international private banking sector in Western Europe. In total 190 advisors, a sample size consistent with other studies involving financial professionals (e.g., Crifo et al., 2015; Pikulina et al., 2017), complete the survey in late 2019 and early 2020, yielding in total

[6] The portfolios are based on The Personal Investment Management & Financial Advice Association's benchmark portfolios at the time the survey was constructed: https://www.pimfa.co.uk/.

1863 observations. Advisors were 39 years old on average, with a gender distribution of 77% male (N = 146) and 23% female (N = 44), reflecting the male dominance in the broader advisor population (Croxson et al., 2019). With an average experience of 10-11 years, our advisors manage relationships with 39 clients each. Over one third (37%) of advised clients own personal wealth exceeding $30 million with average assets under each advisor's management of $623 million in total (Table 1). Our advisors therefore work with much wealthier clients than the retail samples investigated by other scholars, e.g., (Foerster et al., 2017), relevant since the wealthy population engage financial advisors to a much greater extent than the less affluent general retail population.[7] Access to this unique sample was possible due to previous industry experience of one of the researchers. Ethical clearance was obtained prior to the survey being administered with answers recorded anonymously to ensure confidentiality. Consent was obtained to use their anonymized data as part of academic research that may be published.

Advisors' own investment attitudes, including their investment risk tolerance, investment confidence and knowledge were assessed using a Likert-type answer scales. Advisors were also asked to indicate the % of equities and % fixed income held in their personal 'self-owned' portfolios as well as the maximum investment loss they had incurred. Advisors' personal portfolios are equity heavy, with a mean of 41.35%, with a lower fixed income mean of 7.45%. Alternatives, and "other" assets are held by few, but holdings are highly skewed.

We then asked about the asset allocation in advisors' current 'real-life' client base. Notably advisors report their clients holding on average, 28.42% of equities and 21.51% fixed income, i.e., a 12.93 pp lower allocation to the risky assets, equities, than their advisors. See Tables 1 and A3.

Finally, advisors were asked to read each of the ten vignettes in turn, after which we asked: (1) 'Which portfolio would you recommend to this client?' (selecting one of the seven portfolios

[7] Statistics show that, while 70-80% of USD millionaires use financial advisors, less than 10% of the general population does. See: https://www.fca.org.uk/news/press-releases/more-people-have-bank-accounts-one-ten-have-no-cash-savings; https://www.forbes.com/sites/forbes-research/2025/07/07/the-wealthy-are-dissatisfied-with-their-financial-advisors/

described in Appendix A2); (2) 'Relative to the average investor, how much control do you think this client is likely to have over their investments?' (1 = a lot less than the average investor, 5 = a lot more than the average investor); (3) 'Relative to the average investor, how much financial risk do you think this client is willing to take with their investments?' (1 = a lot less than the average investor, 5 = a lot more than the average investor); (4) 'On a scale from 1 to 10 how knowledgeable would you rate this client to be about investments?' (1 = not at all knowledgeable, 10 = extremely knowledgeable). Table 1 shows moderately distributed portfolio recommendations with a mean recommendation of 3.68, average control and risk ratings of 3.15 and an average knowledge judgement of 5.72.

We create two composite indices to measure risk orientation: A.*RiskIndex* (advisor's personal holdings) and C.*RiskIndex* (client portfolios). Each is a weighted sum of the asset-class proportions of advisors and their real-life clients, with risk weights assigned as follows: equities (5), alternatives (4), investment property (3), fixed income (2), cash (1), and "other" (1.5). Higher scores indicate portfolios with higher risky asset shares. This preserves the original allocation structure while simplifying the model and mitigating collinearity. A.RiskIndex has a large mean of 314.92 and moderate variability, while less dispersed and lower *client risk index* with mean=266.80, reflecting lower risk client versus advisor portfolios. Their skewness values (approximately -0.25 and -0.20, respectively) suggest relatively symmetrical distributions after accounting for scale differences.

Table 1 Vignette and human professional advisor descriptive statistics

	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
					Vignette descriptives						
V.Age	47.72	13.34	45	47.27	10.38	25	74	49	0.31	-0.41	0.31
V.Gender	0.50	0.50	0	0.50	0.00	0	1	1	0.00	-2.00	0.01
V.NW	11.77	13.37	5	9.62	4.45	0.8	40	39.2	1.29	-0.02	0.31
V.Income	262.60	273.80	125	228.76	185.33	0	800	800	0.74	-0.91	6.34
V.Out	209.98	184.70	100	169.62	74.13	50	700	650	1.65	2.03	4.28
V.Invamt	2.35	1.54	2	2.26	1.48	0.5	5	4.5	0.66	-0.88	0.04
V.RecPort	3.68	1.58	4	3.66	1.48	1	7	6	0.14	-0.69	0.04
V.Control	3.15	1.04	3	3.13	1.48	1	5	4	-0.05	-0.68	0.02
V.Risktol	3.15	0.97	3	3.14	1.48	1	5	4	-0.08	-0.64	0.02
V.Knowledge	5.72	2.09	6	5.76	2.97	1	9	8	-0.14	-0.84	0.05
					Human professional advisor descriptives						
A.Age	39.15	8.74	39	38.91	10.38	21	60	39	0.22	-0.76	0.20
A.Gender	0.23	0.42	0	0.16	0.00	0	1	1	1.31	-0.30	0.01
A.Yearsexp	11.49	8.52	10	10.81	8.90	0	33	33	0.58	-0.56	0.20
A.Noclients	39.32	46.04	30	32.58	22.24	0	470	470	5.62	44.36	1.07
A.Investssetabove 30M	37.02	36.05	20	33.79	29.65	0	100	100	0.58	-1.23	0.84
A.totalAUM	623.14	737.73	375	472.39	407.72	0	5000	5000	2.39	7.48	17.09
A.Noinstitutions	2.61	1.53	2	2.42	1.48	1	8	7	0.98	0.61	0.04
A.Married	0.63	0.48	1	0.66	0.00	0	1	1	-0.54	-1.71	0.01
A.Eduaction	2.85	1.04	3	2.88	1.48	1	5	4	-0.01	-1.36	0.02
A.Children	0.58	0.49	1	0.60	0.00	0	1	1	-0.31	-1.90	0.01
A.Equities	41.35	31.42	40	39.56	37.07	0	100	100	0.33	-1.05	0.73
A.FI	7.45	12.47	0	4.93	0.00	0	100	100	3.07	15.08	0.29
A.Cash	29.45	30.46	20	24.63	22.24	0	100	100	1.14	0.13	0.71
A.Prop	17.10	26.95	0	11.68	0.00	0	100	100	1.37	0.62	0.62
A.Alternat	2.22	8.18	0	0.25	0.00	0	80	80	6.03	46.00	0.19
A.Other	2.44	8.05	0	0.29	0.00	0	55	55	4.24	19.35	0.19
A.maxloss	3.05	1.23	3	3.06	1.48	1	5	4	0.16	-0.89	0.03
A.Risktol	3.45	1.07	4	3.48	1.48	1	5	4	-0.32	-0.62	0.02
A.knowl	3.99	0.83	4	4.05	1.48	2	5	3	-0.44	-0.50	0.02
A.Conf	3.79	0.83	4	3.83	1.48	1	5	4	-0.48	0.06	0.02
C.Equities	28.42	16.56	28	27.96	17.79	0	70	70	0.27	-0.57	0.38
C.FI	21.51	14.48	20	20.46	14.83	0	90	90	1.20	3.11	0.34
C.cash	32.18	22.40	30	30.17	22.24	0	100	100	0.88	0.81	0.52
C.alternat	9.07	11.14	8	7.25	4.45	0	100	100	4.48	28.89	0.26
C.other	8.82	17.42	1	4.84	1.48	0	100	100	3.62	14.83	0.40
A.RiskIndex	314.92	108.72	310.00	318.96	111.20	100.00	500.00	400.00	-0.25	-0.59	2.52
C.RiskIndex	266.80	71.30	265.00	268.25	74.13	100.00	433.00	333.00	-0.14	-0.33	1.65

Notes: Table 1 presents an overview of descriptive statistics for all key variables used in the analysis for the fictional client vignettes, professional advisors and their real-life client bases. The table includes measures of central tendency (mean, median, trimmed mean), dispersion (standard deviation [sd], median absolute deviation [mad], range), and distribution shape (skewness, kurtosis). Additionally, minimum, maximum, and standard error (se) values are provided to understand the spread and reliability of the estimates. Male=0, female=1. V.NW=Net worth, V.Out=outgoings, V.Invamt=Investment amount.

Appendix Table A2 shows the portfolios recommended to the vignettes, V.RecPort, correlating positively with their assumed risk tolerance, V.Risktol (0.40***), and weakly with V.Knowledge (0.10***), indicating recommendations are more influenced by how risk tolerant, than financially literate, advisors perceive the fictional clients to be. V.RecPort correlates negatively with vignette age V.Age (-0.41***) and their investment amount V.Invamt (-0.16***), but positively with outgoings V.Out (0.20***) and the number of clients advisors manage (0.048*). Demographic correlations are small and mostly insignificant. "Investsset" variables show mixed, minor associations; self-reported and client portfolio measures are also small and generally non-significant. Overall, portfolio recommendations appear to reflect many small influences acting cumulatively. As such advisors make good use of all the information presented to them about the fictional clients.

2.3. AI Advisor Data

We create four AI "advisor" conditions by querying two state-of-the-art large language models, OpenAI GPT-4o and Google Gemini 2-Flash-Lite, under two prompt regimes. In the "biased" regime each vignette is prefaced with the exact demographic profile of our professional advisor respondents (advisor variables), Table 1. The "unbiased" regime includes the same vignette omitting all advisor variables. Both prompts instruct the model to recommend a portfolio based solely on the client's financial profile. One prompt was added to each model for every (advisor x vignette) combination per regime, collecting the model's single-line choice for (i) recommended portfolio, (ii) perceived client control, (iii) perceived risk tolerance, and (iv) perceived knowledge. Following Holtzman et al. (2020), and OpenAI guidance we set the temperature parameter to 0.5 to balance consistency and human-like variability in the AI-generated responses.[8]

The following demographic context was included in the biased regime: "*You are a financial advisor with the following profile: Age: {'A.Age'}, "Gender: {'A.gender}, Marital Status: {'A.Marital'},*

[8] In large language models, low temperatures (0.1–0.2) produce deterministic outputs, while high values (0.7–1.0) encourage diversity. This is especially suitable for behavioral or survey experiments, where moderate randomness simulates diverse decision-making without sacrificing structure or intent. See: (Holtzman et al., 2020).

{'A.Children'}, Education: {'A.Eduaction'}. You have {'A.Yearsexp'} years of professional experience and serve {'A.Noclients'} clients. Moreover, {'A.Investssetabove30M'}% of your clients have assets exceeding US$30M, and your current total assets under management amount to US${'A.totalAUM'} million. You have worked for {'A.Noinstitutions'} different financial institutions throughout your career. In your personal portfolio, approximately {'A.Equities'}% is allocated to equities, {'A.FI'}% to fixed income, {'A.Cash'}% is kept in cash, {'A.Invprop'}% is invested in additional property, and {'A.alternat'}% is allocated to alternative investments. As a professional financial advisor, please answer concisely using only your selection (e.g., 'Portfolio 3' or a single number) based solely on the client's financial information and asset allocation; disregard any potential biases from your personal background. Do not provide additional commentary or repeat any parts of the vignette. Here is your client case: ": (Vignette)". Although the instruction requests that any potential biases be disregarded, as is also the case with human advisors, the mere presence of the respondent's demographics in the prompt provides a test condition for potential bias.

The unbiased prompt includes only the vignettes and survey questions, with explicit instructions to answer based only on financial information and asset allocation: "*As a professional financial advisor, please answer concisely using only your selection (e.g., 'Portfolio 3' or a single number) based solely on the client's financial information and asset allocation. Do not include additional commentary or repeat any parts of the vignette. Here is your client case: (Vignette)*".

Our design yields four AI sources, i.e., ChatGPT-Biased, ChatGPT-Unbiased, Gemini-Biased, Gemini-Unbiased that are perfectly aligned to the human sample. The AIs are confronted with the identical ten fictional client vignettes and, under the biased condition, inherit the human respondent's personal profile through the inclusion of advisor variables. The set-up lets us (a) test whether LLMs replicate or diverge from professional advice on the same facts, and (b) isolate the incremental effect of removing demographic cues, i.e., the common "fairness-through-unawareness" debiasing strategy (Dwork et al., 2012; Barocas and Selbst, 2016), on the models' recommendations.

3. Advice-giving behavior: Professional humans versus AI

We combine two complementary mixed-effects designs, one absorbing vignette-level heterogeneity (random intercept by vignette) with the other absorbing respondent-level heterogeneity (random intercept by advisor), using Deming regression for method comparison (Deming, 1943; Linnet, 1993) and concordance metrics (Lin's CCC; ICC). This hybrid approach (i) cleanly separates case difficulty from judge leniency, (ii) yields slope and agreement measures that place professionals and LLMs on a common scale, and (iii) is easily reused whenever human and algorithmic judgments are elicited on the same test bed.

3.1. Advisor portfolio mirroring effects

The drivers of financial advisors' personal portfolio allocations are quantified by estimating two sets of hierarchical OLS models with polynomial contrasts for confidence (linear, quadratic, cubic, quartic terms) and selected interaction terms to capture moderation effects: (1) the proportion of equities (*A.Equities*) and (2) the proportion of fixed income (*A.FI*).

The baseline specification is:

$$A.Assets\ i = \beta 0 + \beta 1 A.Gender + \beta 2 A.Age_i + \beta 3 sA.conf.Li + \beta 4 A.conf.Qi + \beta A.conf.Ci + \beta 6 Aconf4i + \beta 7 sA.maxloss_i + \beta 8 A.risktol_i + \beta 9 sA.knowl_i + \beta 10 A.Investssetabove30M_i + \beta 11 A.totalAUM_i + \beta 12 A.Yearsexp_i + \beta 13 A.Noinstitutionsi + \beta 14 A.Eduactioni + \beta 15 AMarriedi + \beta 16 A.Childreni + \beta 17 (AGender \times A.conf.L)i + \beta 18 (A.Gender \times A.conf.Q)i + \beta 19 (A.Gender \times A.conf.C)i + \beta 20 (A.Investssetabove30M \times A.conf.L)i + \beta 21 (A.Investssetabove30M \times A.conf.Q)i + \beta 22 (A.Investssetabove30M \times sA.conf.C)i + \beta 23 (A.Yearsexp \times A.conf.L)i + \beta 24 (A.Yearsexp \times A > conf.Q)i + \beta 25 (A.Yearsexp \times A.conf.C)i + \varepsilon i \quad (1)$$

where **A. Assets i** A.Equities and A.FI for models 1 and 2, respectively. The variable A.Conf is an ordered categorical measure (1-5) of self-assessed confidence. Treating it as purely linear risks overlooking non-monotonic relationships, while treating it as a set of dummy variables inflates model complexity and collinearity. We therefore apply orthogonal polynomial contrasts, which decompose the effects into independent linear (A.Conf.L), quadratic (A.Conf.Q), cubic

(A.Conf.C), and quartic (A.Conf4). This approach preserves the ordinal nature of the variable, allows detection of complex non-linear patterns (e.g., U-shaped or S-shaped relationships), and ensures that each polynomial term is uncorrelated with the others.

We create five specifications per dependent variable, starting from a minimal model (gender, age, confidence polynomials), progressively adding controls and then interaction terms. We include interactions between confidence polynomials and key advisor characteristics, i.e., gender, client wealth concentration (Investment above $30M), and experience, to test whether the shape and strength of the self-confidence effect varies across these subgroups.

We examine whether advisors' personal portfolio composition predicts the composition of the portfolios they recommend to the vignettes or those they manage for their current clients in three sets of projection models. The dependent variable varies by model: (1) proportion of equities in clients' portfolios (*C.EquitiesC.equities*); (2) proportion of fixed-income securities in clients' portfolios (*C.FIC.FI*); and (3) composite *C.risk_index*. In all models, the primary predictor is the corresponding personal allocation measure: *A.Equities*, *A.FI*, or *A.RiskIndex*. Continuous variables are z-standardised prior to estimation. Categorical variables retain their original coding. The baseline bivariate specification is:

$$\text{C.Allocation}_i = \beta_0 + \beta_1 \, \text{A.Allocation}_i + \varepsilon_i \quad (2)$$

Where C.Allocation_i is *C.Equities* and *C.FI,* A.Allocation_i is the matched professional advisor allocation measure.

Model 2 adds advisor-level controls:

$$\text{C.Allocation}_i = \beta_0 + \beta_1 \, \text{A.Allocation}_i + \text{X}_{\text{i}\gamma} + \varepsilon_i \quad (3)$$

where X_i = [A.Gender, A.Age, A.A.Investssetabove30M, A.totalAUM, A. Yearsexp, A. Noinstitutions, A. Noclients, A.Education, A.Married, A.Children].

Models 3–5 sequentially add interaction terms to test moderation by gender, age, and professional experience:

$$\text{C.Allocation}_i = \beta_0 + \beta_1 \, \text{A.Allocation}_i + \text{X}_{\text{i}\gamma} + \beta_k (\text{A.Allocation}_i \times \text{Z}_{ik}) + \varepsilon_i \quad (4)$$

where $\mathrm{Z_{ik}} \in$ {A.Gender, A.Age,A.Yearsexp}.

Additionally, we specify the models using aggregated Client/Advisor Risk Index:

$$\text{C.riskIndex}_i = \beta_0 + \beta_1 \text{ A.RiskIndex}_i + X_{i\gamma} + \varepsilon_i \quad (5)$$

This model also is sequentially extended with the same interactions, i.e., A.RiskIndex×Gender (Female), A.RiskIndex×A.Age, A.RiskIndex×A.Yearsexp.

All models are estimated via OLS with heteroskedasticity-robust (HC1) standard errors. Standardized slopes for continuous predictors are interpreted as "SD change in the client measure for a one SD change in the predictor", and dummy variable coefficients represent SD differences relative to the reference category.

3.2. Fictional client vignette portfolio recommendations

We analyze advisors' portfolio recommendations in response to ten client vignettes using a generalized ordered logit (partial proportional odds) model. The dependent variable, *V.RecPort*, is an ordinal outcome with J=7 ordered categories (1 = most conservative, 7 = most aggressive). Unlike the standard ordered logit, this specification relaxes the proportional odds assumption by allowing each explanatory variable to have a distinct slope at each cut point j=1,…,J-1. For each cut point j, the model estimates:

$$\log\left(\frac{\Pr(Y_i \le j)}{\Pr(Y_i > j)}\right) = \alpha_j + \beta_{1j} \text{ A.RiskIndex}_i + \beta_{2j} \text{ Female}_i + \beta_{3j} \text{ A.Age}_i + \beta_{4j} \text{ A.Yearsexp}_i +$$

$$\beta_{5j} \text{ A.Noinstitutions}_i + \beta_{6j} \text{ A.Noclients}_i + \beta_{7j} \text{ A.Education}_i \quad (6)$$

where is α_j intercept at cut point j; β_{kj} is slope for predictor k at cut point j (negative values shifts probability mass above cut point j suggesting more aggressive choice); Y_i is vignette portfolio choice for decision *i*. The key predictor, *A.RiskIndex (Advisor Risk Index)*, is a continuous composite measure of the advisor's personal portfolio risk, standardized prior to estimation; other continuous covariates are also z-standardized, while binary indicators retain initial coding. Models are estimated via maximum likelihood, with heteroskedasticity-robust standard errors clustered at the advisor level to account for multiple vignette decisions per respondent. Interpretation is threshold specific.

3.3. AI vs Human models

To test whether AI portfolio recommendations reflect the human advisor's own portfolio risk posture (AI projection), we re-estimate the vignette projection models from Section 2.3, replacing the dependent variable "human *V.RecPort*" with the AI-generated portfolio recommendations under four prompt/model conditions: ChatGPT-Biased, ChatGPT-Unbiased, Gemini-Biased, Gemini-Unbiased. The dependent variable Yi is ordinal (1 = most conservative, 7 = most aggressive). For each condition we fit a generalized ordered logit model (partial proportional odds), allowing the slope for each predictor to vary across the J-1 = 6 cut-points:

$$\log\left(\frac{\Pr(Y_i \le j)}{\Pr(Y_i > j)}\right) = \alpha_j + \beta_{1j}\,\text{A.RiskIndex}_i + \beta_{2j}\,\text{A.Gender}_i + \beta_{3j}\,\text{A.Age}_i + \beta_{4j}\,\text{A.Yearsexp}_i + \beta_{5j}\,\text{A.Noinstitutions}_i + \beta_{6j}\,\text{A.Noclients}_i + \beta_{7j}\,\text{A.Education}_i + \epsilon_{ij} \quad (7)$$

where: α_j is cut-point-specific intercept; $\beta_{kj}$is slope for predictor k at cut-point j. Maximum likelihood is applied, with robust standard errors clustered by advisor ID to account for multiple vignettes per source. Therefore, prompt- and model-specific runs identify whether AI outputs "project" human self-risk patterns, and whether this projection survives removal of demographic cues.

To formally test whether the difference between advisors' and AIs' vignette recommended risk profile is negligible, we use the Two One-Sided Tests (TOST) procedure. It tests the null hypothesis that the difference is larger than this threshold (against the alternative that the difference is within the acceptable range). To compare mean portfolio recommendations and vignette perceptions (control, risktol, knowledge) across six sources (Advisors, Students, ChatGPT-Biased, ChatGPT-Unbiased, Gemini-Biased, Gemini-Unbiased), we apply one-way ANOVA with "Advisor" as the reference group, followed by Tukey's Honest Significant Difference test to adjust p-values and CIs for multiple comparisons.[9] Output is reported as mean differences (Δ) and 95 % Tukey-adjusted CIs. Negative Δ means lower ratings than advisors.

To quantify agreement between human advisors and AI outputs for identical vignette–advisor pairs, we compute three complementary statistics per outcome. First, we estimate Deming Regression (errors-in-variables, slope β and intercept α):

$$Y_{AI,i} = \alpha + \beta\, Y_{Human,i} + \epsilon_i \tag{8}$$

where both axes are measured with error; slope = 1 and intercept = 0 indicate perfect agreement. This type of errors-in-variables model is appropriate when both X and Y variables are subject to measurement error, unlike an ordinary least squares mode which assumes the independent variable is measured without error. By using Deming regression, we obtain an unbiased estimate of the slopes[10]. The Deming regression results, combined with confidence intervals, allow to test hypotheses like "is the slope = 1?" directly.

Second, to assess the agreement between human and AI recommendations, we calculate Lin's Concordance Correlation Coefficient (CCC):

$$\rho_c = \frac{2\rho\sigma_x\sigma_y}{\sigma_x^2+\sigma_y^2+(\mu_x-\mu_y)^2} \tag{9}$$

where ρ is the Pearson correlation, σ_x and σ_y are the variances, and μ_x and μ_y are the means of the two sets of recommendations. This is a more stringent test than a simple Pearson correlation, because CCC penalizes systematic differences between the two measures.

Third, we estimate Intraclass Correlation Coefficient (ICC, two-way mixed, consistency) is estimated as follows, assessing reliability of absolute scores:

---

[9] We create a control sample of non-professional advisors by collecting data from 114 international summer school students at a UK business school during the summer of 2019. Participating students had an average age of 21 and were 72% participants female. Apart from questions relating to work as professional advisors, we asked students similar questions as those posed to our advisor sample.

[10] The slopes are adjusted using the error variance ratio $\lambda = \frac{\sigma_\epsilon^2}{\sigma_\delta^2}$. This allows hypothesis tests such as β=1 (perfect proportionality) and α=0 (no systematic bias).

$$\text{ICC}(2,1) = \frac{\sigma^2_{\text{between}}}{\sigma^2_{\text{between}} + \sigma^2_{\text{within}}} \quad (10)$$

Finally, we fit linear mixed-effects models predicting four outcomes (*RecPort*, *Control*, *RiskTol*, *Knowledge*) from source type, vignette features, and their interactions. First, we test "level bias", i.e., mean displacement between AI and human ratings on the same vignette (Model A - Random Intercept by Vignette):

$$Y_{iv} = \beta_0 + \sum_k \beta_k X_{ik} + \sum_m \beta_m (\text{Source}_i \times Z_{vm}) + u_v + \epsilon_{iv} \quad (11)$$

where $X_{ik}$is fixed effects for source type, i.e., ChatGPT-Biased, ChatGPT-Unbiased, Gemini-Biased, Gemini-Unbiased (Advisors ref.); Z_{vm} are vignette-level attributes gender, net wealth, age, income, spending, investable assets); $u_v \sim N(0, \sigma_v^2)$ is random intercept for vignette v; and $\epsilon_{iv} \sim N(0, \sigma^2)$.

And Model B (Random Intercept by Respondent ID) that tests "slope bias", i.e., whether AI outputs track within-advisor adjustments to client cues.

$$Y_{ip} = \beta_0 + \sum_k \beta_k X_{ik} + \sum_m \beta_m \left(\text{Source}_i \times Z_{pm}\right) + u_p + \epsilon_{ip} \quad (12)$$

where u_p is random intercept for respondent p, capturing individual baselines. Other terms as in Model A. We apply maximum likelihood, Satterthwaite df corrections for fixed-effect SEs. All continuous predictors z-standardized; binary predictors coded without additional transformation.

4. Main Results

4.1. Drivers of personal asset allocation of financial advisors

First, we analyze whether advisors reporting higher self-assessed risk-taking propensity (self-risk) and greater over-confidence allocate a larger share of their personal portfolios to equities and fixed income (A.Equities and A.FI), see Table 4. Across all models, A.Gender advisors (Female) is negative and statistically significant, indicating female advisors holding less equities and fixed income than males, consistent with differences in risk preferences/investment style. Age is positive, implying older advisors allocate more to equities and fixed income. Overconfidence, linked to greater risk-taking and excessive trading, especially among professionals and men (Aristei and Gallo, 2022; Jackson, 2005; Gao et al., 2021; Pikulina et al., 2017), raises self-equity non-linearly and lowers self-bonds, with experience tempering the confidence–equity link.

The composite A.RiskIndex is a strong, stable predictor of equities (0.38***-0.39***) but near-zero for fixed income, indicating that a risk-seeking personal posture manifests mainly as higher equity, not lower bonds. Several controls have small, interpretable signs: more institutional clientele correlates with lower equities (-0.12***) and higher fixed income (0.03*-0.04*), and larger books/AUM slightly raise fixed income. "Max loss tolerated" enters modestly negative in both sets; knowledge is weak.

For equities, the linear confidence contrast (A.Conf.L) is positive and significant in Models 1-4 but turns negative and nonsignificant. Once interactions enter (Model 5), the quadratic term (A.Conf.Q) is negative and significant in Models 2-4; while the highest-order term (A.Conf4), is robustly positive throughout. This pattern implies a non-monotonic relation in which very high self-confidence is associated with more equity, but intermediate confidence can flatten or dip. Experience moderates this: in Model 5, Yearsexp × Conf.L = -0.994*** (dampening the average linear effect) and Yearsexp × Conf.Q = +0.743*** (amplifying curvature), with non-significant cubic interaction is, i.e., experience flattens the linear slope while accentuating tail non-linearities. The linear confidence effect is weak or ambiguous for fixed income with the quadratic term consistently negative (-0.39* to -0.82**, concave), positive quartic term (0.22**-0.31***). This yields an S-shaped response: higher confidence is associated with lower bond shares, especially away from the center; with confidence interactions mostly null.

Table 4 Drivers for professional human advisor equity and fixed income allocations

	Proportion of Advisor equities (A.Equties)					Proportion of Advisor fixed Income (A.FI)				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
A.Gender(Female)	-0.613***	-0.503***	-0.545*	-0.536***	-0.518***	0.016	-0.162**	0.050	-0.156**	-0.146**
	(0.053)	(0.052)	(0.221)	(0.052)	(0.052)	(0.053)	(0.050)	(0.210)	(0.050)	(0.050)
A.Age	0.084***	0.134***	0.129***	0.088*	0.122***	0.278***	0.197***	0.210***	0.205***	0.211***
	(0.022)	(0.037)	(0.037)	(0.037)	(0.037)	(0.022)	(0.035)	(0.035)	(0.036)	(0.035)
A.ConfA.Conf.L	0.917***	0.712***	0.730***	0.837***	-0.152	-0.027	0.163	0.154	0.196	0.478
	(0.195)	(0.184)	(0.185)	(0.186)	(0.294)	(0.194)	(0.176)	(0.176)	(0.180)	(0.282)
A.ConfA.Conf.Q	-0.304	-0.494**	-0.478**	-0.467**	0.287	-0.387*	-0.587***	-0.594***	-0.516**	-0.816**
	(0.166)	(0.162)	(0.163)	(0.167)	(0.260)	(0.166)	(0.154)	(0.155)	(0.162)	(0.250)
A.Conf.C	0.186	0.195	0.173	-0.214	-0.301	-0.311**	-0.128	-0.103	-0.122	0.028
	(0.113)	(0.107)	(0.112)	(0.123)	(0.164)	(0.113)	(0.102)	(0.106)	(0.119)	(0.157)
A.Conf4	0.185**	0.255***	0.241***	0.357***	0.472***	0.258***	0.269***	0.313***	0.277***	0.224**
	(0.065)	(0.060)	(0.068)	(0.082)	(0.077)	(0.065)	(0.057)	(0.065)	(0.079)	(0.074)
A.maxloss		-0.039*	-0.040*	-0.050**	-0.050**		-0.035*	-0.033	-0.035*	-0.029
		(0.018)	(0.018)	(0.018)	(0.018)		(0.018)	(0.018)	(0.018)	(0.018)
A.RiskIndex		0.392***	0.394***	0.378***	0.376***		-0.005	-0.006	-0.001	0.004
		(0.022)	(0.022)	(0.022)	(0.022)		(0.021)	(0.021)	(0.021)	(0.021)
A.knowl		-0.052	-0.049	-0.066*	-0.058		0.017	0.014	0.015	0.009
		(0.030)	(0.030)	(0.030)	(0.030)		(0.029)	(0.029)	(0.029)	(0.029)
A.Investssetabove30M		-0.001	-0.001	0.006*	-0.002**		-0.003***	-0.003***	-0.002	-0.003***
		(0.001)	(0.001)	(0.003)	(0.001)		(0.001)	(0.001)	(0.003)	(0.001)
A.totalAUM		0.000	0.000	0.000	0.000*		0.0001*	0.0001	0.0001*	0.0001*
		(0.000)	(0.000)	(0.000)	(0.000)		(0.00003)	(0.00003)	(0.00003)	(0.00003)
A.Yearsexp		0.115**	0.120***	0.121***	0.439***		0.005	-0.004	0.001	-0.102
		(0.036)	(0.036)	(0.036)	(0.090)		(0.034)	(0.034)	(0.035)	(0.087)
A.Noinstitutions		-0.120***	-0.121***	-0.113***	-0.118***		0.034*	0.040**	0.033*	0.032*
		(0.015)	(0.015)	(0.015)	(0.015)		(0.014)	(0.014)	(0.014)	(0.014)
A.Eduaction		-0.055**	-0.058**	-0.056**	-0.050*		-0.033	-0.025	-0.029	-0.033
		(0.020)	(0.020)	(0.020)	(0.020)		(0.019)	(0.019)	(0.019)	(0.019)
A.Married(Yes)		-0.199**	-0.186**	-0.272***	-0.168**		0.084	0.054	0.096	0.085
		(0.061)	(0.062)	(0.062)	(0.061)		(0.058)	(0.059)	(0.060)	(0.059)
A.Children		0.003	-0.009	0.118	0.051		-0.133*	-0.106	-0.152*	-0.146*
		(0.065)	(0.067)	(0.067)	(0.065)		(0.062)	(0.063)	(0.064)	(0.063)
)A.GenderFemale:A.Conf.L			0.121					-0.747		
			(0.676)					(0.644)		

A.GenderFemale:A.Conf.Q			-0.258					0.687		
			(0.506)					(0.482)		
A.GenderFemale:A.Conf.C			0.222					-0.567*		
			(0.269)					(0.256)		
A.GenderFemale:A.Conf4										
A.Investssetabove30M:A.Conf.L				-0.017					-0.002	
				(0.009)					(0.009)	
A.Investssetabove30M:A.Conf.Q				0.011					-0.0005	
				(0.007)					(0.006)	
A.Investssetabove30M:A.Conf.C				0.004					-0.001	
				(0.003)					(0.003)	
A.Investssetabove30M:A.Conf4										
A.Yearsexp:A.Conf.L					-0.994***					0.253
					(0.256)					(0.246)
A.Yearsexp:A.Conf.Q					0.743***					-0.376
					(0.204)					(0.196)
A.Yearsexp:A.Conf.C					-0.134					0.152
					(0.112)					(0.107)
A.Yearsexp:A.Conf4										
Constant	-0.093	-0.445*	-0.454*	-0.429*	-0.169	-0.118	0.229	0.229	0.226	0.112
	(0.065)	(0.192)	(0.193)	(0.192)	(0.210)	(0.064)	(0.184)	(0.184)	(0.185)	(0.202)
Observations	1,863	1,830	1,830	1,830	1,830	1,863	1,830	1,830	1,830	1,830
R^2	0.113	0.287	0.288	0.308	0.304	0.120	0.140	0.144	0.141	0.146
Adjusted R^2	0.110	0.280	0.280	0.300	0.296	0.118	0.132	0.134	0.131	0.136
Residual Std. Error	0.943	0.847	0.847	0.836	0.838	0.939	0.808	0.807	0.808	0.806
F Statistic	39.337***	40.556***	34.894***	38.302***	37.564***	42.355***	16.405***	14.505***	14.182***	14.697***

Notes: Table 4 presents how five hierarchical regressions predict A.Equities and A.FI (advisor equity and fixed income share). Models add controls and then interaction terms to capture both direct and moderating effects. A.Conf (ordered, 5 levels) is modeled with polynomial contrasts: A.Conf.L (linear trend across levels 1–5), A.Conf.Q (quadratic; U- or inverted-U-shape), A.Conf.C (cubic; higher-order "wavy" pattern), and A.Conf4 (quartic; fourth-order pattern). This specification allows non-linear responses to confidence beyond a straight line. Robust (heteroskedasticity-consistent) standard errors are reported. Significance: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

4.2. Human advisor projection

4.2.1. Real-life clients

Table 5 verifies large and stable advisor-to-client projection whereby the percentage of equities and fixed incomes (FI) securities in advisors' personal portfolios influence larger holdings in their real-life client portfolios. Advisors pass a sizable fraction of their own asset mix to clients in both sleeves (maybe asset classes? Unless sleeves is used a lot for this), roughly one-quarter to one-third of their own tilt. For equities, the self-to-client passthrough is 0.34 SD in the bivariate model, remaining approximately 0.27-0.28 SD after adding eleven controls and interactions (cols. 1-5) with a similarly large and stable passthrough at approximately 0.23-0.31 SD (cols. 6-10) for FI. Thus, roughly one-quarter to one-third of an advisor's personal tilt is transmitted to client portfolios in each asset class.

Female advisors recommend safer equity (gender ≈ -0.29 SD), but similar FI fixed income to men. Education shifts risk in opposite directions, i.e., lower client equities (approximately -0.09 SD) but higher client FI (approximately +0.11 SD). For fixed income specifically, marriage (+0.21–0.28 SD), greater AUM (small +), fewer clients (-0.12 to -0.13 SD), and higher age (+0.11 to +0.16 SD) are salient with insignificantly small corresponding effects for equities. Equity self-to-client passthrough is not reliably different by gender (A.Equities×Female ≈ +0.07, ns), but declines with advisor age and experience (A.Equities×Age = -0.087***; A.Equities×Experience = -0.069**). Fixed-income projection shows stronger heterogeneity: female advisors project markedly less of their own bond share (A.FI×Female = -0.139***) and projection falls with age (-0.108*) and tenure (-0.178***). Hence, older and more experienced advisors exhibit lower self-to-client passthrough, and female advisors, specifically for bonds, pass through about half as much as their male peers.

Table 5 Advisor real-life client equity and fixed income portfolio projections

	Dependent variable:									
	Proportion of Equities in Clients' Portfolios (C.Equities)					Proportion of Fixed income in Clients' Portfolios (C.FI)				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
A.Equities / A.FI	0.338***	0.280***	0.264***	0.268***	0.272***	0.267***	0.233***	0.285***	0.312***	0.291***
	(0.022)	(0.023)	(0.026)	(0.023)	(0.023)	(0.022)	(0.023)	(0.028)	(0.027)	(0.024)
A.Gender(Female)		-0.294***	-0.269***	-0.301***	-0.293***		0.004	0.002	0.022	-0.001
		(0.054)	(0.057)	(0.054)	(0.054)		(0.053)	(0.053)	(0.053)	(0.052)
A.Age		-0.016	-0.014	-0.036	-0.019		0.114**	0.115**	0.160***	0.091*
		(0.038)	(0.038)	(0.038)	(0.038)		(0.040)	(0.039)	(0.040)	(0.039)
A.Investssetabove30M		0.0002	0.0001	0.0002	0.0002		-0.001	-0.001	-0.001	-0.001
		(0.001)	(0.001)	(0.001)	(0.001)		(0.001)	(0.001)	(0.001)	(0.001)
A.totalAUM		-0.002	-0.001	-0.012	-0.007		0.070**	0.066**	0.073**	0.084***
		(0.025)	(0.025)	(0.025)	(0.025)		(0.025)	(0.025)	(0.025)	(0.025)
A.Yearsexp		0.137***	0.136***	0.163***	0.151***		-0.068	-0.070	-0.111**	-0.068
		(0.037)	(0.037)	(0.038)	(0.037)		(0.038)	(0.038)	(0.038)	(0.038)
A.Noinstitutions		-0.022	-0.024	-0.022	-0.021		0.018	0.018	0.014	0.023
		(0.016)	(0.016)	(0.016)	(0.016)		(0.016)	(0.016)	(0.016)	(0.016)
A.Noclients		-0.006	-0.008	-0.001	-0.008		-0.127***	-0.123***	-0.131***	-0.124***
		(0.023)	(0.023)	(0.023)	(0.023)		(0.023)	(0.023)	(0.023)	(0.023)
A.Eduaction		-0.086***	-0.087***	-0.090***	-0.093***		0.111***	0.118***	0.112***	0.092***
		(0.021)	(0.021)	(0.021)	(0.021)		(0.021)	(0.021)	(0.021)	(0.021)
A.Married(Yes)		0.032	0.035	0.060	0.045		0.278***	0.275***	0.267***	0.209**
		(0.064)	(0.064)	(0.065)	(0.064)		(0.065)	(0.065)	(0.065)	(0.066)
A.Children		0.127	0.128	0.117	0.124		0.015	0.016	0.001	0.066
		(0.069)	(0.069)	(0.069)	(0.069)		(0.071)	(0.070)	(0.070)	(0.070)
A.Equities / A.FI:A.Gender(Female)			0.070					-0.139**		
			(0.054)					(0.045)		
A.Equities / A.FI:A.Age				-0.087***					-0.108***	
				(0.023)					(0.020)	
A.Equities / A.FI:A.Yearsexp					-0.069**					-0.178***
					(0.021)					(0.028)
Constant	-0.000	0.270**	0.281**	0.281**	0.288**	-0.000	-0.506***	-0.520***	-0.472***	-0.432***
	(0.022)	(0.089)	(0.089)	(0.089)	(0.089)	(0.022)	(0.090)	(0.090)	(0.089)	(0.090)
Observations	1,863	1,863	1,863	1,863	1,863	1,863	1,863	1,863	1,863	1,863
R^2	0.114	0.164	0.165	0.171	0.169	0.071	0.132	0.136	0.145	0.150
Adjusted R^2	0.114	0.159	0.160	0.165	0.164	0.071	0.127	0.131	0.140	0.145
Residual Std. Error	0.941	0.917	0.917	0.914	0.914	0.964	0.935	0.932	0.927	0.925
F Statistic	240.286***	33.092***	30.485***	31.726***	31.400***	142.470***	25.554***	24.320***	26.217***	27.247***

Notes: Table 5 reports five regression models that examine how financial advisors' personal investment behaviors and characteristics predict the proportion of equities / fixed income in their clients' portfolios (C.Equities / C.FI). All continuous predictors were standardized, except the dependent variable (client risk index), which remained in raw points. The A.RiskIndex is in raw points; one SD ≈ 108 (Table 3); hence every slope can be read as "SD change in the client's equity weight for a one SD change in the predictor"; Categorical dummies keep their original 0/1 coding, so their β's remain "differences in SDs" relative to the reference group; and the intercept represents the the predicted value for the baseline category with average(0) continuous covariates. (ii) Categorical variables (e.g. Female = 1, Male = 0) were left in their original coding.Model 1: client equities / client fixed income=β0+β1A.Equities / A.FI+ε; Model 2: Model 1 + eleven advisor level controls (gender, A.Age, assets under advice, years' experience, no. of institutions & clients, education, marital status, children); Model 3: Model 2 + interaction A.Equities / A.FI×Female; Model 4: Model 2 + interaction A.-equities / A.-FI×A.Age; Model 5: Model 2 + interaction A.Equities / A.FI×Experience. Heteroskedasticity robust (HC1) errors were computed. The table reports standard errors in parentheses; *$p<0.05$; **$p<0.01$; ***$p<0.001$. In terms of interpretations, for instance, Model (6) has an intercept ≈ 0.000 and residual SE ≈ 0.96. If the DV were in raw pp, the intercept would be near the sample mean (≈21.51 pp from Table 1). The slope on A.FI is in SD of client fixed income per 1 SD of advisor fixed income. To express the effect in percentage points, we multiply the standardized slope by SD(C.FI) = 14.48 pp (Table 1): 0.233 × 14.48 = 3.37 pp; 0.267 × 14.48 = 3.87 pp; 0.285 × 14.48 = 4.13 pp; 0.312 × 14.48 = 4.52 pp; 0.291 × 14.48 = 4.21 pp. Hence ~0.23–0.31 SD translates to ~3.4–4.5 pp per +1 SD of the advisor's own bond share.

Finally, we use the risk indexes (A.RiskIndex and C.RiskIndex). The estimated projection from the advisor's own risk posture to that recommended for clients is large and significant across all five specifications, Table 7. The effect of the Advisor Risk Index (A.RiskIndex) on the client risk index runs from 24.6 to 17.7 points across Models 1-5. In standardized terms, that is ~0.35 - ~0.26 SD of client risk per 1 SD of advisor risk (24.6/71.3 to 18.8/71.3; slope/SD). Women start from safer baselines (-21*** pts) but project more of their own risk: the interaction is +8.1*** pts, so the slope is ~26.9 for women vs 18.8 for men, implying ~43% stronger effect. Projection weakens with both age and experience (A.Age × A.RiskIndex = -5.69*** pts/SD age; Experience × A.RiskIndex = -3.33 * pts/SD exp, with SD(exp)=8.52 yrs). More clients (+3.3* pts) associate with higher client risk. Age, education, AUM, and number of institutions are not reliably related here.

Table 7 Regression results for advisor projection: Client Risk Index

	Dependent variable:				
	Client Risk Index (client				
	(1)	(2)	(3)	(4)	(5)
A.A.RiskIndex	24.622***	19.617***	17.684***	17.942***	18.764***
	(1.551)	(1.617)	(1.851)	(1.686)	(1.667)
A.Gender(Female)		-21.278***	-18.965***	-21.619***	-21.546***
		(3.779)	(3.927)	(3.769)	(3.778)
A.Age		-3.386	-3.504	-4.274	-3.762
		(2.687)	(2.685)	(2.692)	(2.691)
A.Investssetabove30M		0.084	0.080	0.083	0.085
		(0.049)	(0.049)	(0.049)	(0.049)
A.totalAUM		0.280	0.506	-0.204	0.098
		(1.744)	(1.746)	(1.745)	(1.745)
A.Yearsexp		12.999***	13.333***	14.615***	13.941***
		(2.643)	(2.645)	(2.678)	(2.679)
A.Noinstitutions		-0.675	-1.104	-0.726	-0.544
		(1.115)	(1.132)	(1.112)	(1.116)
A.Noclients		3.282*	3.230*	3.558*	3.305*
		(1.595)	(1.593)	(1.592)	(1.593)
A.Eduaction		-2.005	-2.411	-2.445	-2.451
		(1.468)	(1.478)	(1.469)	(1.482)
A.Married(Yes)		-12.253**	-11.195*	-12.329**	-12.130**
		(4.563)	(4.585)	(4.550)	(4.559)
A.Children		17.635***	16.813***	17.739***	17.657***
		(4.903)	(4.914)	(4.889)	(4.899)
A.RiskIndex:A.Gender(Female)			8.101*		
			(3.787)		
A.RiskIndex:A.Age				-5.694***	
				(1.670)	
A.RiskIndex:A.Yearsexp					-3.327*
					(1.597)
Constant	266.802***	273.545***	276.026***	276.088***	274.932***
	(1.551)	(6.280)	(6.380)	(6.306)	(6.309)
Observations	1,863	1,863	1,863	1,863	1,863
R^2	0.119	0.175	0.177	0.180	0.177
Adjusted R^2	0.119	0.170	0.171	0.175	0.171
Residual Std. Error	66.936	64.970	64.908	64.785	64.912
F Statistic	251.945***	35.614***	33.091***	33.802***	33.067***

Notes: Table 7 presents five regression models assessing whether financial advisors' own portfolio risk (captured by A.RiskIndex) predicts the risk embedded in the asset allocations of their clients' portfolios (client_risk_index). The models examine the strength and stability of this projection effect, while controlling for advisor demographics and professional characteristics. The continuous predictors, were z standardised prior to estimation; Categorical dummies keep their original 0/1 coding, so their β's remain "differences in SDs" relative to the reference group; and the intercept represents the the predicted value for the baseline category with average(0) continuous covariates. Categorical variables (e.g. Female = 1, Male = 0) were left in their original coding. Model 1: client equities=β0+β1A.equities+ε; Model 2: Model 1 + eleven advisor level controls (gender, age, assets under advice, years' experience, no. of institutions & clients, education, marital status, children); Model 3: Model 2 + interaction A.Equities×Female; Model 4: Model 2 + interaction self-equities×Age; Model 5: Model 2 + interaction A.Equities ×A.Experience. Heteroskedasticity robust (HC1) errors were computed; (v) the table reports standard errors in parentheses; *p<0.05; **p<0.01; ***p<0.001

4.2.2. Vignette projection

To delve into the vignette projections, we build a generalized ordered-logit model (a partial-proportional-odds model), in which every explanatory variable is allowed to have a different slope at every cut-point, Table 8. Projection is threshold-specific, the Advisor Risk Index (A.RiskIndex) is negligible at low and mid cut-points/portfolios (>1- >3), marginal at >4 (β = -0.10†), and significantly negative at the top (>5: -0.27*; >6: -0.40*), implying approximately31–49% higher odds of recommending very high-equity portfolios per +1 SD in A.RiskIndex. Experience tempers aggressiveness at the top (>5: +0.29*; >6: +0.31†), more institutions are associated with a greater willingness to recommend riskier portfolio (>6: -0.29*). Gender, age, and education effects are small or sporadic with a modest safer tilt for women and higher education at mid thresholds. We show similar equity projection for men and women (0.27-0.34 SD), while women project less in fixed income (gender interaction ≈ -0.139 SD).

In our sample, women have safer composite risk index baselines (-21 pts) yet project more of their own risk (+8.1 pts) than men, illustrating that levels and slopes can move in opposite directions, consistent with context-dependent, smaller gender gaps among professionals (Filippin and Crosetto, 2016; Bollen and Posavac, 2018), while remaining compatible with settings where advisor gender can flip the client gender tolerance gap (Baeckström et al, 2021b). Age results reconcile mixed findings (Berger et al., 2014; Bucciol and Miniaci, 2011; Dimmock and Kouwenberg, 2010; Guiso et al., 2008; Trabert, 2023): projection falls with age and tenure and older/seasoned advisors thus impose less self-risk even when their portfolios remain risky. Beyond shaping advisors' own portfolios (Section 4.1), confidence does not robustly predict client allocations. Instead, personal risk chiefly predicts entry into the highest-equity tiers (6/7), with tenure pulling back at the top.

Table 8
Generalised ordered-logit model results: vignette projection

	> 1	*> 2*	*> 3*	*> 4*	*> 5*	*> 6*
A.RiskIndexA.RiskIndex	-0.14	0.01	-0.02	-0.10 †	-0.27*	-0.40*
	(0.09)	(0.06)	(0.05)	(0.05)	(0.07)	(0.12)
A.Gender (Female)	0.05	0.13	0.20 †	0.14	0.04	0.12
	(0.20)	(0.13)	(0.12)	(0.13)	(0.18)	(0.30)
A.Age	-0.05	0.11	0.05	0.11	-0.12	0.09
	(0.14)	(0.09)	(0.08)	(0.08)	(0.11)	(0.18)
A.Experience	0.24 †	-0.01	-0.03	0.03	0.29*	0.31 †
	(0.14)	(0.09)	(0.08)	(0.09)	(0.12)	(0.19)
A.Number of Institutions	-0.05	-0.00	0.02	-0.07 †	-0.15	-0.29*
	(0.06)	(0.04)	(0.03)	(0.04)	(0.05)	(0.07)
A.Number of Clients	-0.09	-0.09	-0.10**	-0.06	-0.13**	0.25
	(0.11)	(0.06)	(0.05)	(0.05)	(0.05)	(0.16)
A.Education	-0.01	0.08	0.08 †	0.10 †	0.07	-0.10
	(0.08)	(0.05)	(0.05)	(0.05)	(0.07)	(0.11)

Notes: Table 8 presents results from a generalized ordered logit model (also known as a partial proportional odds model) analyzing financial advisors' portfolio risk recommendations in response to client vignettes. Unlike standard ordered logit models, this approach allows the effects of predictors to vary across thresholds of the ordered outcome, relaxing the proportional odds assumption. The outcome variable is the vignette portfolio recommendation, categorized from 1 (most conservative) to 7 (most aggressive). The columns represent the log odds of choosing a recommendation greater than a given threshold (e.g., >1, >2, ..., >6), with separate coefficients at each level. at cut-point j, a negative βj shifts probability mass *above* j (i.e., toward riskier choices), while a positive βj shifts mass at/below j (i.e., safer). The odds multiplier for being above j per +1 unit of x is exp(-βj). Key predictor: A.RiskIndex (Advisor Risk Index). Robust standard errors in parentheses. † p < .10, * p < .05, ** p < .01, *** p < .001.

4.3. AI generated portfolio recommendations: AI advisors vs human advisors

4.3.1. First impressions

We compare six respondent sources: Human professional advisors, Students, ChatGPT (Biased Prompt), ChatGPT (Unbiased Prompt), Gemini (Biased Prompt), and Gemini (Unbiased Prompt), on four vignette-based outcomes: portfolio recommendation (rec_port), perceived client control over investments (control), perceived client risk tolerance (risktol), and perceived client investment knowledge (knowledge).[11] The descriptive comparison (Table 9) shows that professional human advisors make the most aggressive recommendations (recommended portfolio M=3.68) with student advisors slightly more conservative (3.42). AI advisors are more conservative than professional advisors across the board, especially Gemini-Unbiased (2.53). The average portfolio recommendation of 2.81 by ChatGPT-Biased, increases to 3.50, i.e., closest to human advisors, with human variables stripped in the ChatGPT-Unbiased model. The same action pushes Gemini toward the conservative recommendations further removed from those by human advisors (from 3.02 to 2.53). Regarding perceptions, humans rate client control and risk tolerance around the midpoint (~3), whereas AIs rate both lower (most means <3), aligning with their safer portfolio recommendations. For knowledge, ChatGPT tracks humans (5.65-6.08), but Gemini understates sophistication (by approximately 4.7-5.0).

4.3.2. AI advisor vignette projection

Table 10 demonstrates that projection depends on AI and prompt design. Gemini-Biased shows clear, top-loaded projection with A.RiskIndex (Advisor Risk Index) negative and highly significant from thresholds >2 to >6 (β≈-0.002…-0.016 per risk-index point). Using SD(A.RiskIndex)=108, this implies roughly 1.24× to 5.6× higher odds of recommending above those thresholds per +1 SD in advisor risk, strongest at the riskiest cut-point (>6). ChatGPT shows no material projection

[11] We create a control sample of non-professional advisors by collecting data from 114 international summer school students at a UK business school during the summer of 2019. Participating students had an average age of 21 and were 72% participants female. With the exception of questions relating to work as professional advisors, we asked students similar questions as those posed to our advisor sample.

under either prompt (coefficients ~0). Removing demographics collapses projection (Gemini–unbiased ≈ 0; ChatGPT–unbiased ≈ 0), indicating prompt content is the main leakage channel. Other covariates are small or inconsistent; in Gemini–biased, experience and institutions become negative at upper cut-points, nudging choices away from the very riskiest portfolios.

4.3.3. Human advisor versus AI advisor vignette projection

A one-way ANOVA is conducted, followed by Tukey's HSD to identify pairwise differences. Across the six respondent sources, one-way ANOVAs with Tukey-HSD contrasts (Table 11) show systematic differences in vignette outcomes. Relative to advisors (mean recommended portfolio equals 3.68), portfolio recommendations are more conservative for students (Δ = -0.26***), ChatGPT-Biased (-0.87***), ChatGPT-Unbiased (-0.19***; closest to advisors), Gemini-Biased (-0.67***), and Gemini-Unbiased (-1.15***). AIs assign significantly lower scores on perceived control (e.g., ChatGPT-Biased 2.15; ChatGPT-Unbiased 1.95; Gemini 2.38–2.84 vs. advisors 3.15) and perceived risk tolerance (ChatGPT 2.59-2.74; Gemini 2.62-2.78 vs. advisors 3.15), indicating they view clients as having less control and lower risk tolerance than advisors do. For knowledge (1-10), ChatGPT-Biased approximately matches advisors, ChatGPT-Unbiased is higher (+0.36***), and Gemini is lower (approximately -0.78 to -1.00***). This perception pattern aligns with AIs' more conservative portfolio recommendations relative to advisors. Therefore, AI systems, especially Gemini-Unbiased, recommend safer portfolios than advisors while simultaneously inferring lower client control and risk tolerance; removing demographics narrows these gaps for ChatGPT but not for Gemini.

Table 9

Descriptives of AI generated portfolio recommendations vs human responses

Panel A: Professional human financial advisors	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
V.RecPort	1863	3.68	1.58	4	3.66	1.48	1	7	6	0.14	-0.69	0.04
V.Control	1863	3.15	1.04	3	3.13	1.48	1	5	4	-0.05	-0.68	0.02
V.Risktol	1863	3.15	0.97	3	3.14	1.48	1	5	4	-0.08	-0.64	0.02
V.Knowledge	1863	5.72	2.09	6	5.76	2.97	1	9	8	-0.14	-0.84	0.05
Panel B: Students												
V.RecPort	1110	3.42	1.79	3	3.34	1.48	1	7	6	0.28	-0.99	0.05
V.Control	1109	3.22	1.01	3	3.21	1.48	1	5	4	-0.16	-0.67	0.03
V.Risktol	1108	2.97	1.03	3	2.96	1.48	1	8	7	0.14	-0.36	0.03
V.Knowledge	1106	5.96	2.01	6	6.01	1.48	1	10	9	-0.27	-0.45	0.06
Panel C: AI advisor: ChatGPT_Biased												
V.RecPort	1900	2.81	1.24	3	2.77	1.48	1	5	4	0.21	-0.71	0.03
V.Control	1900	2.15	0.95	2	2.06	0	1	4	3	0.89	-0.08	0.02
V.Risktol	1900	2.74	1.02	2	2.76	1.48	1	4	3	0.17	-1.5	0.02
V.Knowledge	1900	5.65	2.36	5	5.55	2.97	1	10	9	0.26	-0.88	0.05
Panel D: AI advisor: ChatGPT_Unbiased												
V.RecPort	1900	3.5	1.31	3	3.57	1.48	1	6	5	-0.18	-0.77	0.03
V.Control	1900	1.95	0.91	2	1.81	0	1	4	3	1.13	0.72	0.02
V.Risktol	1893	2.59	0.92	2	2.5	0	1	4	3	0.85	-1.21	0.02
V.Knowledge	1900	6.08	1.74	6	6.25	1.48	1	9	8	-0.61	-0.35	0.04
Panel E: AI advisor: Gemini_Biased												
V.RecPort	1900	3.02	1.14	3	2.98	0	1	7	6	0.41	0.44	0.03
V.Control	1898	2.84	0.85	3	2.91	1.48	1	4	3	-0.48	-0.3	0.02
V.Risktol	1727	2.78	0.94	2	2.74	0	1	4	3	0.32	-1.56	0.02

V.Knowledge	1900	4.72	2.87	4	4.53	1.48	1	10	9	0.85	-0.34	0.07
Panel F: AI advisor: Gemini_Unbiased												
V.RecPort	1900	2.53	1.02	3	2.54	1.48	1	5	4	-0.16	-1.08	0.02
V.Control	1900	2.38	0.87	2	2.29	0	1	4	3	1.05	-0.16	0.02
V.Risktol	1900	2.62	0.93	2	2.53	0	2	4	2	0.81	-1.35	0.02
V.Knowledge	1900	4.95	2.09	5	4.57	2.97	2	10	8	1.18	0.67	0.05

Notes: Table 9 presents descriptive statistics for portfolio recommendations (A.RecPort) and related perceptual evaluations (A.Control, A.Risktol, A.Knowledge) made by different groups: human financial advisors, students, and several configurations of AI models (ChatGPT and Gemini, each in biased and unbiased variants) to compare mean levels, distribution characteristics, and alignment or divergence between human and AI-generated judgments in response to standardized investment vignettes. Panel A – Financial Advisors, Panel B – Students, Panel C – ChatGPT (Biased Version), Panel D – ChatGPT (Unbiased Version), Panel E – Gemini (Biased Version), Panel F – Gemini (Unbiased Version). Advisors are the most aggressive recommenders. Students lean slightly more conservative. AI models are consistently more conservative, especially Gemini-Unbiased. ChatGPT-Unbiased comes closest to human benchmarks. Control and Risk Perception: Humans (advisors and students) perceive vignette clients as having moderate control and risk tolerance. AI systems consistently perceive clients as having much less control and lower risk tolerance, which likely drives their conservative portfolio recommendations.

Table 10
Generalized ordered-logit model results: AI projection

Term	*ChatGPT (biased)*	*ChatGPT (unbiased)*	*Gemini (biased)*	*Gemini (unbiased)*
(Intercept):1	-1.942 (0.433)***	-1.874 (0.594)**	-2.260 (0.539)***	-1.234 (0.410)**
(Intercept):2	-0.513 (0.346)	-1.585 (0.420)***	-0.458 (0.378)	-0.117 (0.339)
(Intercept):3	1.863 (0.402)***	0.013 (0.337)	2.204 (0.389)***	1.434 (0.433)***
(Intercept):4	2.520 (0.493)***	1.104 (0.370)**	3.651 (0.467)***	-1.875 (5.535)
(Intercept):5	—	2.435 (0.969)*	6.530 (0.001)***	—
(Intercept):6	—	—	11.768 (0.01)***	—
A.RiskIndex:1	0.001 (0.001)\·	-0.000 (0.001)	0.000 (0.001)	-0.000 (0.001)
A.RiskIndex:2	-0.001 (0.000)	0.000 (0.001)	-0.002 (0.001)***	0.000 (0.000)
A.RiskIndex:3	-0.001 (0.001)	-0.000 (0.000)	-0.003 (0.001)***	-0.000 (0.001)
A.RiskIndex:4	-0.001 (0.001)\·	0.000 (0.001)	-0.005 (0.001)***	-0.001 (0.006)
A.RiskIndex:5	—	0.002 (0.001)\·	-0.007 (0.000)***	—
A.RiskIndex:6	—	—	-0.016 (0.000)***	—
A.gender:1	0.020 (0.147)	-0.167 (0.206)	-0.312 (0.197)	-0.032 (0.139)
A.gender:2	0.040 (0.117)	-0.059 (0.142)	-0.029 (0.128)	-0.001 (0.114)
A.gender:3	-0.044 (0.134)	0.015 (0.114)	0.154 (0.131)	0.019 (0.147)
A.gender:4	0.012 (0.167)	0.147 (0.127)	0.072 (0.152)	2.877 (3.644)
A.gender:5	—	0.499 (0.369)	-0.125 (0.001)***	—
A.gender:6	—	—	-5.543 (0.001)***	—
A.Age:1	0.005 (0.011)	-0.006 (0.016)	0.007 (0.014)	0.003 (0.011)
A.Age:2	-0.000 (0.009)	0.002 (0.011)	-0.004 (0.010)	-0.001 (0.009)
A.Age:3	-0.008 (0.010)	0.001 (0.009)	0.004 (0.010)	0.002 (0.011)
A.Age:4	-0.010 (0.013)	-0.014 (0.010)	-0.002 (0.011)	0.242 (0.143)
A.Age:5	—	-0.014 (0.025)	-0.003 (0.001)***	—
A.Age:6	—	—	0.084 (0.001)***	—
A.Yearsexp:1	-0.014 (0.012)	0.006 (0.017)	-0.019 (0.015)	-0.000 (0.011)
A.Yearsexp:2	-0.004 (0.010)	0.002 (0.012)	0.004 (0.011)	0.005 (0.009)
A.Yearsexp:3	0.004 (0.011)	-0.005 (0.009)	-0.002 (0.010)	0.004 (0.012)

A.Yearsexp:4	0.009 (0.013)	0.003 (0.010)	0.005 (0.012)	-0.279 (0.180)
A.Yearsexp:5	—	-0.004 (0.027)	-0.018 (0.0001)***	—
A.Yearsexp:6	—	—	-0.195 (0.00001)***	—
A.Noinstitutions:1	0.042 (0.043)	-0.039 (0.060)	-0.005 (0.055)	-0.035 (0.042)
A.Noinstitutions:2	0.038 (0.035)	-0.017 (0.042)	-0.031 (0.038)	-0.021 (0.034)
A.Noinstitutions:3	-0.005 (0.040)	-0.025 (0.034)	-0.011 (0.038)	-0.023 (0.043)
A.Noinstitutions:4	0.008 (0.049)	-0.014 (0.037)	-0.002 (0.044)	0.143 (0.487)
A.Noinstitutions:5	—	0.006 (0.093)	-0.033 (0.0001)***	—
A.Noinstitutions:6	—	—	-0.511 (0.001)***	—
A.Noclients:1	0.001 (0.001)	-0.001 (0.002)	-0.001 (0.002)	0.000 (0.001)
A.Noclients:2	-0.000 (0.001)	-0.000 (0.001)	-0.001 (0.001)	-0.000 (0.001)
A.Noclients:3	-0.001 (0.001)	0.001 (0.001)	-0.001 (0.001)	0.000 (0.001)
A.Noclients:4	-0.001 (0.001)	0.000 (0.001)	0.000 (0.000)	0.065 (0.042)
A.Noclients:5	—	0.008 (0.005)	-0.004 (0.001)***	—
A.Noclients:6	—	—	0.002 (0.001)***	—
A.Education:1	-0.026 (0.057)	-0.034 (0.078)	0.063 (0.072)	-0.014 (0.054)
A.Education:2	0.037 (0.046)	0.016 (0.055)	0.063 (0.051)	-0.026 (0.045)
A.Education:3	-0.031 (0.053)	0.072 (0.045)	-0.010 (0.050)	0.018 (0.058)
A.Education:4	-0.014 (0.065)	0.072 (0.049)	0.004 (0.058)	0.299 (0.606)
A.Educacion:5	—	0.188 (0.129)	-0.063 (0.001)***	—
A.Education:6	—	—	0.827 (0.001)***	—

Notes: Coefficient estimates (standard errors in parentheses) are from four generalized ordered-logit models predicting the AI-recommended portfolio rating in each condition: ChatGPT with demographic prompts ("biased"), ChatGPT unbiased, Gemini with demographic prompts ("biased"), and Gemini unbiased. Each model relaxes the parallel-slopes assumption, allowing coefficients to vary across the cut-points shown (e.g. (Intercept):1-(Intercept):6). Slopes for each predictor at a given threshold are labeled by suffix (e.g. A.RiskIndex:3 applies to threshold 3). Predictors include the composite A.RiskIndex, respondent gender, A.Age, marital status, number of children, education, years of experience, number of institutions advised, and number of clients. A dash (—) indicates that a threshold or slope does not apply in that model. Significance: *** p < .001, ** p < .01, * p < .05, · p < .10.

Table 11
Comparison AI vs Human

Outcome	*Comparison*	*Mean Difference (95% CI)*
V.RecPort	Students – Advisors	-0.259 [-0.403, -0.115]***
	ChatGPT_Biased – Advisors	-0.870 [-0.994, -0.746]***
	ChatGPT_Unbiased – Advisors	-0.187 [-0.311, -0.063]***
	Gemini_Biased – Advisors	-0.667 [-0.791, -0.543]***
	Gemini_Unbiased – Advisors	-1.150 [-1.274, -1.026]***
C.Control	Students – Advisors	0.072 [-0.029, 0.174]
	ChatGPT_Biased – Advisors	-1.005 [-1.092, -0.917]***
	ChatGPT_Unbiased – Advisors	-1.199 [-1.286, -1.112]***
	Gemini_Biased – Advisors	-0.307 [-0.394, -0.220]***
	Gemini_Unbiased – Advisors	-0.774 [-0.861, -0.687]***
C.Risktol	Students – Advisors	-0.180 [-0.285, -0.076]***
	ChatGPT_Biased – Advisors	-0.402 [-0.492, -0.313]***
	ChatGPT_Unbiased – Advisors	-0.554 [-0.644, -0.464]***
	Gemini_Biased – Advisors	-0.370 [-0.462, -0.277]***
	Gemini_Unbiased – Advisors	-0.521 [-0.611, -0.431]***
C.Knowledge	Students – Advisors	0.238 [-0.004, 0.480]
	ChatGPT_Biased – Advisors	-0.072 [-0.280, 0.136]
	ChatGPT_Unbiased – Advisors	0.360 [0.152, 0.567]***
	Gemini_Biased – Advisors	-0.999 [-1.207, -0.791]***
	Gemini_Unbiased – Advisors	-0.777 [-0.985, -0.569]***

Notes: Table 11 presents pairwise contrasts for portfolio recommendation, control, risk tolerance, and knowledge comparing Students, ChatGPT (Biased/Unbiased), and Gemini (Biased/Unbiased) against Advisors. Reported values are mean differences (group − Advisors) with Tukey–HSD–adjusted 95% CIs (multiple-comparisons controlled). Negative = lower than Advisors; positive = higher. Significance codes: ***** p adj < .001; no symbol = p adj ≥ .05.

To further quantify how closely each large language model simulation matches human advisor recommendations, we compared every advisor's response with the answer generated for the same advisor ID and vignette under four AI conditions. For each AI-human pair, we compute (i) Deming regression (symmetric errors in X and Y) with intercept and slope that tell us the baseline bias and the proportional bias of the AI relative to humans (Carstensen 2010; Deming, 1943; Linnet, 1993); (ii) Lin's Concordance Correlation Coefficient (CCC; Lin, 1989); and (iii) Intraclass Correlation Coefficient (ICC, two-way-mixed, consistency), to show the ? reliability of absolute levels (Koo and Li, 2016; Shrout and Fleiss, 1979).

Benchmarking AI advisors against human advisors with Deming regression, Lin's CCC, and ICC shows systematic disagreement rather than noise (Table 12). For portfolio recommendations, Deming slopes are well below 1 (for ChatGPT the slopes are 0.54-0.56; for Gemini - 0.37-0.38) with positive intercepts (~0.8-1.6), indicating scale compression toward the center and safer AI portfolios. For perceived control and risk tolerance, slopes are negative (for perceived control the slopes are -0.48 to -0.01; for perceived risk they range from -1.18 to -0.77) with large positive intercepts (≈ 2.4-6.5), implying an inverted mapping relative to human ratings. AI to human agreement is highest for knowledge (Gemini-Unbiased CCC ≈ 0.43; ICC ≈ 0.46). However, slopes deviate from 1: the Gemini-Biased slope equals 2.19, the ChatGPT-Unbiased slope is 0.61, showing over- and under- dispersion, respectively. Overall concordance is poor to fair (CCC/ICC typically ≤ 0.35), evidencing stable directional biases in LLM judgments.

Table 12
Human–AI Agreement on Vignette Judgements

Outcome (advisor scale)	*AI model*	*Deming Intercept †*	*Deming Slope*	*Lin's CCC (95 % CI)*	*ICC(2,1)*
V.RecPort	ChatGPT-Biased	0.76	0.56	.33 (.29-.36)	.39
	ChatGPT-Unbiased	1.51	0.54	.28 (.24-.32)	.28
	Gemini-Biased	1.61	0.38	.25 (.22-.29)	.28
	Gemini-Unbiased	1.15	0.37	.26 (.23-.29)	.35
V.Control	ChatGPT-Biased	3.66	-0.48	-.07 (-.10 to -.04)	-.11
	ChatGPT-Unbiased	2.58	-0.20	-.03 (-.06 to -.01)	-.06
	Gemini-Biased	4.02	-0.38	-.16 (-.20 to -.12)	-.17
	Gemini-Unbiased	2.40	-.01	-.00 (-.04 to .03)	-.00
V.Risktol	ChatGPT-Biased	6.47	-1.18	-.29 (-.32 to -.25)	-.31
	ChatGPT-Unbiased	5.02	-0.77	-.18 (-.22 to -.14)	-.21
	Gemini-Biased	5.50	-0.87	-.27 (-.31 to -.23)	-.29
	Gemini-Unbiased	5.34	-0.86	-.28 (-.32 to -.24)	-.32
V.Knowledge	ChatGPT-Biased	-2.58	1.44	.33 (.29-.37)	.33
	ChatGPT-Unbiased	2.62	0.61	.34 (.30-.38)	.35
	Gemini-Biased	-7.81	2.19	.33 (.29-.37)	.36
	Gemini-Unbiased	-0.83	1.01	.43 (.39-.46)	.46

Notes: This table evaluates how closely various AI models (ChatGPT and Gemini, both biased and unbiased versions) align with human financial advisors in their judgements on hypothetical clients (vignettes). Alignment between AI models (ChatGPT/Gemini; biased/unbiased) and human advisors on four vignette outcomes (portfolio, control, risk tolerance, knowledge) is assessed via Deming regression (Intercept = bias offset; Slope = scaling agreement, 1 = perfect), Lin's CCC (−1…1; precision + accuracy), and ICC(2,1) (reliability; < .40 = poor–fair, (McBride, 2005). Portfolio Deming slopes ~ 0.54-0.56 (ChatGPT) and 0.37–0.38 (Gemini) mean compressed, safer recommendations vs. humans. Control and Risk show negative slopes (AIs infer less control and lower risk tolerance; note reversed scales). Knowledge slopes > 1 (e.g., Gemini-Biased = 2.19) mean over-dispersion. Intercepts for Control/Risk > 2 indicate an overall upward shift (AIs anchor near "A.Age/low control"). CCC/ICC are generally < .40 (poor–fair agreement); only Knowledge with Gemini-Unbiased reaches moderate concordance (CCC~.43; ICC~.46). Control and Risk exhibit negative concordance (rank order inversions).

† Deming adjusts for random error in both human and AI measures.

Finally, we estimate two complementary LMMs with full vignette covariates and source×vignette-feature interactions to ensure the robustness of our results (Tables A3-A4 in the Appendix). Model A uses a random intercept by vignette, benchmarking level-bias on identical cases (how human versus AI view the same client). Model B uses a random intercept by respondent (advisor/student), benchmarking advisor-anchored deviations (within-person, repeated measures). Overall, Model A (Table A3) quantifies cross-source level shifts on the same case. Model B (Table A4) quantifies within advisor tracking of case cues with both showing a stable, conservative tilt in AI judgments paired with lower inferred client control and risk tolerance. For knowledge, ChatGPT-Biased and Gemini-Biased score lower than advisors, ChatGPT-Unbiased is close to advisors, and Gemini-Unbiased is higher. Removing demographics narrows ChatGPT's portfolio gap but intensifies conservatism for Gemini. Switching to Model B shifts variance to the respondent level (ICCs ~0) and amplifies vignette effects, indicating that, once individual baselines are removed, client attributes drive most remaining variation.

These findings verify our prior results that (i) AIs provide less risky recommendations than human advisors even when holding vignette fixed (Model A); (ii) AIs do not track advisor baselines (Model B), providing recommendations that still show a stable offset from that advisor; (iii) interclass correlation (ICC) drop shows within-person, scenario-driven variance implying, that the vignette details drive the rating, and (iv) asymmetric prompt effects are present (ChatGPT narrows responses while Gemini widens them).

4.4.4. Economic consequences of AI advice

Finally, using the survey's seven risk tier portfolio asset weights (Appendix A2) and capital-market assumptions (Table A5), we map the observed tier gaps (Table 11) into expected return, volatility, and Sharpe ratio differences. Portfolio outcomes are computed via weighted averages for return and (diagonal) weighted variances for volatility; average per-tier step effects come from Table A8.

Tables 13a and 13b present the converted gaps between AI and professional human advisors into changes in return, volatility and Sharpe ratios of the recommended portfolios, showing the potential lifetime wealth loss, when one follows AI recommendations.

Table 13a shows that moving one tier toward caution lowers expected return by roughly 0.98 pp (1-yr) and 1.07 pp (10-yr). Long-run volatility changes little (+0.35 pp per tier). Because our short-horizon inputs imply declining 1 yr volatility at higher tiers, shifting to more conservative tiers can raise 1-yr volatility and yield large percentage changes in 1-yr/3-yr Sharpe (denominators near zero). Short-horizon Sharpe ratios are therefore unstable, with a robust takeaway of return shortfall with little, long-run risk reduction. Untuned LLMs recommend portfolios with annualized expected 10 year returns of 0.2–1.2 pp less than professional human advisors without commensurate volatility relief, implying meaningful expected wealth losses if adopted at scale. ChatGPT with "unbiased" prompts narrows this cost, while Gemini (especially unbiased) widens it.

Using the professional human advisor mean recommended portfolio (3.68) and the Tier 3/Tier 4 10-year return points from Tables A6-A7, the implied advisor baseline is approximately 4.27% p.a. (Table 13b). Linear interpolation at that tier lets us convert each AI to human tier gap into an annual return shortfall and subsequent financial losses via compounding. Because a one-tier move toward caution trims ~1.07 pp from the 10-yr expected return with little volatility relief (Table A7), AI conservatism turns into sizable wealth gaps over time. A 100,000 investment in the mean portfolio recommended by professional human advisors would be expected to grow to about 230,600 over 20 years. Relative to this benchmark, Gemini-Unbiased leaves investors with approximately 17,000 less after 10 years and roughly 49,000 less after 20 years per 100k invested, equivalent to an annualized return shortfall of about 1.23 percentage points. By contrast, ChatGPT-Unbiased forgoes only approximately 1,900 (10y) and 7,900 (20y), corresponding to an annualized shortfall of about 0.2 percentage points (Fig A1, Appendix).

Table 13a
Δ return, Δ volatility and Δ Sharpe ratio per financial advisor pair, AI or Students versus Prof. Advisors

Group comparison of differences between AIs/Students minus Prof. advisors	Mean	Return loss 1y	Return loss 3Y	Return loss 5y	Return loss 10y	Volatility Δ 1y	Volatility Δ 3y	Volatility Δ 5y	Volatility Δ 10y	Sharpe ratio 1y	Sharpe ratio 3y	Sharpe ratio 5y	Sharpe ratio 10y
Students – Prof. advisors	-0.259	-0.25%	-0.42%	-0.40%	-0.28%	0.21%	0.04%	-0.06%	-0.09%	31.86%	287.93%	-173.59%	-79.89%
ChatGPT_Biased – Prof. Advisors	-0.87	-0.85%	-1.40%	-1.36%	-0.93%	0.69%	0.13%	-0.20%	-0.30%	107.03%	967.18%	-583.11%	-268.37%
ChatGPT_Unbiased – Prof. Advisors	-0.187	-0.18%	-0.30%	-0.29%	-0.20%	0.15%	0.03%	-0.04%	-0.06%	23.00%	207.89%	-125.33%	-57.68%
Gemini_Biased – financial Advisors	-0.667	-0.65%	-1.07%	-1.04%	-0.71%	0.53%	0.10%	-0.16%	-0.23%	82.05%	741.50%	-447.05%	-205.75%
Gemini_Unbiased – Prof. Advisors	-1.15	-1.13%	-1.85%	-1.80%	-1.23%	0.92%	0.17%	-0.27%	-0.40%	141.47%	1278.45%	-770.77%	-354.74%

Note: This table quantifies the financial impact of risk-tier deviations between each group (Students or AI models) and professional financial advisors. The table reports group–advisor differences ("AI/Students minus Prof. Advisors") in expected return, volatility, and Sharpe ratio after translating the average risk-tier gap for each group into portfolio outcomes at 1-, 3-, 5-, and 10-year horizons calculated as $\Delta\mu_i = \text{gap}_i \times \text{average return change per step}$, $\Delta\sigma_i = \text{gap}_i \times \text{average volatility change per stepm}$ $\Delta Sharp_i = \text{gap}_i \times \text{average Sharpe ratio change per step}$. Prof. Advisors=professional human advisors. Negative tier gaps indicate more conservative recommendations than advisors. Return and volatility entries are percentage-point (pp) differences in annualized values; Sharpe entries are percent changes relative to advisors' baseline Sharpe at each horizon. "Tier gap" is the mean difference in recommended risk tier (group minus advisors).

Table 13b
Economic impact of AI advisor conservatism - terminal wealth shortfalls vs. Prof. Advisors

Group	ΔTier	Δr (pp/y)	AI Terminal	Loss	Shortfall	AI Terminal	Loss	Shortfall
			10-year horizon			20-year horizon		
ChatGPT (Biased)	-0.870	0.93	138,840	13,014	8.6%	192,766	37,832	16.4%
ChatGPT (Unbiased)	-0.187	0.20	149,949	1,905	1.3%	222,740	7,858	3.4%
Gemini (Biased)	-0.667	0.71	141,784	10,070	6.6%	201,026	29,571	12.8%
Gemini (Unbiased)	-1.150	1.23	134,862	16,992	11.2%	181,877	48,719	21.1%
Advisor baseline†	—	—	151,854	—	—	230,596	—	—

Note: Initial wealth 100,000. "ΔTier" means risk-tier gap (group minus advisors; Table 12). Annual return shortfall Δr = ΔTier × average 10-year per-tier return step (1.0696 pp; Table A8). Terminal wealth is computed by compounding at (r_advisor − Δr) with r_advisor ≈ 3.62%+0.68×(4.57%−3.62%)≈4.266%≈4.27%. Advisor terminals are 151,854 (10y) and 230,596 (20y). No volatility/covariance effects; results isolate return drag from more conservative tiers. Percent shortfall = loss / advisor terminal.

5. Summary and conclusions

We use ten richly specified client vignettes to whom financial advisors recommend one of seven portfolios with varied risk and return profiles. Matched one-for-one to 190 professional human advisors and mirrored to AI advisors, GPT-4-o and Gemini under biased vs. unbiased prompts, we deliver four core findings.

First, professional human advisor projection is large and structured. Advisors' own portfolios spill into client portfolios by about one quarter to one-third SD for both equities and fixed income. Passthrough declines with age and experience; women set safer baselines and, for fixed income, project roughly half as much as men. In ordered logit vignette models, projection concentrates at the top tiers (6/7); tenure tempers the jump into the riskiest portfolios. Second, AI advisors exhibit conservatism–compression. Across identical cases, all large language model (LLM) settings recommend safer, less dispersed portfolios than professionals. Third, prompt design matters but doesn't cure it. Supplying human advisor variables in our biased-prompt settings induces "AI projection" in Gemini while removing them collapses projection across both models. Yet the conservatism–compression band persists under unbiased prompts, indicating a structural model property rather than mere demographic leakage. Fourth, because our seven portfolios pay almost linearly with increased risk, a one-tier de-risking costs ~0.98 pp over a 1-year holding period. Relative to professional human advisors, expected shortfalls run from -65 to -113 bp per annum for some AI settings which compound to sizable wealth gaps over time.

Thus, adopting a "machine-psychology" lens (Hagendorff et al., 2024) on standardized client vignettes unearths three regularities. First, AI projection is model and prompt dependent. Advisor traits pass through to recommendations for Gemini-Biased but not for ChatGPT (in either regime); removing demographics collapses projection, consistent with prompt sensitivity and cross-model heterogeneity in LLMs (Ouyang et al., 2022). Alignment and safety tuning steer models toward safer defaults (Argyle et al., 2022), and the resulting projection is systematic and model-specific (Jia et al., 2024). The upside is that it can protect novices from extreme bets and promote baseline diversification when client inputs are noisy or strategically framed, but the downside is that - it can misfit atypical but legitimate profiles (e.g., long horizons, high risk capacity, urgent liquidity), effectively trading personalization for safety and consistency. Second, the results show conservatism with compression: relative to professionals, LLM recommendations are safer and less dispersed. Under "fairness-through-unawareness" (Dwork et al., 2012; Barocas and Selbst, 2016), ChatGPT-Unbiased shows stronger negative wealth elasticity (more conservative as net worth rises), while Gemini-Unbiased shows a sign-reversing positive age interaction (riskier as age rises). Adding demographics narrows ChatGPT's spread but widens Gemini's, both are still below human variability, consistent with variance attenuation (Argyle et al., 2022) and stable directional biases, not mere noise (Hagendorff et al., 2024). Prior work shows that redacting protected terms dampens but does not remove stereotyping (Lucy and Bamman, 2021) and that credit/mortgage ML models retain demographic skews even without protected features (Fuster et al., 2022; Jagtiani and Lemieux, 2019), reinforcing critiques of "debiasing-by-omission" (Dwork et al., 2012; Barocas & Selbst, 2016; Obermeyer et al., 2019): formal neutrality may reduce overt bias yet also mute legitimate cue responsiveness and personalization. Third, building on evidence that algorithms struggle to replicate expert nuance (Dietvorst, Simmons and Massey, 2015), we propose the conservatism–compression hypothesis: LLM advisors systematically exhibit lower mean aggressiveness and reduced variability than human advisors because model training and safety layers over-smooth contextual signals.

Practically, clients gravitate to robo-advice (D'Acunto, Rossi & Weber, 2019) with humans remaining more variable and context adaptive (Foerster et al., 2017; Bhattacharya et al., 2012). In our data, advisor demographics shape allocations (women safer; older tilt to equities and fixed income; overconfidence raises equity nonlinearly), whereas AI (biased or unbiased) shows lower variance and weaker cue-responsiveness; stripping demographics can improve formal fairness yet further dampens sensitivity, clustering around conservative averages. Behavioral finance views human choices as noisy but context-responsive (Kahneman and Tversky, 1979; Barberis, 2018);

current LLMs act as conservative, homogenizing copilots. Hence, LLMs are not black-box humans: used alone they narrow and de-risk advice; in hybrid workflows they may curb human projection and unwarranted variance but risk under-tailoring atypical clients.

Human advice is not bias free and requires monitoring and tempering of projections. Fairness-through-unawareness (Dwork et al., 2012; Barocas and Selbst, 2016) is insufficient for AI. Therefore, we conclude that the mere availability of AI-assisted financial advice cannot solve the current problematic and biased human financial advice process. In its current stage of development, LLMs do not secure better portfolio performance than that provided by professional human advisors, with growth-oriented clients chronically under-risked. Financial market regulators, the financial advice profession, and LLM developers need to master the evolution of the next generation of AI advisors to ensure their suitability to administer investment recommendations to the underinvested mass market, and the growing cohort of advised investors who are disgruntled with the performance of their human advisors.

References

Abid, A., Farooqi, M., & Zou, J. (2021). Persistent anti-Muslim bias in large language models. *Nature Machine Intelligence*, 3, 461–463.

Aher, G. V., Arriaga, R. I., & Kalai, A. T. (2023). Using large language models to simulate multiple humans and replicate human subject studies. In A. Krause, E. Brunskill, K. Cho, B. Engelhardt, S. Sabato, & J. Scarlett (Eds.), Proceedings of the 40th International Conference on Machine Learning (Vol. 202, pp. 337–371). PMLR. https://proceedings.mlr.press/v202/aher23a.html

Aguinis, H., & Bradley, K. J. (2014). Best practice recommendations for designing and implementing experimental vignette methodology studies. *Organizational research methods*, 17(4), 351-371.

Ambuehl, S., & Ockenfels, A. (2017). The ethics of incentivizing the uninformed: A vignette study. *American Economic Review*, 107(5), 91-95.

Argyle, L. P., Busby, E. C., Fulda, N., Gubler, J., Rytting, C., & Wingate, D. (2022). Out of one, many: Using language models to simulate human samples. *Political Analysis, 31*(3), 337–352. https://doi.org/10.1017/pan.2023.2

Atzmüller, C., & Steiner, P. M. (2010). Experimental vignette studies in survey research. Methodology.

Baeckström, Y., Marsh, I. W., & Silvester, J. (2021a). Financial advice and gender: Wealthy individual investors in the UK. *Journal of Corporate Finance*, 71, 101882.

Baeckström, Y., Marsh, I. W., & Silvester, J. (2021b). Variations in investment advice provision: A study of financial advisors of millionaire investors. *Journal of Economic Behavior & Organization*, 188, 716–735.

Barberis, N. (2018). The Psychology of Tail Events: Progress and Challenges. *American Economic Review*: Papers & Proceedings, 108, 563–567.

Barocas, S., and Selbst, A. D. (2016). Big data's disparate impact. *California Law Review*, 104(3), 671–732.

Berger, A. N., Kick, T., & Schaeck, K. (2014). Executive board composition and bank risk taking. *Journal of Corporate Finance*, 28, 48–65.

Bhattacharya, S., & Pfleiderer, P. (1985). Delegated portfolio management. *Journal of Economic Theory*, 36(1), 1–25.

Bhattacharya, U., Hackethal, A., Kaesler, S., Loos, B., and Meyer, S. (2012). Is unbiased financial advice to retail investors sufficient? Answers from a large field study. *Review of Financial Studies*, 25(4), 975–1032.

Bollen, N. P. B., and Posavac, S. (2018). Gender, risk tolerance, and false consensus in asset allocation recommendations. *Journal of Banking & Finance*, 87, 304–317.

Bucciol, A., & Miniaci, R. (2011). Household portfolios and implicit risk preference. *Review of Economics and Statistics*, 93(4), 1235–1250.

Calcagno, R., & Monticone, C. (2015). Financial literacy and the demand for financial advice. *Journal of Banking & Finance*, 50, 363-380.

Campbell, J. Y., Chan, Y. L., & Viceira, L. M. (2003). A multivariate model of strategic asset allocation. *Journal of financial economics*, 67(1), 41-80.

Carstensen, B. (2010). Comparing Clinical Measurement Methods: A Practical Guide. John Wiley & Sons.

Choi, J.J. and Robertson, A.Z. (2020), "What matters to individual investors? Evidence from the horse's mouth", *The Journal of Finance*, Vol. 75 No. 4, pp. 1965-2020, doi: 10.1111/jofi.12895.

Crifo, P., Forget, V. D., & Teyssier, S. (2015). The price of environmental, social and governance practice disclosure: An experiment with professional private equity investors. *Journal of Corporate Finance*, 30, 168–194

Croxson, K., Mittendorf, D., Ng, C., & Robertson, H. (2019). Research Note: Gender diversity in UK financial services. Financial Conduct Authority. Available on https://www.fca.org.uk/publication/research/research-note-gender-diversity-in-uk-financial-services.pdf.

D'Acunto, F., Prabhala, N., & Rossi, A. G. (2019). The promises and pitfalls of robo-advising. *Review of Financial Studies*, 32(5), 1983–2020

Deming, W. E. (1943). Statistical Adjustment of Data. Wiley.

Dietvorst, B. J., Simmons, J. P., and Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, 144(1), 114–126.

Dimmock, S. G., & Kouwenberg, R. (2010). Loss-aversion and household portfolio choice. *Journal of Empirical Finance*, 17(3), 441–459.

Dwork, C., Hardt, M., Pitassi, T., Reingold, O., & Zemel, R. (2011). Fairness through awareness. arXiv preprint arXiv:1104.3913. https://doi.org/10.48550/arXiv.1104.3913

Elton, E. J., & Gruber, M. J. (1997). Modern portfolio theory, 1950 to date. *Journal of banking & finance*, 21(11-12), 1743-1759.

Filippin, A., & Crosetto, P. (2016). A reconsideration of gender differences in risk attitudes. *Management Science*, 62(11), 3138–3160.

Financial Conduct Authority. (2015/2023). Knowing your customer and assessing their needs. https://www.fca.org.uk/firms/knowing-customer-assessing-needs

Foerster, S., Linnainmaa, J. T., Melzer, B. T., & Previtero, A. (2017). Retail financial advice: does one size fit all?. *The Journal of Finance*, 72(4), 1441-1482

Fuster, A., Goldsmith Pinkham, P., Ramadorai, T., & Walther, A. (2022). Predictably unequal? The effects of machine learning on credit markets. *The Journal of Finance*, 77(1), 5 47.

Gao, H., Shi, D., & Zhao, B. (2021). Does good luck make people overconfident? Evidence from a natural experiment in the stock market. *Journal of Corporate Finance*, 68, 101933.

Gaudecker, H.-M.V. (2015), How Does Household Portfolio Diversification Vary with Financial Literacy and Financial Advice?*The Journal of Finance*, 70: 489-507.

Gennaioli, N., Shleifer, A., & Vishny, R. (2015). Money doctors. *The Journal of Finance*, 70(1), 91–114.

Giannetti, M., & Simonov, A. (2006). Which investors fear expropriation? Evidence from investors' portfolio choices. *The Journal of Finance*, 61(3), 1507-1547.

Guiso, L., Sapienza, P., & Zingales, L. (2008). Trusting the stock market. *The Journal of Finance*, 63(6), 2557–2600.

Hagendorff, T., Dasgupta, I., Binz, M., Chan, S. C. Y., Lampinen, A., Wang, J. X., Akata, Z., & Schulz, E. (2024). Machine psychology (arXiv:2303.13988v6). arXiv. https://doi.org/10.48550/arXiv.2303.13988

Hoechle, D., Ruenzi, S., Schaub, N., & Schmid, M. (2017). The impact of financial advice on trade performance and behavioral biases. *Review of Finance*, 21(2), 871–910.

Holtzman, A., Buys, J., Forbes, M., & Choi, Y. (2020). The curious case of neural text degeneration. arXiv preprint arXiv:1904.09751.

Horton, J. J. (2023). Large language models as simulated economic agents: What can we learn from Homo Silicus? (arXiv:2301.07543). arXiv. https://doi.org/10.48550/arXiv.2301.07543

Inderst, R., Ottaviani, M., 2012. How (not) to pay for advice: a framework for consumer financial protection. *Journal of Financial Economics*, 105 (2), 393–411.

Jackson, A. R. (2005). Trade generation, reputation, and sell-side analysts. *The Journal of Finance*, 60(2), 673–717.

Jagtiani, J., and Lemieux, C. (2019). The Roles of Alternative Data and Machine Learning in Fintech Lending: Evidence from the LendingClub Consumer Platform. Federal Reserve Bank of Philadelphia, Working Paper 18-15R

Jia, J., Yuan, Z., Pan, J., McNamara, P., & Chen, D. (2024). Decision-making behavior evaluation framework for LLMs under uncertain context. *arXiv preprint arXiv:2406.05972*.

Kahneman, D., and Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2), 263–291.

Koo, T. K. & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability. *Research Journal of Chiropractic Medicine*, 15, 155-163

Lin, L. I-K. (1989). A Concordance Correlation Coefficient to Evaluate Reproducibility. *Biometrics*, 45, 255–268.

Linnainmaa, J. T., Melzer, B. T., & Previtero, A. (2021). The misguided beliefs of financial advisors. *The Journal of Finance*, 76(2), 587-621.

Linnet, K. (1993). Evaluation of Regression Procedures for Method Comparison Studies. *Clinical Chemistry*, 39(3), 424-432.

Lucy, L., & Bamman, D. (2021). Gender and representation bias in GPT-3 generated stories. In N. Akoury, F. Brahman, S. Chaturvedi, E. Clark, M. Iyyer, & L. J. Martin (Eds.), Proceedings of the Third Workshop on Narrative Understanding (pp. 48–55). Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.nuse-1.5

Markowitz, H. (1952). Modern portfolio theory. *The Journal of Finance*, 7(11), 77-91.

McBride, G. B. (2005). A proposal for strength-of-agreement criteria for Lin's concordance correlation coefficient (NIWA Client Report: HAM2005-062). National Institute of Water & Atmospheric Research Ltd. Prepared for the Ministry of Health.

Obermeyer, Z., Powers, B., Vogeli, C., and Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464), 447–453.

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P., Leike, J., & Lowe, R. (2022). Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems, 35*, 27730–27744.

Parker, J. A., Schoar, A., Cole, A., & Simester, D. (2022). Household portfolios and retirement saving over the life cycle. *The Journal of Finance*.

Pikulina, E., Renneboog, L., & Tobler, P. N. (2017). Overconfidence and investment: An experimental approach. *Journal of Corporate Finance*, 43, 175–192.

Rossi, A. G., and Utkus, S. (2024). The diversification and welfare effects of robo-advising. *Journal of Financial Economics, 157*, 103869.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86(2), 420–428.

Stoughton, N. M. (1993) Moral hazard and the portfolio management problem, *The Journal of Finance*, 48, 2009-2028

Youyou, W., Kosinski, M., & Stillwell, D. (2015). Computer-based personality judgments are more accurate than those made by humans. *Proceedings of the National Academy of Sciences*, 112(4), 1036–1040.

Zhao, J., Wang, T., Yatskar, M., Ordonez, V., & Chang, K.-W. (2017). Men also like shopping: Reducing gender bias amplification using corpus-level constraints. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (pp. 2979–2989). Association for Computational Linguistics. https://doi.org/10.18653/v1/D17-1323.

Appendix

A1. Sample Vignette

1: Seven years ago, following a successful banking career Adam, 45, set up a hedge fund together with his business partner Carol. The business has had its ups and downs, but they are now making healthy profits. Adam managed to draw a salary of £125,000 last year, but spends at least £230,000 per year. He dates regularly but isn't interested in settling down or having children. Adam knows exactly what he wants to invest in and arranges all the research on coverage. You believe he is worth about £8 million and may consider investing up to 10% in his scheme. He intends to retire at 50 to pursue his interest in vintage cars.

Table A2: Portfolios 1–7, asset allocation

Asset class	Portfolio 1	Portfolio 2	Portfolio 3	Portfolio 4	Portfolio 5	Portfolio 6	Portfolio 7
UK Equities	11%	19%	27%	35%	37%	40%	42%
International Equities	8%	11%	14%	18%	28%	38%	44%
Bonds	51%	45%	39%	32%	20%	7%	3%
Cash	9%	5%	5%	5%	4%	2%	0%
Commercial Property	6%	5%	5%	5%	5%	5%	3%
Hedge Funds / Alternatives	18%	15%	10%	5%	6%	8%	8%
Total	100%	100%	100%	100%	100%	100%	100%

Table A3. Human Professional Advisor Variables

Variable Name	Description
	Demographics
A.Age	Age
A.Gender	0=male, 1=female
A.Married	0=no, 1=yes
A.Education	No degree=1 to PhD=5
A.Children	0=no, 1=yes
A.Experience	Years of experience
A.Risktol	Higher-than-average: 1=Strongly disagree ... 5=Strongly agree
A.Knowl	Higher-than-average: 1=Strongly disagree ... 10=Strongly agree
A.Conf	Higher-than-average: 1=Strongly disagree ... 5=Strongly agree
A.Maxloss	Largest personal portfolio loss: 1=Nothing … 5=More than 50%
	Advisor personal self-owned portfolio
A.Equity	% equities
A.FI	% fixed income
	Advisor personal investment behaviour
A.Equities	Proportion of personal portfolio in equities, %
A.FI	Proportion in fixed income, %
A.selfcash	Proportion in cash, %
A.selfinvprop	Proportion in additional property, %
A.selfalternat	Proportion in alternative investments, %
A.selfother	Proportion in other asset classes, %
A.maxloss	Largest personal portfolio loss: 1=Nothing … 5=More than 50%
A.RiskIndex	Higher-than-average: 1=Strongly disagree ... 5=Strongly agree
A.Knowledge	Higher-than-average: 1=Strongly disagree ... 5=Strongly agree
A.Conf	Higher-than-average: 1=Strongly disagree ... 5=Strongly agree
	Real-life client base and portfolios
A.Noclients	Number of clients
A.Investssetabove30M	Proportion of clients with >US$30M in investable assets
A.totalAUM	Total assets under management, US$M
A.Noinstitutions	Number of financial institutions worked at
C.Equities	Client equity holdings, %
C.FI	Client fixed income holdings, %
C.cash	Client cash holdings, %
C.alternat	Client alternative investments, %
C.other	Client other investments, %

Notes: The table shows the information collected about the professional human advisors and their real-life client base.

Table A2. Correlation matrix

	V.Gender	V.RecPort	V.Control	V.Risktol	V.Knowledge	V.NW	V.A.Age	V.Income	V.Out	V.Invamt	A.Age	RespGender	A.Yearsexp	A.Noclients	A.Investssetabove30M	A.totalAUM	A.Noinstitutions	A.Married	A.Education	A.Children	A.Equities	A.FI	A.cash	A.prop	A.alternat	A.other	A.maxloss	A.RiskIndex	A.knowl	A.Conf	C.Equities	C.FI	C.cash	C.ternat	C.other	A.RiskIndex	C.RiskIndex
V.Gender		0.05*	-0.05*	-0.042	-0.068**	0.01	0.002	-0.006	-0.005	0.011	0.001	-0.002	0.001	0	0	0	0	0.002	-0.001	0.002	0.001	0	-0.001	0	0	0	0.003	0	0.001	0.001	0.001	0.001	0.001	0.001	-0.002	0.001	0.002
V.RecPort	0.05*		0.039	0.401***	0.1***	-0.023	-0.407***	-0.042	0.196***	-0.157***	-0.035	-0.044	-0.035	0.048*	-0.054*	-0.023	0.02	-0.018	-0.044	-0.003	0.054*	-0.008	-0.035	-0.012	-0.055*	0.029	0.035	0.08**	0.037	0.006	0.058*	-0.111***	0.035	-0.054*	0.027	0.044	0.009
V.Control	-0.05*	0.039		0.373***	0.581***	-0.073**	0.085***	0.155***	0.167***	0.06*	0.046*	-0.004	0.061**	0.016	0.021	0.012	0.011	-0.008	-0.04	0.053*	0.066**	-0.019	-0.031	-0.038	0.016	0	0.012	0.052*	0.026	0.067**	0.044	-0.031	-0.035	0.052*	-0.004	0.059*	0.058*
V.Risktol	-0.042	0.401***	0.373***		0.397***	-0.038	-0.289***	0.096***	0.287***	-0.079**	0.013	0.004	0.015	0.014	-0.024	-0.008	0.015	0.019	-0.027	0.025	0.068**	-0.015	-0.013	-0.034	-0.074**	-0.003	0.041	0.028	0.043	0.002	0.031	-0.042	-0.018	-0.033	0.049*	0.043	0.011
V.Knowledge	-0.068**	0.1***	0.581***	0.397***		-0.138***	0.073**	0.288***	0.316***	0.004	0.032	0.004	0.041	-0.014	0.001	-0.06**	-0.015	0.039	-0.002	0.031	0.063**	-0.04	-0.014	-0.049*	0.011	0.022	0.036	0.037	0.041	0.07**	0.008	-0.033	-0.045	0.037	0.054*	0.047*	0.024
V.NW	0.01	-0.023	-0.073**	-0.038	-0.138***		0.006	0.006	-0.081***	0.743***	-0.002	-0.001	-0.001	0.001	-0.001	0.002	0.001	-0.001	-0.001	-0.001	0.002	-0.008	0.002	-0.001	-0.001	0.001	0	0	0.003	0.001	0	-0.003	0.001	0.001	0.001	0.001	0
V.A.Age	0.002	-0.407***	0.085***	-0.289***	0.073**	0.006		0.118***	-0.363***	0.482***	0	0.004	0	-0.001	0.001	0	-0.002	0.001	-0.001	0.001	0.001	-0.002	-0.001	0.002	0	0	-0.001	0.002	0	0	0.001	0.002	0	0	-0.003	0.001	0.001
V.Income	-0.006	-0.042	0.155***	0.096***	0.288***	0.006	0.118***		0.65***	-0.126***	-0.003	-0.003	-0.001	0.004	0	0.003	0.002	-0.001	-0.008	-0.002	0.005	-0.012	-0.001	0	0.001	0.003	0.001	-0.001	-0.002	0.003	0.003	-0.007	0.007	-0.003	-0.003	0.005	-0.001
V.Out	-0.005	0.196***	0.167***	0.287***	0.316***	-0.081***	-0.363***	0.65***		-0.217***	-0.002	-0.001	-0.001	0.002	0	0.002	0.001	-0.001	-0.004	-0.001	0.003	-0.009	0	0	0	0.002	0	-0.001	0	0.002	0.001	-0.005	0.004	-0.001	-0.001	0.003	-0.001
V.Invamt	0.011	-0.157***	0.06*	-0.079**	0.004	0.743***	0.482***	-0.126***	-0.217***		-0.003	0.003	-0.001	-0.001	-0.002	0.001	-0.002	0	0.002	0	0	-0.009	0.002	0.002	-0.002	-0.001	0	0.003	0.003	0	-0.002	0	0	0.002	0	0	-0.001
A.Age	0.001	-0.035	0.046*	0.013	0.032	-0.002	0	-0.003	-0.002	-0.003		-0.113***	0.793***	0.028	0.191***	0.26***	0.323***	0.432***	-0.004	0.589***	0.129***	0.251***	-0.254***	0.051*	-0.081***	-0.019	0.198***	-0.045	0.155***	0.048*	0.177***	0.196***	-0.175***	0.033	-0.128***	0.184***	0.204***
A.Gender	-0.002	-0.044	-0.004	0.004	0.004	-0.001	0.004	-0.003	-0.001	0.003	-0.113***		-0.038	-0.028	-0.014	-0.097***	-0.097***	-0.072**	0.045	-0.064**	-0.266***	-0.005	0.136***	0.143***	0.041	0.011	-0.294***	-0.184***	-0.109***	-0.082***	-0.206***	-0.02	0.085***	-0.025	-0.12***	-0.227***	-0.193***
A.Yearsexp	0.001	-0.035	0.061**	0.015	0.041	-0.001	0	-0.001	-0.001	-0.001	0.793***	-0.038		-0.013	0.229***	0.251***	0.411***	0.386***	-0.021	0.512***	0.113***	0.147***	-0.206***	0.047*	-0.023	-0.021	0.255***	0.017	0.199***	0.08**	0.188***	0.131***	-0.204***	0.081***	-0.077**	0.164***	0.23***
A.Noclients	0	0.048*	0.016	0.014	-0.014	0.001	-0.001	0.004	0.002	-0.001	0.028	-0.028	-0.013		-0.213***	0.077**	-0.007	0.058*	-0.113***	0.074**	0.132***	0.008	-0.001	-0.108***	-0.112***	-0.047*	-0.016	0.047*	-0.013	-0.072**	0.046*	-0.112***	0.047*	0.118***	-0.087***	0.073**	0.065**
A.Investssetabove30M	0	-0.054*	0.021	-0.024	0.001	-0.001	0.001	0	0	-0.002	0.191***	-0.014	0.229***	-0.213***		0.394***	0.081***	0.18***	0.093***	0.194***	0.073**	-0.004	-0.137***	0.054*	0.055*	0.003	0.132***	0.157***	0.108***	0.023	0.061**	0.058*	-0.046*	0.114***	-0.119***	0.123***	0.107***
A.totalAUM	0	-0.023	0.012	-0.008	-0.06**	0.002	0	0.003	0.002	0.001	0.26***	-0.097***	0.251***	0.077**	0.394***		0.061**	0.007	0.001	0.13***	0.128***	0.03	-0.103***	0.006	-0.084***	-0.092***	0.072**	0.127***	0.174***	0.105***	0.084***	0.067**	-0.078**	0.085***	-0.09***	0.132***	0.121***
A.Noinstitutions	0	0.02	0.011	0.015	-0.015	0.001	-0.002	0.002	0.001	-0.002	0.323***	-0.097***	0.411***	-0.007	0.081***	0.061**		0.075**	-0.052*	0.11***	-0.069**	0.042	0.04	-0.002	0.028	0.031	0.127***	0.08**	0.16***	0.105***	0.024	0.053*	-0.082***	0.033	0.017	-0.069**	0.05*
A.Married	0.002	-0.018	-0.008	0.019	0.039	-0.001	0.001	-0.001	-0.001	0	0.432***	-0.072**	0.386***	0.058*	0.18***	0.007	0.075**		-0.003	0.714***	0.01	0.108***	-0.068**	0.035	-0.074**	0.011	0.184***	0.002	0.166***	0.123***	0.116***	0.175***	-0.042	-0.102***	-0.138***	0.025	0.079**
A.Education	-0.001	-0.044	-0.04	-0.027	-0.002	-0.001	-0.001	-0.008	-0.004	0.002	-0.004	0.045	-0.021	-0.113***	0.093***	0.001	-0.052*	-0.003		-0.019	-0.069**	-0.026	-0.062**	0.085***	0.131***	0.126***	-0.028	-0.019	0.145***	0.025	-0.116***	0.119***	-0.045	0.082***	0.017	-0.006	-0.043
A.Children	0.002	-0.003	0.053*	0.025	0.031	-0.001	0.001	-0.002	-0.001	0	0.589***	-0.064**	0.512***	0.074**	0.194***	0.13***	0.11***	0.714***	-0.019		0.069**	0.12***	-0.134***	0.02	-0.009	-0.008	0.193***	0.017	0.096***	0.001	0.16***	0.156***	-0.113***	0.029	-0.155***	0.102***	0.175***
A.Equities	0.001	0.054*	0.066**	0.068**	0.063**	0.002	0.001	0.005	0.003	0	0.129***	-0.266***	0.113***	0.132***	0.073**	0.128***	-0.069**	0.01	-0.069**	0.069**		0.008	-0.564***	-0.449***	-0.146***	-0.129***	0.146***	0.392***	0.057*	0.148***	0.338***	0.023	-0.142***	-0.002	-0.158***	0.896***	0.299***
A.FI	0	-0.008	-0.019	-0.015	-0.04	-0.008	-0.002	-0.012	-0.009	-0.009	0.251***	-0.005	0.147***	0.008	-0.004	0.03	0.042	0.108***	-0.026	0.12***	0.008		-0.263***	-0.169***	0	-0.019	-0.002	-0.117***	-0.047*	-0.106***	0.006	0.267***	-0.102***	0.053*	-0.13***	0.039	0.069**
A.cash	-0.001	-0.035	-0.031	-0.013	-0.014	0.002	-0.001	-0.001	0	0.002	-0.254***	0.136***	-0.206***	-0.001	-0.137***	-0.103***	0.04	-0.068**	-0.062**	-0.134***	-0.564***	0.263***		0.298***	-0.059*	-0.117***	-0.213***	-0.245***	-0.001	-0.088***	-0.273***	-0.156***	0.201***	-0.092***	0.19***	-0.848***	-0.305***
A.prop	0	-0.012	-0.038	-0.034	-0.049*	-0.001	0.002	0	0	0.002	0.051*	0.143***	0.047*	-0.108***	0.054*	0.006	-0.002	0.035	0.085***	0.02	-0.449***	-0.169***	0.298***		-0.129***	-0.072**	0.07**	-0.11***	-0.021	0.014	-0.062**	0.052*	0.044	-0.055*	-0.006	-0.075**	-0.073**
A.alternat	0	-0.055*	0.016	-0.074**	0.011	-0.001	0	0.001	0	-0.002	-0.081***	0.041	-0.023	-0.112***	0.055*	-0.084***	0.028	-0.074**	0.131***	-0.009	-0.146***	0	-0.059*	-0.129***		0.211***	-0.065**	-0.071**	-0.158***	-0.078**	-0.058*	-0.025	-0.122***	0.4***	-0.024	0.001	0.126***
A,other	0	0.029	0	-0.003	0.022	0.001	0	0.003	0.002	-0.001	-0.019	0.011	-0.021	-0.047*	0.003	-0.092***	0.031	0.011	0.126***	-0.008	-0.129***	-0.019	0.117***	0.072**	0.211***		0.072**	0.016	0.087***	-0.044	-0.031	-0.061**	-0.072**	0.051*	0.14***	-0.102***	0

A.maxloss	0.003	0.035	0.012	0.041	0.036	0	-0.001	0.001	0	0	0.198***	-0.294***	0.255***	-0.016	0.132***	0.072**	0.127***	0.184***	-0.028	0.193***	0.146***	-0.002	-0.213***	0.07**	-0.065**	0.072**		0.254***	0.147***	0.156***	0.228***	-0.026	-0.079**	-0.003	-0.092***	0.191***	0.194***
A.RiskIndex	0	0.08**	0.052*	0.028	0.037	0	0.002	-0.001	-0.001	0.003	-0.045	-0.184***	0.017	0.047*	0.157***	0.127***	0.08**	0.002	-0.019	0.017	0.392***	-0.117***	-0.245***	-0.11***	-0.071**	0.016	0.254***		0.244***	0.393***	0.162***	-0.137***	0.058*	-0.024	-0.1***	0.37***	0.099***
A.knowl	0.001	0.037	0.026	0.043	0.041	0.003	0	-0.002	0	0.003	0.155***	-0.109***	0.199***	-0.013	0.108***	0.174***	0.16***	0.166***	0.145***	0.096***	0.057*	-0.047*	-0.001	-0.021	-0.158***	0.087***	0.147***	0.244***		0.483***	0.01	0.02	0.035	-0.041	-0.045	0.018	-0.011
A.Conf	0.001	0.006	0.067**	0.002	0.07**	0.001	0	0.003	0.002	0	0.048*	-0.082***	0.08**	-0.072**	0.023	0.105***	0.105***	0.123***	0.025	0.001	0.148***	-0.106***	-0.088***	0.014	-0.078**	-0.044	0.156***	0.393***	0.483***		0.015	-0.041	0.025	-0.073**	0.035	0.146***	-0.025
C.Equities	0.001	0.058*	0.044	0.031	0.008	0	0.001	0.003	0.001	-0.002	0.177***	-0.206***	0.188***	0.046*	0.061**	0.084***	0.024	0.116***	-0.116***	0.16***	0.338***	0.006	-0.273***	-0.062**	-0.058*	-0.031	0.228***	0.162***	0.01	0.015		0.082***	-0.523***	-0.059*	-0.309***	0.347***	0.88***
C.FI	0.001	-0.111***	-0.031	-0.042	-0.033	-0.003	0.002	-0.007	-0.005	0	0.196***	-0.02	0.131***	-0.112***	0.058*	0.067**	0.053*	0.175***	0.119***	0.156***	0.023	0.267***	-0.156***	0.052*	-0.025	-0.061**	-0.026	-0.137***	0.02	-0.041	0.082***		-0.41***	-0.079**	-0.332***	0.075**	0.202***
C.cash	-0.001	0.035	-0.035	-0.018	-0.045	0.001	0	0.007	0.004	0	-0.175***	0.085***	-0.204***	0.047*	-0.046*	-0.078**	-0.082***	-0.042	-0.045	-0.113***	-0.142***	-0.102***	0.201***	0.044	-0.122***	-0.072**	-0.079**	0.058*	0.035	0.025	-0.523***	-0.41***		-0.289***	-0.263***	-0.184***	-0.737***
C.alternat	0.001	-0.054*	0.052*	-0.033	0.037	0.001	0	-0.003	-0.001	0.002	0.033	-0.025	0.081***	0.118***	0.114***	0.085***	0.033	-0.102***	0.082***	0.029	-0.002	0.053*	-0.092***	-0.055*	0.4***	0.051*	-0.003	-0.024	-0.041	-0.073**	-0.059*	-0.079**	-0.289***		-0.145***	0.069**	0.38***
C.other	-0.002	0.027	-0.004	0.049*	0.054*	0.001	-0.003	-0.003	-0.001	0	-0.128***	0.12***	-0.077**	-0.087***	-0.119***	-0.09***	0.017	-0.138***	0.017	-0.155***	-0.158***	-0.13***	0.19***	-0.006	-0.024	0.14***	-0.092***	-0.1***	-0.045	0.035	-0.309***	-0.332***	-0.263***	-0.145***		-0.2***	-0.3***
A.RiskIndex	0.001	0.044	0.059*	0.043	0.047*	0.001	0.001	0.005	0.003	0	0.184***	-0.227***	0.164***	0.073**	0.123***	0.132***	-0.069**	0.025	-0.006	0.102***	0.896***	0.039	-0.848***	-0.075**	0.001	-0.102***	0.191***	0.37***	0.018	0.146***	0.347***	0.075**	-0.184***	0.069**	-0.2***		0.345***
C.RiskIndex	0.002	0.009	0.058*	0.011	0.024	0	0.001	-0.001	-0.001	-0.001	0.204***	-0.193***	0.23***	0.065**	0.107***	0.121***	0.05*	0.079**	-0.043	0.175***	0.299***	0.069**	-0.305***	-0.073**	0.126***	0	0.194***	0.099***	-0.011	-0.025	0.88***	0.202***	-0.737***	0.38***	-0.3***	0.345***	

Notes: Table A2 reports pairwise correlations across advisor demographics, experience, client background, self-reported investment attributes, and portfolio recommendations. Stars indicate statistical significance (*p < .05, **p < .01, ***p < .001). Several patterns stand out: (i) strong positive associations among psychological constructs (control, risk tolerance, knowledge) and with recommended portfolio risk; (ii) negative correlations of A.Age and experience with risk tolerance and equity allocations, consistent with safer preferences later in career; (iii) client portfolios correlate with advisors' own holdings, indicating projection effects.

Table A3. Model A: Mixed-Effects Models predicting vignette responses with random intercepts by vignette

	RecPort (riskier portfolio = higher)	Control 1 = "a lot *more* control than the averageinvestor"; 5 = "a lot *less* control"	RiskTol 1 = "a lot *more* willing to take risk"; 5 = "a lot *less* willing"	Knowledge 1 = "not knowledgeable"; 10 = "extremely knowledgeable"
(Intercept)	5.365***	2.951***	5.019***	8.725***
	(0.643)	(0.394)	(0.603)	(1.664)
Students	-1.952*	0.673	-1.765	-2.659
	(0.995)	(0.600)	(0.974)	(2.796)
ChatGPT_Biased	-0.814***	-0.470**	-4.861***	-3.515***
	(0.222)	(0.160)	(0.144)	(0.264)
ChatGPT_Unbiased	-0.194	-0.827***	-3.039***	-0.063
	(0.222)	(0.160)	(0.144)	(0.264)
Gemini_Biased	-1.019***	-1.883***	-4.519***	-3.460***
	(0.222)	(0.160)	(0.147)	(0.264)
Gemini_Unbiased	-1.836***	-3.018***	-5.264***	0.960***
	(0.222)	(0.160)	(0.144)	(0.264)
v_gender	0.155***	-0.101**	-0.077**	-0.274***
	(0.046)	(0.033)	(0.030)	(0.054)
v_nw	-0.025	-0.035**	-0.038	-0.101
	(0.020)	(0.012)	(0.020)	(0.058)
v_age	-0.049**	-0.009	-0.053**	-0.088
	(0.017)	(0.010)	(0.017)	(0.048)
v_income	-0.000	0.001	0.001	0.005*
	(0.001)	(0.000)	(0.001)	(0.002)
v_out	0.001	0.001	-0.001	-0.003
	(0.001)	(0.001)	(0.001)	(0.003)
v_invamt	0.312	0.353**	0.392	0.772
	(0.199)	(0.118)	(0.203)	(0.598)
Students:v_gender	-0.166*	-0.095	-0.017	0.035
	(0.075)	(0.054)	(0.049)	(0.089)
ChatGPT_Biased:v_gender	-0.200**	0.136**	0.096*	0.321***
	(0.064)	(0.046)	(0.042)	(0.076)
ChatGPT_Unbiased:v_gender	-0.163*	0.062	0.074	0.267***
	(0.064)	(0.046)	(0.042)	(0.076)
Gemini_Biased:v_gender	-0.158*	0.083	0.124**	0.298***
	(0.064)	(0.046)	(0.043)	(0.076)
Gemini_Unbiased:v_gender	-0.177**	0.121**	0.068	0.264***
	(0.064)	(0.046)	(0.042)	(0.076)
Students:v_nw	0.011	0.011	0.027	0.056
	(0.028)	(0.017)	(0.028)	(0.083)
ChatGPT_Biased:v_nw	-0.021***	0.036***	0.100***	-0.025***
	(0.006)	(0.004)	(0.004)	(0.007)
ChatGPT_Unbiased:v_nw	-0.072***	0.031***	0.082***	-0.060***
	(0.006)	(0.004)	(0.004)	(0.007)
Gemini_Biased:v_nw	0.005	0.066***	0.077***	-0.113***
	(0.006)	(0.004)	(0.004)	(0.007)
Gemini_Unbiased:v_nw	-0.007	0.064***	0.073***	-0.129***
	(0.006)	(0.004)	(0.004)	(0.007)
Students:v_age	0.039	0.002	0.039	0.090
	(0.026)	(0.015)	(0.025)	(0.074)
ChatGPT_Biased:v_age	-0.004	-0.008*	0.124***	0.098***
	(0.005)	(0.004)	(0.004)	(0.007)
ChatGPT_Unbiased:v_age	-0.002	-0.009*	0.076***	0.033***
	(0.005)	(0.004)	(0.004)	(0.007)

Gemini_Biased:v_age	0.013*	0.059***	0.115***	0.030***
	(0.005)	(0.004)	(0.004)	(0.007)
Gemini_Unbiased:v_age	0.028***	0.060***	0.132***	-0.058***
	(0.005)	(0.004)	(0.004)	(0.007)
Students:v_income	-0.000	-0.000	-0.001	-0.004
	(0.001)	(0.001)	(0.001)	(0.003)
ChatGPT_Biased:v_income	-0.001*	0.000*	-0.001***	-0.005***
	(0.000)	(0.000)	(0.000)	(0.000)
ChatGPT_Unbiased:v_income	-0.001**	-0.000	0.000*	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
Gemini_Biased:v_income	0.000	-0.001***	-0.002***	-0.004***
	(0.000)	(0.000)	(0.000)	(0.000)
Gemini_Unbiased:v_income	-0.000	-0.001***	-0.004***	-0.002***
	(0.000)	(0.000)	(0.000)	(0.000)
Students:v_out	0.001	-0.001	0.001	0.002
	(0.002)	(0.001)	(0.002)	(0.005)
ChatGPT_Biased:v_out	0.002***	0.001***	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
ChatGPT_Unbiased:v_out	0.001***	0.002***	-0.002***	-0.002***
	(0.000)	(0.000)	(0.000)	(0.000)
Gemini_Biased:v_out	0.001**	0.001***	0.000*	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)
Gemini_Unbiased:v_out	0.001***	0.003***	0.003***	-0.001*
	(0.000)	(0.000)	(0.000)	(0.000)
Students:v_invamt	-0.092	-0.209	-0.245	-0.617
	(0.284)	(0.169)	(0.288)	(0.847)
ChatGPT_Biased:v_invamt	0.068	-0.411***	-0.999***	0.057
	(0.059)	(0.042)	(0.038)	(0.070)
ChatGPT_Unbiased:v_invamt	0.417***	-0.299***	-0.791***	0.105
	(0.059)	(0.042)	(0.038)	(0.070)
Gemini_Biased:v_invamt	-0.228***	-0.917***	-0.860***	1.218***
	(0.059)	(0.042)	(0.040)	(0.070)
Gemini_Unbiased:v_invamt	-0.316***	-0.831***	-0.908***	1.309***
	(0.059)	(0.042)	(0.038)	(0.070)
SD (Intercept vignette_number)	0.337	0.197	0.348	1.031
SD (Observations)	0.987	0.708	0.640	1.170
R2 Marg.	0.410	0.505	0.395	0.438
R2 Cond.	0.471	0.541	0.533	0.684
ICC	0.1	0.1	0.2	0.4
RMSE	0.99	0.71	0.64	1.17

Note: This table reports mixed-effects models in which a random intercept is included for each vignette. This specification isolates level-bias, or systematic displacement between AI and human judgments on identical client scenarios. By controlling for vignette-level heterogeneity, the analysis ensures that fixed effects (such as source and source × vignette feature interactions) reflect differences in how each judge, human or AI, interprets the same client case. Lower Control / RiskTol scores = advisors/AI think the client holds more control or is more risk tolerant.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ All models include a random intercept by vignette. SE in parentheses

Table A4. Model B: Mixed-Effects Models Predicting Vignette Responses (a random intercept by respondent)

	RecPort	Control	RiskTol	Knowledge
(Intercept)	6.318***	3.192***	4.608***	5.492***
	(0.167)	(0.118)	(0.125)	(0.264)
Students	-2.923***	0.443*	-1.346***	0.598
	(0.273)	(0.193)	(0.204)	(0.433)
ChatGPT_Biased	-0.801***	-0.461**	-4.870***	-3.569***
	(0.233)	(0.166)	(0.175)	(0.372)
ChatGPT_Unbiased	-0.180	-0.818***	-3.038***	-0.117
	(0.233)	(0.166)	(0.176)	(0.372)
Gemini_Biased	-1.005***	-1.871***	-4.651***	-3.514***
	(0.233)	(0.166)	(0.179)	(0.372)
Gemini_Unbiased	-1.822***	-3.009***	-5.272***	0.906*
	(0.233)	(0.166)	(0.175)	(0.372)
v_gender	0.157**	-0.102**	-0.081*	-0.278***
	(0.048)	(0.034)	(0.036)	(0.076)
v_nw	-0.030***	-0.036***	-0.036***	-0.085***
	(0.004)	(0.003)	(0.003)	(0.007)
v_age	-0.067***	-0.014***	-0.045***	-0.028***
	(0.004)	(0.003)	(0.003)	(0.007)
v_income	0.000*	0.001***	0.001***	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)
v_out	-0.000	0.000	-0.000	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)
v_invamt	0.319***	0.351***	0.388***	0.763***
	(0.044)	(0.031)	(0.033)	(0.070)
Students:v_gender	-0.165*	-0.093	-0.017	0.050
	(0.079)	(0.056)	(0.059)	(0.125)
ChatGPT_Biased:v_gender	-0.201**	0.136**	0.097	0.325**
	(0.067)	(0.048)	(0.051)	(0.108)
ChatGPT_Unbiased:v_gender	-0.164*	0.062	0.075	0.271*
	(0.067)	(0.048)	(0.051)	(0.108)
Gemini_Biased:v_gender	-0.159*	0.083	0.131*	0.302**
	(0.067)	(0.048)	(0.052)	(0.108)
Gemini_Unbiased:v_gender	-0.178**	0.121*	0.069	0.269*
	(0.067)	(0.048)	(0.051)	(0.108)
Students:v_nw	0.016*	0.012*	0.025***	0.039***
	(0.007)	(0.005)	(0.005)	(0.011)
ChatGPT_Biased:v_nw	-0.022***	0.036***	0.100***	-0.023*
	(0.006)	(0.004)	(0.005)	(0.010)
ChatGPT_Unbiased:v_nw	-0.072***	0.031***	0.081***	-0.058***
	(0.006)	(0.004)	(0.005)	(0.010)
Gemini_Biased:v_nw	0.004	0.066***	0.080***	-0.111***
	(0.006)	(0.004)	(0.005)	(0.010)
Gemini_Unbiased:v_nw	-0.007	0.063***	0.073***	-0.127***
	(0.006)	(0.004)	(0.005)	(0.010)
Students:v_age	0.057***	0.006	0.031***	0.029**
	(0.007)	(0.005)	(0.005)	(0.011)
ChatGPT_Biased:v_age	-0.004	-0.009*	0.124***	0.099***
	(0.006)	(0.004)	(0.004)	(0.009)
ChatGPT_Unbiased:v_age	-0.003	-0.009*	0.075***	0.035***
	(0.006)	(0.004)	(0.004)	(0.009)
Gemini_Biased:v_age	0.013*	0.059***	0.121***	0.032***
	(0.006)	(0.004)	(0.004)	(0.009)
Gemini_Unbiased:v_age	0.027***	0.060***	0.132***	-0.057***
	(0.006)	(0.004)	(0.004)	(0.009)
Students:v_income	-0.001***	-0.000*	-0.001*	-0.002***
	(0.000)	(0.000)	(0.000)	(0.000)
ChatGPT_Biased:v_income	-0.001*	0.000*	-0.001***	-0.005***

	(0.000)	(0.000)	(0.000)	(0.000)
ChatGPT_Unbiased:v_income	-0.001**	-0.000	0.000	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
Gemini_Biased:v_income	0.000	-0.001***	-0.002***	-0.004***
	(0.000)	(0.000)	(0.000)	(0.000)
Gemini_Unbiased:v_income	-0.000	-0.001***	-0.004***	-0.002***
	(0.000)	(0.000)	(0.000)	(0.000)
Students:v_out	0.002***	-0.001*	0.001	-0.002**
	(0.000)	(0.000)	(0.000)	(0.001)
ChatGPT_Biased:v_out	0.002***	0.001***	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.001)
ChatGPT_Unbiased:v_out	0.001***	0.002***	-0.002***	-0.002***
	(0.000)	(0.000)	(0.000)	(0.001)
Gemini_Biased:v_out	0.001*	0.001***	0.001*	0.002***
	(0.000)	(0.000)	(0.000)	(0.001)
Gemini_Unbiased:v_out	0.001***	0.003***	0.003***	-0.001
	(0.000)	(0.000)	(0.000)	(0.001)
Students:v_invamt	-0.102	-0.203***	-0.237***	-0.601***
	(0.072)	(0.051)	(0.054)	(0.115)
ChatGPT_Biased:v_invamt	0.069	-0.408***	-0.998***	0.040
	(0.062)	(0.044)	(0.046)	(0.098)
ChatGPT_Unbiased:v_invamt	0.418***	-0.295***	-0.789***	0.088
	(0.062)	(0.044)	(0.046)	(0.098)
Gemini_Biased:v_invamt	-0.227***	-0.912***	-0.915***	1.201***
	(0.062)	(0.044)	(0.048)	(0.098)
Gemini_Unbiased:v_invamt	-0.314***	-0.828***	-0.907***	1.292***
	(0.062)	(0.044)	(0.046)	(0.098)
SD (Intercept id)	0.225	0.034	0.000	0.050
SD (Observations)	1.035	0.736	0.778	1.651
R2 Marg.	0.430	0.507	0.378	0.482
R2 Cond.	0.455	0.508		0.482
ICC	0.0	0.0		0.0
RMSE	1.03	0.74	0.78	1.65

Notes: This table presents models with a random intercept for each respondent ID, aligning the analysis to individual-level baselines. This specification captures slope-bias, or the degree to which AI judgments track or deviate from the advisor's own preferences across different client profiles. Unlike Model A, where the intercept represents average judgment on a vignette, here the intercept represents the respondent's typical rating, and fixed effects reflect deviations relative to that baseline. The shift in structure has two key effects: (1) ICCs drop to near-zero, suggesting most variation is now within-person, not across respondents; and (2) predictor estimates (e.g., vignette features like gender, wealth, age) become larger and more significant, indicating that once individual-level preferences are removed, scenario features regain explanatory power.

All models include a random intercept by respondent SE in parentheses; * p<.05, ** p<.01, *** p<.001

Table A5. Calculated return and volatility assumptions

Asset Class (proxy index)	1-Yr Return	3-Yr Return	5-Yr Return	10-Yr Return	1-Yr Vol	3-Yr Vol	5-Yr Vol	10-Yr Vol
UK Equities (MSCI United Kingdom IMI, GBP)	**11.11%**	**10.44%**	**10.93%**	**6.62%**	**n/a**	**10.80%**	**11.62%**	**12.18%**
International Equities (MSCI ACWI ex UK, GBP)	**7.54%**	**13.35%**	**11.84%**	**12.37%**	**n/a**	**11.28%**	**11.31%**	**11.86%**
Bonds [12]	**-0.80%**	**-3.79%**	**-4.45%**	**-0.24%**	**8.76%**	**15.02%**	**10.96%**	**10.41%**
Cash (custom BoE-rate index)	***1.30%***	**1.19%**	**0.72%**	**0.36%**	**0.11%**	**0.41%**	**0.50%**	**0.52%**
Commercial Property (MSCI UK IMI Liquid Real Estate, GBP net)	**-4.21%**	**-7.87%**	**0.55%**	**-0.08%**	**n/a**	**15.61%**	**13.72%**	**12.35%**
Hedge-Funds / Alternatives (custom long/short factor index: [Custom Index of 100% Long MSCI World Diversified Multiple Factor Index + 70% Short MSCI World Index]	***7.09%***	**3.91%**	**3.46%**	**1.84%**	**5.71%**	**5.71%**	**5.70%**	**5.66%**

Note: The table presents historical returns and volatility estimates for six key asset classes over multiple investment horizons (1-, 3-, 5-, and 10-year windows). Return figures represent annualized performance, while volatility is measured as the standard deviation of returns. These figures are used to calculate return, volatility and Sharpe ratios for both human advisors' and AIs' portfolio recommendations.

Table A6. Return and volatility per hypothetical (vignette) portfolio

Return per portfolio	1-Yr Return	3-Yr Return	5-Yr Return	10-Yr Return	1-Yr Vol	3-Yr Vol	5-Yr Vol	10-Yr Vol
Portfolio (Tier) 1	2.52%	0.59%	0.58%	1.94%	5.50%	11.74%	9.65%	9.39%
Portfolio 2	3.50%	2.00%	1.96%	2.80%	4.80%	11.71%	9.95%	9.80%
Portfolio 3	4.31%	3.27%	3.28%	3.62%	3.99%	11.73%	10.28%	10.22%
Portfolio 4	5.20%	4.70%	4.77%	4.57%	3.09%	11.70%	10.61%	10.66%
Portfolio 5	6.33%	6.73%	6.73%	5.99%	2.10%	11.30%	10.71%	10.89%
Portfolio 6	7.64%	8.93%	8.88%	7.48%	1.07%	10.90%	10.87%	11.19%
Portfolio (Tier) 7	8.40%	10.22%	9.96%	8.36%	0.72%	10.88%	11.05%	11.47%

Note: The table shows the estimated performance of seven model portfolios, corresponding to the "risk tiers" from the original survey. Portfolio Tier 1 is the most conservative, with a high bond allocation, while Tier 7 is the most aggressive, heavily weighted toward equities. Portfolio returns were calculated using historical capital-market assumptions from Table A5, applying the standard weighted-return and weighted-volatility formulas: $\mu_p = \sum_a w_{a,p}\mu_a$, $\sigma_p = \sqrt{\sum_a w_{a,p}^2 \sigma_a^2}$

[12] FTSE Russell as at 30 June 2025.

Table A7. Sharpe ratio per hypothetical (vignette) portfolio

	Sharpe 1y	Sharpe 3y	Sharpe 5y	Sharpe10y
Portfolio (Tier) 1	0.4578	0.0498	0.0599	0.2069
Portfolio 2	0.7282	0.1706	0.1968	0.2858
Portfolio 3	1.0787	0.2785	0.3194	0.3546
Portfolio 4	1.6804	0.4020	0.4496	0.4291
Portfolio 5	3.0154	0.5957	0.6288	0.5498
Portfolio 6	7.1202	0.8187	0.8171	0.6687
Portfolio (Tier) 7	11.6699	0.9399	0.9011	0.7289

Notes: Sharpe ratios (return-to-risk) in the table illustrate the efficiency of risk-taking across the same portfolios. The short-horizon Sharpe (1-year) increases dramatically as portfolios become more equity-heavy, i.e., from 0.46 in Tier 1 to an extreme 11.67 in Tier 7. This is driven primarily by decreasing 1-year volatility rather than sharply rising returns. In contrast, long-run Sharpe ratios (10-year) rise more modestly, from 0.21 in Tier 1 to 0.73 in Tier 7, suggesting that over strategic horizons, extra return comes with proportionally more risk.

Table A8. Average per step change from portfolios 1 to 7

	Average return change per step 1 to 7	Average volatility change per step 1 to 7	Average Sharpe ratio change per step 1 to 7
1-Yr Return	0.009804	-0.00797	-1.23019
3-Yr Return	0.016061	-0.00144	-11.117
5-Yr Return	0.015636	0.002333	6.702375
10-Yr Return	0.010696	0.003467	3.084669

Notes: This table shows the average one-tier step changes when one moves from Portfolio 1 to Portfolio 2….. to Portfolio 7: $\frac{1}{6}\sum_{p=1}^{6}(\mu_{p+1} - \mu_p)$. It demonstrates that return rises almost linearly: on average +0.98 pp (1-yr) and +1.07 pp (10-yr) per step. By Tier 7 one can earn ~8 % p.a. versus ~2 % in Tier 1. 1-year volatility falls (heavy bond sleeve is replaced by equities but cash share also shrinks), so short-horizon Sharpe explodes from 0.46 to 11.7. At strategic (10-yr) horizon σ climbs slowly but Sharpe still improves but only from 0.21 to 0.73. Step-by-step Sharpe gains slow after Portfolio 5 because extra equity brings proportionally more risk. Short-run Sharpe is dominated by declining σ; long-run Sharpe by rising μ.

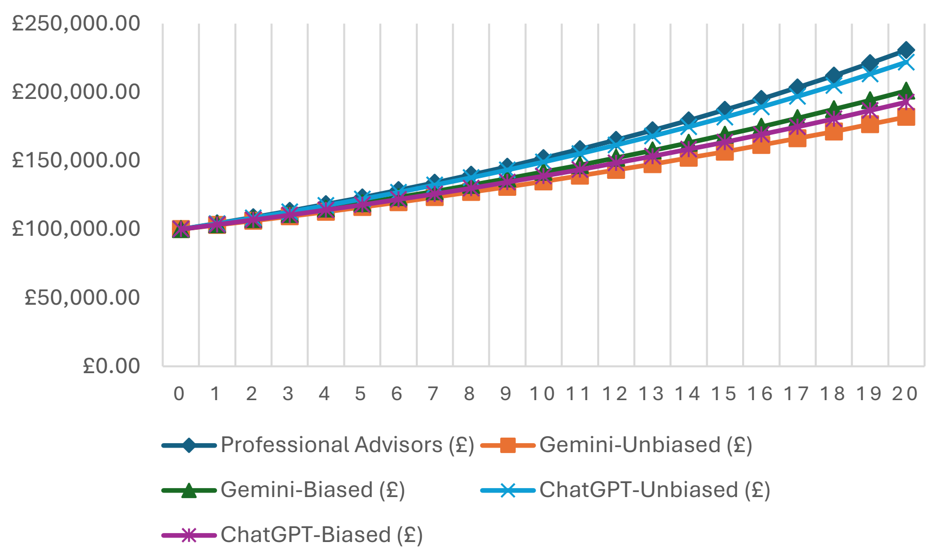

Fig A1. Wealth trajectories, growth of £100,000 over 20 years.