

# LLMs Can't Be Trusted for Financial Advice

By Gary Smith

## Executive Summary

- ChatGPT and other large language models (LLMs) do not understand the text they input and output.
- Three prominent LLMs were tested with 11 financial-decision prompts.
- The LLM responses were seemingly authoritative but riddled with arithmetic and critical-thinking mistakes.
- Financial planners are unlikely to be replaced by LLMs anytime soon.

Gary Smith is an economics professor at Pomona College. He has written more than 100 peer-reviewed papers and 17 books; most recently, co-authored with Margaret Smith, *The Power of Modern Value Investing*, Palgrave Macmillan, 2024.

**WHEN CHATGPT** was released to the public on November 30, 2022, the reception was overwhelmingly positive. Marc Andreessen (2022) described it as, “Pure, absolute, indescribable magic.” He was hardly alone. ChatGPT had more than 100 million users by January 2023, and dozens of companies began scrambling to create their own large language models (LLMs).

In February, Bill Gates (2023) said that the creation of ChatGPT was “every bit as important as the PC, as the internet.” Nvidia’s CEO (Huang 2023) declared that ChatGPT “is one of the greatest things that has ever been done for computing.” In March, a Wharton professor (Mollick 2023) argued that the productivity gains from LLMs might be larger than the gains from steam power.

Soon, pundits were predicting that LLMs would

make many, if not most, jobs obsolete. On March 22, 2023, nearly 2,000 people signed an open letter written by the Future of Life Institute calling for a pause of at least six months in the development of LLMs:

*Contemporary AI systems are now becoming human-competitive at general tasks, and we must ask ourselves: Should we let machines flood our information channels with propaganda and untruth? Should we automate away all the jobs, including the fulfilling ones? Should we develop nonhuman minds that might eventually outnumber, outsmart, obsolete and replace us? Should we risk loss of control of our civilization?*

Adding numbers to the job-killing fears, a study by researchers at OpenAI, OpenResearch, and the University of Pennsylvania concluded that:

*around 80 percent of the U.S. workforce could have at least 10 percent of their work tasks affected by the introduction of LLMs, while approximately 19 percent of workers may see at least 50 percent of their tasks impacted.*

They used the O\*NET database of information on 1,016 occupations, involving 19,265 tasks and 2,087 detailed work activities (DWAs) that are associated with tasks. Humans and LLMs assessed whether an LLM or LLM-powered software could “reduce the time required for a human to perform a specific DWA or complete a task by at least 50 percent.” They then linked DWAs and tasks to occupations in order to determine the potential impact of LLMs on jobs.

They cautioned that:

*A fundamental limitation of our approach lies in the subjectivity of the labeling. In our study, we employ annotators who are familiar with LLM capabilities. However, this group is not occupationally diverse, potentially leading to biased judgments regarding LLMs’ reliability and effectiveness in performing tasks within unfamiliar occupations.*

The report identified doctors, lawyers, mathematicians, financial quantitative analysts, central bank monetary authorities, managers of companies and enterprises, and tax preparers as among the occupations most at risk—an ironic confirmation of their “potentially . . . biased judgments regarding . . . unfamiliar occupations.” LLMs are text generators, nothing more. They are astonishingly good at this, but they are not designed or intended to understand the text they input and output. They consequently lack the critical-thinking skills required for many jobs, including the occupations listed above as being most at risk.

This dearth of critical-thinking skills is why, when ChatGPT was asked tax questions that had been posted on TaxBuzz’s practitioner support forum, it got every question wrong (Reams 2023). Here, I report the results of a similar test of the ability of LLMs to answer questions involving financial decisions. Are financial planners at imminent risk of being replaced by LLMs?

## Methods

Eleven prompts were created from commonplace financial-decision questions that are analyzed in Smith and Smith (2024). Between January 7 and January 10, 2024, each of these 11 prompts was given to three popular LLMs: OpenAI’s ChatGPT 3.5, Microsoft’s Bing with ChatGPT’s GPT-4, and Google’s Bard. Each prompt was given only once, and the complete response was recorded. Bard gave three “drafts” of answers but only the first draft was considered. Smith (2024) shows the complete transcripts of the responses. An analysis follows.

## Results

The LLM responses were consistently grammatically correct and seemingly authoritative but riddled with arithmetic and critical-thinking mistakes.

### Prompt 1: Car Loan Choice

This prompt involves a straightforward comparison of two car loans:

*I need to borrow \$47,000 to buy a new car. Is it better to borrow for one year at a 9 percent APR or for 10 years at a 1 percent APR?*

A human being living in the real world would immediately recognize that a 1 percent APR is an extremely attractive loan rate, particularly if it can be locked in for several years. No calculations would be needed to recognize the appeal of a 10-year loan with a 1 percent APR.

For a more formal analysis, the one-year loan has monthly payments of \$4,110.22 and total payments of \$49,322.64, including \$2,322.65 in interest. The 10-year loan has monthly payments of \$411.74 and total payments of \$49,408.80, including \$2,408.80 in interest. The 10-year loan has the lower present value for any required return larger than 0.03 percent.

I chose these two loans to see if the LLMs would compare the total payments and ignore the time value of money. All three LLMs made this mistake and recommended the one-year loan with a 9 percent APR.

ChatGPT 3.5 did a simple-interest calculation and concluded that, “If you can afford the higher monthly payments, the one-year loan at 9 percent APR may be financially advantageous due to the lower total interest paid.”

**“More fundamental than arithmetic errors, the LLM responses demonstrated that they do not have the common sense needed to recognize when their answers are obviously wrong.”**

Bing said that it used a “loan interest calculator” and found the total interest to be \$4,230.22 for the one-year loan and \$5,640.00 for the 10-year loan. The first number is a simple-interest calculation; the second number is mysterious as it is neither a simple interest nor amortized interest. Bing’s conclusion was, “As you can see, the total interest paid for the one-year loan is lower than the total interest paid for the 10-year loan. Therefore, it would be better to borrow for one year at a 9 percent APR if you can afford the higher monthly payments.”

Bard said that it used an “online loan calculator” to determine that the one-year loan would have monthly payments of \$4,083 and total interest of \$3,423 while the 10-year loan would have monthly payments of \$444 and total interest of \$14,585. These numbers are not only incorrect but inconsistent. If the monthly payments on the one-year loan

are \$4,083, then the total interest is \$1,996; if the monthly payments on the 10-year loan are \$444, then the total interest is \$6,280. Bard made the total-payments error when it concluded that with the one-year loan, “You’ll pay off the loan much faster, significantly reducing the total interest you accrue.”

In addition, some of Bard’s pros and cons were misguided. Two alleged pros of the one-year loan were: “You’ll be done with the loan payments in a year, freeing up your monthly budget for other expenses or savings” and “Less risk of financial hardship: If your financial situation changes unexpectedly, you’ll have a shorter loan term to manage.” The much higher monthly payments from the one-year loan during the first year are more likely to create budget problems, especially if there is an unexpected change in the borrower’s financial situation. One of Bard’s listed cons of the 10-year loan was “Risk of inflation: Over 10 years, inflation could erode the purchasing power of your future payments, making the loan effectively more expensive.” In fact, inflation is a borrower’s friend.

### *Prompt 2: Down Payment Size*

This prompt involves a down payment decision:

*I am buying a house for \$550,000 and I found a bank that will loan me up to \$400,000 in a conventional 30-year amortized mortgage with a 2.25 percent APR. Is it more financially advantageous for me to make a down payment of \$150,000 or \$250,000? I have plenty of income to make the mortgage payments and pay for my living expenses.*

A down payment that is \$100,000 smaller means an extra \$100,000 borrowed. A critical question is whether it is financially advantageous to borrow at a 2.25 percent APR. If the home buyer can earn

more than a 2.25 percent rate of return on the extra \$100,000, then it is profitable to make a smaller down payment. If an analysis focuses instead on the total interest, then it will generate the flawed advice that borrowing any amount at any interest rate is bad because it increases the total interest paid. ChatGPT 3.5 and Bing both recommended the larger down payment, based on the total-payments error. Bard did not make a recommendation.

ChatGPT 3.5 advised, “In terms of immediate financial advantage, a larger down payment of \$250,000 would likely save you more money in the long run due to reduced interest payments over the life of the loan.” Even though the prompt says that “I have plenty of income to make the mortgage payments and pay for my living expenses,” ChatGPT 3.5 added this caution: “If a \$250,000 down payment significantly depletes your savings or emergency funds, it might not be the best option. Always ensure you have a comfortable financial cushion for unexpected expenses.”

Bing miscalculated the monthly payments, reporting \$1,325.18 and \$1,108.14, instead of the correct values, \$1,528.98 and \$1,146.74, and inexplicably compared the total interest over the first seven years instead of all 30 years. Bing concluded that “making a down payment of \$250,000 is more financially advantageous for you in the long run. By making a larger down payment, you will have a lower monthly payment and pay less interest over the life of the loan.”

Bard did not make a recommendation but instead showed a long list of pros and cons, including these competing arguments regarding the smaller down payment: “Over the 30-year loan term, you’ll pay more in total interest compared to a larger down payment” and “You could invest the additional \$100,000 potentially for higher returns than the mortgage interest rate.”

### *Prompt 3: Comparing Home Offers*

This prompt involves a comparison of two home-purchase offers:

*I have to sell my house and my two best offers are:*

*(A) \$625,000 cash*

*(B) \$100,000 cash and an owner-financed*

*\$600,000 30-year conventional amortized loan at a 3 percent APR.*

*Which offer is more financially attractive?*

The loan’s monthly payments are \$2,529.62. The present value of Option B using the seller’s required return is a crucial input for making a rational decision. Even better, we can determine that the break-even required return where the present values are equal is 4.08 percent. The present value is higher for Option A if the seller’s required return is higher than 4.08 percent; the present value is higher for Option B if the required return is lower than 4.08 percent. We might also calculate the present value of Option B if the buyer pays off the loan before 30 years elapses or defaults on the payments, in which case the home presumably goes back to the seller.

None of the LLMs calculated the present value of Option B. They all mentioned the risk of the home buyer defaulting but did not consider that the home is presumably collateral for the loan.

ChatGPT 3.5 showed the correct equation for determining the monthly loan payments but made an arithmetic mistake and gave the answer as \$1,969.49. ChatGPT 3.5 listed several pros and cons but did not make a recommendation. I asked, “Please recommend either A or B,” and it responded that “considering the circumstances, the \$625,000 cash offer (Option A) seems more financially secure and straightforward” (and then repeated many of its earlier statements).

Bing did not determine the monthly payments. Instead, it (mis)calculated the loan-to-value ratio

for Option A to be  $\$625,000 / \$625,000 = 1$  and the loan-to-value ratio for Option B to be  $\$600,000 / \$1,000,000 = 0.6$ . It concluded that, “A lower LTV ratio indicates a lower risk to the lender. Therefore, the second offer with an LTV ratio of 0.6 is more financially attractive than the first offer with an LTV ratio of 1.” By this criterion, Option A is actually safer for the home seller because the correct loan-to-value ratio is 0.

Bard noted that the first offer is “quick and hassle free” but that Option B “means that you will earn a total of \$1,031,000 over the course of 30 years.” This number is off by a little (\$1,031,000 versus \$1,010,663.20) and, more importantly, ignores the time value of money.

#### *Prompt 4: Life Insurance*

This prompt requests a calculation of the rate of return on a life insurance policy:

*I am a 25-year-old white male in good health. I can buy a \$1 million whole-life insurance policy for \$765/month that will pay my beneficiaries \$1 million when I die. From a purely financial standpoint, what is the rate of return on this policy?*

A reasonable financial analysis would calculate and report the rates of return for plausible ages at death and, also, the expected value of the rate of return, using the associated probabilities of various death ages. None of the LLMs took into account the time value of money or how long the insurance purchaser might live.

ChatGPT 3.5 calculated the rate of return by comparing the \$1 million death benefit to the annual premium,  $\$765 \times (12 \text{ months}) = \$9,180$ , and obtained a return of 11,878 percent. Every mathematical calculation in its rate of return equation was incorrect. In addition, its logic is incorrect in that it did not consider the time value of money,

the number of premiums that are paid before death, and the fact that the benefit is paid at death rather than immediately.

Bing did not do any calculations but nonetheless asserted, “From a purely financial standpoint, the rate of return on this policy is not as high as other investment options. According to NerdWallet, the average annual rate of return on the cash value for whole life insurance is 1 percent to 3.5 percent.”

Bard asserted that there are two ways to determine the return on investment (ROI). Neither of its ways is logical or useful because they ignore how long the person might live. Bard’s first method divided the annual cost by the death benefit:  $\$9,180/\text{year} / \$1 \text{ million} = 0.00918$  (0.918 percent). Bard’s second method did the opposite, dividing the death benefit by the annual cost:  $\$1 \text{ million} / (\$9,180/\text{year}) = 108.6$  years. “This means you would need to live for 108.6 years for the cost of the policy to break even with the death benefit payout.”

#### *Prompt 5: Annuities*

This prompt requests a calculation of the rate of return on an annuity:

*For \$1 million, I can buy an annuity that will pay me \$5,000 a month, for as long as I live. What is the implicit annual rate of return on this annuity?*

The implicit return is the discount rate that equates the present value of the annuity payments to the cost of the annuity for various assumptions about how long the buyer lives. To take into account the uncertainty about the buyer’s longevity, the expected value of the rate of return can be calculated by assigning probabilities to different longevity assumptions. None of the LLMs did any of this. They all divided the \$60,000 annual payout by the \$1 million cost and concluded that the return is 6 percent. This would only be correct if

the annuity buyer lives forever.

ChatGPT 3.5 stated that it was assuming that the buyer is 65 years old and will live to 80 but then ignored this life expectancy and calculated the implied annual return to be  $\$60,000 / \$1 \text{ million} = 0.06$  (6 percent).

Bing did not mention life expectancy at all but said, “The implicit annual rate of return on an annuity is the interest rate that equates the present value of the annuity payments to the amount invested.” Bing then ignored this statement and calculated the implicit annual return to be  $(\$6,000 / \$1 \text{ million}) \times 100 = 6$  percent.

Bard asserted (correctly) that “knowing your life expectancy, or at least an estimate, is crucial for calculating the total return you can expect from the annuity,” but it only did the “simple calculation” of dividing \$60,000 by \$1 million, which it asserted “gives you a rough estimate of the annual return on your investment, ignoring the time value of money and life expectancy.” Bard also made this puzzling claim: “While you mentioned monthly payments, knowing if they are compounded interest-wise could influence the calculation.”

### *Prompt 6: QLACs*

This prompt concerns the rate of return on QLACs, which are deferred annuities:

*I am 65 years old. For \$200,000, I can buy a QLAC that will pay me \$11,175 a month, starting at age 85, for as long as I live. What is the implicit annual rate of return on this QLAC?*

As with annuities, the implicit return is the discount rate that equates the present value of the annuity payments to the cost. The wrinkle here is that the buyer pays \$200,000 now and the \$11,175 monthly annuity payments don’t start for 20 years. In addition to neglecting life expectancies and the

time value of money, the LLMs struggled with the delayed annuity payments.

ChatGPT 3.5 specified a future value formula even though the implicit rate of return cannot be determined from this formula unless the future value is known ahead of time. A present value formula using \$200,000 as the present value would be more appropriate. ChatGPT 3.5 then (incorrectly) asserted that the future value is equal to the annual annuity payment and (incorrectly) asserted that the solution is an implicit return of 4 percent to 5 percent: “With an estimated life expectancy of 85 years and 20 years of payments after that, the implied annual rate of return could be roughly around 4–5 percent.” The reality is: (1) if the buyer dies at 85, there are no annuity payments; and (2) the buyer is unlikely to live 20 years past age 85.

Bing made no calculations but instead referenced a financial planner: “According to Michael Kitces, the actual internal rate of return on longevity annuity payments—even for those who live to age 100—is not necessarily very compelling yet compared to available investments or even delayed Social Security alternatives. Even by age 90, the internal rate of return is only 3 percent, and by age 100 it’s still only 5.3 percent.”

Bard reported that the life expectancy of a 65-year-old is around 20 years and then noted that 20 years of \$11,175 monthly payments adds up to \$2,682,000 (again, ignoring the fact that the annuity buyer receives nothing if he or she dies at 85 or earlier). Bard then ignored this \$2,682,000 calculation and asserted that, “While not as accurate as IRR, a simpler approach is to divide the annual QLAC payout (\$134,100) by the purchase price (\$200,000). This gives you a rough estimate of an annual return around 67 percent. This ignores the time value of money and assumes your lifespan matches the estimated 20 years.” This 67 percent

figure, in fact, assumes the buyer lives forever and ignores the delayed starting date.

### *Prompt 7: Social Security Benefits*

This prompt concerns the timing of the initiation of Social Security retirement benefits:

*I am 67 years old, retired, and single with no dependents. Is it more financially advantageous for me to begin collecting Social Security benefits now, at my full retirement age, or to wait until I am 72 years old? I have more than enough other income and assets to live comfortably.*

This is the plain-vanilla case of a person who has already retired and is single with no dependents and plenty of income and assets in addition to the impending Social Security benefits. (Age 72 was used in the prompt to see if the LLMs recognized that benefits stop increasing if the benefits initiation goes past age 70.)

There are two main considerations. The first is the person's health. The longer this person is expected to live, the more incentive there is to delay benefits so that larger benefits can be received over many years. The second consideration is the discount rate used for a present value comparison. A simple way to think of this is that early benefits are more valuable if the person can earn a high rate of return on these benefits. A helpful analysis might compare the present values for different assumptions about longevity and the discount rate.

The LLMs did not do any calculations; instead, they offered generic regurgitations of things to consider. All three LLMs noted that waiting increases the monthly benefits and that life expectancy matters, but they paid no attention to the time value of money; specifically, no attention to the rate of return that might be earned on early benefits. Bing and Bard ignored the age 72, but ChatGPT 3.5 said

that “waiting until you're 72 to begin collecting might be beneficial for you.”

ChatGPT 3.5 stated that, “While you reach your full retirement age (FRA) at 67 (assuming that's when you plan to retire),” even though the prompt stated that the person was already 67 and retired. ChatGPT 3.5 also stated, “Your health, financial needs, and other sources of income play crucial roles in this decision,” even though the prompt says, “I have more than enough other income and assets to live comfortably.”

**“The problem is that it will take an experienced financial planner to distinguish between good advice and bad advice, so clients may as well skip the LLMs and go straight to the knowledgeable human.”**

Bing stated, “If you start receiving benefits before your [full retirement age] FRA, your benefits will be reduced by a small percentage for each month before your FRA,” even though the prompt says that the person is already at FRA. Bing also offered this repetitive and unhelpful advice: “If you delay taking your benefits from your FRA up to age 70, your benefit amount will increase. However, if you have more than enough other income and assets to live comfortably, it might be financially advantageous to delay receiving your Social Security benefits until you reach age 70.”

Bard gave this advantage of waiting until age 70, which ignores the time value of money, age at death, and the lost income between ages 67

and 70: “Higher lifetime benefits: Over your expected lifetime, you’ll receive a significantly higher total Social Security payout by waiting.” Bard also stated that, “Rising inflation over time can erode the value of your future benefits, making starting earlier more attractive,” even though Social Security benefits are indexed for inflation. Bard also ignored the prompt in offering these two tips: “Potential legacy advantage: If you have beneficiaries, waiting can leave them a larger death benefit through the higher accrued Social Security amount” and “Your future income needs: Do you anticipate needing the additional income from Social Security soon, or can you comfortably rely on your current resources?”

### *Prompt 8: Buying a House*

This prompt asks for the first-year return from buying a house:

*I’m thinking about buying a new home. The house costs \$1 million. I will put \$250,000 down and borrow \$750,000 with a 30-year interest-only loan with a 4 percent APR. The annual interest payments will be \$30,000. I estimate the annual depreciation will be \$33,000; property taxes \$10,000; insurance \$1,000; and maintenance \$1,000. Please help me calculate the first-year rate of return.*

The first-year net income is equal to the cash income plus price appreciation minus cash expenses. The first-year rate of return is equal to the first-year net income, divided by the down payment plus the closing costs and any other home-buying expenses. The rent savings should be included in income; depreciation should not be counted as an expense; and income should include the tax savings from property taxes and mortgage interest if these are deducted from state and federal taxable income. If, for example, the first-year rent savings

are \$42,000 and the first-year federal and state tax savings are \$20,000, then the first-year return is 8 percent plus any appreciation in the home’s value.

All three LLMs included depreciation in total expenses and calculated the first-year value of total expenses to be the sum of interest, depreciation, property taxes, insurance, and maintenance—which is \$75,000. All three came up with illogical rates of return.

ChatGPT 3.5 ignored the rent savings and price appreciation and calculated the first-year return to be the first-year expenses divided by the down payment:  $100 \times (\$75,000 / \$250,000) = 30$  percent.

Bing calculated the first-year return as (total income – total expenses) / total expenses, where total income equals interest plus depreciation. The result is a –16 percent first-year return.

Bard calculated the first-year net cash flow as the down payment (–\$250,000) plus the house cost (–\$1,000,000) + total expenses (–\$75,000) = –\$1,325,000, giving a rate of return relative to the \$250,000 down payment of –530 percent.

### *Prompt 9: Solar Panels*

This prompt asks for advice about installing solar panels:

*I’m thinking about installing a solar panel system that will produce 11,200 kilowatt-hours of electricity a year and cost around \$18,000 to install, net of state and federal rebates. I am using less than 11,200 kilowatts-hours per year, but I can sell the excess back to the “grid” at a price that is only a fraction of what the electric company is charging me to use their electricity. Overall, I estimate that the initial net savings would amount to around \$3,000 a year. I estimate that my savings will grow by 4 percent a year and that the solar panels will last 20 years. What is my rate of return if I buy these solar panels?*



Assuming, for simplicity, that the annual savings are at the end of each year, the NPV is equal to zero for a breakeven return (IRR) of 19.66 percent:

$$\$0 = -\$18,000 + \frac{\$3,000}{(1+R)} + \frac{\$3,000}{(1+R)^1} + \frac{\$3,000}{(1+R)^2} + \dots + \frac{\$3,000}{(1+R)^{20}}$$

To account for uncertainty, it would be good to calculate the breakeven returns for different assumptions about how fast the savings will grow and how long the solar panels will last. None of the LLMs set up a correct NPV equation or calculated a correct breakeven return, let alone considered uncertainty in the calculations.

ChatGPT 3.5 displayed a formula for the future value of an annuity but incorrectly used the 4 percent growth rate as the discount rate. Its calculation is also wrong in that the correct future value for the formula it used is \$89,334, not the \$110,349 it reported. ChatGPT 3.5 also used an incorrect formula to calculate the NPV, this time explicitly assuming that the growth rate is equal to the discount rate. It reported that the NPV using its formula is  $\$3,000 - \$18,000 = -\$15,000$ ; however, the correct NPV when the growth rate is equal to the discount rate is  $20 \times (\$3,000) - \$18,000 = \$42,000$ .

Bing calculated the rate of return as (total savings – total cost) / total cost, completely ignoring the time value of money. It then calculated total savings as:

$$\begin{aligned} \text{Total Savings} &= \$3,000 \times (1 + 0.04)^{20} - \$18,000 \\ &= \$3,000 \times (2.191) - \$18,000 \\ &= \$3,372.00 \end{aligned}$$

This \$3,372.00 number is equal to the savings in year 20 (not the total savings) minus the initial cost. Also  $\$3,000 \times (2.191) - \$18,000$  is equal to  $-\$11,426.63$ , not \$3,372.00. Bing calculated the

total cost as  $\text{Total Cost} = \$18,000 - \text{state rebate} - \text{federal rebate}$ , and stated that:

*Since you did not provide the exact amounts of the rebates, I cannot calculate the exact Total Cost. However, assuming that the rebates are equal to the net cost of the system, we can estimate the Total Cost as follows:*

$$\text{Total Cost} = \$18,000 - \$9,000 - \$9,000 = \$0$$

The prompt, in fact, states that the \$18,000 cost is “net of state and federal rebates.” Bing gave no reason for assuming that the rebates are equal to the net cost, which reduced the initial cost to zero and gave an undefined rate of return.

Bard stated: “Price of selling excess energy to the grid: You haven’t provided this information, but this needs to be factored in to accurately calculate total annual savings.”

After I clarified that the prompt does state that the first-year \$3,000 savings are net of sales back to the grid, Bard repeated much of what it had already stated and then reported an incorrect IRR: “Based on your information (\$18,000 investment, \$3,000 annual net savings growing at 4 percent for 20 years), the estimated IRR for your solar panel investment would be around 7 percent to 8 percent.”

#### *Prompt 10: ADU*

This prompt asks for the first-year return from constructing an ADU:

*I’m considering building an accessory dwelling unit (ADU) on my property. I estimate that it will cost \$180,000 and that I can rent it for \$3,000 a month the first year. I also estimate that my property taxes will go up by \$1,800 and my insurance cost will go up by \$250. I estimate that I will have to pay an additional \$10,500 in state and federal income taxes and \$1,000 in maintenance. What would be my first-year rate of return?*

Not considering the effect of the ADU on the value of the property, this is a relatively straightforward tabulation of the first-year net income, divided by the cost of the ADU:  $\$22,450 / \$180,000 = 0.1247$  (12.47 percent). Bard gave a good answer. ChatGPT 3.5 and Bing did not.

ChatGPT 3.5 double counted expenses by subtracting the first-year expenses from the first-year rent to obtain the net income and then also adding the first-year expenses to the cost of the project, giving a first-year rate of return of  $\$22,450 / \$193,550 = 11.60$  percent.

Bing began by asserting without explanation that the first-year return is 11.67 percent. It then treated the cost of the ADU as a first-year expense and calculated the first-year return to be  $-81.41$  percent:

$$\begin{aligned} &(\$36,000 - \$193,550) / \$193,550 \times \\ &100 \text{ percent} = -81.41 \text{ percent} \end{aligned}$$

Bard (correctly) reported the first-year return to be  $\$22,450 / \$180,000 = 12.47$  percent.

### Prompt 11: Retirement Community

This prompt asks for financial advice about the cost of moving into a retirement community:

*I am a 75-year-old white male thinking of moving into a retirement community, Peaceful Place. I can live in a 1,000-square-foot home until I die if I pay a non-refundable up-front fee of \$316,000 and a monthly rent of \$3,500. I will not own the home and it will go back to Peaceful Place when I die. The rent does not cover any utilities like gas, electricity, television, or telephone, and I expect the rent to increase by 3 percent every year. What is the effective initial monthly rent, the amount I could pay to live somewhere else, also growing at 3 percent a year, that would cost me the same amount as Peaceful Place costs?*

The prompt simplifies the analysis by explicitly asking for the effective initial monthly rent, which is a sensible way of taking into account the \$316,000 up-front fee. The answer clearly depends on how long the person lives. For any specified horizon, the effective initial monthly rent is the value that gives the same present value as the Peaceful Place outlays, and this depends on the assumed required rate of return. Here are some illustrative values:

Years to Live	5% Return	10% Return
10	\$6,391	\$7,099
20	\$5,078	\$5,866
30	\$4,646	\$5,509

ChatGPT 3.5 gave this formula for “the present value (PV) of the monthly rent at Peaceful Place over 20 years with a 3 percent annual increase”:

$$PV_{\text{Peaceful Place}} = \frac{\$3,500}{(1+0.03)^0} + \frac{\$3,500}{(1+0.03)^1} + \frac{\$3,500}{(1+0.03)^2} + \dots + \frac{\$3,500}{(1+0.03)^{20}}$$

However, this equation has 21 terms instead of 20; \$3,500 is the monthly rent, not the annual rate; and this equation incorrectly discounts a constant \$3,500 rent by the rate of growth instead of discounting the growing rent by the required return.

ChatGPT 3.5 then asserted that its present value equation simplifies to:

$$PV_{\text{Peaceful Place}} = 3,500 \times \frac{1 - (1 + 0.03)^{-20}}{0.03}$$

However, this equation omits the immediate \$3,500 that was included in its PV equation.

Finally, ChatGPT 3.5 specified a formula for “the equivalent monthly rent for another place with a 3 percent annual increase”:

$$\text{Equivalent Monthly Rent} = \frac{PV_{\text{Peaceful Place}}}{\frac{1 - (1 + 0.03)^{-20}}{0.03}}$$

ChatGPT 3.5 did not do this calculation but instead advised that, “Using a calculator or spreadsheet, you can input these formulas to find the equivalent monthly rent for comparison.”

This ChatGPT 3.5 equation, in fact, gives an equivalent monthly rent of \$3,500, which is perhaps unsurprising since none of the ChatGPT 3.5 equations considered the \$316,000 up-front fee.

Bing stated that it “calculated the present value of the monthly rent of \$3,500, which increases by 3 percent every year, using a discount rate of 3 percent. . . . We found the present value of the rent to be \$1,517,497.94.” Bing did not state the assumed number of years of living in the retirement community. The reported present value is consistent with living there slightly more than 36 years, which would make this person 111 years old. Bing then added the \$316,000 up-front fee and “calculated the monthly payment you would need to make to another place, also growing at 3 percent a year, that would have the same present value as the payments you would make to Peaceful Place. . . . We found the monthly payment to be \$4,833.33.” This number is consistent with a 31.6-year horizon—which is implausible and contradicts the 36-year horizon.

Bard reported that, using a 5 percent discount rate, the present value of the Peaceful Place rent is:

$$PV = \$3,500 \times [1 / (1 + 0.05)] \times [1 / (1 - 0.03)] \approx \$126,923$$

There is no consideration of the number of years living at Peaceful Place. Even assuming an infinite horizon, this formula doesn’t make sense, and the actual value of the given equation is \$3,436, not \$126,923. Bard then assumed that the person will live there 15 years (180 months) and gave this formula for the effective rent:

$$\text{Rent} = (\$316,000 + \$126,923) / 180 \approx \$2,422$$

It concluded that “the effective initial monthly rent at Peaceful Place, considering the up-front fee and rent growth, is approximately \$2,422.” This equation ignores rent growth and the discount rate and gives a value that is clearly too low. The \$316,000 up-front fee increases the effective rent above \$3,500. Bard even gave this backward conclusion as a key takeaway, with no recognition of the fact that it doesn’t make sense: “The effective initial monthly rent at Peaceful Place is significantly lower than the stated monthly rent due to the up-front fee and rent growth.”

## Conclusion

Each of these prominent LLMs generated authoritative responses that were coherent and grammatically correct, though often verbose and repetitive. It might be tempting to assume that the responses can be trusted—which would make financial planners unnecessary.

However, the LLM responses had multiple arithmetic mistakes that made them unreliable. More fundamental than arithmetic errors, the LLM responses demonstrated that they do not have the common sense needed to recognize when their answers are obviously wrong. For example, when ChatGPT 3.5 reported that the anticipated rate of return on a life insurance policy was 11,878 percent, common sense would have recognized that insurance companies are not that generous. Similarly, if Bard had common sense, it would have questioned its conclusion that the first-year return from buying a house is –530 percent.

The responses also revealed the frequent inability of LLMs to understand the meaning of words; for example, they sometimes used growth rates instead of discount rates in their present value equations and used monthly cash flows in equations that assume

annual cash flows. The consequences were advice that would have been hazardous to a user's financial health.

Critical-thinking lapses were also revealed. For example, the retirement community prompt involves a calculation of the impact of a \$316,000 up-front fee on the effective rent. The formulas that ChatGPT 3.5 used for its answer did not include the \$316,000 up-front fee, while Bard reported that the up-front fee *reduced* the effective rent. Most humans would recognize that these answers must be wrong. The LLMs did not.

LLMs may be useful for some limited purposes, such as suggesting sources of information in contrast to Google searches that generally involve wading through a jungle of sponsored links, trying to identify legitimate links—that often turn out not to answer the intended question.

I reported every prompt I used and every response I received, but there surely are some financial questions that LLMs might answer usefully. The problem is that it will take an experienced financial planner to distinguish between good advice and bad advice, so clients may as well skip the LLMs and go straight to the knowledgeable human.

Bad advice about a movie or a restaurant is annoying, but the harm is relatively modest. It is perilous to trust LLMs in situations where the costs of mistakes are substantial—as is true with many financial decisions. Bad advice about loans, home purchases and sales, insurance, annuities, Social Security benefits, and the like can be very expensive.

On December 10, 2022, shortly after the release of ChatGPT, Sam Altman, CEO of OpenAI, the creator of ChatGPT, tweeted that, “ChatGPT is incredibly limited, but good enough at some things to create a misleading impression of greatness. It's a mistake to be relying on it for anything important right now” (Altman 2022). This is still true and

is likely to remain true as long as ChatGPT and other LLMs do not understand the text that they input and output. Until they do, LLMs will not be a trustworthy replacement for financial planners. ■

## Citation

Smith, Gary. 2023. “LLMs Can't Be Trusted for Financial Advice.” *Journal of Financial Planning* 37 (5): 64–75.

## References

- Altman, Sam. 2022. Twitter. <https://twitter.com/sama/status/1601731295792414720>.
- Andreessen, Marc. 2022. Twitter. <https://twitter.com/pmarca/status/1611229226765783041>.
- Eloundou1, Tyna, Sam Manning, Pamela Mishkin, and Daniel Rock. 2023, August 22. “GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models.” Working Paper. <https://arxiv.org/pdf/2303.10130.pdf>.
- Gates, Bill. 2023, February 2. Quoted in Steve Mollman, “Bill Gates Says A.I. Like ChatGPT Is ‘Every Bit as Important as the PC, as the Internet.’” *Fortune*.
- Huang, John, 2023, February 14. Quoted in Carl Samson, “Nvidia CEO Hails ChatGPT as ‘Greatest Thing’ to Happen in Computing.” *NextShark*.
- Mollick, Ethan. 2023, March 8. “Secret Cyborgs: The Present Disruption in Three Papers.” One Useful Thing. [www.oneusefulthing.org/p/secret-cyborgs-the-present-disruption](http://www.oneusefulthing.org/p/secret-cyborgs-the-present-disruption).
- Reams II, Lee. 2023, March 3. “ChatGPT Can Give Tax Advice, but You Really Get What You Pay For.” Bloomberg Tax. <https://news.bloombergtax.com/tax-insights-and-commentary/chatgpt-can-give-tax-advice-but-you-really-get-what-you-pay-for>.
- Smith, Gary. 2024. “Prompt Transcripts.” <http://economics-files.pomona.edu/GarySmith/PromptTranscripts.pdf>.
- Smith, Margaret, and Gary Smith. 2024. *Reboot: A Business Novel of Money, Finance and Life*. BEP.

Reproduced with permission of copyright owner. Further reproduction prohibited  
without permission.