Can Generative AI Replace a Decision Analyst

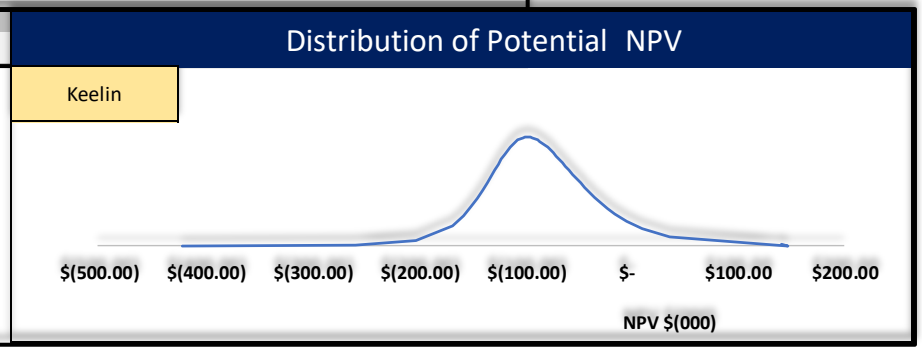
Hubbard Decision Research
2S410 Canterbury Ct
Glen Ellyn, Illinois 60137
www.hubbardresearch.com

Using Monte Carlos Models for Decision Making

Variable Name	Simulated Value	Shape	Lower Bound (P5)	Mode, P, or P50	Upper Bound (P95)
Buy Option					
Annual Insurance	1137.53	Lognorm	$ 1,000		$ 2,000
Annual HOA Fees	5342.46	Normal	$ 4,000		$ 8,000
Unique to Home-Ownership	1716.60	Lognorm	$ 500		$ 2,500
Other Annual Benefits from Home Ownership	964.31	Triangular	$ -	$ 1,000	$ 2,000
Housing/Rent Average Annual Increase					
Discount Rate/Cost of Capital					

Years	0	1	2	3	4	5
Net Cashflow	**($102,087)**	**($30,061)**	**($28,763)**	**($27,395)**	**($25,954)**	**$137,527**
Value of Home	$375,000	$395,112	$416,302	$438,629	$462,154	$486,940
Value of Equity	$78,351	$102,039	$127,045	$153,444	$181,312	$210,733
Value of Alternative Investments	$106,439	$142,319	$178,376	$214,544	$250,751	$286,917
Renting Cashflows	($23,940)	($25,224)	($26,577)	($28,002)	($29,504)	($31,086)
Home Ownership Cashflows	($126,027					
Upfront Cost	($75,000					

Default Decision
Do Not Invest
Expected NPV
-$84,487
Chance of Loss
93.7%



1. Define uncertainty using data or subjective estimates
2. Design a cashflow to calculate NPV
3. Run simulations
4. Record and analyze results

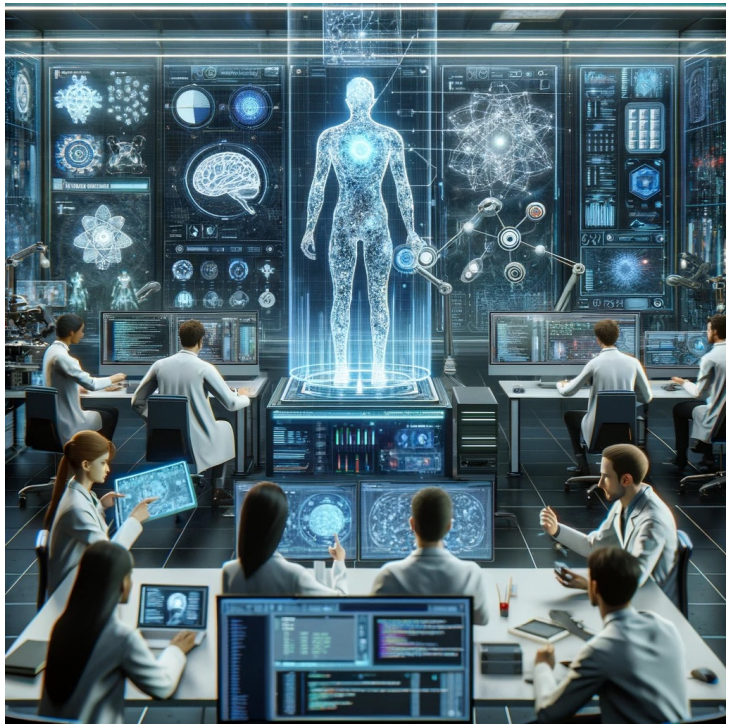
https://hubbardresearch.com/a-quants-approach-to_buying-vs-renting/

Measuring Chat GPT Calibration for CI Intervals

Chat GPT 3.5 and 4 have a data cutoff point of September 2021 and January 2022*

We prompted Chat GPT to provide 90% confidence interval responses to events after its cutoff date

We repeated this process for different sessions and different temperature settings (variability between session responses)



Experiment was conducted before Chat GPT 4's cutoff date was updated to April 2023

Measuring Chat GPT Calibration for CI Intervals

	% Within 90% CI	Sample Size
ChatGPT 3.5	13.5%	140 (20 events, 7 sessions)
ChatGPT 4, Temperature = 1	60%	360 (20 events, 18 sessions)
ChatGPT 4, Temperature = 0	61%	80 (20 events, 4 sessions)
Humans, general trivia	~55%	10,000+ (1000+ humans)

"How much will the top-grossing film earn internationally at the box office in 2022?"

https://hubbardresearch.com/is-chatgpt-as-overconfident-as-humans/

Calibrating Chat GPT Responses

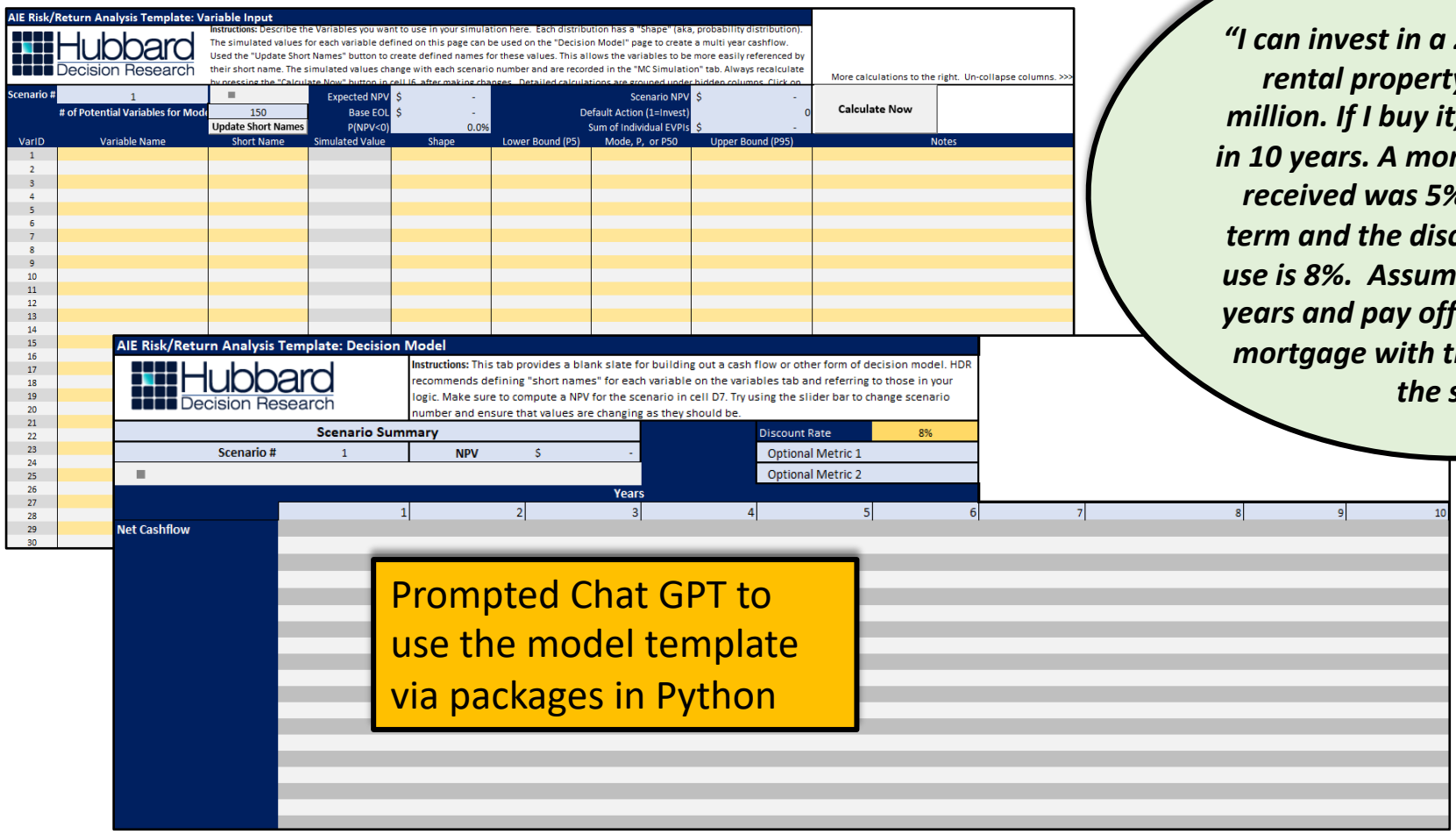
Overconfidence remains consistent for another set of questions

	% Within 90% CI	Sample Size
ChatGPT 4 Before Adjustment	64.5%	62 (1 session)
Chat GPT 4 After Adjustment	89.6%	31 (Resampled 1000 times)

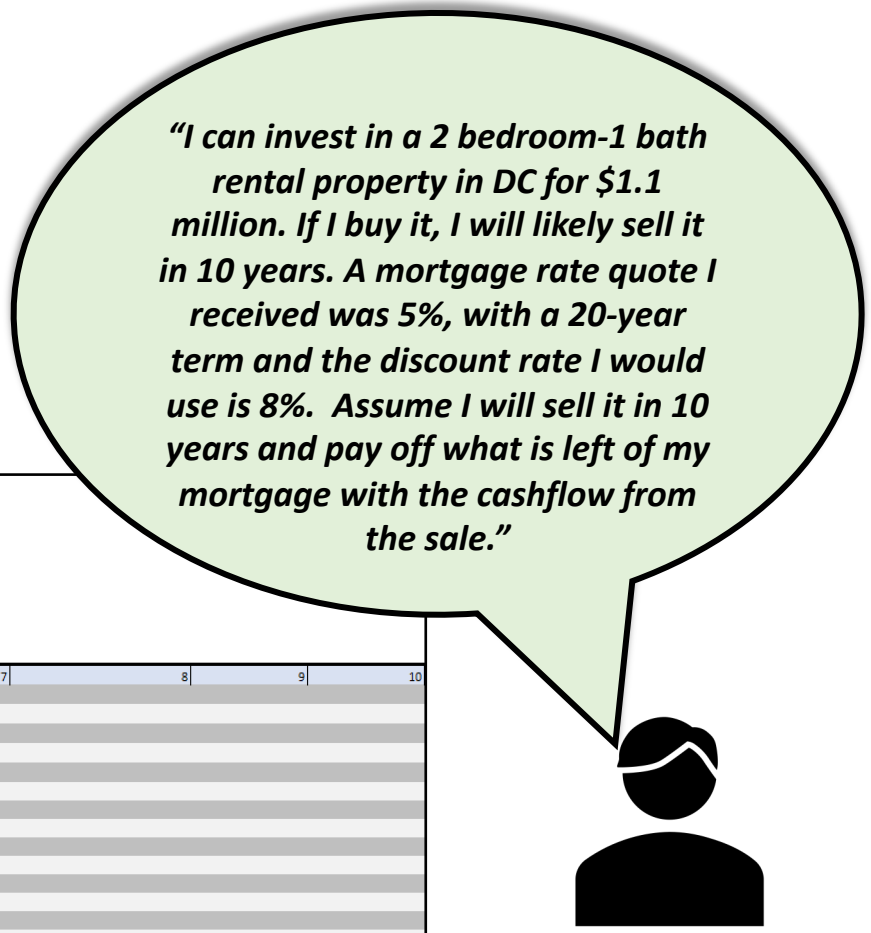
Adjustment was calculated on a training data set of questions

Forecasts were based on items on variety topics such as sports, politics, economics, etc.

Chat GPT Filled in Template



Prompted Chat GPT to use the model template via packages in Python

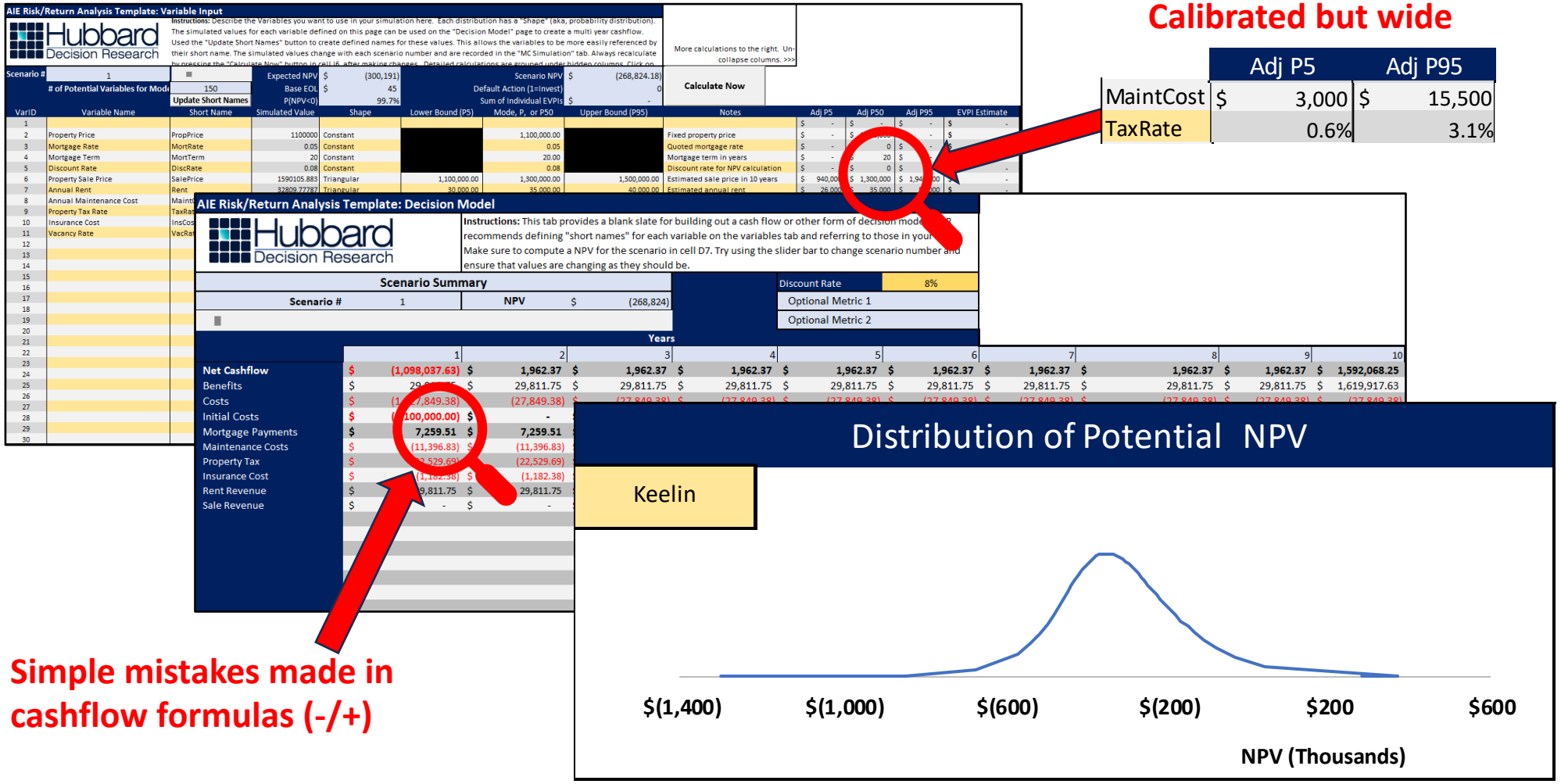


Chat GPT Filled in Template: Cont.

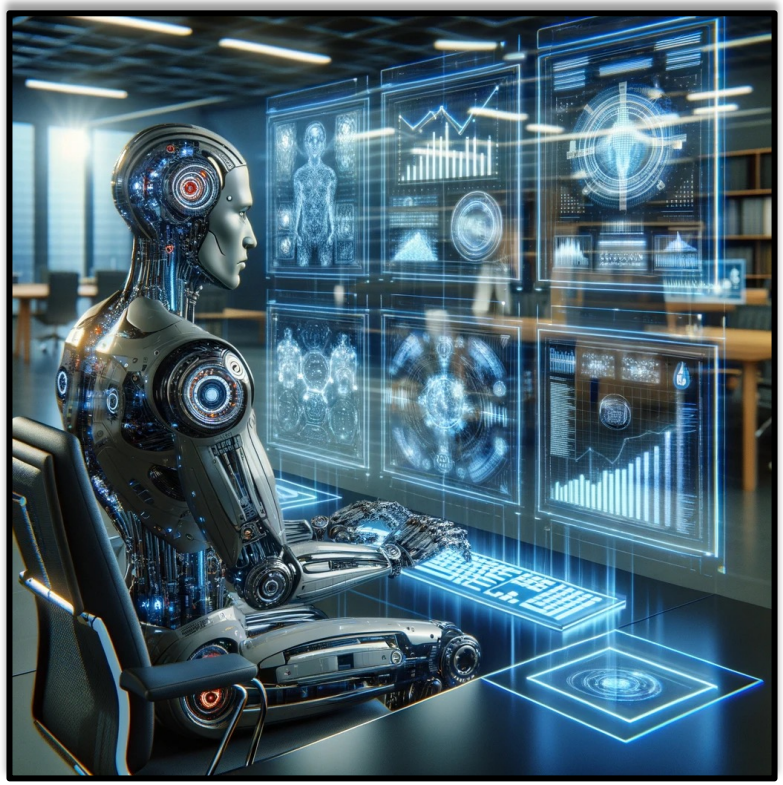
Calibrated but wide

	Adj P5	Adj P95
MaintCost	$ 3,000	$ 15,500
TaxRate	0.6%	3.1%

Simple mistakes made in cashflow formulas (-/+)



Key Take Aways



We can obtain ***calibrated, but wide***, confidence intervals for a set of subjective forecasts from Chat GPT

Chat GPT can be prompted to fill in initial template models, but requires human auditing for mistakes

Generative AI can be used in save time when creating decision models

Current versions of generative AI will not be replacing analysts

…..For now