

First Prize Winner (tied)

Experience Studies Harnessing an AI Agent – A Proof-of-Concept Lightyears Past Code Generation

Mark Spong, FSA, MAAA, CERA

Any views and ideas expressed in the essays are the author's alone and may not reflect the views and ideas of the Society of Actuaries, the Society of Actuaries Research Institute, Society of Actuaries members, nor the author's employer.

INTRODUCTION

At the risk of generalization, I see generative AI being used by actuaries in the following use cases:

Use Case	Examples
Helping to code	<ul style="list-style-type: none"> Generating code from natural language Debugging code that does not work
Synthesizing text	<ul style="list-style-type: none"> Uploading a big pdf and asking specific questions Processing meeting transcripts to generate meeting notes
Writing companion	<ul style="list-style-type: none"> Editing user text for style, grammar, brevity, spelling, or tone Generating outlines Documenting something automatically
Learning tool	<ul style="list-style-type: none"> I'd like to learn more about ... What is the difference between ... ?
Advanced searching	<ul style="list-style-type: none"> Finding or summarizing internal documents Aggregating publicly available information

These use cases all share a common basic transactional structure with the AI relying on user input. While the AI might pull in a basic calculator plug-in or search the internet, most of the time it lacks enough context to lead any sort of complex evidence-based inquiries that require agency. The AI ultimately relies on the user to do something with the information. Responses need to be checked for hallucinations, and facts need to be checked for accuracy. Coders and experience study actuaries like me still have to copy/paste, execute, debug, and ultimately interpret results.

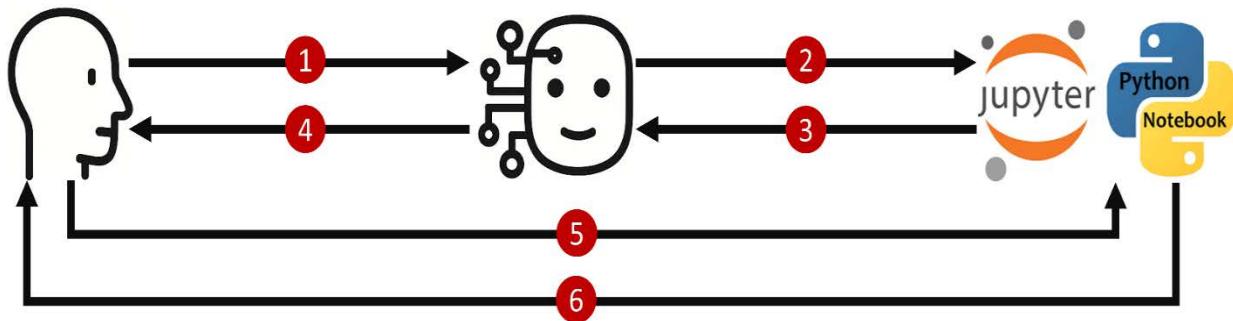
When using generative AI in these ways, the job of an actuary does not change. There may be a few more tools to learn and some tricks to speed things up, but no actuary is worrying about being automated away.

THE GROUNDWORK FOR A NEW USE CASE: COLLABORATING WITH AI ON DATA ANALYSIS

At least, that is what I believed until 2023 when I had the opportunity to collaborate with AI on a data analysis project under fairly unique circumstances:

Unique Circumstance	Implication
The dataset was simulated by me in 2023 as part of a volunteering activity for the SOA.	<ul style="list-style-type: none"> The data is not proprietary and is in the public domain as part of the 2023 SOA Case Study Challenge. There was no personally identifiable information. There were no concerns sharing the data with a Large Language Model.
The dataset was purposefully designed to illustrate relationships often found within life insurer experience data.	<ul style="list-style-type: none"> The data was sufficient to serve as a proof-of-concept for an AI-led experience study analysis. Since I hand-crafted the relationships used to simulate the dataset, I could objectively assess the quality of the AI's analysis and attempts to uncover those relationships.
There was a direct connection between the AI and a Python notebook.	<ul style="list-style-type: none"> There were no hallucination issues because I could inspect and run the code directly. There was no copy/pasting, as the AI could generate and execute the code directly and review and interpret the output without me as an intermediary. The AI could read the previous code and output, so it had much broader context across prompts.

This last point is powerful and is a massive leap forward from how most actuaries currently see AI being used. It represents the difference between using AI as a tool that generates text to partnering with an AI agent that analyzes data (nearly) autonomously. Consider the following diagram, which summarizes the interactions between me, the AI, and the Python notebook:



Interaction	Description
1	<ul style="list-style-type: none"> • I used natural language to ask one big overarching question of the AI. • I provided guidance based on what I saw in the code and the AI responses.
2	<ul style="list-style-type: none"> • AI generated Python code, inserted it directly into the Python Jupyter notebook and executed the code.
3	<ul style="list-style-type: none"> • AI could read the results directly (for that run and the rest of the notebook for context). • If there was an error, the AI looped back to (2), debugged the code and reran it automatically without any sort of intervention.
4	<ul style="list-style-type: none"> • AI produced a response in natural language describing what it did, sharing any summaries or charts, and proposing next steps.
5	<ul style="list-style-type: none"> • I actively reviewed the Python workbook to see the code pop-up in real-time. • I intervened if AI got caught in an infinite loop of going back and forth with the Python notebook.
6	<ul style="list-style-type: none"> • When I chose to, I let AI take a rest, worked in the Python workbook like a normal user, and edited, debugged, and reran code on my own (no AI interaction).

Effectively, this let me put AI in the driver's seat to perform experience study analysis with me following along at each step of the way. I did not copy and paste code piecemeal or use AI in a small transactional conversation. I did not use it as an SQL generating back-end that would just produce queries and return tables. Rather, I **unleashed AI to solve a big, complicated data analysis problem** involving millions of rows of data, semi-supervised machine learning, and half-formed vague questions that require inquiry, visualization, and intermediate judgement and decisions.

THE PROOF OF CONCEPT: WHAT COLLABORATING WITH AI CAN LOOK LIKE

Although condensed and edited, the following dialogue among me, AI, and the Python notebook **represents an actual conversation** that I had all the way back in 2023. It is a proof of concept which shows that actuaries are massively underutilizing the AI tools that are available.

Human Prompt	AI Response	Python Notebook
Can you confirm you have access to the mortality dataset?	<p>Confirmed.</p> <ul style="list-style-type: none"> • Here are the first ten rows • Here is a summary of the structure of the dataset, list header names <p>What would you like to do?</p>	<p>AI inserted clearly documented code to summarize first ten rows and produce the summary of the structure of the dataset.</p> <p>AI executed code and interpreted results.</p>
I've heard mortality has increased. Determine the cause of the increase and prepare a communication of the findings to an audience of actuaries.	<p>Interesting project!</p> <p>First, let us confirm the premise:</p> <p><line graph showing mortality increasing by calendar year></p> <p>Confirmed. This is generally increasing over time.</p> <p>Should we do a more detailed exploratory data analysis on some of the 10 variables in the dataset? Perhaps see how they change in composition over time?</p>	<p>AI inserted code to produce graph.</p> <p>AI executed code and interpreted results.</p>
Yes, please proceed.	<p>Let us look at the change in composition of the dataset over time.</p> <p>The following six variables (e.g., attained age bucket, face amount) exhibit clear shifts in distribution over time.</p> <p><stacked bar charts showing the mix of each variable changing over time></p> <p>However, these four variables appear to have little to no change in distribution over time (e.g., issue age).</p> <p>Possible next steps:</p> <ol style="list-style-type: none"> 1. Explore which variables have subgroups that have varying mortality patterns. 2. Investigate the collinearity of the subgroups. <p>How would you like to proceed?</p>	<p>Code to produce graphs for each variable</p> <p>AI executed code and produced error (missing a certain Python library).</p> <p>AI added code to install the library.</p> <p>AI reruns the code and produces the graphs.</p>
Let's do both, starting with the first.	<p>How thorough!</p> <p>Starting with univariate data analysis:</p>	<p>For each variable, AI writes and runs code that graphs the mortality split by the unique entries within that variable over time.</p>

Human Prompt	AI Response	Python Notebook
	<p>1. Mortality is higher for older Attained Ages, confirming that an increase in average Attained Age in the study period may contribute to the overall increase in mortality observed.</p> <p>2. Similarly, mortality is higher for Males than for Females, confirming that a shift towards male lives over the study period may contribute to the overall increase in mortality observed.</p> <p>3. ... Continues across all variables...</p> <p>Now for investigating collinearity:</p> <p><scatterplot matrix of all variables></p> <p>There does appear to be very strong correlation between several of the key variables that do appear to contribute to the overall increase in mortality.</p> <p>What an interesting problem! Typically, in problems of this nature, we need to be careful when using regression which can help us control for the changing variable mix. Is that how you would like to continue?</p>	<p>For variables with many unique entries, AI chooses to bucket the data into reasonable groups (e.g., Issue Year is grouped into 5-year buckets).</p> <p>AI writes and executes code to produce a well-formatted scatterplot matrix.</p>
Yes. Please continue. In particular, I'm interested in the drivers for the increase in mortality after Attained Age is controlled for.	<p>Excellent! Feature selection, i.e., choosing which variables go into the model, is critical so let's be thorough and use two approaches to stepwise regression and methodically test each variable using metrics like VIF and AIC.</p> <p>Forward selection (adding features one at a time and testing for inclusion) results:</p> <ul style="list-style-type: none"> ...Many model results with commentary including comparison on VIF and AIC... Final model using Attained Age + 4 other variables. <p>Backward elimination (start with a full model and simplify it) results:</p> <ul style="list-style-type: none"> ...Many model results with commentary including comparison on VIF and AIC... Final model using Attained Age + 4 other variables. <p>The final model in this case is the same between the two methods, which is very good news!</p>	<p>AI writes code and executes code in 10-20 individual steps, with each step requiring interpretation and decision-making based on criteria that it chooses:</p> <p>Forward Selection Method:</p> <ul style="list-style-type: none"> Code for dozens of regressions that build up methodically from an Attained Age only model Code that is run after each regression summarizing the VIF and AIC for 'best current model' with the 'model current under review.' Code creating a table for all models summarizing the variables included and the VIF and AIC to demonstrate the best model. <p>Backward Elimination Method:</p>

Human Prompt	AI Response	Python Notebook										
	<ul style="list-style-type: none"> • Attained Age (block getting older) accounts for most of the mortality increase. • Sex (block becoming more male) accounts for some of the mortality increase. • Risk Class (block becoming more Standard) accounts for some of the mortality increase. • Calendar year has the opposite effect. Essentially after controlling for the variables above, mortality seems to go down over the study period. <p>Would you like me to continue and prepare the communication of the findings to an audience of actuaries?</p>	<ul style="list-style-type: none"> • Code for dozens of regressions that start with a model with all variables and methodically removes them. • Code that is run after each regression summarizing the VIF and AIC for 'best current model' with the 'model current under review.' • Code creating a table for all models summarizing the variables included and the VIF and AIC to demonstrate the best model. 										
<p>Very interesting. So, the initial premise 'mortality is increasing' wasn't taking the natural aging of the block into account.</p> <p>To clarify, these other variables aren't statistically driving mortality increase? Please share the evidence for that in the final report, but not on this step.</p> <p>For now, is there a way to attribute the mortality increase across the variables? For example, I'd like to say that Attained Age shifts accounted for X% of the mortality increase.</p>	<p>Absolutely!</p> <p><detailed explanation on Shapely values></p> <p><exhibit of Shapely values of the variables in the regression></p> <p>Attribution analysis</p> <table border="1"> <thead> <tr> <th>Variable</th><th>Drives X% of mortality increase</th></tr> </thead> <tbody> <tr> <td>Attained Age</td><td>75%</td></tr> <tr> <td>Sex</td><td>20%</td></tr> <tr> <td>Risk Class</td><td>10%</td></tr> <tr> <td>Calendar Year</td><td>(5%)</td></tr> </tbody> </table> <p>Would you like me to continue and prepare the communication of the findings to an audience of actuaries?</p>	Variable	Drives X% of mortality increase	Attained Age	75%	Sex	20%	Risk Class	10%	Calendar Year	(5%)	AI writes and executes code to develop Shapely values and prepare the results in a table.
Variable	Drives X% of mortality increase											
Attained Age	75%											
Sex	20%											
Risk Class	10%											
Calendar Year	(5%)											
Yes, please continue.	<p>Executive style written report with the following sections:</p> <ul style="list-style-type: none"> • Executive Summary • Data – with exhibits for the number of lives, number of deaths, study period, available data fields, univariate data analysis over calendar year, and scatterplot matrix. • Methods – Description of the two methods used and the model evaluation metrics. • Analysis – Summary of final model, including statistical output, and tables summarizing the path to choosing the 	No coding written or run.										

Human Prompt	AI Response	Python Notebook
	<p>final model (each row has a ‘Variable added’ or a ‘Variable discarded’ conclusion).</p> <ul style="list-style-type: none"> ○ Demonstration that when certain other variables that are not included are added, the model metrics get worse or the overall fit not statistically significant. ● Conclusions <ul style="list-style-type: none"> ○ Clear conclusion that, even though on a univariate basis certain variables appear to be contributing to mortality getting worse, they are not a real driver. 	

BARRIERS TO ACCESS AND CONCLUSION

I embarked on this challenge without knowing any Python, but I was aided by AI and a powerful plug-in called Noteable that connected the AI directly with a Python notebook. Although Noteable is no longer available as a standalone product as it was acquired by Confluence, it still proved to me just how close we are to Q&A with an AI to do complex experience studies in minutes. However, actuaries haven't yet figured out how to get these tools online and brought to bear on our business problems. I reached out to Pierre Brunelle, the Co-Founder of Noteable, to ask why not.

Mark: Why isn't this sort of tool widespread? Why am I stuck using AI to generate little snippets of code when it could be facilitating (and running) advanced experience studies and allowing me to interact with my data with natural language questions?

Pierre: From my point of view, the remaining issues are definitely:

- **Data Governance & Security Posture:** The primary blocker is data-related risk. Exposing proprietary data to external, third-party APIs is a non-starter for most regulated industries. The only viable path forward is processing data within a company's own secure perimeter (e.g., in their VPC), which requires a different class of tools with local models.
- **Workflow Transformation vs. Tool Integration:** Most companies are trying to bolt AI onto existing linear workflows, yielding only marginal gains. Value comes from redesigning the workflow entirely around a human-AI interaction, which is a massive organizational and cultural undertaking that most are not ready for.
- **Quantifiable ROI:** Agentic loops are token-intensive and generate unpredictable costs.
- **Auditability:** For a field like actuarial, the entire analytical process must be auditable and reproducible for professional sign-off. Black-box processes, even if effective, are not viable.

I completely agree with Pierre that these are the biggest barriers. Unfortunately, the skillsets of actuaries are not well suited to solving these data engineering and development challenges. That just means we'll have to be vocal internal advocates, help set the vision, and partner with folks like Pierre who can make it happen. I sincerely hope that this proof of concept showcases what is possible with AI, and that we can all get access to these types of tools soon.

Special thanks to Pierre Brunelle for being a sounding board, providing the quote, and pointing me towards his new open-source project, Pixeltab, which is designed to solve issues like these.

* * * * *

Author Byline: Mark Spong, FSA, MAAA, CERA, is an actuary who works with data and predictive models, and studies mortality. He can be reached at markspong@gmail.com.



Give us your feedback!

Take a short survey on this report.

[Click Here](#)

 **SOA**
Research
INSTITUTE