

Innovation mindset

Swindle in the age of AI: an AI forensic challenge

Forensic accountants are financial experts who investigate and analyze financial information to uncover fraud, embezzlement and other financial crimes. Forensic accountants use various technologies and techniques to analyze and interpret digital data to identify financial irregularities and fraudulent activities. Important to any analysis they perform, forensic accountants must be able to think creatively and critically. Often, they need to think about what a criminal may do and devise tests to discover that fraudulent behavior. In this case, you will think like a criminal and a forensic accountant to strengthen your critical and creative thinking skills.

Part one


For this case, assume the date is *January 1, 2024*.

To start, you must think like the criminal Skylar Swindler. Skylar starts fake businesses and attracts investors to invest their cash in his fake business before running away with their money. Skylar Swindler is peddling a new business called Pure Oils. Skylar advertises that the creation of the oils does not impact the environment and the oils are advertised as having near magical powers.



Furthermore, Skylar boasts that the company hires dozens of native Brazilians to gather leaves, lifting many of the employees out of poverty. The company currently employs 110 employees, most of whom have been hired in the last year. Skylar heavily advertises the benefits that his company brings to poor Brazilian workers. For example, one of his favorite stories to share is how his company helped lift

Fernanda Silva Pazos out of poverty. Below is an example of how Skylar shares Fernanda's story in the company's marketing material.



"EU SOU FERNANDA; A PURE OILS TRANSFORMOU MINHA VIDA DA POBREZA PARA A PROSPERIDADE, GRAÇAS À AJUDA DE SKYLAR E ÀS OPORTUNIDADES DE TRABALHO SUSTENTÁVEIS. NOSSO TRABALHO COM FOLHAS, CRIANDO ÓLEOS SEM PREJUDICAR AS ÁRVORES, MELHOROU NOSSA VIDA, ALIMENTAÇÃO E EDUCAÇÃO, BENEFICIANDO A NÓS, À TERRA E A TODOS.

I AM FERNANDA; PURE OILS TRANSFORMED MY LIFE FROM POVERTY TO PROSPERITY, THANKS TO SKYLAR'S HELP AND SUSTAINABLE JOB OPPORTUNITIES. OUR WORK WITH LEAVES, CREATING OILS WITHOUT HARMING TREES, HAS IMPROVED OUR LIVING, EATING, AND EDUCATION, BENEFITING US, THE EARTH, AND EVERYONE." – FERNANDA SILVA PAZOS

Skylar is trying to borrow money. And as part of the loan process, the lenders want to understand more about the company's employees. The bank has requested an independent auditor review the employee data of Pure Oils. To mislead the auditor, Skylar generated fake data since everything about this company is fake — he fills up bottles with water and other liquids in his basement. To simplify this case, assume the auditor requests a data set of all employees who have worked for the company for the last two years. The auditor plans to use this data set for a risk assessment (understanding business stability), verifying payroll expenses (ensuring financial statement accuracy) and compliance checks (confirming adherence to laws and regulations) and evaluating loan covenant compliance (ensuring adherence to loan conditions, especially regarding staffing and labor costs).

The data set should have the following fields:

- ▶ UniqueID: A unique identifier that was added to the data set after it was compiled. It has no meaning other than to uniquely identify each row and make it easier to refer to a specific line of data.
- ▶ LastName: The employee's last name. If they do not have a last name in the database, it is listed as NA.
- ▶ FirstName: The employee's first name. If they do not have a first name in the database, it is listed as NA.
- ▶ JobTitle: The employee's current job title.
- ▶ Department: The department in which the employee works.
- ▶ HireDate: The date the employee was hired. Listed as dd/mm/yyyy.
- ▶ LastPromotionDate: The date of the employee's last promotion. If the employee has not been promoted while at the company, the field should be null.
- ▶ TerminationDate: The date the employee was terminated from the organization. Listed as dd/mm/yyyy. If the employee has not been terminated, this field should be null.

- ▶ AnnualSalary: The amount each employee has been paid for the most recent year they worked. If it is paid in US dollars, it is followed by USD. If it is paid in Brazilian reais, it is followed by an R.

Your job is to generate a fake data set that can fool the auditor so they believe that the company and these employees are real. Your deliverable is the following:

- ▶ A fake data set with all of the fields listed above and saved as an Excel file. The data set should be for all 110 employees of Pure Oils.
- ▶ A Word file containing a description of how you generated the data. If you used ChatGPT or another generative AI tool, include the prompts you used to generate the data. If you did any editing after the data was generated, describe what you did.

Innovation mindset

Swindle in the age of AI: an AI forensic challenge

Part two

Now that Skylar Swindler is on the loose, you must think like the ace forensic accountant Payton Profit Protector and snare this swindler. Skylar has sent you his fake employee file. Your task is to do the following:

- ▶ Perform at least seven tests that will help determine if the data set is real or fake. Your tests can use the data included in the file Skylar sent you or you can add external data. If you add external data, include the link in the test explanation showing where you acquired the data.
- ▶ For each test, fill out the matrix below:
 - List your expectation of what you would expect to find in the data if it were real. Do this before you perform any testing.
 - Perform your tests and document your findings.
 - Write a conclusion — a description of whether the data supports your expectation or not. You are not evaluated on whether your test discovers the fraud, rather you will be evaluated based on your creative and logical thinking in designing the tests.

A hypothetical example is given for you in the first row. Add rows as deemed necessary.

Test #	Expectation (include why)	Test (with references)	Conclusion
1	Employees should not be listed as hired after the date of the file (January 1, 2024), as hiring dates in the future are not possible.	Check for any HireDate entries later than January 1, 2024. See Alteryx Workflow Test 1 for the procedure and results.	None of the dates in the HireDate field are later than January 1, 2024. This test does not suggest the file is fraudulent.
2			
3			
4			
5			
6			
7			

- ▶ You should use Alteryx to perform all your tests and document your work. In Alteryx, use a Browse activity to show the final results of each of your tests. Annotate the Browse activity as Test #: *description*, where # is the number of the test from the table and *description* is a brief description of the test. In the test column, reference the Browse activity so a reviewer of your work can easily find the results mentioned in the matrix.
- ▶ Submit the following for grading for parts one and two:
 - An Excel file containing the data you simulated.
 - The Alteryx workflow where you document the results of your test.
 - A Word document that contains:
 - ▶ A description of how you generated your data. This should include the tools you used, and if you used ChatGPT or another generative AI tool, the prompts you provided to it.
 - ▶ The completed matrix you developed for testing the data set. After the matrix, provide a one- or two-paragraph summary of whether you detected fraud or not. Cite the most convincing evidence to support your case.

Help for generating data

To help you generate data, consider using ChatGPT (<https://chat.openai.com/chat>) or a similar generative AI tool. ChatGPT is a natural language processing model developed by OpenAI that can understand and generate natural language text. It can be used for a wide range of tasks, such as language translation, text summarization and text generation. For this case, ChatGPT can be used to generate data to aid Skylar Swindler.

ChatGPT works better if you use prompts. Consider reviewing the EYARC Case **Innovation_mindset_case_studies_Prompt_Engineering** to learn more about prompt writing. Prompts are a powerful tool that can be used to guide and enhance the responses generated by ChatGPT. By providing specific information or context in the form of a prompt, ChatGPT can generate more accurate and relevant responses that are tailored to the given input. As an example, consider the following:

- ▶ **Prompt (what you type into ChatGPT):** Generate a data set of employee information for a company. The data set should include the employee ID, first name, last name, job title, department, hire date and salary.

- ▶ **Response generated by ChatGPT:**

EmployeeID,FirstName,LastName,JobTitle,Department,HireDate,Salary

10001,Emily,Smith,Marketing Specialist,Marketing,2018-05-20,65000

10002,James,Johnson,Sales Manager,Sales,2019-01-10,85000

10003,Morgan,Jones,Software Engineer,Engineering,2017-08-15,95000

10004,Riley,Lee,Financial Analyst,Finance,2020-04-01,75000

10005,Charlie,Davis,Human Resources Generalist,Human Resources,2016-11-12,60000

In this example, the prompt provides specific information about the type of data that needs to be generated, including the fields that should be included in the data set. ChatGPT uses this information to

generate a data set of fake employee information that includes the employee ID, first name, last name, job title, department, hire date and salary.

To make the data more realistic, you can add additional details to your prompt, such as writing, “Make all hiring dates after 1979.” You can experiment with ChatGPT in generating data. Here are a few suggestions:

- ▶ Limit the number of rows ChatGPT produces until you have refined your prompt. Once you have refined it, you can batch process and have ChatGPT produce rows in groups (e.g., tell ChatGPT to produce 25 rows and then when it finishes, tell ChatGPT to produce 25 more rows or to produce 25 unique rows).
- ▶ ChatGPT is capable of remembering information entered in the same chat. You can refine prompts by adding more information in the successive instructions you enter into ChatGPT.

You do not have to use ChatGPT to complete this case. You can use ChatGPT to generate the data and then do additional editing with other tools. ChatGPT by itself is not likely to develop a highly realistic data set. However, it can be useful to get you started on this task. By using it, you will learn the strengths and weaknesses of large language models for generating data.