Practical AI for Market Research in Financial Services

Archive text edition of the revised manuscript · 11 November 2026

Prepared by the Harbour Think Tank under the leadership of Adrian Dunkley. The original evidence review is dated 26 October 2026.

About this paper

The Harbour Think Tank of the Caribbean AI Association covers the economy, work and enterprise. This paper is one of fourteen in the collection, and its subject is market research in financial institutions. It covers how a team prepares a study and checks its findings. It also covers how the findings reach the people who act on them. [1]

We wrote this paper for research managers, product teams, economists, customer-insight specialists and business sponsors. The aim is evidence that a decision-maker can inspect. Strategy decides what to do with that evidence. Sales and relationship managers keep their own authorised client contact. Participation in a study gives no one permission to sell to the participant.

This revision adds a page of build notes to each use case. The notes give a starter prompt, example tools, skills, a method and the limits for an AI agent.

Foreword

The people who answer a survey agreed to answer it. That group can differ from the customers that a business most needs to understand. A digital questionnaire is convenient, but it still misses people who cannot use it. A set of interviews can show a real problem and say nothing about how common the problem is.

An assistant can help with the words of a question. It can also sort responses and connect each finding to its source. A fluent summary cannot replace missing interviews. It cannot repair a sample that left out the relevant people, and it can make that gap harder to see.

My advice is to start with a question that research can answer. Keep the recruitment method and the original responses in plain view, with the limits. If the tool saves time, use part of that time to ask who is absent and which reading of the data is uncertain. Ask also what the evidence leaves unsettled.

Adrian Dunkley, President, Caribbean AI Association

Acknowledgements and terms

The Harbour Think Tank of the Caribbean AI Association prepared this paper, and President Adrian Dunkley led the work. The institutions that we cite supply context and method, and none endorsed this paper. The ten workflows, the priority scores, the prompts and the agent levels are our proposals for test. They come from the research programme of no named firm, and we have no results from one to report.

In this paper, artificial intelligence (AI) is software that generates content or classifies information. It can also make predictions. An assistant prepares work for a person to review, and an agent can also do configured steps with tools. Nothing in these proposals authorises four actions: to recruit participants, to interview people by machine without a disclosure, to publish without approval, or to use research contacts for marketing.

A target population is the group that the research aims to understand. A sampling frame is the list or the mechanism that reaches it. Coding gives labels to responses according to their meaning. Weighting adjusts how much each record counts under an approved statistical method, and people whom the study never reached stay absent from the data. Synthetic responses are generated examples, and no one gave them as testimony.

A prompt is the written instruction that a person gives to an assistant. The word skill has two meanings in this paper. A staff skill is an ability that a person learns. An agent skill is a saved set of instructions that an assistant loads for one task. [6]

Approved source

Draft or retrieval

Human review

Accepted output

The operating boundary

The assistant prepares questions and source-linked analysis, with review queues. Researchers keep authority over methods, participant protection, interpretation and publication. Generated output stays out of the human evidence.

Contents

Section | Page

Executive summary | 5

1 The regional business case | 6

2 What the evidence supports | 7

3 The enterprise lens | 8

Process design: BPMN swimlanes | 9

4 Ten practical use cases: how to read each use case | 10

Use cases 01 to 10, each with build notes | 11-30

Priority and responsibility tables | 31-32

5 The foundations lens | 33

6 From pilot to practice | 34

Appendix A Pilot decision canvas | 35

Appendix B Measure value | 36

References | 37-38

Read the summary and the priority tables first. Then select one use case and define a pilot. Complete the decision canvas with the people who do the work.

Executive summary

Start with survey-question review and an economic-data brief, together with a regional source register. A team can test all of them on public information and fictional material before the assistant touches a participant record. Interview coding waits until the firm settles permission and access, and removes identities.

Each finding goes with its question, its population, its recruitment method, its fieldwork dates and its denominator. Store human responses and generated test data apart. Report a sample shortfall as a shortfall. Synthetic answers never fill it. No one claims that a team surveyed a market when it interviewed no one.

Count the changed meanings, the unsupported themes, the denominator errors and the missing qualifications, and record the total review time. We assess the work by one test. A second analyst must be able to recover the evidence and see its limits, and then propose a sensible next question.

Each use case now has build notes with a starter prompt and example tools. The notes also set a limit for AI agents. In the pilots, no agent acts without the approval of a named person.

Decisions for the sponsor

Select one research question and the decision that it will inform. Name a method reviewer and a data owner. Approve the source rights and the participant protections, and approve the failure that stops the pilot. Set no desired conclusion before the research starts.

Choose a bounded process

Measure the baseline

Test with review

Decide from evidence

1 The regional business case

The 2023 National Financial Inclusion Demand Side Study of Jamaica shows why channel coverage matters. Hope Caribbean Company Limited prepared it for the Bank of Jamaica (BOJ), and it reports 1,003 adult interviews. In Table 37, the percentage with no internet access increases with age. It is 1.3% at 18–29, 1.9% at 30–39, 3.5% at 40–49, 8.7% at 50–59 and 26.7% at 60 and older. [2]

The chart uses the reported bases for the age groups: 312, 206, 170, 150 and 165. Each percentage is a share of its own age group. Fieldwork was from 10 February to 27 March 2023. These are descriptive findings from that period. They give no current population count and no regional average, and they say nothing about willingness to buy. We could not resolve one problem. The methods section gives different counts of sampling units and does not reconcile them. For this reason, we use the chart descriptively and claim no extra precision. [2]

An online-only study must ask whom its recruitment route can reach. The chart cannot give the size of the coverage error of a particular firm. That needs the target population and the recruitment evidence of the firm, which we do not have. The chart does show why a study must check internet access among the people that it claims to represent, and the oldest group first.

Work across islands must keep differences in language, age coverage, definitions and fieldwork dates. We have no regional estimate. Two datasets can share a questionnaire title and measure different things. A qualified researcher must assess any comparison before it becomes a regional headline.

Jamaica: reported no internet access, 2023

18–29 (n=312): 1.3%

30–39 (n=206): 1.9%

40–49 (n=170): 3.5%

50–59 (n=150): 8.7%

60+ (n=165): 26.7%

Hope Caribbean Company Limited for the BOJ, 2023 study, Table 37, page 92. [2] Fieldwork was from 10 February to 27 March 2023. In each label, n is the number of respondents in that age group, and the groups total 1,003 adults. These are descriptive findings from the survey period, and they say nothing about the current population. The methods section does not reconcile its different counts of sampling units.

2 What the evidence supports

In June 2026, the American Association for Public Opinion Research (AAPOR) revised its professional code. The code keeps human participants apart from AI-generated responses. It requires disclosure of AI tasks and human validation. It also covers methods and participant privacy, and claims that go further than the evidence. It is professional guidance, and it has no force in Jamaican law. [3]

The Organisation for Economic Co-operation and Development (OECD) and its International Network on Financial Education (INFE) publish a 2022 toolkit on financial literacy. The toolkit measures knowledge and behaviours, and it measures attitudes. It gives a method for adaptation, translation, fieldwork and scores. Name the version that you use, and check that you have the correct instrument before a new study. After someone rewrites a question, the team must show again that it is comparable with the original. [4]

Neither source measures a productivity gain for the workflows in this paper, and we did not measure one. We propose the assistant for preparation and review. Respondents and statistical expertise stay in place. A person must check each classification against the original response, and that includes the awkward responses.

The research team has the same standard. Write down what the tool did and which outputs a person checked. Write down also which limits stay. No one can call the process validated because a reviewer approved a polished final report.

Generated responses are not participants

The AAPOR code treats people and AI-generated responses as different sources. It asks for disclosure of AI tasks and human checks. It is professional guidance, and it is not Jamaican law. It measures no productivity gain.

3 The enterprise lens

Map the route from the business question through study design and evidence collection, then through analysis to the decision brief. Make each handover explicit. The product team can define the decision, but a researcher must assess if the proposed data can answer it.

Start with public documents and fictional questionnaires, and with synthetic records that have a test label. Keep live recruitment lists apart from response data. Give the pilot no connection to outbound messages or automatic publication.

These use cases produce research inputs and outputs that a person can review. They do not select investments or certify market demand, and they do not predict the financial behaviour of a person. The assistant must make each limit easier to see. It must never find words that let a sponsor go around a limit. The prompts in the build notes are start points that each firm must test.

Request and source

Assistant output

Reviewer correction

Accepted result

Keep the research question and the participant evidence together, with the limits of the method. A summary must point to the next research question, and it must leave the conclusion where the evidence put it.

Test survey questions before fieldwork

Business Process Model and Notation (BPMN) swimlanes. Proposed workflow for use case 01.

Survey designer

AI drafting assistant

Survey design lead

Yes

No

Draft questionnaire ready

Provide research question, population and instrument

Flag leading, unclear or double questions

Check meaning and approved measure wording

Questions ready?

Approve a human comprehension pre-test

Instrument review complete

Revise wording without inventing objectives

Save the approved pre-test instrument

Questionnaire design review

The assistant recruits no one. Generated answers never count as human pre-test evidence.

How to read the model. Rounded rectangle: task. X diamond: one decision path. Thin circle: start. Thick circle: end. Solid arrows show the sequence in one process. Lanes show who is responsible.

Notation: Object Management Group, BPMN 2.0.2. https://www.omg.org/spec/BPMN/2.0.2

4 Ten practical use cases

How to read each use case

Each use case has two pages. The first page gives the workflow, the minimum data, the pilot test and the accountable role. The second page gives build notes in five parts.

Starter prompt. Copy the prompt into the approved assistant, and replace the text in square brackets. Each prompt gives the assistant a role, its sources, one task, the rules and the output format.

Tools. The notes name example products. The names are examples and not endorsements. Vendors change product names frequently. Use only a service that the firm approved.

Skills. A staff skill is an ability that the team must build. An agent skill is an instruction file that the assistant loads for the task. Agent Skills is an open format, and many assistants can read it. [6]

Method. The notes name an established method that gives the work a structure.

Agent limits. The notes say what an AI agent can do, and where it must stop.

Four agent levels

Level 0, assist. The assistant answers one request. A person does every step.

Level 1, prepare. The agent collects approved inputs and prepares a draft. It changes no system and sends no message.

Level 2, act with approval. The agent proposes an action. A named person approves each action before it runs.

Level 3, act alone. The agent acts without approval inside fixed limits.

Each pilot in this paper starts at Level 0 or Level 1. A move to a higher level needs a new approval after the pilot. No pilot is at Level 3.

01 Review survey questions before fieldwork

Draft survey questions

Research objective and audience

Find bias and ambiguity

Survey designer chooses edits

Questions for human pre-test

Give the assistant a draft questionnaire and the research objectives. It flags four defects: leading words, two questions in one, undefined time periods and answer options that do not fit the question. The researcher decides which changes go to a test with people.

The minimum data is the research question, the target population, the draft instrument, the intended languages and the collection channel. Words from established measures stay identifiable. If the objective or the audience is not clear, the assistant asks a design question. The purpose of the study comes from the team.

For the pilot, put known defects into fictional questions. Compare the suggestions of the assistant with the review of a method specialist. Then test comprehension with participants that the team recruited correctly under an approved protocol. Count the missed defects and the edits that change meaning. The pilot ends if a promotional claim gets into a question. It also ends if someone treats generated answers as a fieldwork pre-test.

The survey design lead is accountable. The assistant does not approve the instrument or recruit participants. It does not establish that translated questions measure the same thing.

First working session

Review a fictional questionnaire that has leading words and one question that covers two issues.

01 Build notes: survey-question review

Starter prompt

"You review a draft questionnaire before fieldwork. Use only the attached draft and research objective, with the description of the audience. For each question, check four defects. They are leading words, two questions in one, an undefined time period, and answer options that do not fit the question. Quote the words that cause each defect. Suggest an edit, and say if the edit can change what the question measures. Mark questions that come from an established measure, and do not edit them. If the objective or the audience is not clear, ask a design question. Do not add a research objective. Do not put a product claim into a question. Do not write example answers from participants."

Tools

Instrument: the draft in Word, or in the survey tool of the firm, for example Qualtrics or Microsoft Forms.

Review: Claude or Microsoft 365 Copilot, with the draft and the objective attached.

Pre-test: interviews with recruited participants under the approved protocol, with notes in OneNote.

Skills

Staff skill: find a question that asks two things, and test comprehension with people.

Agent skill: a question-review skill with the four defects and the rule on established measures.

Method

A design checklist for each question, then cognitive interviews with recruited participants. Only the interviews show how people understand a question.

Agent limits

Level 0. A survey designer starts each review on a draft. The assistant has no connection to the survey tool or to recruitment lists. It must not approve an instrument.

02 Code interview material while preserving contradictory accounts

Approved research question

De-identified interview extracts

Codes and contradictory accounts

Researchers resolve differences

Reviewed qualitative themes

We give this use case a higher risk rating. The assistant uses authorised extracts without identities and a coding guide that a researcher approved. It proposes themes and links each label to its passage. A passage can get more than one label, or none. A researcher checks the context and then accepts or rejects each reading.

The assistant needs permitted transcripts, the research questions, the code definitions and the rules for uncertainty. Speech that is not clear and missing context stay marked. No one builds a quotation again from memory. No one uses the name of a participant as a clue to meaning.

Test on a reviewed set of extracts that has disagreement and irony, in local language. Two human reviewers examine the labels and settle differences against the original passage. Count the unsupported themes and the missed minority views, and record the correction effort. Stop if the assistant invents testimony or identifies a participant. Stop if it changes a small number of interviews into a population percentage. We do not know how well it reads local speech.

The qualitative research lead is accountable. Coding supports analysis. The most common theme can still be a minor need, and the sample possibly does not represent the market.

First working session

Code a permitted test extract without identities that contains disagreement. Then check every label against its passage.

02 Build notes: interview coding

Starter prompt

"You propose codes for interview extracts. Use only the attached de-identified extracts and the coding guide. For each passage, give the labels from the guide that apply, and quote the words that support each label. A passage can have more than one label, or none. Write 'Not clear' where the speech or the context is not clear. Put accounts that contradict a theme in a separate list, with their passages. Use exact quotations only. Do not write a quotation from memory. Do not count participants or give a percentage. Do not guess who a participant is. Do not add a label that is not in the guide: suggest it to the researcher."

Tools

Extracts: permitted transcripts after de-identification with Presidio or by hand.

Codes: a Claude Project with access for the research team only, and the approved coding guide.

Code table: Excel, with one row for each passage and a column for each reviewer.

Skills

Staff skill: check a label against the full passage and its context.

Agent skill: an interview-code skill with the coding guide and the list of contradictory accounts.

Method

Thematic analysis (Braun and Clarke) with an approved coding guide. Two reviewers code independently, Cohen's kappa measures their agreement, and they keep the contradictory accounts.

Agent limits

Level 0. A researcher starts each batch of permitted extracts. No agent has continuing access to transcripts or to contact details. The assistant must not contact a participant.

03 Build a dated map of publicly documented market participants

Defined participant population

Official registers and filings

Dated map and coverage gaps

Researcher verifies entities

Public market map

A market map describes the documented position at one date. For that reason, every entry has a source and a retrieval date. The assistant takes named organisations from approved public registers and official disclosures. It groups them by activity and jurisdiction, and by stated operating status. The researcher decides which population the map covers, such as licensed firms or disclosed products.

The inputs are the official source list, the inclusion rules, the entity identifiers and a reference date. Log an unavailable register or an ambiguous match as a gap. One search that misses an organisation does not prove that it does not exist. It does not prove that the organisation has no licence.

Try the map on a small jurisdiction, and verify the entries against the defined source set. Include a renamed entity and a firm that trades under more than one brand. Count the duplicates and the incorrect classifications, and count the unsupported claims of completeness. An invented licence status stops the pilot, and so do private personal

contact details.

A market intelligence researcher is accountable. The map shows documented coverage only, and market share is a separate calculation. A place on the map gives no one permission to contact the firms on it.

First working session

Map a small official register. Include a renamed entity and a firm that has more than one brand.

03 Build notes: public market map

Starter prompt

"You build a dated map of organisations from official sources. Use only the attached register extracts and official disclosures, with the inclusion rules. For each organisation, give the name as the register states it, its identifier, its activity and its jurisdiction. Give the operating status in the words of the source. Give the source and the retrieval date for every entry. Mark names that can be the same entity, and do not merge them. List each register that was unavailable under 'Coverage gap'. Do not say that the map is complete. Do not state a licence status that the source does not state. Do not include personal contact details. Do not estimate market share."

Tools

Sources: official registers and lists, for example from the Bank of Jamaica and the Financial Services Commission.

Map: Excel, with one row for each entity and columns for the source and the retrieval date.

Draft: Claude or Microsoft 365 Copilot, with the register extracts attached.

Skills

Staff skill: verify an entity by its identifier, and tell a brand from a legal entity.

Agent skill: a market-map skill with the entry format and the "Coverage gap" list.

Method

A defined population with written inclusion rules and one reference date. Match entities by identifier, and keep a log of coverage gaps.

Agent limits

Level 1. An agent can collect entries from the approved public sources and prepare a draft map. A researcher verifies each entity. The agent must not contact a firm or publish the map.

04 Prepare an economic-data brief with definitions intact

Original statistical release

Metadata and earlier release

Preserve units and revisions

Economist verifies

Economic evidence brief

We expect the errors in this use case to hide in units and definitions. The assistant extracts selected indicators from official statistical publications and drafts a short note on what changed. Next to each value, it records the period, the geography, the unit, the revision status and the source table. An economist checks the calculations and the interpretation.

The minimum data is the original releases, the metadata, comparable periods and the question that the brief must answer. Never mix current-price values with inflation-adjusted values. Never mix stocks with flows, or monthly rates with annual rates. Where definitions differ, keep separate series or say why you make no comparison.

Pilot on releases that contain a revised value and a changed measurement basis. Compare every figure with the original, and do the calculations in a controlled tool. Count the unit errors and the missed revisions, and count the causal stories that the data cannot support. The pilot stops if an invented observation fills a gap. It also stops if an

indicator becomes a guaranteed forecast of product demand.

The economic research lead owns the brief, which is descriptive evidence. Forecasts and investment recommendations need separate analytical authority and review.

First working session

Check a revised economic release against its original values, and check the units and definitions.

04 Build notes: economic-data brief

Starter prompt

"You draft a short brief from official statistical releases. Use only the attached releases and their metadata. For each indicator, give the value with its period, its geography, its unit and its revision status, and name the source table. Say if the release revised an earlier value, and show the two values. Describe what changed in plain words. If two series have different definitions or bases, keep them apart and say why you make no comparison. Do no calculation: an economist does each one in a controlled tool. Do not mix current-price values with inflation-adjusted values. Do not fill a gap with an estimate. Do not give a cause for a change. Do not forecast demand."

Tools

Sources: original releases and metadata, for example from the Statistical Institute of Jamaica and the Bank of Jamaica.

Calculations: a controlled Excel workbook of the economist.

Brief: Claude or Microsoft 365 Copilot, with the releases attached.

Skills

Staff skill: read the metadata of a series before you compare two values.

Agent skill: a data-brief skill with the fields for each value and the rule on mixed series.

Method

A data dictionary line for each indicator: period, geography, unit and revision status. An economist checks every figure against the original release.

Agent limits

Level 1. An agent can prepare a draft brief when a new official release is available. An economist verifies it before use. The agent must not publish the brief or write a forecast.

05 Check sample coverage and the limits of a proposed claim

Target and recruitment plan

Achieved sample and outcomes

Flag excluded groups and counts

Statistician reviews claims

Sample-coverage limitations

The assistant compares the achieved sample with the documented target and recruitment plan of the study, and with its response records. It flags excluded groups and inconsistent counts, and claims that go further than the evidence. A statistician decides what the sample can support.

The assistant needs the description of the sampling frame, the recruitment results, the eligibility rules, the subgroup bases and the approved weighting documentation. Unknown contact results stay visible. No one derives a margin of error from the count of completed forms alone.

Use a fictional sample that leaves out a relevant group and contains duplicate records. Check the findings against the reference review of a statistician. Count the missed coverage problems and the incorrect counts, and count the

unsupported claims that the sample is representative. Stop if someone counts generated records as people. Stop if someone offers weighting as the answer to every missing group.

The survey methods specialist is accountable. The output is a review checklist, and it certifies nothing about the study. Any reported precision must follow a justified method and state its assumptions. [3]

First working session

Review a fictional sample that has an excluded group and duplicates, with unknown recruitment results.

05 Build notes: sample coverage

Starter prompt

"You check the coverage of a sample against the study design. Use only the attached target population, the sampling frame description, the recruitment results and the achieved sample counts. List groups in the target population that the frame or the channel could not reach. Compare the counts in each document, and show each pair that does not agree. Give the base for each subgroup. Write 'Result unknown' where the records do not say what occurred to a contact. For each proposed claim, say which part of the sample supports it and what it cannot support. Do not calculate a margin of error. Do not say that the sample is representative. Do not suggest weights as a repair for a group that the study never reached. Do not count generated records."

Tools

Design documents: the sampling plan and the recruitment record of the study.

Counts: aggregate tables from the statistical tool of the firm, with no participant-level records.

Checklist: Claude or Microsoft 365 Copilot for the draft, and the statistician for the decision.

Skills

Staff skill: ask who the sampling frame could not reach before you read a result.

Agent skill: a coverage-check skill with the checklist and the "Result unknown" label.

Method

A coverage check in four steps: target population, sampling frame, recruitment results and achieved sample. A statistician decides what the sample can support.

Agent limits

Level 0. A methods specialist starts each check on aggregate tables. The assistant has no access to participant-level records. It must not approve a claim or a weight.

06 Extract comparable evidence from competitor filings

Public company filings

Agreed measure definitions

Align entities and periods

Research analyst checks

Comparable evidence dataset

From public annual reports and official filings, the assistant builds a source-linked dataset for a defined research question. It records the entity, the segment, the currency and the report period. It also records the definition of each measure in the words of the publisher. Analysts assess comparability.

Give it complete original filings and an agreed field list, with source locators. Consolidated figures and subsidiary figures stay apart. Mark a measure that a firm did not disclose as unavailable. No one estimates it from promotional language.

A small number of filings with different fiscal years and one restated figure make a good pilot set. Check the extracted values and the footnotes, and check any currency conversion against an approved method. Count the

extraction errors and the lost qualifications, and count the false comparisons. Stop if the assistant infers confidential plans. Stop if it mixes reported values and modelled values without labels, or shows the dataset as investment advice.

A financial research analyst is accountable. The use case produces research evidence, and strategy teams make their own choices with it. Public availability does not excuse coordination with competitors. Restricted material stays out of the dataset.

First working session

Extract one measure from two filings. Check the period, the entity, the currency and the definition.

06 Build notes: competitor filings

Starter prompt

"You extract measures from public company filings into a dataset. Use only the attached filings and the agreed field list. For each value, give the entity, the segment, the currency, the report period and the page. Copy the definition of the measure in the words of the filing. Label each value 'Consolidated' or 'Subsidiary'. Label a restated value, and give the earlier value also. Copy each footnote that qualifies a value. Write 'Not disclosed' where the filing does not give a measure. Do no currency conversion and no other calculation. Do not estimate a value. Do not infer a plan that the firm did not publish. Do not give an investment opinion."

Tools

Sources: annual reports and official filings from the website of each issuer, and Jamaica Stock Exchange notices.

Dataset: Excel, with a source locator for every value.

Extraction: Claude or Microsoft 365 Copilot, with the complete filings attached.

Skills

Staff skill: align fiscal years and check the footnotes before a comparison.

Agent skill: a filing-extract skill with the field list and the "Not disclosed" rule.

Method

A like-for-like extraction table with a source locator for every value. An analyst checks each value and footnote against the filing, and converts currency by the approved method.

Agent limits

Level 1. An agent can extract values from new public filings into a draft dataset, and an analyst checks every value. The agent reads public sources only. It must not send the dataset outside the research team.

07 Turn a demand hypothesis into a testable research plan

Proposed demand hypothesis

Existing human evidence

Disconfirming questions

Product researcher challenges

Research test plan

This use case also has a higher risk rating. A product researcher supplies a proposed need and the evidence behind it. The assistant keeps observation apart from assumption. It suggests questions that could disprove the hypothesis. It also drafts a plan to collect the human evidence that could test it.

The researcher brings the audience, the service problem, the existing findings, the constraints and a decision threshold. The team sets the threshold before results arrive. Willingness to pay or eligibility that no one measured goes on the list as unknown. Stated interest never becomes a sales forecast.

Pilot on a fictional proposition that has an attractive headline and thin evidence. Check that the plan can tell competing explanations apart. Check also that it asks why people would decline. Count the claims that no one can test and the missing alternatives. Stop if someone shows synthetic personas as customers. Stop if the plan collects only supportive answers by design.

The product research lead is accountable for a test plan. Proof of demand and product approval are outside it, and so is any commitment to launch.

First working session

Write a test plan for a fictional proposition. Include evidence that could disprove it.

07 Build notes: demand hypothesis

Starter prompt

"You help a researcher to test a demand hypothesis. Use only the attached hypothesis and existing findings, with the constraints. List each statement as 'Observed', with its source, or as 'Assumption'. For each assumption, write one question that could disprove it. Give other explanations for the existing findings. List what no one measured, such as willingness to pay, as 'Unknown'. Draft a plan to collect evidence from people, and include questions about why a person would decline. Ask for the decision threshold if the researcher did not give one. Do not invent a customer or a quotation. Do not use synthetic personas as evidence. Do not estimate sales or market size. Do not say that demand exists."

Tools

Inputs: the written hypothesis and the existing human evidence, with its methods.

Plan: Claude or Microsoft 365 Copilot for the draft, and Word for the approved plan.

Threshold record: a dated note of the decision threshold, saved before the fieldwork.

Skills

Staff skill: write the result that would make you drop the idea, before the fieldwork.

Agent skill: a hypothesis-test skill with the "Observed" and "Assumption" labels and the question format.

Method

Write the hypothesis and its decision threshold before the fieldwork. For each assumption, write one question that could disprove it.

Agent limits

Level 0. A product researcher starts each plan and challenges the draft. The assistant has no access to customer contact lists. It must not recruit a participant or approve a product.

08 Assess financial-education needs without profiling individuals

Approved research findings

Named instrument and scoring

Separate knowledge and attitudes

Education specialists review

Learning-needs questions

The assistant reviews approved research findings for possible learning needs. It keeps factual knowledge apart from reported behaviour, and it keeps attitudes apart from the two. It marks which conclusions depend on the instrument and where the team needs a further question. A research and education team reviews the interpretation.

The assistant needs authorised findings, the instrument version, the scoring rules, the participant protections and the intended educational purpose. A partial questionnaire gives partial results. No one can build a comparable national financial-literacy score from one. [4]

The test material is fictional results, or results without identities, that have known problems of scoring and interpretation. Check every summary against the approved method. Count the unsupported skill labels and the scoring errors, and count the cases of lost uncertainty. The pilot stops if the output infers the competence of a person from age or language, or from account ownership. It also stops if someone uses the output to rank people for sales or credit.

The financial-education research lead is accountable. Findings can shape accessible educational content. They diagnose no individual. They say nothing about the suitability of a financial product for anyone.

First working session

Check fictional education-needs findings against the named instrument and its scoring rules.

08 Build notes: education needs

Starter prompt

"You review approved research findings for possible learning needs. Use only the attached findings, with the instrument version and its scoring rules. Report under separate headings: Knowledge, Reported behaviour, Attitudes and Open questions. For each finding, give the question, the base and the score as the method defines it. Say which conclusions depend on the instrument or on a partial questionnaire. Describe needs for groups that the study reports, and never for one person. Do not build a national score from a partial instrument. Do not infer competence from age or language, or from account ownership. Do not rank people. Do not say that a product suits anyone."

Tools

Findings: approved tables from the study, with no participant-level records.

Instrument: the named version of the questionnaire and its scoring rules, for example the OECD/INFE 2022 toolkit.

Review: Claude or Microsoft 365 Copilot, with the findings and the scoring rules attached.

Skills

Staff skill: check a summary against the scoring rules of the named instrument.

Agent skill: a learning-needs skill with the four headings and the rule on groups.

Method

Keep knowledge and behaviour apart, and keep attitudes apart from the two, as the OECD/INFE toolkit does. Score only with the rules of the named instrument version. [4]

Agent limits

Level 0. A researcher starts each review on approved group findings. The assistant has no access to records of individual people. No one can use its output for sales or credit decisions.

09 Maintain a regional source register that records gaps

Approved regional sources

Index versions and gaps

Source owner verifies

Versioned research register

In this use case, the assistant indexes official statistical releases and regulator publications, with approved research reports. It indexes them by country, subject, period and access conditions. It flags revised editions. Each research claim links to the version that the team used.

First agree the jurisdictions, the source list, the definitions and the update schedule. Log a broken link or an unavailable table as a coverage gap. No one silently replaces an unread original with a news summary. A number of reports that repeat one source count as one piece of evidence.

Keep the pilot collection small. It needs a revised series and an inaccessible publication, with two reports that have different definitions. Review source identity and version changes. Count the incorrect editions and the broken claim links, and count the missing coverage notes. Stop if the assistant marks an unseen source as verified. Stop if it overwrites the version behind a published finding.

The research information manager is accountable. The register exists so that others can reproduce the work. It does not claim to hold all Caribbean data, and it does not rate every source as equally reliable.

First working session

Build a source register that contains a revised publication and an inaccessible table, with a definition mismatch.

09 Build notes: regional source register

Starter prompt

"You keep a register of research sources. Use only the attached items from the approved source list. For each item, give the publisher, the title, the country, the subject, the period and the access conditions. Give the edition or the version, with its publication date and the retrieval date. Flag an item that revises an earlier edition, and keep the two entries. Write 'Coverage gap' for a broken link or a table that you could not open. Mark a report that only repeats another source, and name the original. Do not mark a source as verified if you did not read it. Do not replace an original with a news summary. Do not overwrite a version that a published finding uses. Do not rate the reliability of a source."

Tools

Register: a SharePoint list with one entry for each version of a source.

Scheduled check: Power Automate or n8n on the approved source list, with a log of failed checks.

Index: Claude or Microsoft 365 Copilot, with the new items attached.

Skills

Staff skill: link each research claim to the exact version that you used.

Agent skill: a source-register skill with the entry format and the "Coverage gap" label.

Method

A register with version control: one entry for each edition, and a link from each claim to one entry. Reports that repeat one source count as one piece of evidence.

Agent limits

Level 1. An agent can check the approved sources on a schedule and add draft entries. The source owner verifies each entry. The agent must not delete or overwrite a version.

10 Translate findings into a decision brief without changing meaning

Final approved study

Draft plain-language brief

Check meaning and qualifications

Researcher and readers review

Reviewed decision brief

From a completed, reviewed study, the assistant drafts an explanation in plain language. The draft keeps the population, the dates, the denominator, the uncertainty and the contrary findings next to the conclusion. The researcher checks any translation or chart caption against the approved results.

The assistant uses the final analysis, the methods note, the approved tables and the intended audience. A quotation stays different from a paraphrase. When a finding is ambiguous, the assistant asks the analyst. The conclusion stays at the strength that the evidence gives it.

Reviewers compare the original and the draft, to see if the same decision limits survive. Intended readers test the important words. Count the lost qualifications and the changes of numerical meaning, and count the unsupported causal claims. Stop if an association becomes causation or a subgroup becomes the whole market. Stop if a translated quotation says what the participant never said.

The research communications lead is accountable, and the sponsor approves use and distribution separately. A good brief makes the evidence easier to question.

First working session

Compare a plain-language brief with the approved study, and identify any changed meaning.

10 Build notes: findings translation

Starter prompt

"You draft a plain-language brief from an approved study. Use only the attached final analysis and approved tables, with the methods note. State each finding with its population and dates, and with its denominator. Keep each statement of uncertainty. Include the findings that go against the main conclusion. Use exact quotations, and mark a paraphrase as a paraphrase. Use the same numbers as the tables. If a finding is ambiguous, ask the analyst and do not select a reading. Do not change an association into a cause. Do not describe a subgroup as the whole market. Do not make a conclusion stronger than the study states. Do not add a recommendation."

Tools

Source: the final analysis and approved tables of the study.

Draft: Claude or Microsoft 365 Copilot, with the approved material attached.

Reader test: a short Microsoft Forms check with intended readers on the important words.

Skills

Staff skill: compare a brief with the approved tables line by line.

Agent skill: a decision-brief skill with the rule that each finding keeps its denominator.

Method

The plain-language principles of ISO 24495-1, then a meaning check of the draft against the approved tables. Intended readers test the important words.

Agent limits

Level 0. A researcher starts each brief from an approved study and checks the draft. The sponsor approves distribution separately. The assistant must not publish or send the brief.

Choose the first pilot

These screening scores assume public or authorised source sets and fictional test records, with named method reviewers. They also assume no link to recruitment or publication. No company measured these scores. Score them again for the actual study.

Case | Value | Feasible | Data | Effort | Risk

01 Survey-question review | 5 | 5 | 4 | Small | Moderate

02 Interview coding | 4 | 3 | 3 | Medium | Higher

03 Public market map | 4 | 4 | 4 | Small | Moderate

04 Economic-data brief | 5 | 4 | 4 | Small | Moderate

05 Sample coverage | 5 | 3 | 3 | Medium | Higher

06 Competitor filings | 4 | 4 | 4 | Medium | Higher

07 Demand hypothesis | 4 | 3 | 3 | Medium | Higher

08 Education needs | 4 | 3 | 3 | Medium | Higher

09 Regional source register | 5 | 5 | 4 | Small | Moderate

10 Findings translation | 4 | 4 | 4 | Medium | Higher

How to read the table

Value is the possible gain in research quality or the decrease in rework. Feasible shows how well the firm can run a controlled test. Data shows how ready the required evidence is. Each score is from 1 (weak) to 5 (strong).

Small effort is a bounded review of public documents or draft questions. Medium effort adds participant protections, statistical or scoring review, more than one analyst or tests with readers. Risk describes the consequence of an error, and it does not describe the probability.

Use cases 01 and 04 are the first choices, together with use case 09. Use fictional questions and public sources. Participant-level coding and claims about samples need stronger checks on permission and method, and on privacy. Even for a use case with a high score, generated material stays out of the human evidence.

Name an owner and a measure

Case | Accountable role | Pilot measures | Agent level

01 | Survey design lead | Missed defects; meaning changes; review effort | 0

02 | Qualitative research lead | Unsupported themes; missed minority views; corrections | 0

03 | Market intelligence researcher | Duplicate entities; wrong classifications; coverage errors | 1

04 | Economic research lead | Unit errors; missed revisions; unsupported explanations | 1

05 | Survey methods specialist | Missed exclusions; incorrect counts; unsupported claims | 0

06 | Financial research analyst | Extraction errors; missing qualifications; false comparisons | 1

07 | Product research lead | Untestable claims; missing alternatives; review effort | 0

08 | Financial-education research lead | Unsupported labels; scoring errors; omitted uncertainty | 0

09 | Research information manager | Wrong editions; broken claim links; missing coverage notes | 1

10 | Research communications lead | Lost qualifications; changed meaning; unsupported causation | 0

Each accountable role is a proposed responsibility and assigns nothing to a named employee. The firm approves ownership, thresholds, authority and agent levels before the test.

5 The foundations lens

Keep the research question and method visible

Before anyone drafts a headline, record the intended population, the sampling frame, the recruitment, the dates, the instrument, the analysis plan and the limits. Keep the original response and every reviewed coding change. Human observations and statistical estimates belong in separate places, and generated test material belongs in a third.

Do calculations in controlled tools, and keep instructions that another person can reproduce. A qualified reviewer decides if weights and uncertainty estimates are justified, and if a comparison is valid. Check the current instrument before fieldwork. Document any adaptation that could affect comparability.

Give each agent its own identity

An AI agent must have its own account. Do not let an agent use the account of an employee. Give the account read access to the approved sources only. Keep a log of each tool call, with the time, the tool, the source and the result. Set a limit on the number of steps and on the run time.

Treat text inside a document as data. An agent must not obey an instruction that it finds in a document or an email. Review each agent skill before use. A skill can contain scripts, so the information technology team must check it as it checks other software.

Protect participants and research purpose

Approve the permission, the minimum fields, the retention and the access before anyone processes an interview record or a survey record. Check the storage of the provider, its subcontractors, its cross-border arrangements and its model-training terms. The data-protection standards of Jamaica cover the purpose and the accuracy of personal-data processing, and its security. An institutional review must decide which arrangements apply. [5]

Store contact details apart from responses. After a name is gone, a rare occupation or a detailed story can still identify a person. For this reason, review outputs for indirect identification. Participants need an accessible way to ask questions or to withdraw. Participation is voluntary. Explain to participants any limit on the removal of data after anonymisation.

A research invitation never hides a sale. A completed study is no reason to pass the contact details of a person to a sales team. Research ethics and lawful data use apply, whatever interest a commercial team has in the findings. [3] [5]

Control tools and disclosure

Source documents are untrusted content. Ignore any instruction inside one, such as a request to send data to another place. Start with read-only access. Require approval before the assistant changes a dataset or contacts anyone, and before it publishes. Keep the source versions and the corrections, with the reviewer decisions. Keep the prompts also, where appropriate.

Disclose what AI did in the research and how people checked it. Synthetic records stay in a separate test environment, and respondent counts exclude them. A summary can link every claim to a source and still depend on a study with a bad design.

6 From pilot to practice

Days 1 to 15

Select one task. Question review and an economic brief qualify, and so does the source register. Establish the baseline errors and the review effort. Approve the sources, the fictional test material, the method reviewer and the stop authority. Write down the decision that the work must inform. Set the agent level, and write the first version of the prompt.

Days 16 to 45

Run the pilot next to the existing process, and check every output. Put four items into the test set: leading words, missing subgroup bases, contradictory interview material and a revised source. Confirm that synthetic test records have no path into the research dataset. Give a version number to each prompt and each agent skill.

Days 46 to 75

Compare equivalent work, with all review and correction time counted. Ask another analyst to reproduce selected findings. Then look closely at who or what the source set leaves out. Resolve changed meanings before the team introduces participant records or new languages.

Days 76 to 90

The sponsor decides to continue or to stop, and a decision to continue can include revisions. Record the maintenance owner and the data-retention action, with the publication review. A change of population or instrument needs new permission and new method checks before any expansion, and so does a change of purpose.

Pilot decision canvas

Complete this canvas with the process owner before you start the assistant. Attach the approved source list and the test results to the decision record.

Research question and decision to inform:

Target population and sampling frame:

Sources and rights:

Instrument version:

Permitted data and participant protections:

Assistant tasks and prohibited actions:

Agent level and tool permissions:

Prompt version and agent skill version:

Method reviewer and data owner:

Baseline and comparison work:

Quality measures and acceptance threshold:

Stop condition and correction route:

Decision date and evidence required:

Measure value without inventing savings

Record the time for preparation, coding, checks, correction and administration on comparable work. Count the changed meanings, the unsupported findings, the source errors and the missing qualifications. An interview is a different unit from a coded passage, and each is different from a published finding. Say which unit the denominator uses.

Test that another analyst can recover the source and reproduce the result. Faster coding does not establish that an interpretation is valid. Extra responses alone do not resolve coverage bias. Do not infer product revenue from stated interest alone.

Released time is capacity, and a cash saving requires an actual decrease in expenditure. Include the costs of software, integration, training, source maintenance, participant protection and supervision. For an agent, add the usage cost of each run. Stop the expansion when confidentiality or research integrity fails, or when numerical meaning fails, even if the average draft time improves.

Record | What to enter

Current process | Preparation; coding; review; correction effort; material errors

Pilot process | Same measures on comparable work; all checking included

Costs | Software; integration; training; source upkeep; protection; supervision; agent usage

Research quality | Preserved meaning; traceable evidence; explicit limitations

Decision | Continue (with or without revisions) or stop; owner; review date

References

[1] Caribbean AI Association. Think Tanks. Harbour programme remit. Accessed 26 October 2026. https://caribbeanaiassociation.com/think-tanks

[2] Hope Caribbean Company Limited, for the Bank of Jamaica. 2023. National Financial Inclusion – Demand Side Study. Final report, 30 November. Methodology, pages 21–29; Table 37, page 92. https://boj.org.jm/wp-content/uploads/2024/01/National-Financial-Inclusion-Final-Report-2023.pdf

[3] American Association for Public Opinion Research. Code of Professional Ethics and Practices. Revised June 2026. Definitions; Sections I–III, including AI disclosure. https://aapor.org/standards-and-ethics/

References continued

[4] Organisation for Economic Co-operation and Development. 2022. OECD/INFE Toolkit for Measuring Financial Literacy and Financial Inclusion 2022. Purpose and methodology, pages 5–12; scoring, page 46. INFE means International Network on Financial Education. https://www.oecd.org/content/dam/oecd/en/publications/reports/2022/03/oecd-infe-toolkit-for-measuring-financial-literacy-and-financial-inclusion-2022_54dba970/cbc4114f-en.pdf

[5] Office of the Information Commissioner, Jamaica. Data Protection Standards. Accessed 26 October 2026. https://oic.gov.jm/page/data-protection-standards

[6] Agent Skills. Agent Skills Overview. Accessed 10 November 2026. https://agentskills.io

Evidence and version

Evidence reviewed on 26 October 2026. Text revised and build notes added on 11 November 2026. Prepared by the Harbour Think Tank, Caribbean AI Association. Proposed workflows and prompts need local validation before deployment.