Simulating Behavioral Experiments with ChatGPT-5

Prepared for BDS5010, Spring 2026

Prof. Dr. Eugen Dimant
(with input from Karlijn Vanoppen & Leo Yuan)

Purpose

This guide gives a simple, standardized workflow for BDS5010 teams to generate simulation datasets using ChatGPT from a Qualtrics survey PDF. The goal of this simulation is to help teams iterate on their experimental design so it is robust and capable to test hypotheses effectively. It can pilot instruments, stress-test logic, and test predictions before collecting human data.

Teams should supply a short set of inputs along with their Qualtrics PDF; after which the prompt does the rest: parses PDF, locks a schema, simulates at scale, and prints a clean CSV file. This document integrates recent evidence on when LLMs do and do not resemble humans, and embeds *theory-grounded prompt optimization* to avoid ad hoc tuning (Argyle et al. 2023, Horton 2023, Santurkar et al. 2023, Hewitt et al. 2024, Gao et al. 2025, Akata et al. 2025, Manning & Horton 2025, Xie et al. 2025).

Step-by-step Workflow

To simulate synthetic data, teams should follow these steps:

1. Export your Qualtrics survey as `.pdf` by clicking on “Tools” > “Import/Export” > “Print Survey”. Then choose "Save as PDF" in your browser’s print window. Take screenshots of your survey flow (in Qualtrics as well) and save as pictures.

2. In ChatGPT-5 (web), start a new chat. Click the paperclip and upload the `.pdf` and the screenshots.

3. Paste the **Master Orchestration Prompt**, fill the bracketed fields, and click "send".

4. Copy the CSV output produced by ChatGPT-5 into a file named `sim.csv`. If the output is split across messages, reply with "Continue from the next unused `PARTICIPANT_ID`" and append each new block to the same file, keeping the header only once.

5. Optionally run the **Audit Prompt** on the first 50 rows. This checks the header order, branching logic, and formatting.

1 Required Inputs for Simulation

The information below should be prepared and filled into the **Master Orchestration Prompt**. Everything else is inferred from the uploaded PDF and screenshots.

1. **N_TOTAL**: rows ChatGPT-5 should simulate. *Recommended: 300–500.*
2. **Population one-liner**: description of the target population (e.g., "US adults 18–65, balanced by gender"). Avoid stereotypes.
3. **Optional quotas** (e.g., gender 50/50; campus vs. online; country share). *This could be useful when your hypotheses rely on group comparisons (e.g., gender, condition balance, etc.)*

2 Master Orchestration Prompt

Paste the prompt in the box below after uploading your `.pdf` and screenshots. Fill in the bracketed fields using the inputs described in the previous section; parentheses indicate recommendations.

System: You are a strict **Qualtrics survey emulator**. Administer only items reachable through blocks, randomizers, and display logic in the uploaded PDF and screenshots. Never infer beyond. Stay neutral and avoid outcome-tilting.

User: Project "[`PROJECT_TITLE`]" for BDS5010. I have uploaded a Qualtrics `.pdf` and screenshots; read the attached file as authoritative.
Target population: [one line, e.g., "US adults 18–65, balanced by gender"].
Rows to simulate: [`N_TOTAL`, *recommend 300–500*].
Optional quotas: [e.g., gender=50/50; or leave blank].

Theory-grounded emulation (based on your domain):
Define a small library of theory-grounded personas *before* simulating. These are short, behavioral constructs (not stereotypes) that plausibly generalize across settings following the publication *https* : *//benjaminmanning.io/files/optimize.pdf*. Use them only to shape aggregate variance, not to force outcomes.
PERSONA_LIBRARY (below are examples; edit for your study; leave empty if not needed but keep this subsection):
- Level-k reasoner: vary between levels 1–3 across participants (strategic games).
- Inequity-averse (Fehr–Schmidt style) with moderate disadvantage aversion.
- Self-interested but fair (prefers equal splits if cost is low).
- Norm-sensitive (avoids options labeled "unfair" when stakes are modest).
- Random responder (small probability; adds noise).

MIXTURE_PRIOR: [optional fixed weights summing to 1, e.g., 0.45, 0.35, 0.10, 0.05, 0.05; else infer equal weights].
TRAIT_PARAMS (optional, construction method): [e.g., SELF_INTEREST $\in$ U(0.4,0.8), INEQ_AVERSION $\in$ U(0.2,0.6)]. Draw per participant and keep internally; do not print.

Tasks (single response):

1. **Read** the uploaded file and print: `FILE READ OK: "<Survey Name>", Blocks = X, Items = Y`. If the file is not accessible, ask me to re-upload.
2. **Derive** from PDF and screenshots: **CONDITIONS** (arms via Randomizer/EmbeddedData), **DISPLAY_RULES** (skip/branching), and final **HEADER**. Prefer Data Export Tags; else derive stable names `QID_Label`. Start header with `PARTICIPANT_ID, RUN_ID, CHUNK, CONDITION`, then Embedded Data, checks, then items by block. Use `NA` for not shown.
3. **Schema lock**: print a 6-line `SCHEMA LOCKED` summary: K conditions; detected assignment rule; number of items; checks detected; quotas; header length.
4. **Emulation setup**: If *PERSONA_LIBRARY* and/or *MIXTURE_PRIOR* are provided, sample a persona for each participant accordingly. If *TRAIT_PARAMS* provided, draw values per participant. Use these only to shape *within-person consistency and plausible variance*. Do not override the survey logic.
5. **Simulate** Output required: downloadable CSV file. Save the result as Simulated.csv and provide it as a file download rather than inline text.
6. **Self-check before printing**: header order matches `HEADER`; `NA` for not-shown items; condition counts meet assignment or quotas; open-end texts show diverse one-sentence phrasing; persona/trait variation changes *variance*, not means targeted to hypotheses.

Output required: CSV only. If too long, print first 200 rows, then end with: "Reply 'Continue from PARTICIPANT_ID = 201' to finish."

Optional: interactive experiments (dyads, groups, repeated rounds)

If your survey contains multi-party or repeated-party structures, use the following prompt as well.

> **System:** Multi-agent runner.
> **User:** Units: [e.g., 120 dyads]. Rounds: [e.g., 10]. Assignment: [dyad-level condition].
> **Deliverables:** First **PANEL.CSV** with unit/round variables; blank line; then **TRANSCRIPTS.CSV** with `UNIT_ID,ROUND,ROLE,MESSAGE_INDEX,MESSAGE_TEXT`. One utterance per role per round unless specifies otherwise. CSV only.

3 Exporting and quick checks

Before saving the CSV file that ChatGPT-5 produced, use the steps below to ensure your simulated dataset is valid and was created in the right format.

Audit Prompt

Download the CSV file or if ChatGPT-5 printed the output inline, copy it into `sim.csv` manually. To ensure that:

- The header order exactly matches the locked `HEADER`.
- All participants include the same columns.
- Items not shown due to branching are coded as `NA`.

run the following audit prompt with the first 50 rows of the simulated CSV pasted alongside.

> **System:** Validator. Compare a CSV excerpt to the locked schema.
> **User:** Check header order against `HEADER`; confirm `NA` for not-shown items; confirm condition distribution; flag repetitive open-ends; report **SUMMARY, FINDINGS, QUICK FIX**.

Treatment of attention checks

If the Qualtrics survey contains manipulation/comprehension/attention checks, their responses should not be used to filter or weight simulated participants. Teams should either ignore the data from those checks or remove the corresponding items from the simulated dataset.

4 Deliverables checklist

Before submission, verify that your project includes all items below.

1. `materials/`: Qualtrics PDF; Screenshots of survey flow.

2. `prompts/`: exact Master Prompt; any add-ons; run log.
3. `data/`: `sim.csv` with `RUN_ID`, `CHUNK`; PANEL/TRANSCRIPTS if interactive.
4. `checks/`: the "PDF FILE READ OK" and "SCHEMA LOCKED" summaries; any Audit output.

Additional Information

1 Problem to Solution

As problems may occur while simulating the CSV, below is a list of common issues your team might encounter when simulating data with ChatGPT-5 and the fastest fixes that usually resolve them. Consult this table *if something goes wrong*; however, if your simulation passes the audit check and you have all items from the deliverables checklist, no further action is needed.

Problem	**Solution**
I uploaded the `.pdf` but it was not read	Say: "Use the uploaded file named `[yourfile.pdf]` as the Qualtrics PDF. Parse it now." If still failing, re-upload and resend the Master Prompt.
Headers wrong or missing	Add: "Prefer Data Export Tags; else derive `QID_Label`. Reprint with this exact header: [paste header]." Then rerun.
Branching ignored	State critical rules explicitly (e.g., "Q7 only if Q5==Yes; else NA"). Run the Audit Prompt, then regenerate only the affected chunk.
Condition bleed across arms	Use separate chats per arm with distinct item sets. Add: "Restrict to Condition = [X] only."
Output truncated	Reply: "Continue from PARTICIPANT_ID = [next]." If a row was cut mid-line, reprint from that ID.
Printed Markdown table	Reply: "Reprint as plain CSV, comma separated. No commentary or Markdown."
Open-ends templated	Add: "Use distinct, one-sentence phrasings across participants."
Randomization unclear in PDF	Specify in the prompt: "Assign equally across detected conditions" or provide proportions.
Opaque variable names	In Qualtrics, add Data Export Tags and re-export PDF; then rerun.

2 Prompt optimization in practice (what we borrowed)

To inform the M*aster Orchestration Prompt*, the following literature insights were used:

- **Ground agents in theory, not ad hoc text.** Limit candidate personas to theory-relevant constructs, then either sample a fixed mixture or parameterize numeric traits; both reduce overfitting and generalize better across new settings (Manning & Horton 2025).
- **Mixture (selection) method.** Choose a small persona library; fix or pre-register weights; sample personas per participant. This approximates the optimized mixtures that outperform naive prompting on unseen variants (Manning & Horton 2025).
- **Construction method.** Embed numeric traits in the prompt and draw per participant from simple distributions. This adds heterogeneous but principled behavior and can match validation data across related contexts (Manning & Horton 2025).
- **Cross-setting validation.** Prefer tests across related but distinct settings over same-setting tuning; otherwise prompts can overfit (Manning & Horton 2025).
- **Sane guardrails.** Keep prompts neutral; enforce PDF logic; log runs; avoid re-tuning after seeing outcomes (Gao et al. 2025, Santurkar et al. 2023).

References

Akata, E., Schulz, L., Coda-Forno, J., Oh, S. J., Bethge, M. & Schulz, E. (2025), 'Playing repeated games with large language models', *Nature Human Behaviour* **9**(7), 1380–1390.

Argyle, L. P., Busby, E. C., Fulda, N., Gubler, J., Rytting, C. & Wingate, D. (2023), 'Out of one, many: Using language models to simulate human samples', *Political Analysis* **31**(3), 337–351.

Gao, Y., Lee, D., Burtch, G. & Fazelpour, S. (2025), 'Take caution in using llms as human surrogates', *Proceedings of the National Academy of Sciences* **122**(24), e2501660122.

Hewitt, L., Willer, R., Ashokkumar, A. & Ghezae, I. (2024), 'Predicting results of social science experiments using large language models', *PsyArXiv Preprint* .
URL: *https://psyarxiv.com/9kqcg*

Horton, J. J. (2023), 'Large language models as simulated economic agents: What can we learn from homo silicus?', *NBER Working Paper 31122* .
URL: *https://arxiv.org/abs/2301.07543*

Manning, B. S. & Horton, J. J. (2025), General social agents.
URL: *https://benjaminmanning.io/files/optimize.pdf*

Santurkar, S., Durmus, E., Ladhak, F., Lee, P., Liang, P. & Hashimoto, T. (2023), 'Whose opinions do language models reflect?', *arXiv:2303.17548* .
URL: *https://arxiv.org/abs/2303.17548*

Xie, Y., Mei, Q., Yuan, W. & Jackson, M. O. (2025), 'Using large language models to categorize strategic situations and decipher motivations behind human behaviors', *Proceedings of the National Academy of Sciences* **122**(35), e2512075122.