

Agents for Market Research

How useful are they?



The hosts

Simone Lionetti, PhD

- Deep Representation Learning
- Machine Learning for Health
- Model Evaluation
- Computer Vision
- Theoretical Particle Physics

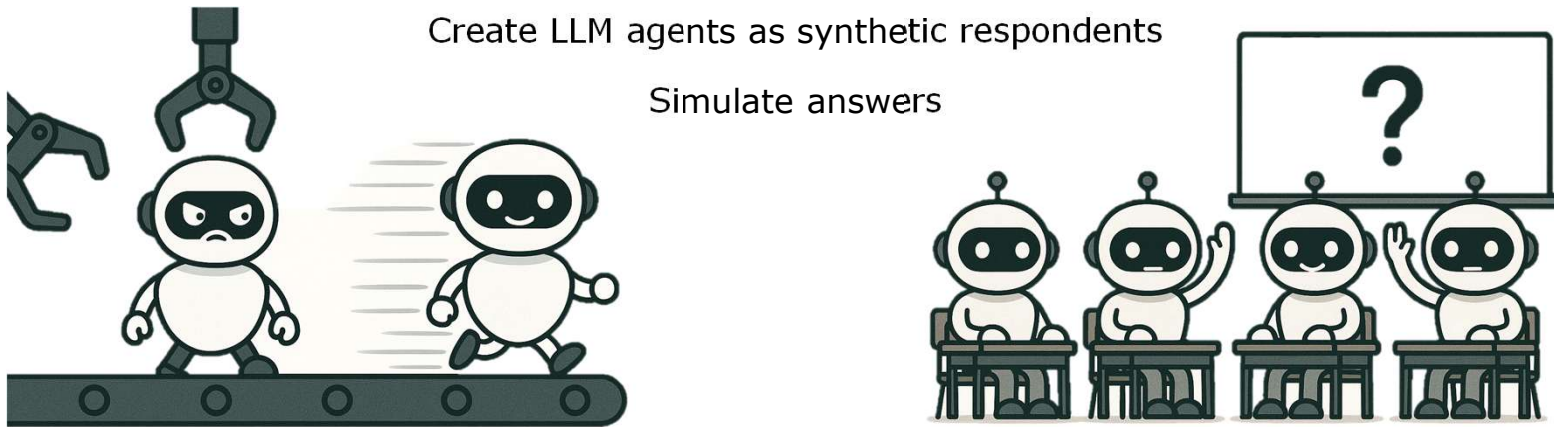


Simone Griesser, PhD



- Marketing Analytics
- Natural-Language-Processing
- Research Methods
- Consumer Behaviour
- Behavioural Economics / Nudging



In-silico Market Research





Who's doing it?


  • YouGov × Yabble: "Virtual Audiences"

 • Largo: "Predicting how target audience reacts to new offering"

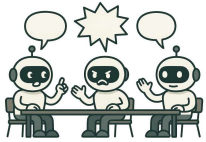
Syntheticusers • Synthetic Users: "AI participants"

 • NielsenIQ: directional screening

 • Qualtrics: "synthetic responses 101"

 • Ipsos: synthetic product testing; Stanford collab

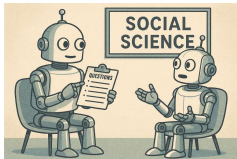
How does it work?



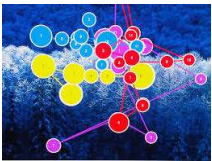
- Sample description in terms of socio-demographics, lifestyle
- Sample size



- Imitate humans responding to survey items --> **quantitative synthetic data**



- Imitate humans responding to semi-structured interviews and focus groups --> **qualitative synthetic data**



- Imitate human eye tracking or tasting data

How does it work?



- Infrastructure:
 - own vs. service provider
 - black vs. white box system
- Type of LLM: local, API, commercial, open-source
- Parameters: prompt, seed, temperature, version
- RAG: trends + client data

How does it work?

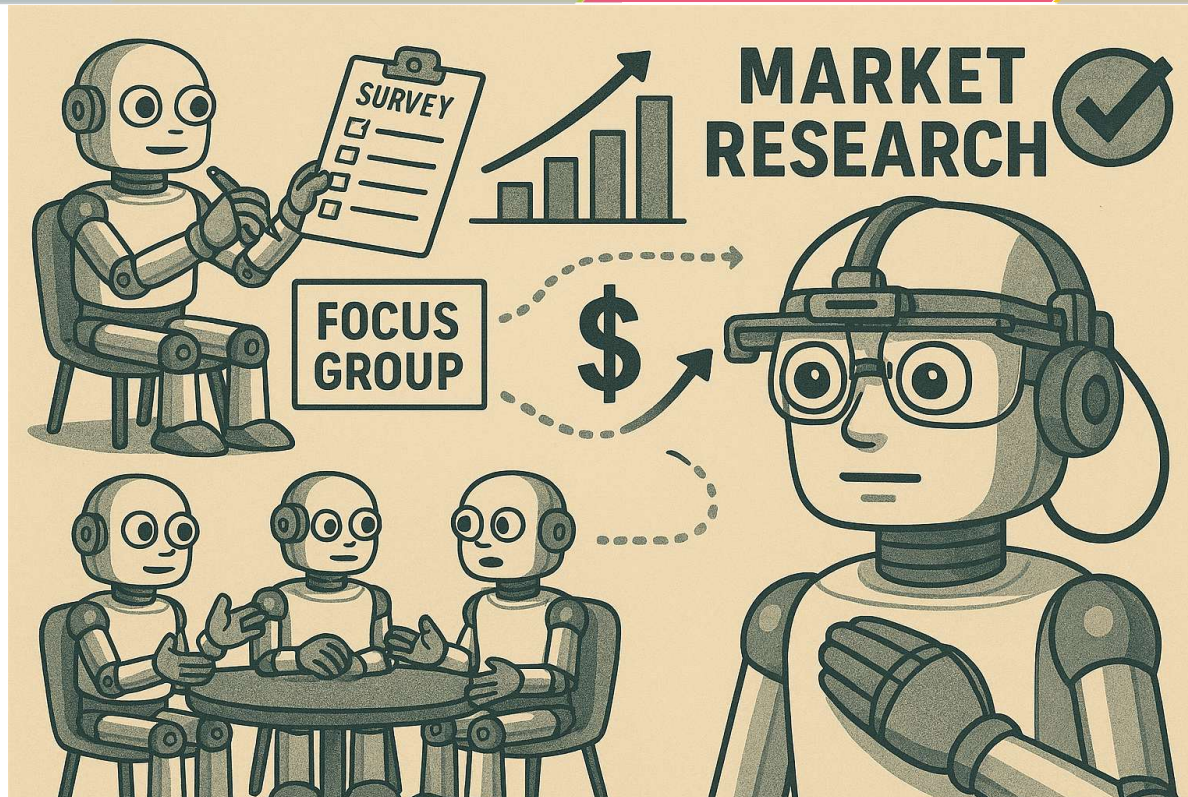
Prompted Personas

LLM Agents

Synthetic Answers

Quality Evaluation

- **Success!**



How does it work? - Quality Evaluation



Evaluation framework for quantitative synthetic data

Criterion	Explanation	Operationalisation
Internal validity	Data collection and analysis choices that rule out alternative explanation	<ul style="list-style-type: none"> Variation in repeated identical executions Systematic prompt variation as sanity checks Quantification of hallucination
External validity	Sampling and setting choices that allow generalizability beyond studied sample	<ul style="list-style-type: none"> Distribution analysis Compare distribution and variability to human responses
Reliability	Measurement should be consistent and stable, i.e., intra and inter-personal consistency	<ul style="list-style-type: none"> Split half reliability / k-fold validation Leave responses out Cronbach's Alpha
Objectivity	Minimising the influence of researchers' opinions and values on data collection, coding, and analysis	<ul style="list-style-type: none"> Prompt review

How does it work? - Quality Evaluation



Evaluation framework for qualitative synthetic data

Criterion	Explanation	Operationalisation
Credibility	Findings are a plausible and faithful representation of participants' realities	<ul style="list-style-type: none"> • Concept coverage within domain
Transferability	Sufficient contextual and methodological detail so that readers can judge whether and how findings may apply to other contexts	<ul style="list-style-type: none"> • Idea density • Semantic elaboration measures
Dependability	Findings are consistent and could be repeated in similar contexts	<ul style="list-style-type: none"> • Compare topics and other psycholinguistic and grammatical markers with human responses
Confirmability	Findings arise from the data rather than researcher's opinions and values	<ul style="list-style-type: none"> • Include self-reflection in prompt, i.e. LLM's positionality and assumptions in data generation process

Key Take Aways

- Human comparison data availability and quality?
- Make evaluation standard
- Quantitative synthetic data is ahead of qualitative synthetic data
- Currently, synthetic personas for qualitative responses are good for direction, not precision, e.g. Griesser et al. 2026
- Augment, don't replace

--> Share methods + metrics

