AI as an individualized economist persona: The case of the “Stephen Littlechild AI Agent”

Festschrift Seminar in honour of Professor Stephen Littlechild

25 March 2026

Cambridge Judge Business School

Based on Bruce Mountain & Shruti Kant VEPC Working Paper 2511 (rev. 4 Jan 2026)

Focus

Development
Assessment
Limitations

A seminar reading of what this paper tells economists about customised AI agents.

Question

Can a general model become a scholar-specific critic?

Why economists should care

> Can a general-purpose AI model be customised to reason in the style of a particular economist?

The paper is interesting not because it claims full imitation, but because it tests whether a large, partly unseen corpus can turn generic AI into a more useful economic critic.

1 **Base models may miss much of the relevant scholarship**

The authors found vanilla ChatGPT could initially identify only 74 Littlechild documents.

2 **Economists often want a particular intellectual frame, not a generic answer**

Here the frame is Littlechild on competition, regulation, negotiation and consumer choice.

3 **The potential use is disciplined critique and synthesis**

The promise is faster understanding of papers, reports and policy questions from a defined perspective.

Today's emphasis: how it was built → how it was assessed → where it still falls short

SCL in plain English: corpus + retrieval + instructions + model

Step 1

Littlechild corpus

Papers, reports, lectures, correspondence

assembled and uploaded

Step 2

Semantic retrieval

find the most relevant passages for each prompt

chunking • embeddings • vector search

Step 3

Configuration

rules on source order, verification and tone

corpus first • verify claims

Step 4

ChatGPT model

turn retrieved evidence into an answer

the response appears as SCL

The key point for non-technical audiences: the model stays general-purpose. The specialisation comes from the retrieved corpus and the instructions that govern how it should use that corpus.

Building the corpus was the hard part

Much of the work looked less like prompt engineering and more like archival reconstruction.

74

documents ChatGPT could initially identify

>350 / 515

author list, later benchmarked against a 515-item full list

~300

searchable PDFs after manual search, OCR and cleanup

~5m

tokens in text files after PDF upload limits were hit

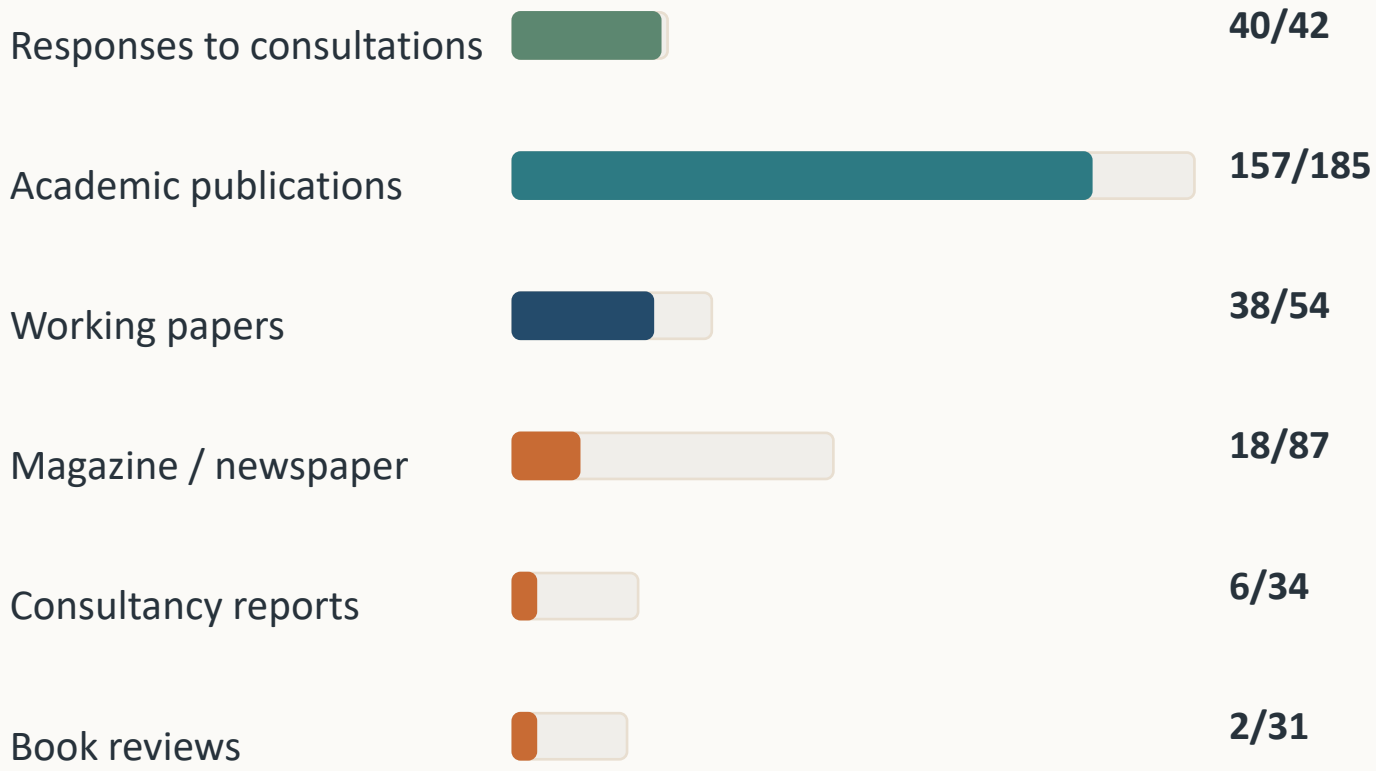
288

documents in the final corpus ingested by SCL

Sources of material included AI-suggested links, manual Google/Google Scholar searches, Internet Archive books and edited volumes, and 11 documents scanned from original paper sources. The eventual workaround was to use text files plus a CSV bibliography rather than only PDFs.

Coverage was strongest in core academic and policy work

Selected categories from the paper's corpus comparison



What this means

The academic and policy core of Littlechild's work is fairly well represented.

Why the gaps matter

Popular writing, consultancy work and shorter pieces are thinly covered — exactly the material that often carries tone, audience awareness and practical style.

Interpretation

Strong reasoning performance is plausible even if "voice" remains incomplete.

Configuration mattered more than one might expect

Reducing hallucination

- Prefer cleaned text files over PDFs when retrieving evidence
- Search the corpus first; only use the web if the corpus is silent
- Force verification of inferences and citations where possible
- Ask precise questions: retrieval became more reliable when prompts were specific

Approximating “voice”

- Explicit do/don’t rules mattered; vague instructions did not
- Model drift was real: GPT-5-Thinking felt more guarded and bureaucratic than o3
- The authors added: “always be critical … don’t be a bureaucrat”
- The final configuration was later trimmed to about three pages

> The system invented “verbatim gems” and even wrote a plausible abstract for a 1966 paper it could not find — including an invented IBM 7090 example.

Assessment design: informed humans, not a benchmark

10 assessors

Six had worked with or for Littlechild.
The rest had long interaction with him and his field.

Design choice

Assessors were free to ask SCL whatever they wished.

No objective machine benchmark for this kind of persona agent was available, so the evaluation is subjective — but deliberately informed.

Six criteria

Completeness

Does it cover the ground Littlechild would likely cover?

Insight

Does it offer innovative critique within his frame?

Fidelity

Is it true to his approach and “voice”?

Usefulness

Would an economist use this tool in their own work?

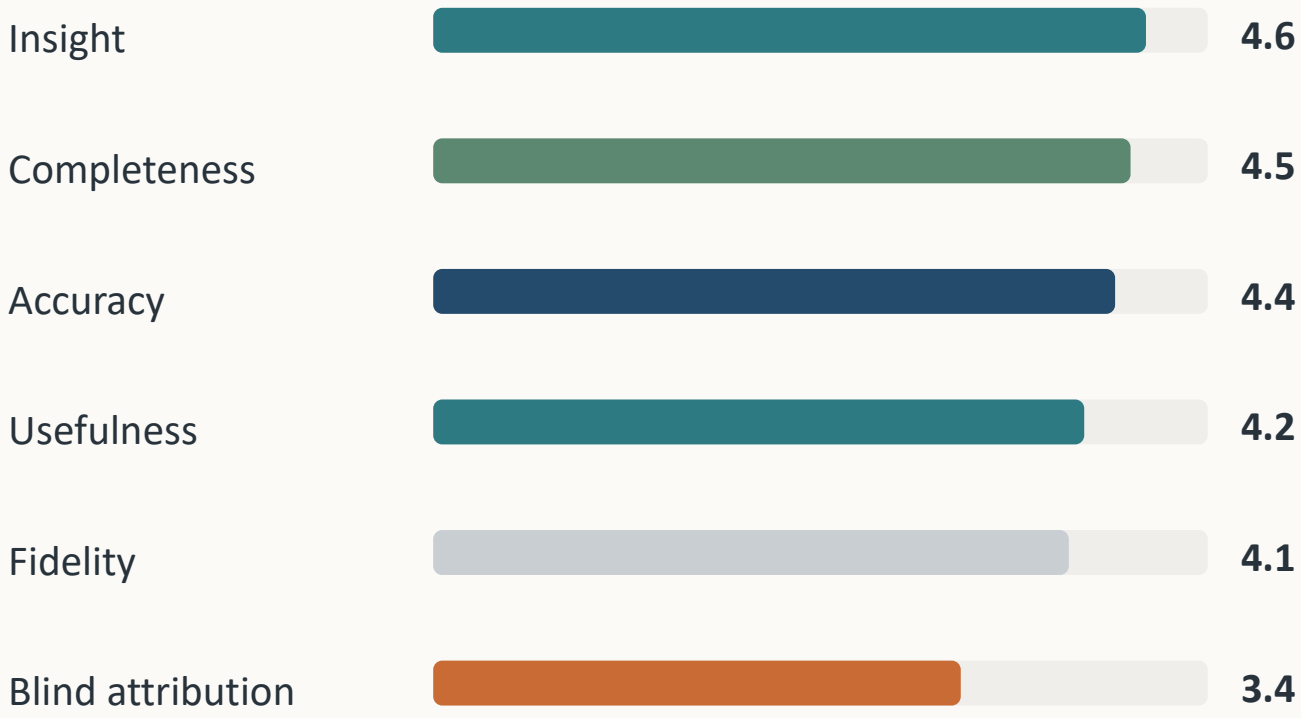
Accuracy

Are facts, dates, quotes and citations correct?

Blind attribution

Would you mistake the answer for Littlechild if the source were hidden?

Human assessment: strong on insight, weaker on imitation



Overall mean

4.2 / 5

Highest score

Insight: 4.6

Seven of ten reviewers gave it 5/5.

Lowest score

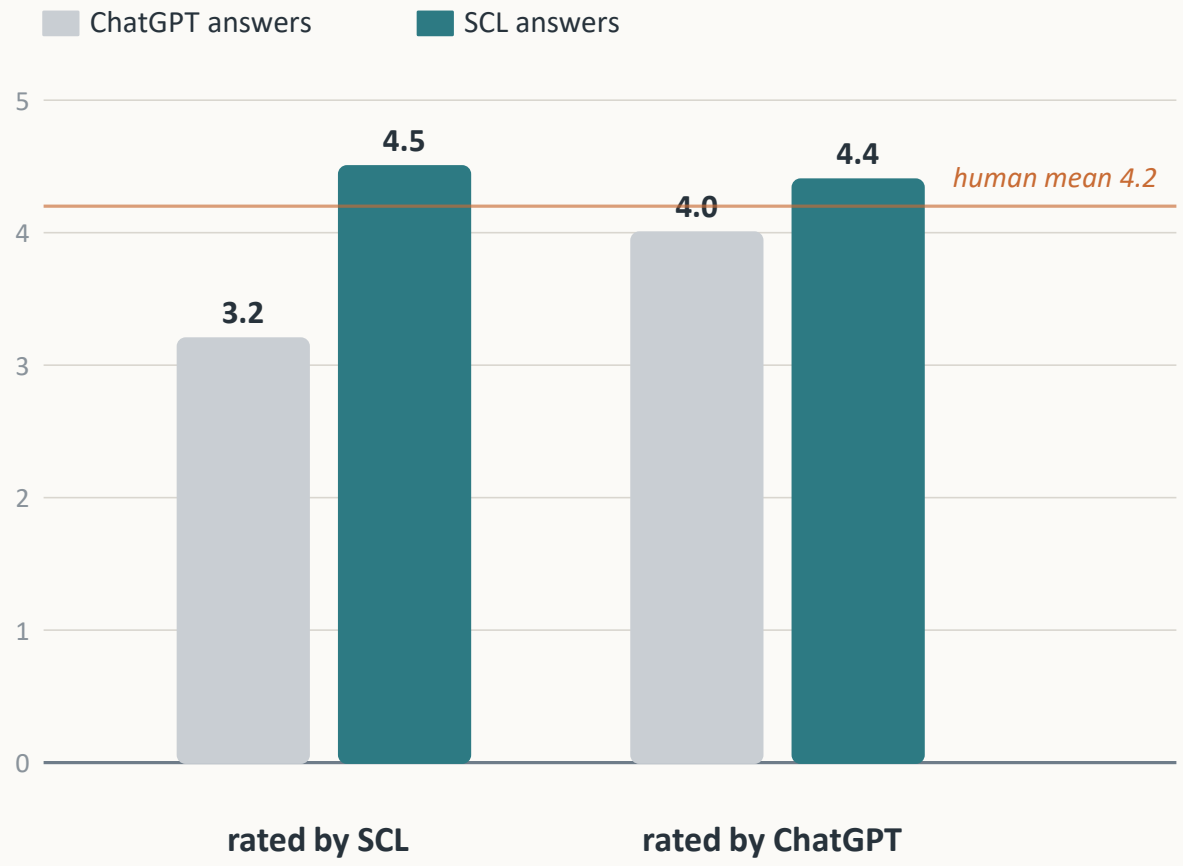
Blind attribution: 3.4

The system sounded less convincing as a stylistic impersonation.

> *"Overall, I felt that the economic content was first rate."*

Reviewers found the system more convincing as a critic than as a stylistic impersonation.

Both SCL and vanilla ChatGPT preferred SCL



Headline result

On the paper's own six-part scoring rubric, both evaluators rated SCL above vanilla ChatGPT.

ChatGPT's own comparison called SCL the "more practical blueprint" for applying Littlechild-style reasoning.

Important caveat

Fidelity and blind attribution were not fully apples-to-apples because standard ChatGPT would not answer in Littlechild's first person.

The remaining limits are about voice, audience and originality

Voice fidelity

Answers often sounded halfway between generic ChatGPT and Littlechild rather than fully like him.

Audience nuance

Responses improved when users specified whether the audience was a minister, official or academic.

Corpus bias

The gaps are largest in popular, lighter and consultancy writing — material that likely carries practical tone.

Biggest open question

Can it stand back, reframe the problem, and find a genuinely different path?

Littlechild's own worry

"I wouldn't start from here."

The deeper question is whether a persona agent can do more than apply a relatively fixed set of principles: can it stand back and ask whether the whole framing is wrong, and whether there is a better way of doing things?

A useful addition to the economist's toolbox — with guardrails

Useful now

- Rapid first-pass critique of papers, reports and consultations
- Faster synthesis of a scholar's corpus
- Perspective-taking from a defined intellectual tradition
- A way to surface questions and lines of argument that generic AI might miss

Guardrails

- Trust the corpus before trusting the answer
- Do not over-read stylistic imitation or first-person fluency
- Expect brittleness when models or configuration change
- Keep humans in the loop for audience judgment and genuinely novel reframing

Future research

The paper ends with a provocative idea: create agents grounded in different economists and let them critique each other.

A B