Artificial Societies vs. Humans & LLMs

**Predicting which LinkedIn post will win –
and why most people guess wrong**

Background

The two most common questions we hear from prospective customers are "Does it work?" and "Can I trust it?"

This report answers these questions by comparing Artificial Societies' predictions head-to-head against human evaluators and LLMs.

We assessed accuracy by analyzing which method could better identify winners between **500 pairs of LinkedIn posts** from the same authors.

Outline

1 The one-glance takeaway

Method (group)	Mean accuracy	Lift above chance
Artificial Societies	**81.8%**	**+32 pp**
Human experts (avg.)	62.9%	+13 pp
Traditional LLMs (avg.)	62.5%	+13 pp
Chance baseline	50.0%	—

2 Why this matters

Tiny creative choices can make an enormous difference in the reach and engagement of social media posts. These small decisions can translate into millions of additional views, likes, shares, or comments. **However, these outcomes only become clear after real audiences interact with the content.**

Traditionally, marketers and creators have relied on limited pre-testing methods, typically involving subjective judgments from individual panelists or evaluations from Large Language Models (LLMs). Such tests analyze copy in isolation, neglecting critical factors like the unique characteristics of the actual audience, how audience members interact with one another, and the speed at which a post gains momentum across networks.

Artificial Societies (AS) addresses these limitations by simulating a highly realistic digital replica of the author's actual audience. AS creates an interactive society of 100-300 AI personas, each carefully constructed to mirror real-world audience members, including their interests, behaviors, and interaction patterns. These personas engage with the content authentically, enabling marketers to observe how initial responses can snowball into widespread engagement through network effects. This realistic simulation helps content creators predict accurately how different creative choices will perform before posting, significantly reducing the uncertainty and guesswork associated with social media campaigns.

3 Even experts struggle

Key question

When two posts differ by at least 20 real-world reactions, how often can even an experienced human judge correctly identify the more successful post in advance?

We recruited five subject-matter experts to rate the same 200 LinkedIn posts. We converted these 200 ratings into 500 pairs of posts to evaluate each expert's success at picking winners. All have first-class pedigrees - yet none cracked the 70% mark.

Expert Background	Agreement with reality
Applied behavioural scientist	69 %
AI engineer & product builder with 10 yrs experience	69 %
Social-science MSc, University of Oxford	61 %
Computational social scientist, University of Cambridge	61 %
Dual-major in Mathematics & Computer Science, MIT	55 %

Key take-aways

Pedigree ≠ clairvoyance.
Whether trained in behavioural theory, machine-learning engineering, or quantitative social science, experts topped out at ~69% - only ~15 percentage points better than flipping a coin.

No single discipline had an edge.
The behavioural scientist narrowly led, but the engineer tied, and the deep-theory academics clustered near 60%.

The human average mirrors frontier LLMs.
Averaging all five experts yields 62.9%, essentially identical to the composite accuracy of eight state-of-the-art language models (62.5%).

This stark ceiling underlines the difficulty of the task and sets a clear benchmark: **Artificial Societies' 81.8% accuracy is not a marginal improvement - it leapfrogs both human and LLM expertise by more than 17 percentage points.**

Top LLMs still fall short

Large language models bring breathtaking scale - trillions of parameters and nine-figure training bills - yet none beat Artificial Societies on our real-world prediction task.

Model (2024–25 vintage)	Est. training cost	Av. Accuracy	Gap vs. AS
Claude 3.7 Sonnet (Anthropic)	"Tens of millions" (TechCrunch)	63.8%	-18 pp
Gemini 2.5 Pro (Google)	$191M (Voronoi App)	63.0%	-19 pp
GPT-4o (OpenAI)	$80–100M (TechRadar)	59.5%	-22 pp
GPT-3.5 Turbo (OpenAI)	$2 - 4M (Forbes)	53.5%	-28 pp

The takeaway

Training cost does not buy social context.
Even models that cost ~$100 M to pre-train evaluate each post in a vacuum. They still miss the complex peer-to-peer cascades that determine virality.

Artificial Societies wins by architecture, not sheer size.
Our simulations piggy-back on whichever foundation model is most efficient for persona reasoning, but the predictive signal comes from networked interaction, not from an ever-larger single brain.

In short, **giant LLMs are extraordinary generalists, but popularity prediction is a *collective* problem.** Artificial Societies succeeds because it treats it that way.

| Human / LLM judgement	Artificial Societies
Single rater / LLM assessment, no network context.	100–300 persona agents read, react, discuss, reshare.
Ignores amplification dynamics.	Models follower graph & cascades.
One score.	Distribution of scores → robust 0–100 rating.

5 What the numbers say

Summary statistics of agreement rate (%) with reality

Statistics	25th percentile	Median	75th percentile	Average
Artificial Societies	**77.1**	**80.0**	**85.4**	**81.8**
Best human expert	56.9	66.3	75.4	69.2
Best LLM (Claude 3.7)	55.4	66.1	81.1	64.0
All humans (avg.)	49.7	62.7	75.1	62.9
All LLMs (avg.)	46.3	63.8	78.1	62.5
Chance	—	—	—	50.0

Overall accuracy by method
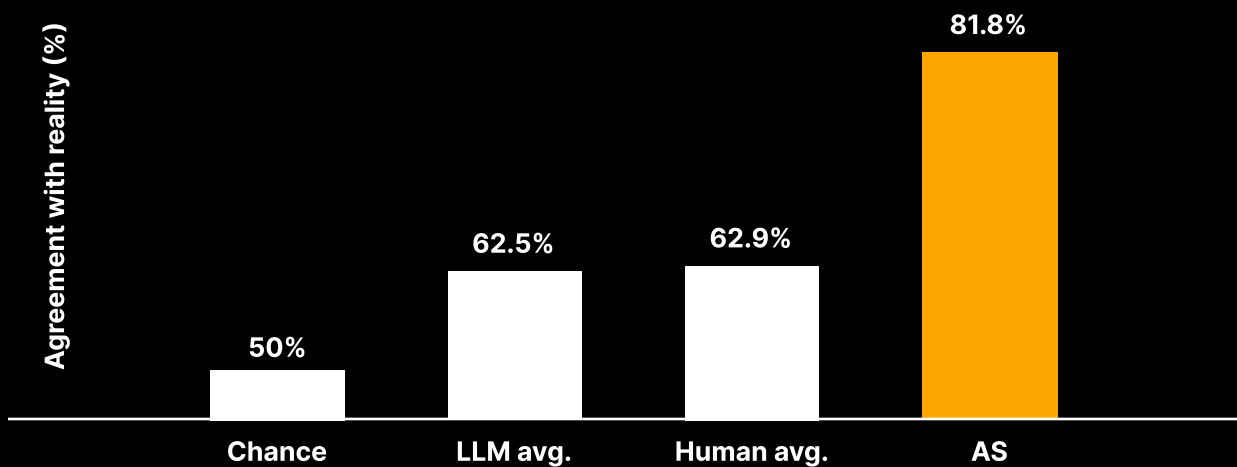
AS nails four out of every five comparisons; the rest of the field barely gets three.

6 Business impact

1. Choose winners up-front
Artificial Societies predicts social media post performance **with over 80% accuracy,** allowing marketers to confidently select the top-performing content. This reduces guesswork, saves valuable time, and prevents costly missteps.

2. Test at scale, cheaply.
Run *hundreds* of content simulations quickly and inexpensively - cheaper than a single hour of expert review. Small teams and budget-conscious companies can now afford thorough, high-quality testing.

3. See the "why."
One of the most useful features is persona conversations. It shows what different types of people might say about your post. You can see where they're confused, what they don't like, and what parts they really connect with. This helps writers and creators improve their messages to get better reactions.

4. Publish faster
Artificial Societies accelerates your content creation cycle by providing immediate, actionable feedback. Less guesswork means quicker refinement, shorter timelines, and a faster response to your audience's needs.

7 Next steps

The sample for this study focused on English-speaking tech creators on LinkedIn who have posted at least ten times. This narrow scope ensures strong data but may limit how well results apply to other types of users.

LinkedIn's algorithm also introduces platform-specific noise. To improve general usefulness, tests on other content types like news and ads are already in progress.

8 Conclusion

Predicting virality isn't a talent contest – **it's a systems problem.**

By simulating the system, **Artificial Societies more than doubles the predictive lift that humans or top LLMs achieve alone**, transforming guesswork into evidence-backed creative decisions.

Test Anything in an Artificial Society

Try free at societies.io

Prepared for prospective clients, May 2025