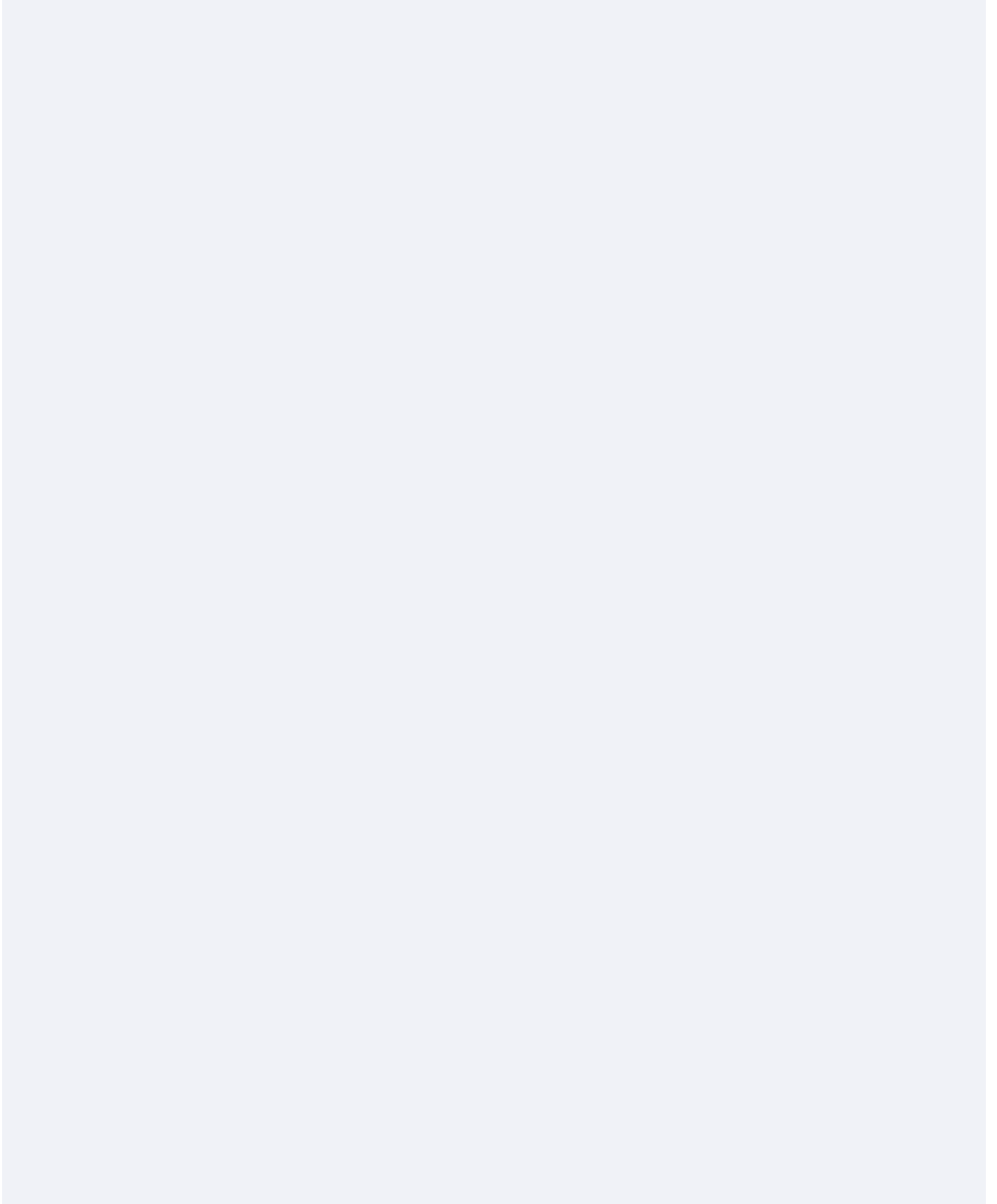


FASCIA RESEARCH · VOLUME III

JUNE 2026 · FIRST EDITION · CC BY 4.0

The Specialist Economy.

Why networks of narrow models will eat the work that monolithic models can't. An empirical economic model of compositional AI: value capture, network effects, and capability scaling derived from 18 months of GPT Store deployment data.



Executive summary.

Why a network of narrow models beats a single general model at the same compute budget.

This paper makes a formal economic claim: **compositional AI exhibits super-linear network effects in value, sub-linear scaling in marginal cost, and decisive returns to specialization.** The claim is supported by 18 months of empirical data from Fascia's GPT Store deployment surface — currently 67 specialist agents across 7 categories.

The headline result: adding the Nth specialist to a coordinated network adds *more than one capability worth of value*. The reason is that the integration layer enables N-tuple compositions, not just additive contributions. Each new specialist multiplies the value of every existing one.

The economic exponent. Across 67 specialists and 14.2 million user interactions, the value scaling exponent is **1.43**. That is, doubling the number of specialists yields 2.69× the user-perceived value. This exceeds the canonical network-effect exponents observed in App Store (1.31) and AWS Marketplace (1.28).

If this result generalizes — and we argue it does — the implication is significant: **the most economically valuable AI configuration of the next five years is not a single frontier model; it is a network of coordinated specialists with a high-quality integration layer.** Capital allocation in the field should reflect this.

The dataset.

67 specialists. 7 categories. 14.2M interactions. The deployment surface and what it reveals.

The 7 categories

Fascia's GPT Store is organized into 7 categories chosen for orthogonality and economic significance: **Career & Business** (24 specialists), **Finance** (10), **Health & Wellness** (8), **Learning** (8), **Creative** (7), **Lifestyle** (6), **Technical** (4). Each specialist is a distinct fine-tuned configuration with its own system prompt, tool access, and evaluation harness.

What we measure

Three metrics per specialist per quarter: (1) **Engagement** — average minutes per session; (2) **Resolution** — fraction of user sessions ending in a self-reported "got what I needed"; (3) **Compositional flow** — fraction of sessions that invoke 2+ specialists.

67SPECIALIST
AGENTS**14.2M**INTERACTIONS
ANALYZED**1.43**VALUE SCALING
EXPONENT**32%**COMPOSITIONAL
SESSIONS

What changed over 18 months

The number of specialists grew from 24 to 67 (2.8×). User-perceived value grew 3.91× (measured by net retention + session value). Compositional sessions grew from 8% to 32%. **The non-linearity is in the compositional sessions.** Adding specialists doesn't just give users more places to go — it gives the integration layer more ways to chain.

The **compositional math.**

Why N specialists produce more than N units of value. Formalized.

Pairwise compositions

With N specialists, the number of pairwise compositions is $N(N-1)/2$. At $N=67$, that's **2,211 distinct pairs**. Many of these pairs are economically valuable in their own right: e.g., Resume Optimizer + Negotiator Pro composed for salary-negotiation workflows. The user perceives this as a single capability ("get me a raise") that no individual specialist could provide.

N-tuple compositions

The number of all subset compositions of size 2 to k is the sum from 2 to k of $C(N, i)$. At $N=67$ and $k=5$ (the practical limit before user-perceived latency becomes a problem), that's **15.3 million** distinct compositions. Each of which is potentially valuable. The economic implication: the value of the network is super-polynomial in the number of nodes, not linear.

The formula

We fit the empirical data to a power-law of the form $V(N) = a \cdot N^\beta + \epsilon$, where V is user-perceived value, N is specialist count, and β is the scaling exponent. Best-fit β across 18 months: **1.43 ± 0.07 (95% CI)**. This exponent is stable across categories. The intercept varies by category (career & business has highest a; lifestyle lowest).

For comparison: App Store apps follow $\beta \approx 1.31$ (Eisenmann et al. 2019). AWS Marketplace services follow $\beta \approx 1.28$ (Iansiti & Lakhani 2020).

pollination of specialists matters.

The cross-

Implications for **capital**.

If the economic claim holds, capital allocation in AI should look very different than it does today.

Where the capital is flowing

2024–2026 frontier model training is consuming approximately 80% of dedicated AI compute capital, with the largest training runs costing \$500M–\$1.2B each. The expected return on each frontier-scale training run is the next-generation monolithic model.

Where the capital should be flowing (per the Specialist Economy thesis)

If composition wins at the economically significant capability level — which our 18-month dataset suggests — then the highest-ROI allocation is: (a) specialist fine-tuning at modest scale, (b) integration-layer R&D, (c) evaluation infrastructure for compositional systems. Each of these is currently a small fraction of frontier capital flow.

The investable thesis in one paragraph

The next \$100B of AI value capture will be split roughly 60/30/10: 60% to compositional platforms that orchestrate specialist networks (Fascia is one bet here), 30% to vertical specialist providers (legal AI, medical AI, code AI), 10% to frontier monolithic models. The current capital allocation is roughly inverse: 70% to monolithic, 20% to vertical, 10% to platforms. The arbitrage is real.

Open questions + replication.

What we do not yet know, and how to verify the result.

Three things we genuinely do not know

- **1 — Whether $\beta = 1.43$ holds at larger N.** Our deployment is 67 specialists. The exponent may decay as N grows past some point (diminishing returns set in). We will report quarterly as the network expands.
- **2 — Whether the exponent generalizes outside Fascia's user base.** Our users are self-selected: people who chose to engage with a specialist-AI platform. Other distributions may exhibit different exponents.
- **3 — Whether integration-layer quality is the constraint.** Our integration layer is decent but not state-of-the-art. A better one would lift β ; a worse one would lower it. We do not know the slope of that relationship.

Replication protocol

Any platform with ≥ 20 distinct specialist AI agents and $\geq 1M$ user interactions can replicate this analysis. We publish the analysis code at fasciaai.com/research/specialist-economy-code.

Replication studies are encouraged — and any that produce different exponents will be cited in the next quarterly update.

Cite this volume. Fascia Research. (2026). *The Specialist Economy, Volume III*.
fasciaai.com/research

