TRENDS – ARTIFICIAL INTELLIGENCE (AI)

May 30, 2025

Mary Meeker / Jay Simons / Daegwon Chae / Alexander Krey
As seen by ben@ibenps.com

Original report https://www.bondcap.com/reports/tai

Anyone will appreciate the depth and foresight in this BOND report. It's a comprehensive snapshot of a hyper-dynamic field, and while much of it confirms what those deep in AI already sense, the aggregated data and specific inflection points highlighted are crucial. Let's dissect the key takeaways and learnings for someone with your expertise.

OVERALL THESIS OF THE REPORT:

The central argument is that Artificial Intelligence is not just another tech cycle; it's a fundamental accelerant of change, developing at a pace and scale that is "unprecedented". This acceleration is built upon decades of digital infrastructure, massive datasets, and recent breakthroughs in Large Language Models (LLMs), notably the user-friendly interface of OpenAI's ChatGPT. The report meticulously charts this through user/usage growth, capital expenditure, cost dynamics, competitive intensity, and the increasing fusion of AI with the physical world and global geopolitics.

Here's a detailed breakdown of the key sections and their implications for an AI expert:

SECTION 1: SEEM LIKE CHANGE HAPPENING FASTER THAN EVER? YES, IT IS

- **Core Learning:** The velocity of AI's evolution and adoption outstrips previous tech waves, including the internet. This isn't just perception; quantitative data on multiple fronts supports this.
- **Key Data Points & Insights for Experts:**
    - **Compounding Effects:** The report emphasizes technology's compounding nature. For AI specifically:
        - Training Dataset Sizes (Words): +260% per year (1950-2025, focused on the last 15 years).

        - Training Compute (FLOPs): +360% per year (1950-2025, focused on the last 15 years).
        - Algorithmic Progress (Effective Compute): +200% per year (2014-2023).
        - AI Supercomputer Performance: +150% per year (2019-2025), enabled by 1.6x annual growth in chips per cluster and 1.6x annual growth in performance per chip.
        - Number of New Large-Scale AI Models: +167% per year (2020-2024).
    - **ChatGPT vs. Google Search Adoption:** ChatGPT hit 365 billion annual searches in 2 years (2024), whereas Google Search took 11 years (2009) to reach the same milestone, making ChatGPT 5.5x faster in this regard.
    - **Knowledge Distribution Evolution:** AI marks a new epoch (Active + Digital + Generative Delivery) after the Printing Press (Static + Physical) and the Internet (Active + Digital). Generative AI is already contributing significantly to scientific publications (7% in 2023) and news articles (6.96% globally by mid-2024).
    - **AI Milestones & Future Capabilities:** The timeline from 1950 to 2025 shows accelerating breakthroughs. ChatGPT's self-assessment of current capabilities (Q2 2025) and its predictions for 2030 (e.g., human-level text/code, full-length film/game generation) and 2035 (e.g., autonomous scientific research, operating autonomous companies) provide a tangible, albeit speculative, roadmap.
    - **AI Development Shift:** Industry surpassed academia in producing notable machine-learning models around 2015, driven by needs for massive data, compute, and financial resources. In 2024, industry produced 55 notable models, while academia produced 0.
    - **Developer Ecosystem Growth:** NVIDIA's ecosystem grew 6x to 6MM developers in seven years (2018-2025). Google's Gemini developer base grew 5x to 7MM Y/Y by May 2025.
    - **AI Performance Surpassing Humans:**
        - On the MMLU benchmark, AI systems reached 92.30% accuracy in 2024, surpassing the human baseline of 89.8%.

        - In Q1 2025, GPT-4.5 (with persona) had a 73% "AI Win Rate" in a Turing Test variation, meaning testers mistook its responses as human.
    - **Benefits vs. Risks:** The report acknowledges Stuart Russell and Peter Norvig's framing: AI could free humanity from menial work and solve major global issues. However, risks from misuse (lethal autonomous weapons, surveillance, bias, job impact) are already apparent or likely.

SECTION 2: AI USER + USAGE + CAPEX GROWTH = UNPRECEDENTED

- **Core Learning:** The metrics for AI adoption, engagement, and the capital pouring into it are unmatched by any previous technology cycle at this stage.
- **Key Data Points & Insights for Experts:**
    - **User Growth & Global Spread:**
        - ChatGPT weekly active users: +8x to 800MM in 17 months (Oct 2022 - Apr 2025).
        - ChatGPT app reached 90% of users outside North America by Year 3; the internet took 23 years.
        - Years to 100MM users: ChatGPT (0.2 years) vs. Instagram (2.5 years), Facebook (4.5 years), Netflix (10.3 years).
        - Time to 50% US household adoption: AI Era projected at 3 years, vs. Mobile Internet (6 years), Desktop Internet (12 years), PC Era (20 years).
    - **Ecosystem Adoption (NVIDIA as Proxy):**
        - Developers: 2.5MM (2021) to 6MM (2025) [+2.4x].
        - AI Startups: 7K (2021) to 27K (2025) [+3.9x].
    - **Enterprise AI Adoption:**
        - S&P 500 firms mentioning "AI" in earnings calls: 50% by Q4 2024, up from <5% in 2015.

        - Global enterprises primarily target GenAI for revenue-focused improvements (Production/Output, Customer Service, Sales Productivity, Revenues) over cost reduction.
        - 75% of Global CMOs were using/testing AI tools in 2024.
        - Specific examples: Bank of America's Erica (2.5B cumulative interactions by Feb 2025); JP Morgan's AI/ML initiatives projected to drive +65% value Y/Y in 2025; Kaiser Permanente's Ambient AI Scribe (over 2.5M uses by Dec 2024).
    - **CapEx Details:**
        - Big Six US Tech CapEx: $212B in 2024 (+63% Y/Y), up from $30B in 2014. CapEx as % of revenue for these companies rose to 15% in 2024 from 8% ten years prior.
        - Data center construction: US annualized private construction value for data centers grew +49% Y/Y by Dec 2024. xAI's Colossus data center (750,000 sq ft) was built in 122 days, housing 200,000 GPUs within seven months of starting.
        - Data centers accounted for ~1.5% of global electricity consumption in 2024, with demand growing 12% per year since 2017.

SECTION 3: AI MODEL COMPUTE COSTS HIGH / RISING + INFERENCE COSTS PER TOKEN FALLING = PERFORMANCE CONVERGING + DEVELOPER USAGE RISING.

- **Core Learning:** A critical dichotomy is emerging: training costs are astronomical and escalating, while inference costs per unit are collapsing. This creates a complex economic environment for AI providers and a boon for developers/users.
- **Key Data Points & Insights for Experts:**
    - **Training Cost Escalation:** Estimated training cost for frontier AI models grew ~2,400x over eight years (2016-2024). Anthropic's CEO Dario Amodei noted (June 2024) training costs are at $100M, with $1B models in training, and $10B models potentially starting training in 2025.
    - **Inference Cost Collapse:**

        - Energy per LLM token (NVIDIA GPUs): -105,000x decline from Kepler (2014) to Blackwell (2024).
        - AI inference price for customers (per 1M tokens): 99.7% lower over two years (Nov 2022 - Dec 2024). This is faster than cost declines for electric power or computer memory historically.
    - **Performance Convergence:** The performance gap (e.g., on LMSYS Chatbot Arena scores) between the top-performing AI models is narrowing.
    - **Developer Proliferation:** This cost collapse and performance accessibility are democratizing AI development.
        - 63% of professional developers were using AI in development processes in 2024, up from 44% in 2023.
        - AI developer repositories on GitHub saw a ~175% increase in 16 months (Nov 2022 - Mar 2024).
        - Google processed 480 trillion tokens monthly across products/APIs by May 2025, up 50x Y/Y. Microsoft Azure AI Foundry processed over 100 trillion tokens in Q1 2025 (ending Apr 2025), up 5x Y/Y.
    - **Jevons Paradox for AI Compute:** Improvements in compute efficiency (cost per token, energy per token) are leading to massively increased overall usage and thus still driving up total compute and energy demand.

SECTION 4: AI USAGE + COST + LOSS GROWTH = UNPRECEDENTED (BUSINESS MODELS & MONETIZATION).

- **Core Learning:** The AI sector exhibits classic tech disruption patterns: early euphoria, massive capital influx, intense competition, and significant early losses for many, with eventual consolidation around winners. The current "make it up on volume" and "monetize later" approach is high-stakes.
- **Key Data Points & Insights for Experts:**
    - **Historical Analogs:** The report references Alasdair Nairn's 'Engines That Move Markets,' highlighting parallels with past tech bubbles (e.g., railways), emphasizing risks of high CapEx and the importance of barriers to entry.
    - **Monetization Ramps (Examples):**

        - NVIDIA: Quarterly revenue +78% Y/Y to $39B (as of Jan 2025); revenue +28x over 10 years.
        - Custom Silicon: Google TPU estimated sales +116% Y/Y to $8.9B (2024); Amazon AWS Trainium estimated sales +216% Y/Y to $3.6B (2025).
        - Compute Services: CoreWeave revenue +730% Y/Y to $1.9B (2024).
        - Foundation Models: OpenAI revenue +1050% annually to $3.7B (2024); Anthropic annualized revenue +20x to $2B in 18 months (by Mar 2025); Perplexity annualized revenue +7.6x to $120MM in 14 months (by May 2025).
        - Specialized AI Software: Anysphere (Cursor AI code editor) ARR from ~$1MM to $300MM in 25 months; Lovable (no-code product building) ARR +13x to $50MM in 5 months.
    - **Cost vs. Revenue (Incumbents & New Entrants):**
        - OpenAI: 2024 revenue est. $3.7B, compute expense est. $5B.
        - Big Tech (MSFT, AMZN, GOOG, META): Showed increased CapEx (2023-2024) alongside decreased or flat Free Cash Flow Margins.
    - **Valuation Multiples:** Private AI model companies show high valuation-to-revenue multiples (e.g., OpenAI 33x, Anthropic 31x, Perplexity 75x as of their latest funding rounds prior to May 2025). OpenAI's EV/NTM Revenue multiple is significantly higher than public tech comparables like Meta or Alphabet.
    - **Tipping Points:** The report identifies ChatGPT (Nov 2022) and NVIDIA's A100 GPU (2020) as key AI tipping points, similar to Netscape's IPO for the Internet or the iPhone for mobile.

SECTION 5: AI MONETIZATION THREATS = RISING COMPETITION + OPEN-SOURCE MOMENTUM + CHINA’S RISE.

- **Core Learning:** The path to sustained AI monetization is challenged by intense and growing competition from various angles: a proliferation of proprietary models, the rapid advancement and adoption of open-source alternatives (often

with lower costs and increasing capability), and a concerted strategic push from China.

- **Key Data Points & Insights for Experts:**
    - **Model Proliferation:**
        - Large-scale multimodal models: +1,150% rise in releases over two years (2023-2025, as of May).
        - Large-scale language models: +420% increase in releases over two years.
        - Website visits show OpenAI's ChatGPT leading significantly, but DeepSeek and xAI's Grok are rising fast.
    - **Open-Source vs. Closed-Source Dynamics:**
        - Closed models (OpenAI, Anthropic) dominate consumer MAU share currently.
        - Closed models generally involve higher compute investment for training.
        - However, the performance gap is closing, with open-source models from China (e.g., DeepSeek R1) showing near-parity with leading closed models on benchmarks like MATH Level 5.
        - Meta's Llama (open-source) downloads: +3.4x in eight months, reaching 1.2B by April 2025, with thousands of derivative models.
    - **China's AI Trajectory:**
        - Leads in cumulative large-scale AI system releases, outpacing the USA in 2024.
        - Rapid relevance of models like DeepSeek R1, Alibaba Qwen 2.5-Max, Baidu Ernie 4.5 Turbo, with claims of competitive or superior performance to Western counterparts and significantly lower costs.
        - Huawei is delivering advanced AI chip "clusters" to Chinese clients, mitigating US export restrictions.
        - China's installed base of industrial robots (300K+ in 2023) is higher than the rest of the world (excluding USA) combined.

        - Desktop user share: Leading USA LLM #1 has 60% share, while a leading China LLM has 21% and USA LLM #2 has 10% as of April 2025.
        - Top Chinese AI platforms (Doubao, DeepSeek, Kimi) have tens of millions of MAUs within China.

SECTION 6: AI & PHYSICAL WORLD RAMPS = FAST + DATA-DRIVEN.

- **Core Learning:** The integration of AI into physical systems is rapidly moving from experimental to revenue-generating deployments, creating vast new datasets and opportunities for AI-native infrastructure.
- **Key Data Points & Insights for Experts:**
    - **Autonomous Vehicles:**
        - Tesla: Cumulative Full Self-Driving (FSD) miles grew from ~40M in June 2022 to ~4B by March 2025 (approx. 100x increase). Tesla FSD v12 is described as an "end-to-end artificial intelligence" system, replacing 330,000 lines of C++ code with neural nets.
        - Waymo: Achieved 27% market share of San Francisco rideshare gross bookings by April 2025, up from 0% in August 2023, competing with Uber and Lyft.
    - **Vehicle Intelligence Software:** Applied Intuition serves 18 of the top global auto OEMs by 2024 (up from 0 in 2016), providing AI-powered tools, autonomy software, and vehicle OS for automotive, trucking, construction, and defense.
    - **Defense:** Anduril is reported to have achieved +2x Y/Y revenue growth for the last two years (F2023, F2024), reaching an estimated $1B in F2024 revenue.
    - **Specialized Industries:** KoBold Metals uses AI to reverse the trend of declining mineral deposit discoveries per $B of exploration spend. Carbon Robotics' AI-driven LaserWeeders have weeded 230K+ acres by May 2025. Halter’s AI-driven intelligent grazing collars saw +150% Y/Y growth in net-new contracts in 2025 (annualized).

SECTION 7: GLOBAL INTERNET USER RAMPS POWERED BY AI FROM GET-GO = GROWTH WE HAVE NOT SEEN LIKES OF BEFORE.

- **Core Learning:** The 2.6 billion people globally still offline represent a massive new user base that will likely be AI-native from their first internet experience, potentially leapfrogging traditional web/app interfaces.
- **Key Data Points & Insights for Experts:**
    - Global internet penetration reached 68% in 2024, up from 16% in 2005. However, disparities remain: 83% urban vs. 48% rural penetration globally. Regions like South Asia and Sub-Saharan Africa have lower penetration rates (below 50%) compared to North America and Europe & Central Asia (above 90%).
    - Global internet users reached 5.5B in 2024, with Y/Y growth accelerating to +6%.
    - ChatGPT mobile app MAUs showed rapid global distribution: 530MM by Apr 2025. Top user countries include India (13.5%), USA (8.9%), Indonesia (5.7%), Brazil (5.4%).
    - DeepSeek mobile app MAUs reached 54MM in 4 months, with 34% of users in China and 9% in Russia.
    - Satellite Internet (SpaceX Starlink): Achieved 5M+ subscribers by Feb 2025, showing +202% annual growth over 3.2 years, expanding global coverage.

Section 8: AI & Work Evolution = Real + Rapid.

- **Core Learning:** AI is not just a tool but a catalyst for rethinking work processes, job roles, and productivity. Human roles may shift towards AI oversight, guidance, and training.
- **Key Data Points & Insights for Experts:**
    - **Corporate Mandates:** CEOs like Tobias Lütke (Shopify) and Luis von Ahn (Duolingo) have issued internal memos making AI proficiency a fundamental expectation, impacting hiring, performance reviews, and resource allocation.
    - **AI Adoption by Firms:** ~7% of US firms were using AI in Q1 2025, a +21% Q/Q increase.

- **Job Market Shifts:**
  - USA AI job postings: +448% (Jan 2018 - Apr 2025), while non-AI IT job postings: -9%.
  - AI-related job titles saw a +200% increase globally over two years (Q2 2022 - Q2 2024).
- **Productivity:** Studies show AI assistance increasing customer support agent productivity by +14% in hourly chats.
- **Long-term Labor Productivity:** Despite technological shifts (PC, Desktop Internet, Mobile Internet), US non-farm employment and labor productivity have both trended upwards over the past 77 years.

Overarching Key Takeaways & Learnings for AI Experts:

1. **Velocity & Compounding are Real:** The speed of development and adoption is not hype. Multiple underlying technological advancements are compounding to create this acceleration. Experts need to recalibrate their expectations for iteration cycles and market penetration.
2. **The Economic Paradox:** The massive capital investment required for frontier model training contrasts sharply with the plummeting cost of inference. This will likely lead to significant market restructuring, favoring those who can manage this dichotomy or find defensible niches. Commoditization at the model level seems plausible for many use cases.
3. **Data as the New Constant:** While models and compute are in flux, the insatiable need for diverse, high-quality data for training and fine-tuning remains a constant driver of value and a potential bottleneck.
4. **Geopolitical AI Race is Definitive:** The US-China dynamic is not a sideshow but a central theater for AI development and deployment. National strategies, sovereign AI initiatives, and efforts to control key technological inputs (chips, data) will profoundly shape the global AI landscape. China's rapid progress, particularly in open-source and specific industrial applications, warrants close attention.
5. **The "Agentic" Shift:** The evolution from responsive chatbots to proactive AI agents capable of multi-step task execution (both digital and physical) represents a significant paradigm shift. Experts should be tracking the development of agentic AI and its potential to reshape user interfaces and automate complex workflows.

6. **Monetization is the Riddle:** Despite explosive user growth and advanced capabilities, sustainable monetization models for many AI companies (especially foundation model providers) are still emerging and face threats from intense competition and open-source alternatives. The "make it up on volume" strategy is being tested at an unprecedented scale.
7. **Physical World Integration is a Major Frontier:** AI's application in robotics, autonomous systems, and specialized industries (agriculture, mining, defense) is creating new value chains and demonstrating tangible ROI, potentially offering more defensible business models than purely digital AI applications.
8. **Human-AI Collaboration is the Future of Work:** The narrative is shifting from AI replacing humans to humans augmented by AI. Expertise will involve leveraging AI tools effectively, and new job categories focused on AI training, oversight, and integration will emerge.

This report serves as a critical data-backed foundation for understanding the current AI inflection point. For an expert, the value lies not just in the individual data points, but in synthesizing them to anticipate the strategic, economic, and technological reconfigurations that AI is driving globally.