# AI Thematic Investing in 2024: The Power of AI-Enabled Patent Analysis as a Factor

Mark Marex, CFA, *Senior Director*, Nasdaq Index Research & Development

The Nasdaq Global Artificial Intelligence and Big Data™ Index (NYGBIG™) was launched on November 12, 2018, using a cutting-edge approach to index construction designed to track a cutting-edge investment theme. By harnessing the power of AI itself to screen millions of patent filings each year – a "big data" undertaking, in its own right – NYGBIG™ acts as a sort of "meta" benchmark. In other words, it leverages some of the same disruptive technologies to assemble its targeted exposures that it aims to ultimately provide to tracking investors. As a result, NYGBIG™ has captured exposure to the latest key evolution in the realm of artificial intelligence: generative AI.

Since the wave of generative AI launches in the fall of 2022 – including the highly impactful launch of ChatGPT – NYGBIG™ has outperformed several leading competitor benchmarks, the broader US equity market, and even the tech-heavy Nasdaq-100 Index®. Upon investigating what's been driving this outperformance story, the high-level takeaway is something that equity investors have been grappling with throughout much of the 21st century. Namely, that in the "new economy" that has emerged since the popping of the Internet bubble, economic growth is being disproportionately driven by technological innovation; innovation-driven growth is fueling stronger fundamental outcomes for companies; and innovation-driven companies are receiving most of the spoils of the equity market in terms of price performance and market cap growth.

Price Return Performance vs. Competitors, Sept '22 - Aug '24
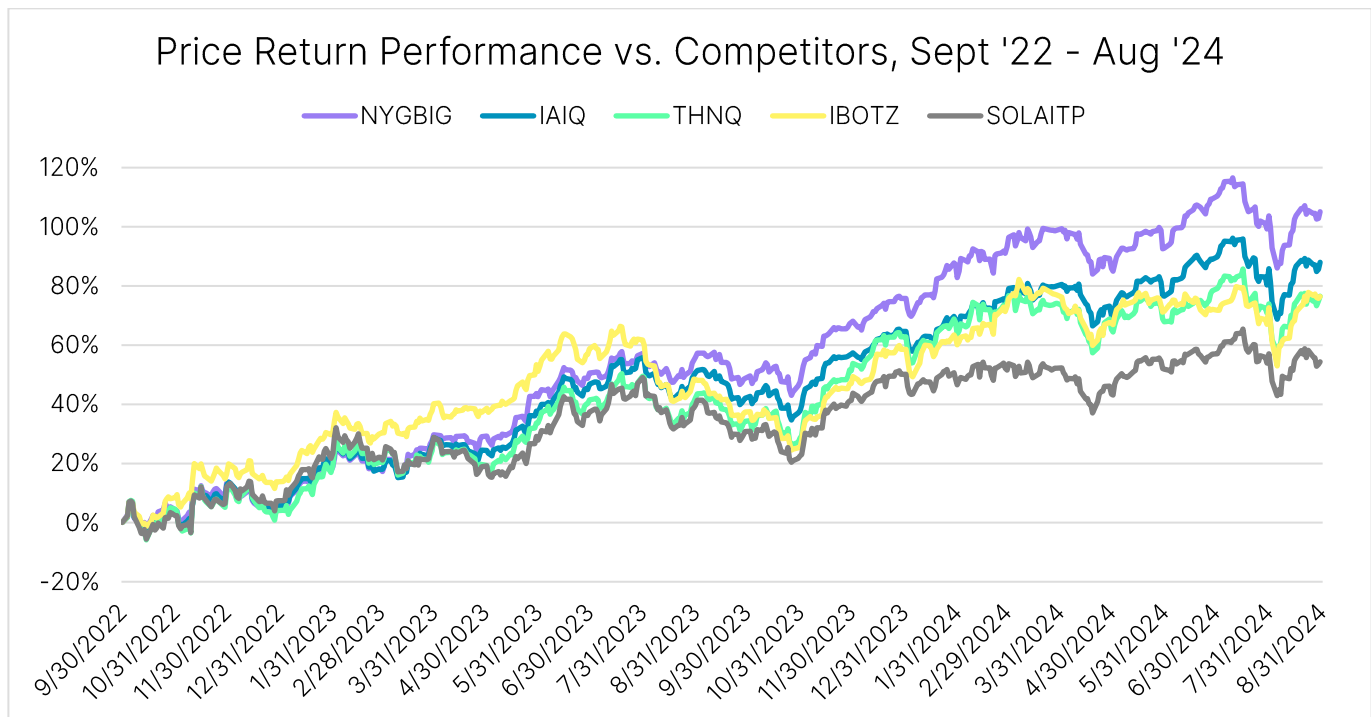
Source: Bloomberg. IAIQ = Indxx Artificial Intelligence and Big Data Index; IBOTZ = Indxx Global Robotics & Artificial Intelligence Thematic Index; THNQ = ROBO Global Artificial Intelligence Index; SOLAITP = Solactive Artificial Intelligence Index

On a more granular level, it is becoming evident that quantitative measures of innovation may offer investors a factor-like explanation of past outperformance, as well as a signaling mechanism for future outperformance. Research & development spending is one metric available to investors to screen companies for their innovation intensity, but it is directly observable on most companies' financial statements and may not offer much of an edge, in the sense of traditional investment alpha. Patent filings, on the other hand, are much more unstructured, can be challenging for investors to classify and analyze, and are expensive to aggregate for large investment universes, such as the one used for NYGBIG™ and its parent index, the Nasdaq Global Disruptive Tech Benchmark™ Index (NYDTB™).

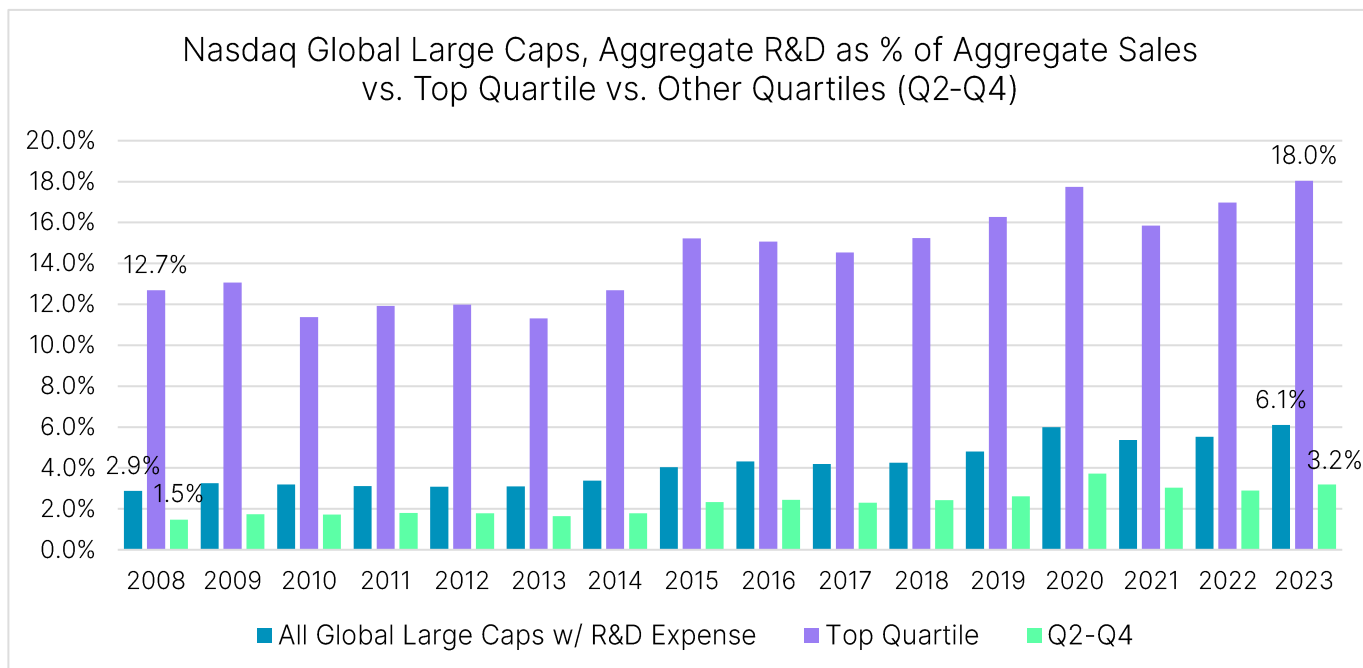## Overview of Nasdaq's AI-Enabled Index Methodology

NYGBIG™ is designed to track the performance of companies that are most active in filing patents relating to: Deep Learning, Natural Language Processing (NLP), Image Recognition, Speech Recognition & Chatbots, Cloud Computing, Cyber Security, and Big Data. These seven disruptive technologies ("sub-themes") were initially identified as most directly relevant to the advancement of the AI & Big Data investment theme. As of the most recent reconstitution in July 2024, there were 1,858 global large and midcap companies in the parent index, NYDTB™, that filed at least one patent in the most recent 12-month period analyzed ending May 31, 2024, relating to one or more of 35 disruptive technologies. Of these, 758 companies filed at least one patent relating to the seven AI & Big Data sub-themes; only 87 of these met all the index criteria listed in the NYGBIG index methodology, including ranking highly on both patent Contribution Scores (i.e., % of all patents filed for a given sub-theme) and patent Pure Scores (i.e., % of a given company's aggregated patent filings that relate to a given sub-theme) [1].

As a first, and arguably most important step in the index construction process, Nasdaq's in-house AI team processes ~500,000 patents filed by ~5,000 companies using a rolling 12-month lookback period, on a quarterly basis, querying an external patent database maintained by IFI Claims. The Nasdaq AI team developed a state-of-the-art NLP model to digest the entirety of each patent's content, estimated to total five million pages of text. In order to properly score patents for each of the 35 disruptive technologies in scope for the model, approximately 18 million scoring assessments are performed each quarter. The NLP model was developed over the course of more than a year of testing, including several years of historical patent data, in order to properly calibrate thresholds to determine patent relevance. Once patents are appropriately classified, individual company Contribution Scores and Pure Scores are calculated to feed directly into a quarterly index evaluation process for NYDTB™, and a corresponding semiannual process for NYGBIG™ which is overlaid with the rest of its unique index methodology criteria.
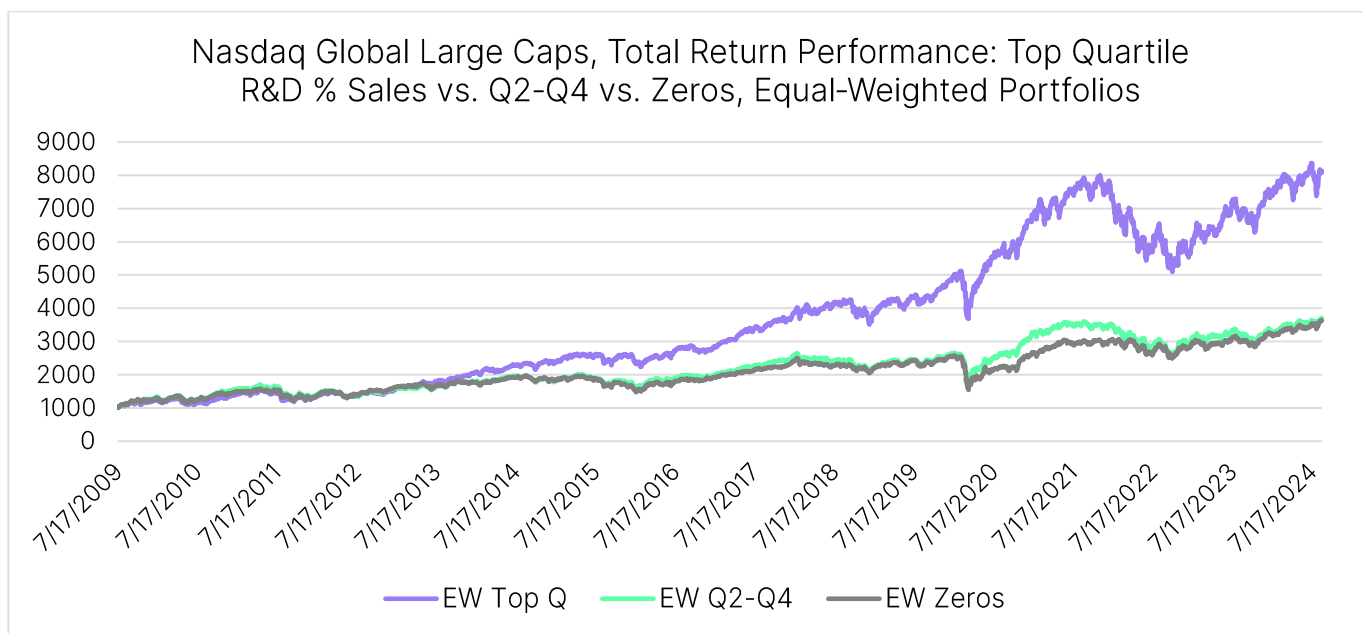
Because patents represent unstructured alternative data, effectively analyzing and classifying them could provide an edge for a systematic investment strategy that is seeking targeted exposure to companies that are leading the innovation race to develop new, groundbreaking technologies. Especially for an area like artificial intelligence, there may be drawbacks to other systematic approaches such as counting mentions of relevant keywords in company earnings calls or financial reports, given an implicit bias for executive management to generate excitement around a level of engagement with the technology that may be more aspirational than tangible. Alternatively, quantifying a company's revenue exposure to the theme may be especially challenging in the crucial early stages of a new breakthrough, when even the company's managers may have difficulty disaggregating the proportion of "AI revenue" derived from a package of software offerings. In yet other examples, such as with a leading AI patent filer like Bank of America, AI tools may be developed purely from a non-commercial perspective to limit fraud, contain cyberthreats, or improve customer service, never hitting the revenue line; they will, however, show up eventually in improved margins, cash flows, and earnings-per-share. From a thematic purity perspective, tracking the companies that are most intensely focused on developing and exploiting the technology, no matter where the impact occurs on their financial statements, seems like a fair and principled approach.

---

[1] See page 3 of index methodology for more info on Score calculations: https://indexes.nasdaqomx.com/docs/methodology_NYGBIG.pdf

Quantifying Innovation: R&D vs. Patents

Nasdaq Global Large Caps, Aggregate R&D as % of Aggregate Sales vs. Top Quartile vs. Other Quartiles (Q2-Q4)

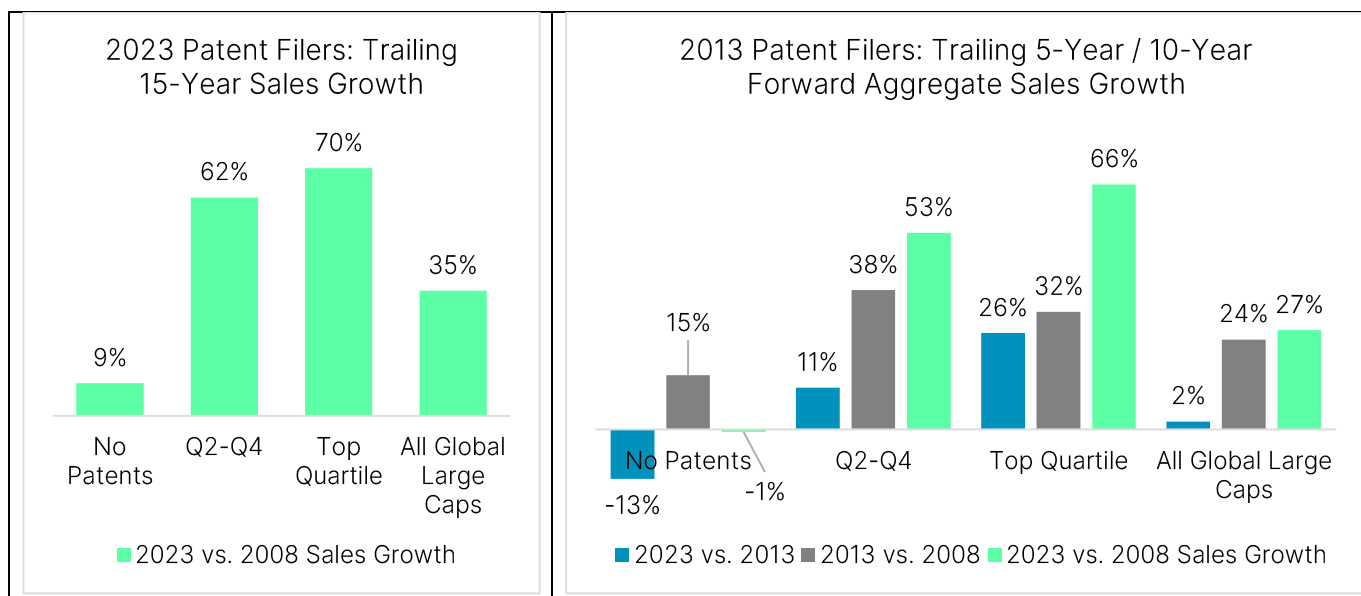Legend: All Global Large Caps w/ R&D Expense | Top Quartile | Q2-Q4

Investors seeking exposure to companies that excel in innovative areas like artificial intelligence should consider just how important research & development (R&D) activities – which include, but are not limited to patent development – have become to the success of large caps in the past decade and a half. In 2008, Nasdaq's benchmark of global large caps (filtered for those with nonzero R&D expense) spent 2.9% of their revenues on R&D; by 2023, this ratio had more than doubled to 6.1%. By scoring these companies on the ratio of their R&D to Sales, we can zoom in to see how concentrated this intensification of R&D has become: the top-ranked quartile of companies increased their R&D as a percent of Sales from 12.7% to 18.0%, while the remaining 75% of companies grew from only 1.5% to 3.2%. Simply put, R&D spending is now one of the largest line items in the income statement for many large global large caps. In 2023, 421 of 1,015 companies in the Nasdaq Global Large Cap™ Index reported some level of R&D expense. Aggregate R&D spending has tripled across the benchmark, from $342B in 2008 to ~$1T in 2023.

Nasdaq Global Large Caps, Total Return Performance: Top Quartile R&D % Sales vs. Q2-Q4 vs. Zeros, Equal-Weighted Portfolios

Legend: EW Top Q | EW Q2-Q4 | EW Zeros

The biggest spenders have demonstrated strong sales growth, with topline revenues up 115% over the period for the top quartile of R&D/Sales companies vs. only 27% for the remaining quartiles (Q2-Q4). This ~4x difference has driven substantial price outperformance as well. An equally-weighted portfolio of the top quartile of R&D/Sales companies, rebalanced annually in July based on prior calendar year financials, would have generated total returns of 713% vs. only 271% for the remaining three quartiles. Interestingly enough, the performance of a portfolio of companies with zero R&D spending underperformed at almost the same level, gaining only 263%. This suggests that it is not enough for companies to spend an average proportion of revenues on R&D to outperform non-spenders in today's global equity markets; they must be above-average.

As mentioned earlier, R&D is directly observable on company financial statements, which may not provide investors much of an edge with a strategy that uses some normalized version of R&D as a ranking and/or weighting factor. R&D spending includes activities such as patent development, but it also includes other intangibles such as non-patented product development and refinement, internally-developed software, literal hiring of research & development staff, cost of materials and facilities, and acquisition of external data resources. Some of this spending may contribute to patent development, but it is impossible to estimate the ratio ex-ante across different companies, as they will differ in their patent output based on a number of variables such as their industry, their legal jurisdiction, their historical experience and success rate, etc. This becomes even more difficult for investors who are not just interested in tracking companies that are broadly innovative, but those who are intensely involved in a particular theme like artificial intelligence.

When ranking the same universe of global large caps by total patent filings instead, a similar dynamic is observed in terms of impact on sales growth in recent years. The top quartile of patent filers in 2023 experienced aggregate sales growth of 70% from 2008 to 2023, double the rate of all global large caps (35%) and more than 7x the rate of those with no patent filings (9%). Unlike the R&D analysis, however, the remaining three quartiles showed nearly identical growth, up 62%, suggesting that the volume of patents is not necessarily the strongest signifier. Rather, the mere presence of patent filing activity is indicative enough of the potential for fundamental outperformance. Looking instead at the top quartile based on 2013 patent filing data, we observe a similar sales growth rate over the entire 15-year period (66%) vs. the other three quartiles (53%). Here we can also see the signaling power of patent screening, as patent filers achieved forward growth rates (11-26%) of ~5-12x vs. the global large cap universe (2%); non-filers actually saw revenues decline by -13%. Of particular significance is the divergence following the initial five-year recovery period post-2008, when non-filers experienced nearly all of the growth they would see in the full 15-year period through 2023.

2023 Patent Filers: Trailing 15-Year Sales Growth

2013 Patent Filers: Trailing 5-Year / 10-Year Forward Aggregate Sales Growth

Putting It All Together: AI-Enabled Patent Analysis + Transparent Patent Ranking

Having established both the explanatory and signalling power of patents as a factor for growth investors, we once again revisit the NYGBIG™ index methodology to highlight how it ranks companies according to the breadth and depth of their patent filings relating to artificial intelligence and big data. NYGBIG™'s universe of eligible constituents are all assigned the aforementioned Pure & Contribution Scores for each sub-theme in which they have patent activity, as well as an Intensity Score, measuring the number of sub-themes with patent filings (max score = 7, min = 1). Potential constituents are divided into comparison groups based on their market cap segment (Large/Mid/Small) and sub-theme, and then filtered on the strength of their Pure & Contribution scores (either score must be in the top 50th percentile to remain in the index, top 35th percentile for additions). Finally, they are sorted by their Intensity Score, Contribution Ratio, and liquidity; in other words, the companies with the strongest breadth (filings across multiple sub-themes), greatest depth (most meaningful contributions to a sub-theme's overall patent filing volumes), and highest liquidity make it into the index.

Preference is given to constituents that belong to 10 Primary ICB Subsectors determined as being most relevant to the theme, while also allowing a few other high-relevance, high-scoring stocks (e.g., Amazon, Bank of America) that are eligible irrespective of industry classification. A maximum of 100 securities may comprise the index at any given time, including at most four non-Primary securities. Constituents are weighted by float-adjusted market capitalization, subject to a cap of 4.5%. The index is rebalanced and reconstituted semiannually. (ESG eligibility criteria are also applied; for a list of all criteria, please see the full [index methodology](#).)

Generative AI represents perhaps the most exciting, sophisticated, impactful, and expensive technological innovation of the 21st century. Considerable evidence exists to support the theory that the most successful companies in the new, highly digitized economy are those that reinvest the greatest percentage of their revenues into research & development activities. Patent development is likely the most important of these when evaluating a company's impact on a specific technological investment theme, such as artificial intelligence. A transparent, systematic investment strategy that leverages AI technology itself to screen and rank companies based on the relevance, breadth, and intensity of their patent filings can offer investors a uniquely powerful approach to finding an appropriate level of exposure to the theme.

ETFs currently tracking NYGBIG™ include the Xtrackers Artificial Intelligence & Big Data ETF (Germany: XAIX / US: XAIX).

Sources: Nasdaq Global Indexes, Nasdaq AI Team, IFI Claims, FactSet, Bloomberg