



Thematic Investment Whitepaper

Thematic Investment Demand

The EquBot AI platform continually analyzes both traditional and alternative investment data around the clock in 13 global languages. The system identified “Thematic Investing” as a key global growth concept for investment management. Rapid economic and geopolitical changes and an increasing rate of industrial evolution has led to an increased demand for specific thematic investment solutions.

Traditional sector classifications are one-dimensional and do not capture the multi-dimensional nature of today’s companies that operate across multiple markets and offer a variety of products. A perfect example is Amazon. Amazon’s traditional sector classification is “Internet Retail”, but that does not capture Amazon’s video streaming and cloud service businesses. Thematic investing recognizes this limitation and the opportunity to look across sectors and regions.

Thematic investments are a financial expression of an investor’s forecast on a key market shift. Many of the most popular investment themes are somewhat ambiguous as in how an “investment theme” should be expressed in a portfolio. Thematic investment appeal continues to be driven by an alignment of investor unique views and values with respect to global market movements as assets growth rates continue at an impressive rate.

EquiBot takes an extraordinary leadership role constructing investment thematic solutions through a completely data driven process. Key components and companies relative to specific evolving investment themes are identified and “bagged” through IBM Watson’s world leading natural language processing and EquiBot’s machine learning algorithms. Traditional and alternative thematic data is aggregated through EquiBot’s knowledge graphs. The EquiBot and IBM Watson technologies have the ability to identify both positive and negative exposures to a specific theme. These novel investment platform features allow for complete customization, a quick turnaround, and a multi-dimensional lens while maintaining a disciplined data driven approach.

Thematic products have been predominantly deployed within investor portfolios as long-term tactical allocations. The thematic exposure is viewed as a strategic source of future alpha given current market dynamics. This allows investors to disregard moderate concentration of risk and apply a forward looking element to their investment portfolios. Investor reception of thematic strategies has been well received as the forecast and investment market views associated with thematic products tend to align with investor interests to innovate on traditional investment practices.



Thematic assets continue to rise for both retail and institutional investor solutions. We anticipate thematic investments to continually evolve but buyer beware. Thematic solutions vary broadly and can sometimes offer very little exposure to a stated theme.

Applying AI to Thematic Products

Investors justify thematic exposures in the risk reward framework through prospective future growth. AI investment technology maintains an innovative forward-looking market perception so the structural aspects of thematic and AI investing align. Currently, there are several shortcomings associated with the construction of thematic products. Providers frequently deploy filtered lists that contain only the Large and Mega Cap names within a dataset. This methodology introduces size bias into the thematic that often dilutes exposure and omits smaller capitalized firms who can have a higher and a more pure exposure to a theme. Investment professionals working on thematic products can be bound by the data they are exposed to, and may have other premonitions influencing investment positions. Additionally, investment professionals may use a subjective list of keywords or miss the inferred indirect connections to a theme. AI investment integration is a tool that can help investment institutions navigate the aforementioned pitfalls.

Using AI technology to identify growth patterns in the data aligns with forward looking thematic investment foundations. Data can be used to identify thematic patterns and forecast with more accuracy companies that will have greater influence in a specific market. Proper integration of AI technology can allow an investment solution to capture key thematic investment opportunities through the powers of natural language processing.

EquiBot Process Overview

Data can be used to connect an investor to a particular AI enriched theme. Thematic investing offers investors a way of categorizing companies not by sector or size but by the exposure they may offer to these mega trends.

EquiBot's thematic platform is a multi-dimensional classification system that provides a profound understanding of a company's exposure to various themes by combining fundamental and quantitative factors. It can be utilized to collect useful data and analyze it in a way that gives investors relevant thematic information to help drive data weighted investment decisions. The AI platform leverages unique technology to find, select, and measure stocks based on their exposures to themes. To do this, we use natural language processing, machine learning algorithms, and powerful knowledge graphs to help measure thematic exposures. This data driven systematic approach helps ensure a stock is selected for its thematic exposure instead of past price performance or fundamental metrics such as earnings growth. Thematic stock portfolios differ from other types of stock selection and dynamically adjust based on the latest market data, financial data and filings, global news, and more.



Our thematic algorithms are predominantly powered by EquiBot’s knowledge graphs and IBM Watson’s natural language processing (NLP). NLP uses syntactic and semantic analysis of text and speech to determine the meaning of a sentence. Syntax refers to the grammatical structure of a sentence, while semantics alludes to its intended meaning. NLP also establishes a relevant ontology: a data structure which specifies the relationships between words and phrases. The combination of our knowledge graph and IBM’s NLP create thousands of company specific investment models. The model capabilities allow our platform to identify, score, and rank companies based on their exposure to a particular investment theme. Taking a theme like “Automation”, the NLP model and knowledge graphs can instantaneously reference historical data signals and formulate a portfolio based on quantified thematic exposure.

As a result, the EquiBot platform offers a way to help identify companies that have significant levels of exposure to a particular investment theme. These portfolios of stocks will continually adjust as new data is released to ensure investors maintain relevant stocks.

EquiBot Platform Technology – Thematic Attribution

EquiBot uses proprietary convolutional knowledge graph technology for thematic investment solutions. Knowledge Graphs are powerful model tools that represent a knowledge base that leverages a graph-structured model to capture the interconnectedness or relationships of the data points. The graphs are a connection of nodes connected by edges. Nodes are represented by concepts extracted from document snippets and edges between nodes are computed based on strong connections calculated over the text sources processed. Concepts can be companies, entities, people, places, technology, an abstract idea, etc. The edges are the levels of association based on the textual database. The graph can make both direct and inferred indirect connections. Knowledge graphs are commonly used in internet search engines, such as Google Search to retrieve the most relevant results from their sources and enrich those results.

Knowledge graphs have the advantage of enriching the search results through a network of ‘concepts’ on the graph rather than just keywords. In the thematic exposure context, it offers a more comprehensive approach by anchoring around concepts connected to a given theme based on text snippets and their edges on the graph. The inferred indirect connections from the knowledge graph are what is missing in the traditional textual analysis where it often relies on a list of keywords.

In connection with the knowledge graphs EquiBot uses an enriched BERT algo that is finance and investment market focused as well as augmented filtering from IBM Watson NLP. BERT, which stands for Bidirectional Encoder Representations from Transformers, is based on Transformers, a deep learning model in which every output element is connected to every input element, and the weightings between them are dynamically calculated based upon their connection. The aforementioned EquiBot approach reduces noise, creates an accurate mapping, and deduces complex thematic ideas leading to an optimal thematic basket.



EquiBot Thematic Exposure Technology uses two key concepts:

1. Financial Relevancy Score (F_{Rel})— Fundamental data such as revenue and R&D are mapped and correlated to a theme. Mapping is done across the entire income statement, balance sheet, and management discussion. The Thematic F_{Rel} has a 50% weight.

2. Information Relevancy Score (I_{Rel})— Unstructured data from news articles, blogs, social media, patent filings and research reports with proximal relevance attribution to a theme. The Thematic I_{Rel} has a 50% weight.

The “Thematic Score” is an aggregate score of both the F_{Rel} and I_{Rel} . The default is a 50% weight for both F_{Rel} and I_{Rel} , but the investor has the ability to change the weights based on personal preferences.

Thematic Investing Methodology

1. Thematic Selection & Definition

Investment themes need to be first identified and defined. Thematic definitions can be broad such as “automation” or more focused such as “Cybersecurity”. EquiBot’s thematic capabilities expand across multiple sectors and geographies.

2. Data Sources & Monitoring

EquiBot utilizes the full extent of its technology platform, a variety of diverse sources, and 13 global languages to capture thematic data. The data sources can include, but not limited to:

- Company Websites: Insights can be gathered from how a company defines itself and what information it decides to put forth on its website.
- Company and Regulatory Reporting and Filings: These filings can offer insights that go beyond the traditional classification methods that limit a company to a single segment. Instead a company can be more appropriately classified based on dynamic changes in its revenue mix, R&D, investments, and new market segment efforts.
- Global News and Social Media: Provides real time information from reporters, management, employees, and customers on a company's thematic exposure and leadership. One million global news articles are analyzed every day. Social media and blogs can capture emerging trends before they go mainstream.
- Syndicate Research: Independent expert opinion on specific themes such as what to watch out for, market size and growth, key players, events, and related concepts.
- Globally Filed Patents: Provides insights to a company’s intellectual property related to a theme. Over 3 million patents captured.
- Other: Any other relevant and alternative datasets on the theme.



Figure 1. Data Sources

58 signals feed are processed from millions of regulatory filings, articles, press releases, social media posts, etc. for each of the stocks in the investible universe.

Financial Relevancy Score

Primary Sources: (i) Company Websites (ii) Company Profiles (iii) Company and Regulatory Reporting and Filings (iv) Financial Data Providers (i.e. S&P)

- | | |
|------------------------------------|---|
| 1. 10K/10Q Management discussions | 27. Paid in Common Capital (Net) |
| 2. Sales/Revenue | 28. Retained Earnings |
| 3. COGS | 29. Common Dividends |
| 4. R&D Expense | 30. EBITDA |
| 5. SG&A Expense | 31. EPS |
| 6. Depreciation & Amortization | 32. Legal Proceedings in 10K/10Q |
| 7. Interest Expense | 33. Risk Factors/Disclosures about Market Risk in 10K/10Q |
| 8. Non-Operating Income/Loss | 34. Cash Flow summary in 10K/10Q |
| 9. Income Taxes | 35. Long form company description from SEC |
| 10. Other Income (Loss) | |
| 11. Ext. Items & Disc. Ops. | |
| 12. Preferred Dividends | |
| 13. Operating Cash and Market Sec. | |
| 14. Receivables | |
| 15. Inventories | |
| 16. Other Current Assets | |
| 17. PP&E (Net) | |
| 18. Investments | |
| 19. Intangibles | |
| 20. Other Assets | |
| 21. Current Debt | |
| 22. Accounts Payable | |
| 23. Other Current Liabilities | |
| 24. Long-Term Debt | |
| 25. Other Liabilities | |
| 26. Preferred Stock | |

Information Relevancy Score

Primary Sources: (i) Global News and Social Media (ii) Blogs (iii) Syndicate Research (iv) Patent Filings

1. Analyst Forecast on Earning (EPS)
2. Analyst Growth forecast
3. News Sentiment
4. Headline Frequency
5. Social Media Sentiment
6. Patent Filings
7. Media Engagement
8. Legal Involvement
9. Production loss
10. Earnings Revision
11. Crisis Response
12. Regulatory Changes
13. Natural Disaster
14. Customer sentiment
15. New Products
16. Strategic Partnership
17. M&A Detail
18. Correlated Company Sentiment
19. ESG
20. Product Launch
21. Innovation spend
22. Customer Satisfaction
23. Leadership Influence

The above list of signals and sources is nonexhaustive. EquiBot’s proprietary crawler is continuously testing new data sources capturing thematic information that is not reflected in financial statements or mainstream media. Ongoing data monitoring and model updates will transpire through vetted AI investment operations.

EquiBot can also ingest proprietary client data to augment the thematic investment processes (ex. internal equity models, company risk scores, etc.).

3. Calculate Theme Exposure - “Thematic Score”

IBM Watson’s NLP advanced algorithm extracts metadata (info on data) on themes, concepts, keywords, entities, relationships and events from the aforementioned data sources and feeds this information into EquiBot’s proprietary knowledge graphs. The knowledge graphs link concepts, entities, relationships, and events. Furthermore, the knowledge graphs put the metadata into context and provide a framework for ingesting and unifying the data. The knowledge graphs along with the deep learning models establish the importance of the different data sources and that information is then used to update the knowledge graphs



further. For example, a mention in quarterly or annual filings will have higher importance than a published blog.

Once the data is unified via knowledge graphs, a particular theme can be queried. A query maps the related themes keywords, concepts, phrases, and other relevant data to specific companies and entities which are “bagged” by the EquBot platform.

Next, the thematic exposure is calculated by its relevance looking at both the thematic financial attribution (F_{Rel}) and the thematic information attribution (I_{Rel}). The thematic financial attribution is calculated based on items such as a company’s revenue mix, R&D, and investments. For example, company’s with a higher percent of its total revenue exposed to a theme will receive a higher F_{Rel} . This approach allows us to select companies with the greatest financial exposure to a theme and not just the largest companies.

The thematic information attribution (I_{Rel}) is calculated based on both the frequency of mentions to keywords, phrases, and concepts and the proximity of these words to the company’s name. For example, a company who is mentioned both frequently and whose name appears right next to a thematic keyword will be assigned a higher I_{Rel} . The F_{Rel} and the I_{Rel} are aggregated together each with a 50% weight to determine the total “Thematic Score”. Each company is assigned a “Thematic Score”.

Themes and companies continuously evolve. As a result, thematic data that is more recent is more heavily weighted than historical thematic data. There is also more data available on larger companies than smaller companies. Smaller companies can potentially offer a more pure exposure to a particular theme than larger companies who may have multiple businesses. Through a systematic normalization process small and large companies can more appropriately be compared relative to a theme.

Figure 2. IBM Watson NLP Cyber Example (Single URL)

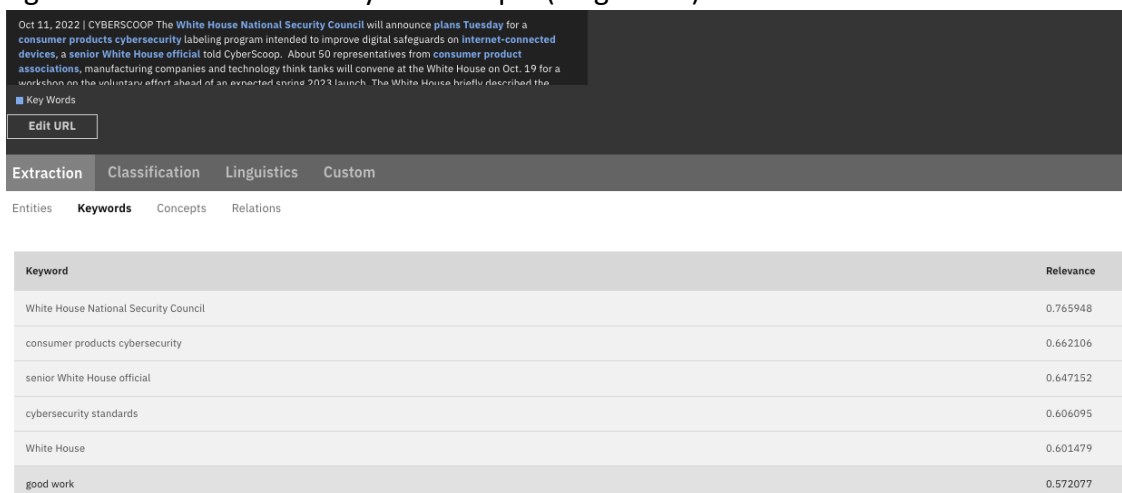
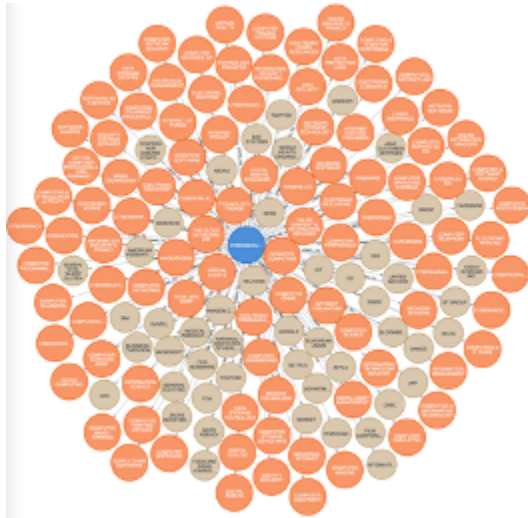


Figure 3. EquBot Knowledge Cyber Graph



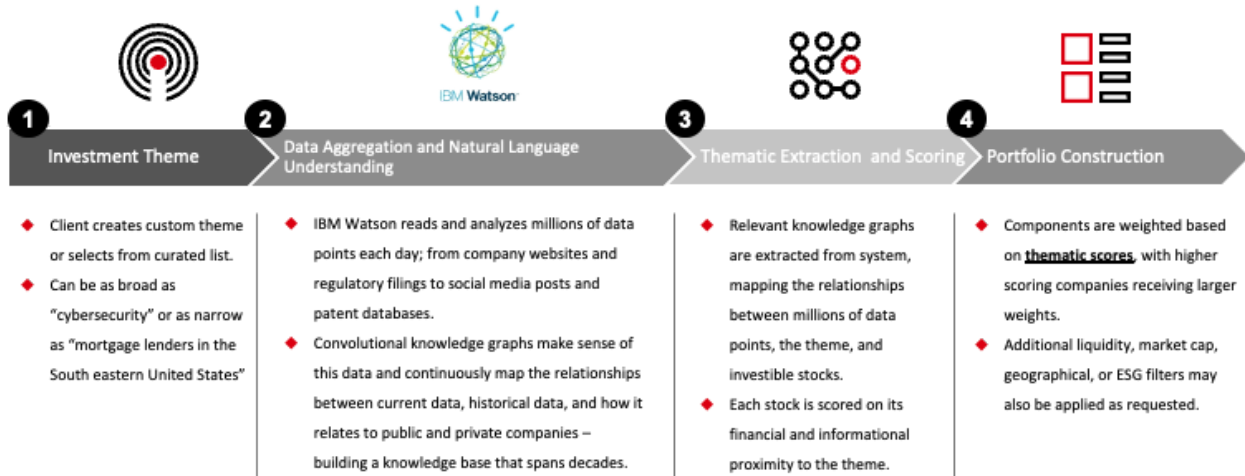
4A. Portfolio Construction

Through the knowledge graph “bagged” companies that have any exposure to a particular theme can be identified. Additional filtering criteria such as liquidity or capitalization size may subsequently be applied based on client input.

Then the companies are systematically ranked by their “Thematic Score” and then the top global stocks are selected. The resulting portfolio is a list of companies that have the highest exposure to a particular theme. Companies with the highest “Thematic Score” are assigned the highest portfolio weight.

The knowledge graphs are continuously updating as qualitative data such as keywords and quantitative data such as revenue mix are reviewed. The goal is to continue to track the themes and related companies as they both evolve to ensure companies with the highest thematic exposure are reflected accordingly in customers’ solutions.

Figure 4. Thematic Investing Methodology



4B. Dynamic thematic strategies:

EquBot’s thematic architecture can be leveraged to detect the most active thematic trends. Most active thematic trends are the trends and associated concepts that demonstrate the highest frequency in the various data sources previously referenced in the Thematic Investing Methodology. The information relevancy of each concept can be measured for a specific duration such as the last 30 days. Within the universe of concepts in the knowledge graph the system has the ability to identify the most active concepts. Additionally, instead of processing all of the concepts located in EquBot’s knowledge graph, the system can instead be selective in processing only the most active concepts which can speed up processing times and improve efficiency. By identifying the hottest trends over specific prior periods, the system can create buckets for the most active concepts. Next, using the BERT model and knowledge graph the system reverse matches the concepts to one or more themes. Lastly, the system can then rotate between themes based on which themes tend to be the most active in the captured data sources via the knowledge graph.

5. Reporting

The EquBot team will discuss client requirements specific to output observability and produce relevant automated thematic exposure reports, investment insight summaries, and scores tabulations.

Conclusion

EquBot’s thematic technology allows for a unique opportunity for custom thematic investment solutions. Utilization of EquBot AI can help investors grow their portfolios by taking advantage in today’s world of disruptive, complex, and rapid changes. Thematic investments can



undoubtedly drive unique exposures not found in traditional index or factor-based investing and seemingly make sense in today's market environment for certain investors.