Mastères Spécialisés® & MSc

MSc in Big Data & Business Analytics

Master Thesis
Class 2024

Master thesis presented by: Baccio Federico Galletti

Under the supervision of: Howard Zhong

Title: The Galletti Index: Unveiling Hidden Strategies in Corporate Technology Investments

Your master thesis is confidential: ☐ yes ☒ no

SPECIALISED MASTER

SWORN STATEMENT

I, the undersigned, Baccio Federico Galletti

Student of the Specialised Master in International Msc in Big Data and Business Analytics, certify that I have plagiarised nothing in the paper attached, which means that I am the sole author of each sentence contained in this text. Any other sentence by another author than myself has been enclosed in quotation marks, with explicit indication of its source. I am aware that by breaking this rule, I am violating the fundamental academic principles and subjecting myself to sanctions that will be determined by the disciplinary board.

I also certify that this paper has never been presented in previous studies at ESCP. If this research has been conducted as part of parallel studies, I have to specify it.

The views expressed in this professional thesis are mine alone.

Place : Paris Date : 06/11/2024

The Galletti Index: Unveiling Hidden Strategies in Corporate Technology Investments

Table of Contents

Abstract ***4***

Introduction ***5***

1.1. Traditional Investment Analysis Approaches **6**

1.2. Why Patents? **7**

1.3. The Gartner Hype Cycle **9**

1.4. The Galletti Index: an innovative, forward looking investment analysis model **13**

Literature review: ***14***

2.1. The role of patents in technology forecasting **14**

2.2. Patent analysis: methodologies proposed **15**

2.3. Discussing the Gartner Hype Cycle: valuable insights or just hype? **15**

2.4. Literature Review conclusion **18**

Methodology: ***19***

3.1. Overview of the Approach **19**

3.2. In-Depth Methodology Analysis **24**

3.2.1. Data Collection 24
3.2.2. Data Pre-processing 24
3.2.3. Term Frequency-Inverse Document Frequency (TF-IDF) 25
3.2.4. Ontology Creation 26
3.2.5. Semantic Analysis 26
3.2.6. Visualization: Hype Cycle and Investment Intensity 27
3.2.7. The Galletti Index 29
3.2.8. Combining the Scores: The Overall Technology Index 33
3.2.9. Calculating the final Galletti Index 33

Results ***35***

4.1. Investment Intensity **35**

4.2. Visualization: an immediate impact **38**

4.3. Calculating the Indexes **40**

4.3.1. Time-Based Investment Intensity Score 41
4.3.2. Technology Benefit Intensity Score 42
4.3.3. Overall Technology Index 43
4.3.4. The Galletti Index 44

Discussion ***45***

5.1. AI Industry Focus 45
5.1.1. Autonomous AI 46
5.1.2. Boosting Developer productivity 46
5.1.3. Empowering With Total Experience: 46
5.1.4. Delivering Human-Centric Security and Privacy 47
5.2. Microsoft investment strategy – Macro and Micro analysis 47
5.2.1. Autonomous AI 47
5.2.2. Booasting Developer Productivity 50
5.2.3. Empowering with Total Experience **– Total Score:** 2.47 51
5.2.4. Human-Centric Security and Privacy 53
5.2.5. Macro & Micro analysis conclusion 58
5.3. Visualization – Gartner's Hype Cycle with Investment Intensities 60
5.4. The Galletti Index 61
5.4.1. The Galletti Index – summary of conclusions 64
5.5. Strengths, Limitations and Future Research 65
Conclusion 66
6.1. Additional remarks 69
References: 69
Appendix 74
8.1. Patent Example 74

Abstract

This thesis examines corporate investment strategies by mapping them onto the Gartner Hype Cycle, a widely recognized framework for tracking the maturity and adoption of emerging technologies. The study uses patent data – which represent a company's public declaration of their R&D efforts - as a proxy for investment intensity, revealing a the firm's medium-to-long-term strategy. This approach is particularly valuable for investors who often lack detailed information on a firms' long-term investment plans, particularly in R&D. Furthermore, the thesis argues that patent data, when analyzed in conjunction with the Gartner Hype Cycle – a "map" of emerging technologies visualizing their journey towards mainstream adoption - can offer even more practical, valuable insights into a company's strategic direction, allowing investors to anticipate firm or market trends and align their portfolios accordingly.

To further enhance the analysis, the thesis introduces the Galletti Index, a composite indicator designed to quantify a company's investment strategy based on three key dimensions: (1) Investment Intensity, (2) Time Horizon and (3) Technology Benefit. The Galletti Index provides a single, quantifiable score that reflects the company's overall investment approach, revealing whether they are prioritizing shorter-term gains or long-term innovations, and whether they are focusing on high-impact, transformative technologies or incremental improvements. The thesis concludes that the Galletti Index, combined with the visualization of investment intensity on the Gartner Hype Cycle, offers a powerful tool for investors seeking to understand a company's long-term strategic direction. By bridging the information asymmetry gap that often exists between companies and investors, this approach empowers stakeholders to make more informed investment decisions, aligning their portfolios with companies that are poised to lead in the future.

Introduction

Understanding corporate technology investments is essential for investors – Venture Capitals, Private Equity firms or individual, private entities - to evaluate with a higher level of confidence a company’s future competitiveness and resilience - which serves as baseline analysis to allocate capital. According to Deloitte, in fact, organizations that strategically allocate their tech budgets toward impactful innovations not only achieve operational efficiency but also drive top-line growth. As the tech sector permeates nearly every industry, investing in technology allows companies to differentiate, adapt to market changes or anticipate evolving customer demands. Investors, therefore, gain a significant advantage by analysing where companies are channelling their technology resources, as it provides insights into a company’s strategic direction and long-term sustainability, which could lead to a higher company valuation. As investors seek to maximize returns, finding untapped market opportunities becomes complex, especially given that public markets are assumed to already price all available, public information. This makes it challenging to identify undervalued investments that do not yet reflect a company’s true potential (Boston Consulting Group, 2021; From Tech Investment to Impact: Strategies for Allocating Capital and Articulating Value, 2024).

1.1. Traditional Investment Analysis Approaches

Traditional investment analysis methods – which entail researching and evaluating opportunities to determine their investment profiles weighting potential risks and returns (Tamplin, 2023) – are mainly based on past-looking data, disregarding precious information on what the future of the company will look like. As a matter of fact, fundamental and technical analysis – two of the most used investment analysis approaches – effectively assess past or current company performances, but often fail to capture early signals of technological innovation. Fundamental analysis consists in evaluating an investment intrinsic value, relying on financial health metrics and analysing a firm's revenues, profits, cash flows and operating expenses. Moreover, it also considers the broader market context, and how that could affect the stock price. Technical analysis, on the other hand, focuses on historical price trends and its fluctuations, using computer calculated graphs and charts (Twin, 2023). Both methods are backward-looking, relying on historical data to assess an asset's current, intrinsic value relative to its market value (thus identifying undervalued investments the market is overlooking). This approach, however, ignores a company's strategic efforts to build future competitive advantage, as traditional investment analysis lacks mechanisms to incorporate forward-looking indicators of innovation or potential market shifts. Additional investment analysis approaches include top-down or bottom-up methods: top-down entails analysing individual stocks focusing on macroeconomic trends, narrowing down to the most promising sectors, while bottom-up focuses directly on individual company metrics (Twin, 2023). While these methods are widely used, they fail in considering either market or company information in their model – both fundamental in correctly evaluating an investment. This limitation is especially relevant for private firms, which provide little to no public data until a formal release - these companies disclose only the information they choose, and once it becomes public, the market rapidly adjusts to reflect the newly available data, often leaving investors with little opportunity to anticipate the major shifts. Overall, all approaches discussed do not take into account *"forward-looking" historical data* – as they fail to capture a company's past and current efforts in investing in their future, namely in nascent technologies that haven't yet impacted financial statements or market metrics. Thus, to address the limitations of traditional methods, we propose a new approach that integrates (public) *forward-looking, historical data*, offering a comprehensive view of a company's future potential. This method goes beyond static financial metrics, incorporating indicators of a firm's strategic investments in technology and innovation that are critical to building competitive advantage, and positioning the firm for future, sustainable growth. Investors can now benefit from a new, predictive model based on patent analysis.

More specifically, despite the availability of financial reports, investors lack a tool to assess firms' medium-to-long-term technological investments, the maturity and lifecycle of those technologies. Publicly available resources, in fact, do not convey the underlying risk level, stage of maturity or market promotion such investments present, which highlights an assymetry between accessible information and a corporations' intent. Given these limitations, this research aims at filling such void, providing investors with a tool – more specifically, a Natural Language Processing model – to map firm investments along the Gartner Hype Cycle using patent data. More specifically, it allows access to information otherwise hidden: on what emerging technologies is a firm directing its financial efforts in the medium-to-long term? Leveraging publicly available patent data, the model will map the technologies a firm is betting on over the Gartner Hype Cycle of Emerging Technologies, in order to enrich the decision-making process of investors and funds. Given the public nature of patents data, the model is agnostic to the ownership structure of a corporation (i.e. private or public), making it particularly appetible for evaluting private firms' investment strategies – which would remain otherwise unclear for investors.

1.2. Why Patents?

In today's competitive business landscape, patents play a critical role fostering a company growth. Patents, which provide a temporary exclusive right granted for an invention, give the patent owner the authority to determine whether, and under what conditions, others may use the invention (World Intellectual Property Organization, n.d.) – providing them with legal protection (What Is Intellectual Property (IP)?, n.d.). Corporations leverage patents not only to safeguard their inventions from imitation, protecting its intellectual property (IP), but also to strengthen their market position by limiting competition and creating barriers to entry (Argente et al., 2020), and serve as valuable assets that can be monetized through licensing agreements or partnerships.

As a matter of fact, patents are considered one of most effective protection tools against competition, increasingly utilized by companies all over the world. Additionally, as Ziedonis (2003) highlights, aggressive patenting strategies are also leveraged to avoid being “fenced-in” by external patent holders whose technologies may inadvertently be incorporating their own products. Lastly, the sharp rise in patent filings can also be attributable to their key role in attracting potential investors, partners or fostering a positive company image (Wściubiak, 2016). Consequently, the number of annual patent applications has, in fact, seen a sharp rise in the last decades, increasing from roughly 650,000 in 1985 (first available data) to almost 2.4 million in 2021 (Multiple sources compiled by World Bank (2024) – processed by Our World in Data).

Especially for the ICT sector, which experienced a remarkable growth and represented approximately 15.5% of the global GDP in 2020 (but expanding at a rate 2.5 times fast than global GDP - Huawei Technologies Co., Ltd., 2017), patents are playing a crucial role in creating and maintaining a company’s competitive advantage. As a matter of fact, in 2021, software-related patents accounted for 63.1%, 49.4% and 40.2% of all patents filings respectively in the United States, Europe and China (U.S. Patent Grants Fell 7% Last Year, But ‘Software-Related’ Grants Remained at 63%, 2022), with Huawei Technologies and Samsung Electronics being the top filers of Patent Cooperation Treaty (PCT) international applications in 2023 (Interactive Charts: Intellectual Property Facts and Figures, n.d.). Furthermore, patents play an increasingly important role for larger corporations, and act as a strategic tool to foster growth. As a matter of fact, market leaders use patents differently from followers, exploiting them mainly to prevent competitors from innovating and protect sales of existing products, rather than to foster product innovation (Argente et al., 2020). In line with this, as noted by Cohen et al. (2014), large corporations have increasingly turned to patents as a strategic tool to safeguard their sales against competitive threats. Moreover, a recent study by Argente et al., (2020) argues that a larger corporation is able to derive higher financial returns from the same patented idea compared to a smaller enterprise, underlying the strong importance of patent filings for larger corporations.

Given the strategic importance of patents for large corporations in the ICT sector, patents can serve as a critical source of information for analyzing, understanding, and predicting a firm's interests and influence in a particular technological field. In fact, utilizing patent data to map or forecast a company's relevance within specific technology fields is a widely recognized practice among researchers. For instance, several empirical studies, such as those by Genelöv and Yun (2018), suggest that patents can act as a reliable proxy for positioning firms within various technological domains, thereby highlighting their strengths and market relevance.

1.3. The Gartner Hype Cycle

Furthermore, this thesis not only harnesses patent data, but also incorporates the Gartner Hype Cycle framework into its analysis. Not to be confused with economic cycles, which fluctuate over time, often moving up and down without stabilizing at a productivity plateau, the Hype Cycle tracks the development of a specific technology, which eventually matures and reaches a productive phase. Unlike fads or market fluctuations (e.g., house prices), the Hype Cycle focuses on innovations with long-term potential (Vern Burkhardt & Jackie Fenn, n.d.), taking into account the human factor and its influence over a technology's adoption. Additionally, the Gartner Hype Cycle follows the idea that the history of innovations repeats itself, with a recognizable pattern of initial hope and strong optimism, soon faded to leave space to the disappointment of reality. As a matter of fact, history teaches numerous examples: from the telephone and automobiles in the early 20th century, to the Internet in the '90s, and – most recently – biotechnology and nanotechnology, all highly anticipated innovations follow the hype cycle trend (Fenn & Raskino, 2008). More specifically, starting off with a technological breakthrough or product demonstration, a new innovation generates press and industry interest, soon transformed into a wave of high notoriety and unrealistic expectations. Consequently, optimism fades and the initial obstacles arise: performance problems, slow adoption or failure to achieve expected financial returns result in growing disillusionment. Despite these challenges, early adopters experience initial technological and financial benefits, which restores hope in the innovation. Lastly, building on the experience of early adopters, proof-of-concepts and demonstrated benefits, an increasing number of corporations feel comfortable in adopting the technology, generating tangible and real value (Gartner, 2018). The core concept of the Gartner Hype Cycle can be summarized with what is now known as the

First Law of Technology: "*we invariably overestimate the short-term impact of a truly transformational discovery, while underestimating its longer-term effects*" (Collins, 2024).

Gartner, the world's leading research and advisory firm, made of the hype cycle one of its mantras: the firm publishes yearly researches mapping innovations over the hype cycle within their technology space, offering a snapshot of market maturity and perceived value, differentiating hype and "buzz" against real and tangible value (Gartner, 2018). Corporations rely on Gartner's Hype Cycles to guide their decision-making process regarding new technologies. More specifically, they get educated over the potential economic and technological impact of an innovation, evaluate its stage of maturity, understand its future directions and inform their investment decisions (Gartner, 2018). In particular, all Hype Cycles follows the same structure, and divide a technology's journey towards adoption in different phases (as shown in the Figure 1):

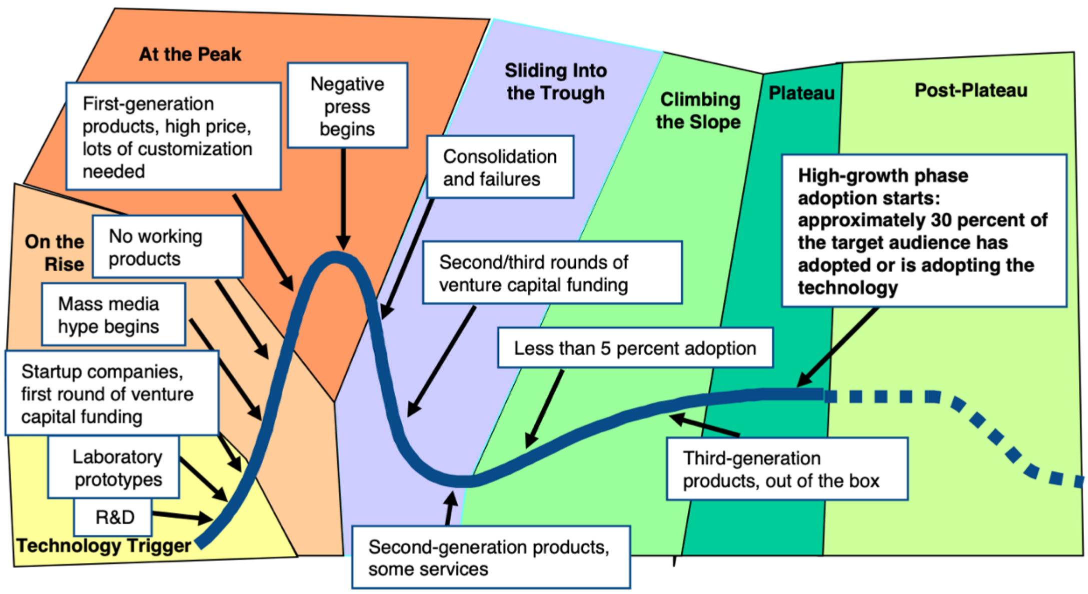
- **Innovation Trigger:** an initial breakthrough, public showcase or major product launch ignites widespread interest across media and industry circles.
- **Peak of Inflated Expectations:** a period of intense enthusiasm and extremely high expectations, marked by notable successes (but even more failures) – mostly contributing to hyped media coverage.
- **Through of Disillusionment:** as the innovation fails to meet its exaggerated expectations, its popularity declines, leaving little to no attention from the media.
- **Slope of Enlightenment:** through focused efforts and experimentation, a more varied group of organizations begins to truly understand the technology's practical use cases, along with its risks and advantages. Standardized tools start hitting the market, facilitating development and first deployments.
- **Plateau of Productivity:** the innovation's practical benefits are demonstrated, becoming widely recognized and accepted. Products evolve to more stable versions, reducing risks. The number of organizations adopting the technology increase, as the rapid growth phase begins. At this stage, around 20% of the target market has begun to adopt the technology.

Figure 1:

(Gartner Hype Cycle and its stages, Linden et al., 2003)

Figure 2 below also provides a more visuale glimpse into the Hype Cycle's stages and their characteristics:

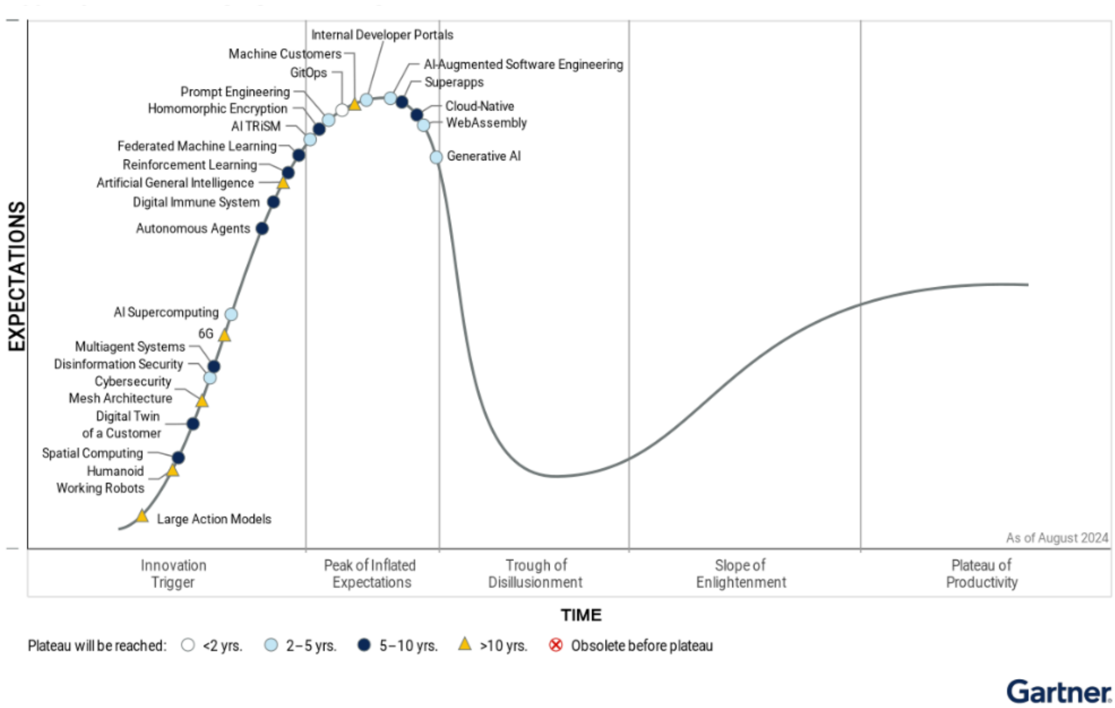
Figure 2:



(Gartner Hype Cycle and its stages, Linden et al., 2003)

Every year, Gartner publishes numerous Hype Cycles, each addressing a specific technology space. These segments can vary in scope, from narrower areas to broader technological landscapes, such as Cyber Risk Management or Emerging Technologies. The latter is particularly significant due to its comprehensive scope, offering a high-level overview of the most transformative technologies – regardless of their belonging sector - that have the potential to disrupt industries and create new markets. By identifying key innovations in their nascent stages, this Hype Cycle enables organizations to anticipate shifts in the technological landscape, often setting the agenda for technological discussions globally. Based on this information, corporations will make investment decisions and pursue one of several strategies. The 2024 version is illustrated in Figure 3 below.

Figure 3:



(Gartner Hype Cycle for Emerging Technologies, Lajoie & Bridges, 2014)

Firms with a higher willingness to take risks may opt for early adoption, potentially generating higher returns. On the other hand, more cautious organizations may require a thorough evaluation of the opportunity, prioritizing detailed cost-benefit analyses, especially when the technology is in an embryonic phase. Consequently, given that a multitude of firms base their investment decisions on the Gartner's Hype Cycle for Emerging Technologies, understanding the latter and leveraging its information becomes equally valuable for investors and funds when evaluating their own investments. In fact, knowing which technologies a firm is investing in, and at what stage of maturity and market visibility those innovations are, provides critical information investors cannot ignore. Alongside a quantitative analysis of a company's financials and overall health, in fact, they need to perform a qualitative analysis on the future growth and investments such firm is pursuing. Consequently, they also must decide whether to invest and back companies pursuing high-risk, early-stage innovations, or focus their financial efforts on those betting on more mature, more widely adopted technologies which, however, might offer lower financial gains. This strategic assessment allows investors to align their portfolios with their risk tolerance and long-term financial goals, combining quantitative and qualitative data to inform their decision making.

1.4. The Galletti Index: an innovative, forward looking investment analysis model

Leveraging the Galletti Index, and using Microsoft as an example, the company's technology investments are evaluated by integrating patent data with the Gartner Hype Cycle, revealing strategic focuses on transformational technologies like AI-Augmented Software Engineering, Machine Customers or Digital Twin of a Customer. These technologies, while in early stages, are poised to reshape industries, with Microsoft positioning itself as an early leader in these disruptive markets. The Galletti Index, which combines investment intensity, technology impact (or benefit), and time horizon, fills information gaps by providing insights into Microsoft's future R&D priorities beyond what traditional analyses reveal. This approach not only highlights Microsoft's potential for medium-term gains, but also offers a scalable framework for investors to assess technology investment strategies across companies. By transforming public patent data into actionable insights, the Galletti Index enables investors to anticipate corporate strategy shifts, aligning their choices with long-term, innovative growth opportunities.

Literature review:

2.1. The role of patents in technology forecasting

Given their public availability, patents data have been leveraged by a multitude of practitioners to pursue studies in the technological field, as they implicitly carry critical information about innovations, as well as the strength of directed investments and market growth (Yang, 2012). More specifically, technology forecasting - both at a company and market level – has been a widely researched topic associated with patents. For example, Trappey et al. (2011) have leveraged Chinese patent data to explore RFID technology development and its future trends. Through the combination of content extraction, clustering and technology forecasting, their research provides valuable insights to evaluate potential market opportunities within the technology sector (Trappey et al., 2011). Kwon et al. (2022) have also proposed a framework to predict technological change to identify untapped market opportunities and inform future product innovation roadmap through patent data analysis within the logistics service. Furthermore, Databases like the World Intellectual property Organization (WIPO) and the United States patent and Trademark Office (USPTO) have been processed through text mining and clustering techniques, in order to investigate respectively the trends of Blockchain technology (Bamakan et al., 2021) and the performance improvement rate of technologies (Singh et al., 2021). However, a broader research by Yoon et al. (2012) argued that patent analysis for technology and trend forecasting is not equally effective across all industries: manufacturing sectors, for instance, proved more beneficial than service sectors, with an emphasis on biotechnology, specialized supplier and information technology sector – which exhibit a high forecasting accuracy. With a narrower focus on one specific company, Jun and Park (2013) have also leveraged patent analysis to examine Apple Inc.'s technological innovations, identifying unexplored markets and core investment fields through a combination of statistical and machine learning techniques. While this empirical study proved useful to understand the firm's current focus as well as its future potential innovation areas, it left unexplored the implied characteristics of latters: is Apple focusing its investment on early-stage, untapped technological spaces, or on more mature, known innovations? Are these investments directed at currently hyped, highly promoted technologies, or already adopted innovations with proven benefits? This research paper aims at bridging this gap, combining technology forecasting and emerging markets analysis to determine the appetibility of a corporation as investment opportunity.

2.2. Patent analysis: methodologies proposed

Furthermore, empirical researches have proposed different methodologies for effective patent analysis. Sharma at al. (2012) have reviewed multiple text clustering techniques identified from publications, journals and reviews, to recommend different models based on the analyst's need and objective. In particular, K-Means Clustering algorithms have proved efficient in visualizing large datasets while maintaining semantic relationships, while Document Clustering (K-Means) combined with Time Series Regressions have performed well in predicting future technological trends. Moreover, Distance Determination Approach (through k-means and k-medoids techniques) and Bayesian Clustering were proposed as optimal algorithms for respectively automatic discovery of data clusters and high-dimensional datasets, while Fuzzy Logic Control is deemed appropriate for automatic interpretation and clustering using an ontology schema. In fact, keywords-based methodologies often struggle in effectively analyse patent documents, as only partial or isolated terms or meanings are used, leading to inaccurate clustering. Introducing an ontology schema and using a fuzzy logic control approach, the authors found optimal results in matching patent documents to relevant clusters based on ontological semantic webs (Sharma et al., 2012). Among all techniques, the latter seems appropriate for this reaserch topic: its ability to effectively map technologies over different categories while allowing for flexible grouping (an important aspect given the overlapping nature of certain technologies and innovations) will prove useful in the analysis. In fact, this method will be one of the leveraged factors in the analysis.

2.3. Discussing the Gartner Hype Cycle: valuable insights or just hype?

The Gartner Hype Cycle has also been widely researched in empirical studies, as critics discuss the framework's reliability and applicability in the real world. In support of its validity, Stikov and Karakuzu (2023) have leveraged the Hype Cycle to map the industrial adoption of relaxometry (i.e. the measurement of relaxation variables in Nuclear Magnetic Resonance and Magnetic Resonance Imaging, or "MRI"), as well as to anticipate its future development. Grundmeyer (2013) leveraged a combination of interviews data and Gartner's Hype Cycle to effectively plan and incorporate new educational technologies, as the author argues that "*the Hype Cycle can ensure money spent on technology is timely*", anticipating potential challenges in adopting new technologies. Thus, the framework helps in allocating budgets, develop new policies and train staff in advance, ensuring stronger support from stakeholders. Weinraub and Bridges (2014) also argue Gartner's Hype Cycle to be an important tool in assisting technology adoption and decision-making, while managing an organization's appetite for risk in the context of library administrations. Oosterhoff and Doornberg (2020) leveraged the framework to map Artificial Intelligence's evolution in orthopaedic surgery, from its hyped and mediatic phase to disillusionment and realization, envisioning its development into mainstream adoption. Lastly, an empirical validation of the Hype Cycle pattern was also provided by Adamuthe et al. (2015), who investigated its applicability to Cloud Computing technologies - through news articles from Google Trends, published research papers and patents filings - revealing that cloud computing, virtualization, and Web 2.0 follow expected hype cycle patterns.

On the other hand, Dedehayir and Steinert (2016) criticize Gartner's hype cycle reliability by comparing the expected technology evolution of 3 innovations - tidal power, IGCC (Integrated Gasification Combined Cycle) and photovoltaic generation (all renewable or clean energy generation technologies) – with their actual progression based on Google News and Google Insight data, highlighting significant incongruences between the two. More specifically, the authors present three major limitations: (i) the framework's combination of two distinct and incompatible evolutionary models (i.e. the human-centric hype expectations model and the technology S-curve), (ii) the ambiguous definition of the dependent variable "visibility" on the y-axis and (iii) the questionable applicability of the expectation pattern across all actor groups – meaning, it is unclear whether the hype cycle pattern represents a universal trend applicable to all stakeholders - entrepreneurial firms, incumbent companies, governments and NGOs - or it is instead focused on only some actors, aggregating different points of view. Furthermore, Järvenpää and Mäkinen attempted to build the framework's theoretical foundation, introducing "technology life cycle indicators" – in order to track the visibility of technologies like DVD, MP3, Bluetooth, and Blu-ray over time. They concluded that - while there are evidence of hype cycle trends in mainstream and technical news for MP3 and Bluetooth - Blu-ray and DVD lack clear hype dynamics (Dedehayir and Steinert, 2016). On the other hand, Dedehayir and Steinert's literature review revisited and completed such conclusions, arguing that only four cases - MP3, Bluetooth, natural gas, and stationary fuel cells - were found to replicate the expected pattern. The study identified other variations, such as nine instances of hype followed by disappointment, six cases of multiple peaks and troughs, and two with fluctuating dynamics. The inconsistency in these empirical findings raised concerns about the general applicability of the Hype Cycle framework (Dedehayir and Steinert, 2016).

Nevertheless, the Gartner Hype Cycle remains one of the most utilized tools companies leverage to set their future strategic direction, in virtually all technology sectors and industries. One of the Hype Cycle's major use cases consists in providing access to a complete mapping of the most important innovations in the medium-to-long term, identify opportunities for business growth, their gap to mainstream adoption and financial benefits, as well as prioritize investments while managing risks. In other words, corporate strategies and innovations are both subject to and explained by the Hype Cycle, which can be seen as "overlapping hype cycles" (Vern Burkhardt & Jackie Fenn, n.d.). As Jackie Fenn - former Gartner fellow and co-creator of the framework – stated: "*the best way we know to (counter the belief that you are immune to the effects of the forces underlying the hype cycle) is through a systematic, disciplined framework that ensures that you incorporate the lessons of the hype cycle into every adoption decision that you make"* (Vern Burkhardt & Jackie Fenn, n.d.). On the other hand, a reactive approach will cause firms to be late on innovations, struggling to distinguish hype and tangible value, being overwhealmed by the force of the hype cycle (i.e. the artificially created "buzz" around a technology).

However, visualizing, identifying and weighting risks and opportunities of a technological innovation is only a small part of what Gartner refers to as the

"STREET" process – a six-stage approach for selecting and implementing an innovation. In particular, (i) Scope, (ii) Track, (iii) Rank, (iv) Evaluate, (v) Evangelize and (vi) Transfer represent Gartner's proposed best practices for technology planning and adoption. While this paper previously discussed and focused on the first four steps, the Hype Cycle also plays a crucial role in the final last stages. In particular, by leveraging the Hype Cycle, the heads of emerging technologies teams in IT organizations – responsible for promoting the adoption of innovations - generate internal momentum, persuade stakeholders as well as inspire decision makers and potential implementers about the functions and benefits of the new technology. Similarly, during the Transfer phase, the Hype Cycle provides valuable insights by suggesting when a technology has reached maturity and is ready for full-scale implementation. It helps organizations avoid premature deployment by guiding them through the trough of disillusionment, reducing risks and facilitating a smoother transition from small-scale pilots to large-scale adoption (Vern Burkhardt & Jackie Fenn, n.d.). Consequently, the Hype Cycle play a comprehensive, pivotal role in all stages of technology selection and implementation.

2.4. Literature Review conclusion

This literature review initially explores the role of patent data in technology forecasting. It highlights empirical studies and researches leveraging patent data for trend analysis, as well as discussed various methodologies for patent analysis, such as fuzzy logic and K-means. Furthermore, empirical evidence questioning the Hype Cycle's general applicability are proposed, highlighting the framework's downfalls and limitations. The review emphasizes the need for further integration between patent analysis and the Hype Cycle to provide deeper insights into firms' innovation strategies. Previous studies have primarily focused on leveraging patent data for technology forecasting and highlighting companies' major investment areas, proposing valuable insights mainly exploitable by individual firms. On the other hand, this research project aims at providing a more complete analysis by integrating market dynamics - such as technology maturity, market size, adoption rates, visibility, and time-to-full-scale adoption. By mapping a firm's technological investments on the Hype Cycle, it allows investors to visualize a company's medium-to-long-term plans, and whether they focus on early-stage technologies or more mature, closer to full-scale adoption once, helping align with their investment strategies.

Methodology:

3.1. Overview of the Approach

This research project aims at building a replicable model to be leveraged periodically upon performing a company analysis. It involves a combination of several steps and techniques, as mapped below:

1. **Data collection:** the accuracy of this thesis model's output heavily relied on the completeness and quality of the chosen dataset. In particular, while patent data are publicly and widely available, most sources impose significant restrictions on visualization and download – such as on fields, timeframes or patent type. For instance, several websites offered *"bulk downloads"* of pre-processed data belonging to a certain technological category or sector. Finally, Lens.org - a free and easily accessible platform offering comprehensive global patent information - was leveraged due to its search capabilities, completeness of fields as well as ease of download. More specifically, two fields were necessary for the subsequent NLP analysis, namely (1) Patent Title and (2) Abstract, which were later combined to provide additional semantic and contextual meaning. An example of patent (3 entries, first 7 columns) from Lens.org is present in the Appendix (1).

2. **Data pre-processing:** a fundamental step for ensuring a strong algorithm performance, it is the process of transforming raw data into a clean, organized and usable format before feeding it to Machine Learning (ML) models or analytical processes. In particular, two approached were usedç
   a. Natural Language Processing (NLP): a subfield of Artificial Intelligence (AI) used to process and analyze large amounts of data encoded in natural language. In this research project, it was

leveraged to clean, lemmatize and identify top keywords present in patents data.

b. Regular Expressions (REGEX): search patterns used to match character combinations in strings. Commonly used to finding patterns within text (*Regular Expressions - JavaScript | MDN*, 2024), it was leveraged to extract technology categories and their definitions from the Gartner Hype Cycle.

3. **Term Frequency-Inverse Document Frequency (TF-IDF):** used in the fields of Information Retrieval (IR) and Machine Learning, it is a statistical technique that quantifies the importance of a word in relation to a document and a corpus (Anirudha Simha, Principle Associate Software Engineer, Kai Chatbot Team, n.d.). While the Term Frequency (TF) measures how often a term appears in a document, the Inverse Document Frequency (IDF) quantifies the rarity of the word in the corpus. The combination of the two techniques ensures only the most frequent but distinctive terms are captured, giving higher weight to those words that repeatedly appear in a patent, but are rare across the whole dataset. For instance, terms like "system" or "method" may appear frequently across all patents, but they would receive a lower score compared to more specific and descriptive terms like "quantum computing" or "homomorphic encryption".

4. **Ontology creation:** a structure knowledge base was necessary to effectively represent the relationship between each technology present in the Gartner Hype Cycle and their definition. Thus, an ontology was created, serving as reference point to the subsequent semantic analysis, ensuring patents were mapped to the correct technological categories, even in cases where the textual descriptions varied significantly.

5. **Semantic Analysis:** in order to map and quantify the similarity between patents and technologies, three different models were chosen – specifically spaCy, FuzzyWuzzy and OpenAI embeddings:

a. **SpaCy:** a free, open-source library capable of processing advanced NLP analysis on large datasets at high speed, making it one of the most used NLP libraries in Python (*What Is spaCy? | Domino Data Lab*, n.d.). In particular, SpaCy leverages pre-trained word vectors to capture the underlying contextual meaning of terms, gather higher semanting meaning. In this research, the cosine similarity between patents and the previously created ontology was calculated – scoring the angle between two vectors in a multi-dimensional space (*Word Embeddings in spaCy*, n.d.).

b. **FuzzyWuzzy**: an AI-based python library that leverages advanced capabilities to identify and match similar (but not identical) terms in a document corpus. Unlike typical search logic, which operates on a binary pattern (i.e. 0/1, yes/no, true/false), fuzzy string matching identifies strings, entries or text within datasets that lie in between these definitive parameters, accounting for partial matches and varying degrees of similarity (Silva, 2023). In particular, this python library leverages the Levenshtein distance algorithm – a mathematical methodology calculating the minimum number of single-character edits (insertions, deletions or substitutions) required to change one word into the other - to compute the similarity between two strings (Stephen M. Walker II, n.d.). For instance, even if a patent abstract contains the terms "AI-driven automation" while the technology definition refers to "automation powered by artificial intelligence", FuzzyWuzzy would recognize the similarity between the two, allowing for a high matching flexibility and ensuring minor wording variations do not result in false negatives.

c. **OpenAI embeddings**: The OpenAI embedding model, specifically the "text-embedding-ada-002", was utilized to generate high-dimensional vector representations of both the patent text and technology definitions, measuring the relatedness of text strings. While embeddings – mathematical representation of data representing real-world instances (i.e. words, images or videos) in a form that ML models can easily process - are commonly known for a series of analysis, namely (1) Search, (2) Clustering, (3) Recommendations, (4) Anomaly Detection, (5) Diversity Measurement and (6) Classification, this research project mainly leveraged (2) Clustering and (6) Classification. As the OpenAI model is trained on a vast corpus of text data, it is able to capture deep semantic relationships among words and phrases through computing the cosine similarity between two vectors. This model was particularly valuable in capturing nuanced semantic relationships that might be missed by simpler models like FuzzyWuzzy. For instance, OpenAI embeddings can recognize that "neural networks" and "deep learning" are closely related concepts, even though not described using the same terms. Thus, the context surrounding each word is also captured, allowing entire sentences, paragraphs, and documents to be searched and analysed based on their meanings. While this process requires significant computational power, storing the context of queries as embeddings enables more efficient handling of future queries by reducing the need for recalculating relationships from scratch (OpenAI, n.d.; *What Are Embeddings?*, n.d.). Furthermore, in order to leverage such powerful model, I had to register and secure OpenAI's API key. Once obtained, I was able to connect to OpenAI's services through API. However, to reduce computational complexity and avoiding the need for repreated API calls during similarity calculations, the embeddings were pre-computed – allowing me to store the vector representations. In fact, pre-computing embeddings is necessary when handling large datasets, in order to save time by caching results, reducing the API load and minimizing redundant calls.

6. **Visualization:** a sixth step used to map the calculated fields and visualize the investment magnitude through colour gradients for each technology category in the Hype Cycle, generating a detailed colour legend as well. In

order to achieve this outcome, the aggregated patent scores were first processed through different techniques, such as logarithmic transformation and quartile scaling. This approach provides a clear and intuitive visualization, enabling a quick and easy identification of the investment intensities across all emerging technologies of the Gartner Hype Cycle.

7. **The Galletti Index:** an innovative, composite indicator developed to quantify a company's investment strategy based on both its impact and time-horizon focus. In particular, it takes into account 3 distinct concepts:

    a. **Investment Intensity**: an aggregated score based on the semantic similarity calculated in previous steps, it reflects the magnitude of a company's investment in each technology.

    b. **Technology Benefit Score**: extracted directly from each technology's "Benefit Rating" classification in the Gartner Hype Cycle, it assesses the potential transformative impact on the market.

    c. **Time Horizon Score**: a combination of different inputs directly extracted from the Hype Cycle, namely (1) "Years to Mainstream Adoption", (2) "Maturity Level" and (3) "Market Penetration", it reflects an approximate timeframe to achieve tangible returns.

Lastly, most of the study was carried out using Python. However, manual inputs were necessary to facilitate the analysis and ensure accurate mapping, particularly in steps (2) and (4), as well as in retrieving patent data.

While the model is designed to be adaptable and applicable to a wide range of companies and investors, Microsoft was selected as the initial focus to demonstrate the model's effectiveness. This choice was due to Microsoft's large patent portfolio and significant investments in emerging technologies. Using Microsoft as a starting point helps validate the model's capability in assessing corporate investment trends, but the model itself is versatile and can be applied to other corporations and sectors over time.

3.2. In-Depth Methodology Analysis

3.2.1. Data Collection

To successfully perform this analysis, a comprehensive patent database was necessary. While several different sources were considered and explored, among which Google Patents, World Intellectual Property Organization (WIPO) and United States Patent and Trademark Office (USPTO), Lens.org was considered the most exhaustive and accessible tool. In particular, the Lens is a widely recognized open-source tool providing comprehensive coverage of global patent data and scholarly works, allowing to find, extract and manage patent data, aggregating over 153 million patent records from more than 95 jurisdictions globally (*Research Guides: Lens Guide: How to Use Lens.org*, n.d.; Kwilliam, n.d). In fact, it not only granted free access to patent data, but it also provided custom search and download capabilities for all necessary fields. The decision to leverage the Lens platform was not only for its free access to a comprehensive patent dataset, but also due to its ability to allow extraction of both patent titles and their abstracts, particularly important in providing enough semantic context during the NLP analysis.

The patent data focused on filings related to Microsoft Corporation (among which Microsoft Corporation, Microsoft Technology Licensing LLC and Microsoft Mobile OY) and its various legal entities, covering the five-year period from February 21st, 2019 - to September 19th, 2024. This time frame was selected to capture Microsoft's recent investment trends, as well as to provide sufficient historical data to capture the firm's underlying investment behaviour.

3.2.2. Data Pre-processing

After having gathered a comprehensive patent dataset, pre-processing techniques were applied to ensure the data was clean and standardized for subsequent analysis. In particular, three main actions were performed: (1) tokenization, (2) stop words, punctuation & numbers removal and (3) lemmatization. This step was critical for reducing noise (i.e. characters that do not contribute to semantic analysis or context understanding), improving consistency and model performance – reducing words to their base form, making data more digestible - and eliminating bias or inaccuracies in the patent-technology mapping process. Also, patents that were missing titles or abstracts were removed - an essential step for the accurate application of natural language processing techniques. In fact, given the subsequent text analysis relies heavily on these two components, any record lacking this information was excluded from the dataset.

Furthermore, the emerging technologies and their corresponding definitions were extracted from the Gartner Hype Cycle (PDF format) using regular expressions (REGEX), which later formed the basis of an ontology — a knowledge repository for the specific domain. This ontology served as the foundation for linking patents to technologies. Due to the complexity of the document formatting (such as the inclusion of symbols and references to Gartner), this process required some manual intervention. Specifically, the technology list and certain mappings had to be manually linked to ensure accuracy and alignment with the extracted definitions.

3.2.3. Term Frequency-Inverse Document Frequency (TF-IDF)

In parallel with the above-mentioned pre-processing approaches, Term Frequency-Inverse Document Frequency (TF-IDF) was also applied - a widely used statistical technique that measures the importance of words within a document relative to a larger corpus. In this study, TF-IDF was applied to the combination of patent titles and abstracts, allowing the model to capture the most important keywords coming from both sources. TF-IDF allowed for the identification of key terms within each patent that were particularly representative of the technology space, giving higher weight to words that were important within a specific patent but uncommon across the larger patent corpus. Thus, it was a necessary step in preparing the text analysis thanks to its ability to capture the relevance of highly specific terms, particularly important in the patents' domain. As a baseline, the number of keywords selected were 10 – in order to balance the model complexity and interpretability. While a larger number of keywords could have provided more detailed insight into each patent, it would have also increased the risk of diluting the most relevant terms, while heavily worsening computational times.

3.2.4. Ontology Creation

As previously mentioned, an integral part of this study was the creation of an ontology to define and categorize the technologies included in the analysis. The ontology served as a structured representation of the different technologies, their relationships to one another, and their respective definitions. Given the wide range of technologies in this study, a well-defined ontology was necessary to allow for consistent and accurate mapping. Without a clear and standardized ontology, it would have been difficult to ensure patents were being accurately mapped to the correct technologies.

3.2.5. Semantic Analysis

Once the data was pre-processed, the patents were mapped to their respective technologies using three distinct scoring models: (1) cosine similarity with spaCy and (2) OpenAI Embeddings model, as well as (3) fuzzylogic through FuzzyWuzzy. These models were designed to capture different aspects of the text data, ensuring that patents were accurately linked to the most relevant technologies.

In particular, SpaCy was employed to compute the semantic similarity between the top keywords of each patent and the technology definitions within the pre-built ontology. First of all, this approach leveraged SpaCy's ability to capture the terms' deep meaning, going beyond simple keyword matching. However, associating a patent to only one technology would be incorrect, given the underlying technology might fall in different categories (i.e. the patent might overlap different emerging technologies). Thus, in order to ensure a higher model flexibility, Fuzzy Logic was also leveraged to allow for slight variations in wording, making the matching process more robust by capturing connections a strict keyword matching might miss. For instance, different patents descriptions might use synonyms, abbreviations or different phrasing to describe the same underlying concept – leveraging Fuzzy Logic improves the model's ability to match patents and technologies even when they used different terminologies. In addition, a third cosine similarity score was calculated through OpenAI API's embeddings – more specifically, the text-embedding-ada-002 model – a numerical representation of text that captures its semantic meaning in a high-dimensional vector space, trained to understand relationships between words and phrases (or even longer text segments) based on the context they appear in. In particular, it calculates the cosine of the angle between two vectors, namely the patent and ontology. The smaller the angle, the higher the similarity between the two elements, proving useful in capturing nuances that traditional methods might miss. Lastly, an overall score was calculated by adding the three values with equal weight, ensuring no technology or technique disproportionately influenced the final results. In the end, a csv file with over 25,850 patents was produced, each with an associated similarity score to all 25 emerging technologies.

While fairly complete, the use of these techniques also presented some challenges. For instance, the computational cost was significant, requiring several hours or even days to complete, given the large size of the dataset. Moreover, it presented the risk of generating "false positives" when patents shared surface-level similarities without truly representing the same technology.

3.2.6. Visualization: Hype Cycle and Investment Intensity

While these scores were the first steps towards the final, necessary outcome for this study to be successful, their current visualization is not user-friendly – decreasing the model's interpretability and accessibility. Without a clear and intuitive representation, investors would struggle to quickly grasp key insights, making the results harder to analyse, share, or act upon. As a matter of fact, visualization plays a pivotal role in enhancing decision making, making the ability to correctly display data paramount and indispensable (Manchekar et al., 2023).

Thus, this research project's last step involved displaying the output directly on the Gartner Hype Cycle. In particular, the objective was to colour each technology "dot" with its respective, scaled & summed score throughout all patents. Consequently, the visualization step allowed for both an accurate and aesthetically aligned representation of the model's results, enabling quick and easy decision making. On the other hand, this phase also presented a major challenge: ensuring the correct positioning of the coloured "dots" on the Hype Cycle. The original Gartner Hype Cycle image was used as a reference, and an interactive tool was leverage to capture the exact coordinates for each technology's position. These coordinates were then scaled to fit a high-resolution version of the image, allowing for the coloured dot mapping and ensuring an accurate final visualization.

Thus, to accurately estimate and represent the investment intensity of each technology on the Gartner Hype Cycle, the total scores associated with each technology were calculated. Also, given the high variability and potential skewness, a logarithmic transformation was applied to normalize the distribution and reduce the impact of extreme values. This transformation ensured a more balanced dataset, facilitating more meaningful comparisons between technologies. Afterwards, two different scaling approaches were considered:

- **Normalization**: this approach would have meant scaling data relative to the minimum and maximum scores within the dataset. More specifically, the emerging technology with the lowest overall similarity score would have gotten the score of 0, while the one with the highest would have been assigned the value of 1. Although this method ensures the entire range of values is captured, it presents two different limitations:

  a. **Extreme values**: as innovations possibly belong to different technologies at the same time, mapping the lowest value to 0 and the highest to 1 would have respectively eliminated or highly augmented their final output, failing to reflect their real scores.

$$Normalized\ Value = \frac{x - \min(x)}{\max(x) - \min(x)}$$

  b. **Indistinguishable differences**: when values are similar to each other - as it is in this case, where most similarity scores lie within the same range – normalization compresses the differences even further, especially for values in the middle. In fact, the relative differences between datapoints gets compressed in the middle of the range, making it harder to both distinguish and visualize the results.

- **Quartile Scaling**: this approach, on the other hand, spreads data more evenly across the distribution's range, addressing the above-mentioned limitations of normalization. As a matter of fact, it involved calculating the Q1, Q2 and Q3 quartiles, and subsequently scaling the values between the minimum and maximum of each quartile. Consequently, this method emphasizes differences between technologies with similar scores, as well as avoids assigning fixed values to any innovation.

Thus, quartile scaling proved to be a better-suited option in providing more flexibility to the patent-to-technology matching, as well as to ease the output visualization amplifying differences between technologies with similar investment intensities. However, due to the boundaries of each quartile range, values at the extreme ends (min and max) are mapped to 0 and 1, respectively. This boundary mapping allows the scale to capture relative investment differences clearly across all levels, especially at the extremes, ensuring that technologies with marginally different investment levels are visually distinct, making the Hype Cycle representation more meaningful.

Lastly, different colour gradients were applied to reflect investment intensities. For instance, a gradient from dark purple (indicating low investment) to bright yellow (indicating high investment) was used to make the distinction between high and low investment technologies immediately apparent, allowing viewers to quickly identify which technologies were receiving the most attention from Microsoft over the past five years.

3.2.7. The Galletti Index

After having mapped the technology investment intensity over the Gartner Hype Cycle, for a quick and intuitive visualization of Microsoft's most significant medium-to-long term investments, the next step was to find a single, unique proxy that could immediately interpret the analysis. As a consequence, the Galletti Index was introduced - an aggregate measure designed to evaluate a firm's technological investment strategy based on both investment intensity as well as technology attributes, such as their potential impact and time horizon for mainstream adoption.

Hence, a final analysis was built upon the previously calculated similarity scores between patents and innovations, aimed at calculating a single index revealing not only the investment magnitude across Gartner's emerging technologies, but also reflect the potential and time horizon of such innovations. The calculation of the Galletti Index involved several steps, as outlined below:

3.2.7.1. Technology Benefit Score

Through a combination of REGEX and manual input, each innovation's Benefit Rating was extracted from the Gartner Hype Cycle. Manual input was necessary to "polish" a few entries not correctly extracted due to the document's complex formatting. Consequently, emerging technologies were divided into four distinct categories (below as defined by Gartner in the Hype Cycle):

- **Transformational**: "*Enables new ways of doing business across industries that will result in major shifts in industry dynamics*". Weighted at 1.0.
- **High**: "*Enables new ways of performing horizontal or vertical processes that will result in significantly increased revenue or cost savings for an enterprise*". Weighted at 0.75.
- **Moderate: "***Provides incremental improvements to established processes that will result in increased revenue or cost savings for an enterprise*". Weighted at 0.5 – not reflect in 2024's Hype Cycle.
- **Low**: "*Slightly improves processes (for example, improved user experience) that will be difficult to translate into increased revenue or cost savings*". Weighted at 0.25 (not reflect in 2024's Hype Cycle).

(Gartner et al., 2024)

This score reflected each technology's potential disruption on businesses, serving as proxy for a firm's focus on more-or-less impactful innovations.

3.2.7.2. Time-Based Investment Intensity

In parallel with assessing the impact of each technology, a proxy for the innovation's time horizon was quantified. More specifically, the time horizon is a key indicator of the firm's strategic focus: whether it is concentrating on shorter-term returns or long-term innovations. The time horizon was scored leveraging three different inputs, namely (1) "Years to Mainstream Adoption", (2) "Maturity Level" and (3) "Market Penetration" – all extracted through REGEX from the Hype Cycle. The higher the score, the closer to market adoption is the technology. In particular:

3.2.7.2.1. Years to Mainstream Adoption: *The time required for the innovation to reach the Plateau of Productivity* (Gartner et al., 2024). It was scored as follows:

- Less than 2 years: 1.0
- 2-5 years: 0.75
- 5-10 years: 0.5
- More than 10 years: 0.25

Given the complex document formatting of the Gartner Hype Cycle, which prevented a clean and frictionless automatic extraction of the variable, Years to Mainstream Adoption had to be manually mapped and input into the model.

3.2.7.2.2. Maturity Level: the level of maturity of the technology, classified and scored as follows (the earlier the stage, the greater risks for deployment - but potentially higher benefits for early adopters):

Maturity Level	Status	Product/Vendors	Weight
Embryonic	In Labs	None	0.1
Emerging	• Commercialization by vendors • Pilots and deployments by industry leaders	• First generation • High price • Much customization	0.25

Adolescent	Maturing technology capabilities and process understanding Uptake beyond early adopters	Second generation Less customization	0.5
Early Mainstream	Proven technology Vendors, technology and adoption rapidly evolving	Third generation More out-of-box methodologies	0.75
Mature Mainstream	Robust technology Not much evolution in vendors or technology	Several dominant vendors	0.8
Legacy	Not appropriate for new developments Cost of migration constrains replacement	Maintenance revenue focus	0.9
Obsolete	Rarely used	Used/resale market only	0.95

3.2.7.2.3. Market Penetration: how much a product or service is already used by the market, scored as follows:

- Less than 1% of target audience: 0.1
- 1% to 5% of target audience: 0.25

- 5% to 20% of target audience: 0.5
- 20% to 50% of target audience: 0.75

Consequently, the Time-Based Investment Intensity Score was calculated as weighted average of the three scores:

$$\boldsymbol{Time - Based\ Investment\ Intensity\ Score} = (0.5\ x\ Years\ to\ Mainstream\ Adoption) + (0.25\ x\ Maturity\ Level) + (0.25\ x\ Market\ Penetration)$$

More weight was given to the "Years to Mainstream Adoption" variable, as it better serves as proxy compared to the other two variables in evaluating the innovation's expected time-horizon.

3.2.8. Combining the Scores: The Overall Technology Index

Afterwards, for each technology, the three components were combined (i.e. the Investment Intensity, Technology Benefit and Time-Based Investment Intensity), in order to calculate the Overall Technology Index. More weight was given to the Investment Intensity, being the most significant proxy for investment focus:

$$\boldsymbol{Overall\ Technology\ Index} = (0.5\ x\ Investment\ Intensity) + (0.25\ z\ Technology\ Benefit\ Score) + (0.25\ x\ Time - Based\ Investment\ Intensity)$$

3.2.9. Calculating the final Galletti Index

Finally, in order to calculate the single, unique & final index that will reflect the whole analysis, the Overall Technology Index was summed across all technologies and scaled to a range 1-10. In particular, scaling was performed by dividing for the theoretical maximum score (i.e. highest possible score in all technologies), and then multiplying by 10:

$$Galletti\ Index = \left(\frac{\sum Overall\ Technology\ Index}{Max\ possible\ Index\ Sum}\right) x\ 10$$

However, as multiple variables were taken into account when calculating the Galletti index, two composite additional scores were calculated to increase the interpretability of results:

- **Time Horizon Score**: The average of all the Time-Based Investment Intensity scores, reflecting how much the firm is investing in long-term vs. short-term technologies. Scaled from 0-1.
- **Technology Benefit Score**: The average of all the Technology Benefit Scores, reflecting how much the firm is investing in high-impact vs. low-impact technologies. Scaled from 0-1.

As a matter of fact, the Galletti Index alone can provide informative results, but it is advised to observe it alongside the above-mentioned scores. The interpretation of results, based on the calculations performed, is as follows:

- **High Galletti Index (7 - 10):**
    - **High investment in high-impact, shorter-term technologies**: The firm is making significant investments in technologies that will likely become mainstream soon (compared to the other innovations highlighted in the Gartner Hype Cycle), and that are likely to bring transformational disruptions or high impact to industries.

- **Mid-Range Galletti Index (4 - 7)**:
    - **Mixed strategy**: in case the score falls into this category, it needs to be interpreted by examining the **Time Horizon** and **Technology Benefit Scores**, in order to extract useful insights. For instance, a mid-range Galletti index could result from a very high Technology Benefit and a low Time Horizon score (or vice versa).

- **Low Galletti Index (0.0 - 4)**:

- **Investment in lower-impact, longer-term technologies**: The firm is investing less in short-term, high-impact technologies and focusing more on longer-term, lower-impact areas.

While the preference stems from an investor's focus and strategy, a High Galletti Index is often preferable, as it reflects the firm has invested in highly transformative technologies close to full-scale adoption.

Results

In the Results section, the output of this research project will be outlined with minimal analysis. An in-depth assessment and review will follow in the Discussion section.

4.1. Investment Intensity

The first, significant output of the model is the Investment Intensity score (calculated during the *visualization step*) associated with each technology, representing the log-transformed and quartile-scaled semantic similarity of patents, showcasing Microsoft's investment focus spanned across Gartner's 25 emerging technologies present in the 2024 Hype Cycle. Below a table (Table 1) displaying the final results from the calculations, categorized in 4 different quartile groups in descending order (highest investment focus to lowest). In particular, the 4 groups are defined as follows:

- **Lower Quartile:** Investment Score ≤ 0.25
- **Lower-Mid Quartile:** $0.25 <$ Investment Score ≤ 0.5
- **Mid-High Quartile:** $0.5 <$ Investment Score ≤ 0.75
- **Higher Quartile:** Investment Score > 0.75

Table 1 (approximated to 2 decimal numbers):

Gartner Emerging Technology	Quartile Group	Investment Intensity

AI-Augmented Software Engineering	High	1.00
Spatial Computing	High	0.89
Digital Twin of a Customer	High	0.83
Machine Customers	High	0.83
Large Action Models	High	0.81
Digital Immune System	High	0.8
Superapps	Mid-High	0.75
Generative AI	Mid-High	0.69
Prompt Engineering	Mid-High	0.64
AI TRiSM	Mid-High	0.58
AI Supercomputing	Mid-High	0.56
Cybersecurity Mesh Architecture	Mid-High	0.51
Disinformation Security	Low-Mid	0.50
Reinforcement Learning	Low-Mid	0.49
Autonomous Agents	Low-Mid	0.44
Multiagent Systems	Low-Mid	0.39
Internal Developer Portals	Low-Mid	0.3
Artificial General Intelligence	Low	0.25
GitOps	Low	0.25
WebAssembly	Low	0.20
Cloud-Native	Low	0.18
Homomorphic Encryption	Low	0.15
Federated Machine Learning	Low	0.10
Humanoid Working Robots	Low	0.10
6G	Low	0.00

Being the scores quartile-scaled, they can be divided into 4 distinct categories:

- High Investment Focus: Investment Intensity Score ≥ 0.75

Microsoft’s investment efforts (Score) are currently concentrated mainly across AI-Augmented Software Engineering (1.00), Spatial Computing (0.89), Digital Twin of a Customer (0.83) and Machine Customers (0.83), followed by respectively Large Action Models (0.81) and Digital Immune Systems. Among other similarities, all share the implementation of some form of Artificial Intelligence, real-time data processing and predictive modeling to automate complex-decision making processes, replacing or supporting human intervention.

- Medium-High Investment Focus: 0.5 ≤ Investment Intensity Score > 0.75

On the other hand, we see some of the currently most hyped technologies - such as Generative AI (0.69) and Prompt Engineering (0.64) with a mid-high ranking score. Furthermore, we also find some security-related technologies, namely Cybersecurity Mesh Achitecture (0.51) and Disinformation Security (0.5) – alongside Superapps (0.75), AI TRiSM (0.58) and AI Supercomputing (0.56).

- Low-Medium Investment Focus: 0.52 ≤ Investment Intensity Score > 0.5

In the third category, we see certain technologies positioned with moderate investment intensities, such as Artificial General Intelligence (0.25), which is classified in an embryonic phase with over ten years until mainstream adoption - indicating it is not an immediate priority for Microsoft. Similarly, Reinforcement Learning (0.49), Internal Developer Portals (0.3), and GitOps (0.25) hold mid-to-lower scores, highlighting a focus that may develop over time.

Additionally, the mid-quartiles feature various agent and security-related systems. Technologies like Autonomous Agents (0.44) and Multiagent Systems (0.39) represent this category, reinforcing Microsoft’s continued investment in emerging solutions for automated and secure environments.

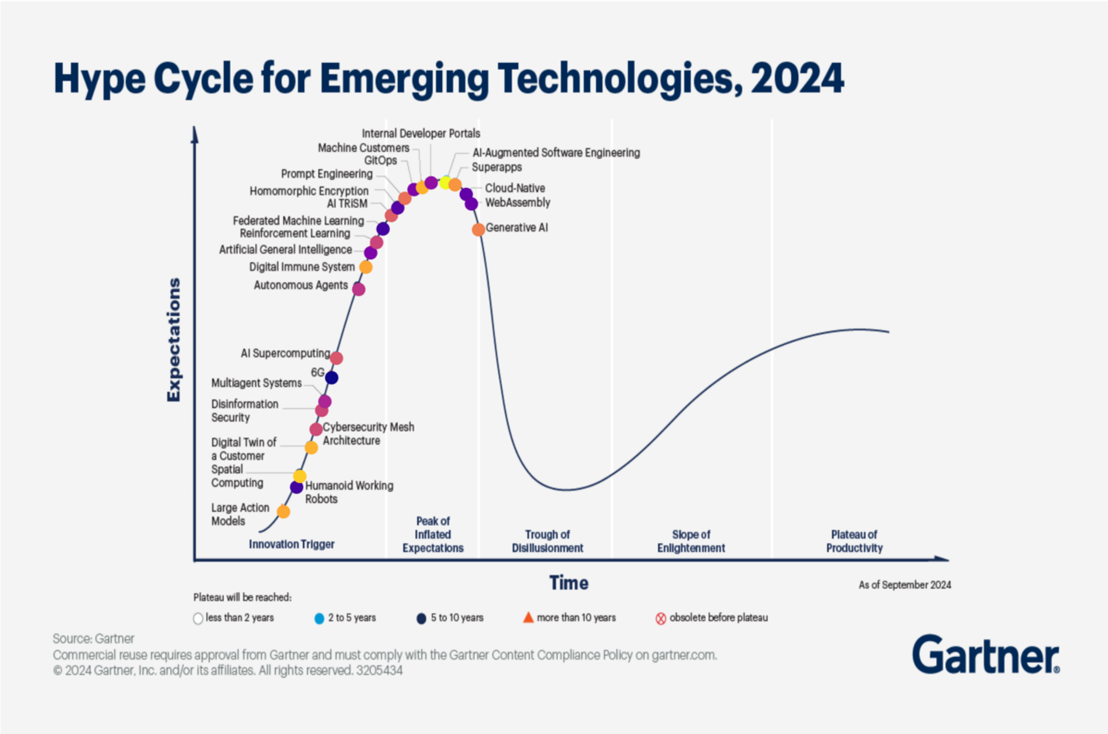
- Low Investment Focus: Investment Intensity Score < 0.25

On the lower-end of the spectrum, on the other hand, we find Cloud-related technologies, such as WebAssembly (0.2) and Cloud-Native (0.18) – where the firm has heavily invested in previous years through Azure -, or innovations not particularly aligned with Microsoft's core business, namely 6G (0.0) and Humanoid Working Robots (0.1). Homomorphic Encryption (0.15) and Federated Machine Learning (0.1) are also among the lowest rated technologies.

4.2. Visualization: an immediate impact

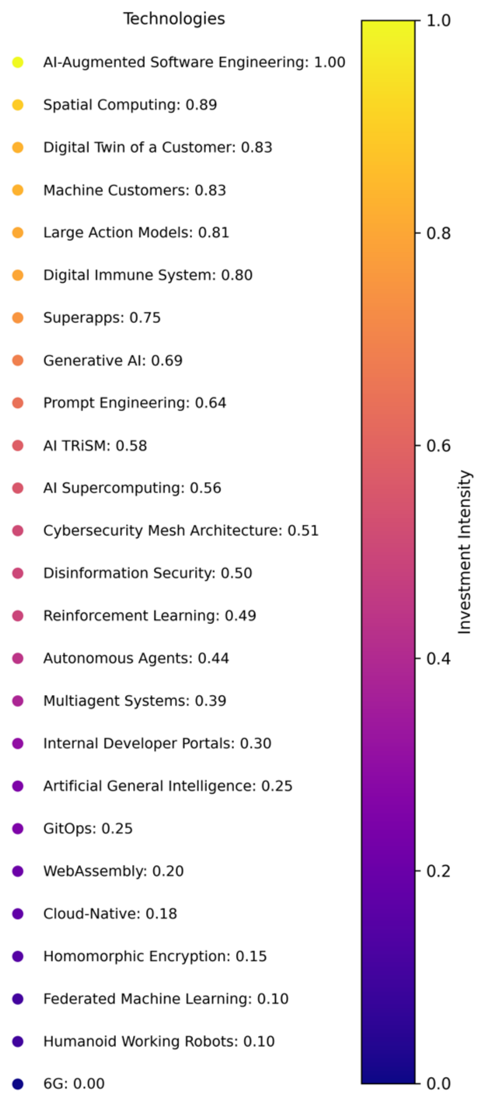
Complementing the quantitative analysis, an intuitive representation (Figure 4) of the investment intensity across all technologies over the 2024 Gartner Hype Cycle was designed. In particular, the image of the framework was manipulated, on which coloured dots – each representing the previously-calculated log-transformed & quartile-scaled investment intensities – were added. Thus, it provides a quick and effective way to assess which technologies are receiving the most attention in terms of R&D efforts. It is also worth mentioning that, while the color was altered, the position of all technologies over the Hype Cycle curve has not been modified.

Figure 4:



The color scale (as represented in Figure 5) ranges from deep blue (representing lower investment) to bright yellow (representing the highest levels of investment). This gradient allows for immediate identification of technologies with high, moderate, and low industry investment at a glance. Thus, technologies with higher investment intensity are displayed in warmer colors (e.g., orange, yellow), while those with lower investment are represented in cooler colors (e.g., purple, blue).

Figure 5:



4.3. Calculating the Indexes

The Indexes Calculation Step provided a comprehensive analysis of emerging technologies by assessing multiple criterias, namely Benefit Rating, Market Penetration, Maturity, Years to Mainstream Adoption and the newly developed Investment Intensity Score. In particular, for each technology, three dimensions were scored and taken into account:

- Time-Based Investment Intensity Score
- Technology Benefit Score
- Patent Investment Intensity (already leveraged in previous step)

4.3.1. Time-Based Investment Intensity Score

In order to assess Microsoft's time-horizon investment focus, the Time-Based Investment Intensity Score was calculated, resulting as shown in Table 2 (sorted in descending order).

Table 2:

Gartner Emerging Technology	Time-Based Investment Intensity
GitOps	0.75
AI TRiSM	0.6875
AI-Augmented Software Engineering	0.6875
AI Supercomputing	0.625
WebAssembly	0.625
Generative AI	0.625
Prompt Engineering	0.5625
Disinformation Security	0.5625
Superapps	0.5
Cloud-Native	0.5
Autonomous Agents	0.4
Spatial Computing	0.375
Digital Immune System	0.375
Reinforcement Learning	0.375
Digital Twin of a Customer	0.3375
Federated Machine Learning	0.3375
Homomorphic Encryption	0.3375
Multiagent Systems	0.3
Internal Developer Portals	0.25
Machine Customers	0.2125
Humanoid Working Robots	0.2125
Large Action Models	0.175
Cybersecurity Mesh Architecture	0.175
Artificial General Intelligence	0.175

6G	0.175

The higher the score, the closer to market adoption is the technology.

4.3.2. Technology Benefit Intensity Score

Serving as proxy for Microsoft's focus on more-or-less impactful innovations, it reflected each technology's potential impact on the market, as outlined in Table 3 below. The higher the score, the higher the forecasted impact of the technology.

Table 3:

Benefit Rating	Score	Technologies
Transformational	1.0	Large Action Models AI-Augmented Software Engineering Generative AI Artificial General Intelligence Digital Twin of a Customer AI Supercomputing Cybersecurity Mesh Architecture WebAssembly
High	0.75	Spatial Computing Disinformation Security AI TriSM

		- Reinforcement Learning - Machine Customers - Superapps - Prompt Engineering - Digital Immune System
Moderate	0.5	- Federated Machine Learning - Cloud-Native - Internal Developer Portals - Multiagent Systems - Autonomous Agents
Low	0.25	- Homomorphic Encryption - 6G - GitOps - Humanoid Working Robots

4.3.3. Overall Technology Index

By weighting all three dimensions together, namely Investment Intensity, Time-Horizon and Technology Benefit Score, the Overall Technology Index was calculated for each technology, as outlined in Table 4.

Table 4 (approximated to 2 decimal numbers):

Gartner Emerging Technology	Overall Technology Index
AI-Augmented Software Engineering	0.92
Generative AI	0.75
Digital Twin of a Customer	0.75
Spatial Computing	0.73
Large Action Models	0.70
Superapps	0.69
AI Supercomputing	0.68
Digital Immune System	0.68
AI TRiSM	0.65
Machine Customers	0.65
Prompt Engineering	0.65
Disinformation Security	0.58
Autonomous Agents	0.57
Cybersecurity Mesh Architecture	0.55
Reinforcement Learning	0.53
WebAssembly	0.51
GitOps	0.50
Artificial General Intelligence	0.42
Homomorphic Encryption	0.41
Cloud-Native	0.40
Internal Developer Portals	0.40
Humanoid Working Robots	0.35
Federated Machine Learning	0.0.32
6G	0.23

4.3.4. The Galletti Index

Lastly, the Galletti Index was calculated by summing the Overall Technology Indexes across all innovations, and scaling by the theoretical maximum score and multiplying by 10. The result was a number between 1 and 10:

Microsoft's Galletti Index: 5.63

However, in order to better interpret the result, the (1) Overall Time Horizon Score and the (2) Overall Technology Benefit Score were also calculated, providing further insights into the Galletti Index:

Overall Time Horizon Score: 0.41

Overall Technology Benefit Score: 0.86

Discussion

Several insights can be gathered from the analysis and output generated from the developed model. Starting from a broader industry perspective, and then narrowing down to Microsoft's strategic investment focus, the following trends immediately catch the eye:

5.1. AI Industry Focus

Gartner's 2024 Hype Cycle innovations clearly reveal a deep focus on AI as the central pillar of its emerging technologies portfolio, as the majority of innovations are powered by or closely tied to AI. Nearly every category, in fact, whether related to cloud-native solutions, cybersecurity, or software development, involves some form of AI integration: from Generative AI, which promises to revolutionize content creation, to AI-Augmented Software Engineering, aimed at automating and optimizing development processes, AI is at the center of nearly all key emerging technologies. Furthermore, technologies such as Reinforcement Learning, Artificial General Intelligence and AI TRiSM are also driven by AI, demonstrating the optimistic market sentiment towards such technology, making AI pivotal to the future of nearly all innovations and indutries – even areas like cybersecurity mesh architecture and disinformation security, traditionally separate from AI, are now evolving with AI-driven enhancements, showing how AI's influence is pervasive across diverse sectors. Thus, AI's consistent presence across Gartner's emerging technologies reflects a broader industry understanding: AI is not just an add-on feature but a transformative force reshaping the entire technological landscape.

In addition to AI's single, core pillar across all technologies, Gartner has identified 4 key distinct Trends that are shaping the future of the technology landscape, namely (1) "Autonomous AI", (2) "Boost Developer Productivity", (3) "Empower With Total Experience" and (4) "Deliver Human-Centric Security and Privacy":

5.1.1. Autonomous AI: refers to the technology category of autonomous and self-improving AI systems that can operate with minimal human oversight. These innovations, such as as Multiagent Systems, Large Action Models, Autonomous Agents and Reinforcement Learning, are enabling a future where AI can perform any task traditionally done by humans.

Full Technology List: AI supercomputing, Artificial general intelligence, Autonomous agents, Generative AI, Humanoid Working Robots, Large action models, Machine customers, Multiagent Systems, Reinforcement Learning.

5.1.2. Boosting Developer productivity: refers to the group of innovations focusing on developers & their role, streamlining collaboration, speeding-up and improving the overall coding experience. Technologies like AI-Augmented Software Engineering, Cloud-Native development, GitOps, and Prompt Engineering are revolutionizing how developers design and deliver software. These technologies help developers focus on creative problem-solving while reducing the overhead associated with tedious coding tasks, significantly improving their productivity.

Full Technology List: AI-augmented software engineering, Cloud-native, GitOps, Internal Developer Portals, Prompt engineering, WebAssembly.

5.1.3. Empowering With Total Experience: refers to the technologies aimed at combining customer, employee, user and multi-experience to create shared

experiences, empowering all stakeholders and addressing critical interactions. Technologies belonging to this group include Digital Twin of a Customer, Spatial Computing, Superapps and 6G.

Full Technology List: 6G, Digital Twin of a Customer, Spatial Computing, Superapps

5.1.4. Delivering Human-Centric Security and Privacy: refers to the technology category focused on building shared trust and awareness among internal and external teams by strenghening security and privacy. Technologies like AI TRiSM (Trust, Risk and Security Management) and Cybersecurity Mesh Architecture belong to this group.

Full Technology List: AI TriSM, Cybersecurity Mesh Architecture, Digital Immune System, Disinformation Security, Federated Machine Learning, Homomorphic encryption.

(Gartner, 2024).

Across all these areas, AI is clearly a key enabler driving innovation. Whether it's enabling autonomy, enhancing developer efficiency, empowering user experiences or securing systems, AI is at the core of technological advancements, reflecting its integral role in shaping the future of all industries.

5.2. Microsoft investment strategy – Macro and Micro analysis

In order to provide a comprehensive and exhaustive analysis, the results section will initially address Microsoft's investment focus both at a grouped (i.e. Gartner's 4 Emerging Trends categories) and individual level. This approach allows for a better understanding of the broader trends Microsoft is betting on (Macro analysis), as well as its most specific and strategic focus (Micro analysis).

5.2.1. Autonomous AI – **Total Score:** 4.56

Given the highest cumulative score, Autonomous AI is the area of investment with the most interest from Microsoft. While it also represents Gartner's widest group, as it includes the highest number of technologies (i.e. 9), its aggregated score more than doubles the second ranked category (Delivering Human-Centric Security and Privacy), making it undoutedly the top investment priority for the American firm. However, the investment is mostly weighted towards Machine Customers (0.83), Large Action Models (0.81) and Generative AI (0.69), rather than non-core or futuristic innovations such as Humanoid Working Robots (0.1) & Artificial General Intelligence (0.25).

Interestingly enough, we find closely-related technologies within the same Gartner category being higher scored compared to others, as they often support each other's development and realization. For instance, taking Machine Customers as an example – defined as nonhuman economic actors that obtain goods or services in exchange for payment (Gartner, 2024) – we can expect other closely-related technologies, such as Generative AI and Large Action Models, to be also a top-priority for Microsoft within the Autonomous AI category. More specifically, we can hint the American firm could integrate the Machine Customers concept into their recently created Microsoft Supply Chain (SC) Center platform, signaling a strong entrance into the industry (while no explicit announcements have been made regarding Microsoft's investment in Machine Customers, the lack of publicly available information makes our research particularly valuable, as it anticipate Microsoft's future strategic developments). This new service integrates with Dynamics 365 and Azure AI to provide a unified interface that allows businesses to manage, automate and optimize supply chain operations through advanced AI (O'Donnell, 2022). Consequently, with over 15 billion of internet-connected machines, combined with AI's advancements, machines are expected to evolve from advisors to decision makers in economic transactions - autonomously buying, selling and requesting services. The Microsoft Supply Chain Center offers the necessary infrastructure to support this industry shift. In fact, by leveraging ML and real-time data analytics, Microsoft is laying the groundwork for AI systems to autonomously manage transactions and supplier relationships, potentially evolving into fully functioning machine customers. While it is still at an Emerging stage, few examples of early machine customers include HP Instant Ink and Tesla cars that order their own supplies. Within this context, Generative AI (Gen AI) and Large Action Models (LAMs) play a crucial role in enabling Machine Customers, accellerating their development and deployment. In particular, Generative AI technologies can create new content, strategies and designs by learning from datasets, while LAMs AI systems are designed not only to provide recommendations, like large language models (LLMs), but also to take actions based on those recommendations to meet a goal, both in digital and physical environments. Thus, in the context of Machine Customers, the synergy with GenAI and LAMs is foundational. Machine customers, which autonomously buy, sell and request services, require both the creativity and content generation capabilities of Gen AI, as well as the autonomous decision-making capabilities of LAMs. For example, while GenAI can generate tailored offers or predict future demands, LAMs can act on those predictions, re-ordering and negotiating supplies or engaging with suppliers autonomously (Gartner, 2024). Other technologies within the same Gartner category, such as Autonomous Agents (0.44), combined systems that achieve goals without human intervention through AI capabilities and Robotic Process Automation (RPA) to trigger actions, or Multiagent Systems (0.39) - a system of multiple, indipendent AI Agents - will also contribute to support Machine Customers development.

However, it is also worth mentioning that, while Generative AI, LAMs and Artificial General Intelligence might not be the highest rated technologies based on patent information, Microsoft has been heavily investing in the technologies through other means, namely the recent 49% stake acquisition in Open AI's profit arm, investing a total of US$13.75B since 2019 (US$ 6.6B only in the latest fundraise closed in October 2024) (Jin & Driebusch, 2024).

Detailed scoring (Table 5):

Gartner Group: Autonomous AI	Total Score: 4.56
Machine Customers	0.83
Large Action Models	0.81
Generative AI	0.69
AI Supercomputing	0.56
Reinforcement Learning	0.49
Autonomous Agents	0.44

Multiagent Systems	0.39
Artificial General Intelligence	0.25
Humanoid Working Robots	0.10

5.2.2. Booasting Developer Productivity – **Total Score:** 2.57

While the "Boosting Developer Productivity" category scored lower compared to "Autonomous AI", the technology on which Microsoft is investing the most – AI-Augmented Software Engineering (AIASE) – belongs to this group. As a matter of fact,

Microsoft's significant investment in AI-Augmented Software Engineering technologies is part of its broader strategy to boost developer productivity and software quality while reducing operational costs. AIASE refers to the use of AI technologies to assist software engineers throughout the software development life cycle, from creation to deployment and maintenance (Gartner, 2024). Investments in this technlogy area can be traced to the augmentation and further development of tools like GitHub Copilot – Microsoft's ML-powered IntelliCode extension – which allows developers to code faster, prevent and identify inconsistencies early in the development lifecycle and automatically set meetings, generate tests and documentation, minimizing repetitive, tedious tasks - leaving engineers more time to focus on creative and problem-solving actions (*Research Talks: AI for Software Development - Microsoft Research*, 2021). As a matter of fact, GitHub Copilot, conceived in 2020 with the aim of ease developers' tedious tasks, has rapidly gained traction, being adopted by over 77,000 organizations worldwide (180% YoY adoption increase at the end of 2023), including industry leaders like BBVA, FedEx and InfoSys. Currently, Copilot drives over 40% of GitHub's $2 billion annual revenue run rate, with users doubling quarter-over-quarter, highlighting its high industry influence (Devdiscourse News Desk, 2024b; Ramel, 2024). Consequently, being Copilot an already established tool with both a high organizational impact and revenue-generating ability, seeing AIASE scoring the highest Investment Intensity seems logical, reflecting the substantial investments Microsoft has made to develop and scale this technology. In fact, as expected, the AIASE's market penetration is currently in the highest range (5% to 20% of target audience), as highlighted by Gartner (Gartner, 2024).

Detailed scoring (Table 6):

Gartner Group: Developer Productivity	**Total Score: 2.57**
AI-Augmented Software Engineering	1.00
Prompt Engineering	0.64
Cloud-Native	0.18
GitOps	0.25
Internal Developer Portals	0.30
WebAssembly	0.20

5.2.3. Empowering with Total Experience – **Total Score:** 2.47

Looking at the third highlighted category - Total Experience – almost all technologies score high in Investment Intensity. First of all, taking a closer look at Digital Twin of a Customer – a virtual representation of a client leveraged to simulate or emulate certain behaviours – we notice it has been on Microsoft's radar for a while. As a matter of fact, the American firm already has a *Digital Twin of a Customer* offering, the "Azure Digital Twins", already virtually operating for real-world systems such as factories, cities and supply chains (*Digital Twins – Modeling and Simulations | Microsoft Azure*, n.d.). Similarly, Microsoft has also been heavily investing into Spatial Computing – a set of innovative technologies, such as Augmented Reality (AR) or Virtual Reality (VR), combining physical and digital objects in a shared frame of reference - through its HoloLens and Mixed Reality VR, both tools developed to break into this nascent market, already in use in sectors like healthcare, education and manufacturing. While Spatial Computing represents the next frontier of Total Experience, Microsoft's investment in the field is not surprising, as its integration into already existing Azure portfolio could allow for the development of more complex simulations and 3D environments (*Microsoft HoloLens | Tecnologia Di Realtà Mista per Le Aziende*, n.d.; (Lolambean, n.d.). As a matter of fact, the American Firm has

also been cited by Gartner as a sample Vendor for this technology in the Gartner Hype Cycle.
Continuing with the Total Experience category, Superapps also exhibit significant investment intensity, which aligns with Microsoft's broader vision of providing a one-stop, holistic and integrated digital solution - Superapps are, in fact, platforms that combine an ecosystem of multiple services and functionalities into a single app, creating an all-in-one experience for users across various sectors, such as payments, messaging and social media. In the context of Microsoft, the company is actively developing products and services around this innovative technology. Microsoft Power Platform, for instance, is one of the firm's first attempts at building an embryonic Superapp - a suite of apps, services, connectors as well as a data platform that allows enterprises to create modular and customized applications with capabilities that span multiple domains, such as data analysis, workflow automation or AI integration (Microsoft Power Platform, n.d.). Consequently, the firm has been already featured as sample vendor for Superapps in the Gartner's Hype Cycle. However, it is rumoured that Microsoft's is eyeing something much more powerful. As a matter of fact, recent reports suggest that Microsoft has been exploring the creation of a consumer-facing Superapp – following WeChat's steps - that would unify its Bing search engine, Teams collaboration tool as well as Azure cloud services into one cohesive platform (Reuters, 2022). Although rumours about Microsoft's entry into the Superapp space circulated in 2022, no formal announcements have been made, leaving speculations unconfirmed. However, patent data suggest otherwise: Microsoft appears to be actively exploring this technology, and a significant move into the Superapp field is likely on the horizon (Ngoma, 2022, Gartner, 2024). Lastly, Microsoft's limited investment in 6G can be explained by its strategic focus on other technologies over telecommunications. Over the past five years, as patent data show, 6G has not emerged as a priority for the company, despite its future potential to revolutionize connectivity with faster speeds, lower latency, and immersive experiences. Instead, Microsoft has chosen to concentrate on cloud, AI and digital transformation technologies, which align more closely with its existing business models and offer immediate returns (Gartner, 2024).

To sum up, Technologies like *Digital Twin of a Customer* and *Spatial Computing* received high scores, reflecting Microsoft's commitment to enhancing customer experiences through Azure Digital Twins and HoloLens. Meanwhile, *Superapps* align with Microsoft's integrated platform strategy, supported by tools like Power Platform. Conversely, *6G* saw has not been a strategic focus in the past 5 years – and will most likely not be one of Microsoft's business drivers in the medium to long term – given the time needed to research and develop such technology.

Detailed scoring (Table 7):

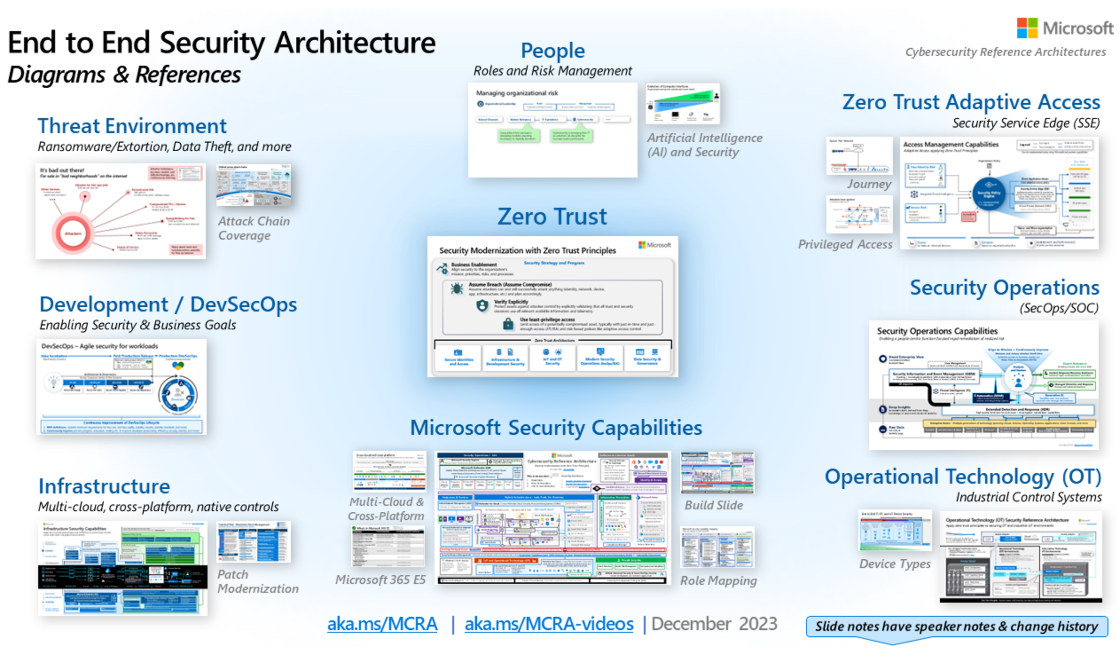
Gartner Group: Total Experience	Total Score: 2.47
Spatial Computing	0.89
Digital Twin of a Customer	0.83
Superapps	0.75
6G	0.00

5.2.4. Human-Centric Security and Privacy – **Total Score:** 2.64

Regarding the fourth and last Gartner category, Human-Centric Security and Privacy, the aggregated Investment Intensity scored similar to Total Experience and Developer Productivity, with some technologies worth analyzing. Starting with the highest-scored innovation, Digital Immune System – a technology that enables a secure software development, making applications more resilient to security breaches in order to quickly recover from failures – represents an area of high interest for Microsoft. As a matter of fact, Gartner highlights that by 2025, companies investing in DIS could see an 80% reduction in downtime, directly improving customer satisfaction and system resilience (Gartner, 2024). Consequently, it does not come as a surprise for the American firm to be investing in such technologies, given Digital Immune Systems create self-healing, resilient infrastructures, enabling continuous monitoring and proactive protection against security vulnerabilities - Microsoft aims at reducing system downtime, improve operational resilience and security measures in its Azure Cloud environment. Furthermore, we also see AI TRiSM, Cybersecurity Mesh Architecture and Disinformation Security with a mid-range score. Starting with AI TRiSM (Trust, Risk and Security Management), an ecosystem of digital solutions (risk management, governance, data protection etc…) aimed at ensuring model and application transparency, content anomaly and protection, providing compliance with EU regulations, plays a crucial role in handling the security challenges presented by AI. While not only associated with Generative AI-related technologies, such as GPT models, which falls under the broader AI umbrella and introduces specific risks related to bias, privacy and misinformation, it does represent one of its major use cases. In fact, given that Gen AI systems autonomously generate content or make decisions, AI TRiSM becomes significantly important in helping address transparency and security challenges by implementing robust monitoring and bias mitigation mechanisms (Gartner, 2024, *AI TRISM: What It Is & Why It's Important | Splunk*, n.d.; *AI Trust, Risk, and Security Management (AI TRISM) - Navikenz*, 2024). Consequently, we see AI TRiSM's Investment Intensity Score as reasonable, as it fairly aligns with the investment efforts Microsoft has made for Generative AI, Prompt Engineering or LAMs – as security investments go hand-in-hand with technology development. Similar to AI TRiSM, Disinformation Security also closely relates to content-generation technologies, given the increasing capabilities of AI systems in creating highly convincing text, images, videos, and even deepfakes, which can be misused to propagate false information and deceive the public. Thus, Disinformation Security – a suite of technologies aimed at addressing misinformation challenges to help enterprises discern trust, protect their brand and secure their online presence – also sees an Investment Intensity score close to AI TRiSM (Gartner, 2024). Over the past years, in fact, Microsoft has been publicly releasing software and promoting initiatives to contrast the spread of deepfakes or false information, such as Microsoft Video Authenticator, a tool designed to detect deepfake content by analysing photos and videos for subtle manipulations that may not be visible to the human eye. Other tools have also been released across Azure services that can detect media manipulated content (Burt & Horvitz, 2021). Furthermore, in collaboration with NewsGuard, Microsoft is also helping users navigate online information by identifying credible news sources and flagging potential misinformation, and it has sealed multiple partnerships – such as with AI Foundation and media outlets (i.e. BBC, CBC, Radio-Canada etc..) to create a standardized approach to verifying the legitimacy of media content. Additionally, Microsoft and OpenAI recently launched a $2M fund aimed at combatting election deepfakes, in order to contracst deepfakes used to deceive votes (Sawers, 2024).

In addition, Cybersecurity Mesh Architecture (CSMA) – an emerging security approach combining a wide array of tools for architecting a composable, adaptable framework - stands out with a mid-range Investment Intensity score in the Human-Centric Security and Privacy category, reflecting Microsoft's ongoing efforts to invest in and develop this area. Alongside AI TRiSM and Disinformation Security, combined with Microsoft's recent announcements and initiatives in these fields, indicate the firm's consistent and continuous investments efforts (Gartner, 2024). As a matter of fact, Microsof has already been implementing an embryonic CSMA approach, as described by the firms's Cybersecurity Reference Architecture (MCRA), illustrating Microsoft's cybersecurity capabilities and technologies – aligning with CSMA principles and focusing on integrating diverse security solutions into a cohesive, composable architecture (Figure 6) that allows for a decentralized and adaptive security across Microsoft's complex digital infrastructure. Services like Microsoft 365 and Azure benefit from this approach (MicrosoftGuyJFlo, 2024).

Figure 6:



(Microsoft Cloud Security Reference Architecture - MicrosoftGuyJFlo, 2024)

Lastly, scoring quite low compared to other innovations within (and outside) belonging to the same Gartner category, we find Federated Machine Learning (FedML) and Homomorphic Encryption (HE) – both within the lowest quartile. Starting with FedML, an innovative technology for compliance aimed at training an ML algorithm across multiple decentralized devices (i.e. silos), to later aggregate the partial models into an orchestrator, without the need of sharing data - resolving data-transfers bottlenecks and training better quality models – we notice it is ranked low. While Microsoft has not heavily invested in the technology through patents in the past 5 years (as highlighted by the low Investment Intensity Score), we do notice the firm has taken its first steps in integrating such approach in its Azure Machine Learning (Gartner, 2024; Federated Learning With Azure Machine Learning: Powering Privacy-Preserving Innovation in AI, n.d.). In fact, its *AzureML SDK v2* offers capabilies to implement FedML pipelines, as illustrated by Figure 7. However, despite these advancements, the relatively low Investment Intensity score (0.10) shows that Microsoft has not heavily prioritized this technology. This could be due to several factors, such as being a technology not widely known within an enterprise, lacking marketing both on the vendor and research side, or by the possibly lower investments required compared to other technologies. However, as LLM models evolve (and with it also the need for high computing power and capabilities), FedML and FedLLM (Federated Large Language Models) research becomes increasingly relevant, to allow for a group of organizations to train LLMs (and similar, complex models) together, allowing for a high degree of collaboration without sharing sensitive information, increasing model generalibility and performance as well as reducing computation costs (Gartner, 2024). Thus, it can be expected to see Microsoft increasing its investment efforts on this topic in the upcoming years.

Figure 7:

(AzureML SDK v2 FedML pipeline, Federated Learning With Azure Machine Learning: Powering Privacy-Preserving Innovation in AI, n.d.)

Homomorphic Encryption, on the other hand, refers to the practise of leveraging cryptographic algorithms to perform computations on encrypted data (without decrypting it). While little to no investments have been made in the past 5 years to support this innovation, Microsoft has released a few open-source tools, such as Microsoft Simple Encrypted Arithmetic Libray (SEAL), to enable the first, emerging use cases. For instance, organizations often send, receive and store cloud data in an encrypted state for security. However, to leverage the full potential of cloud computing, they usually need to either decrypt the data or share encryption keys with service providers, which poses further security risks. Homomorphic encryption addresses this challenge by allowing computations to be performed directly on encrypted data. This enables companies to maintain the confidentiality of their sensitive information while still utilizing Azure's cloud-based processing, offering a secure solution for privacy-critical data (Jocontr, n.d.). Consequently, by making Microsoft SEAL open source, Microsoft has contributed foundational work to the technology field, without committing significant resources to further develop it themselves. This suggests that while they recognize the importance of this technology, it may not align with their current core priorities or strategic focus areas. By releasing SEAL to the public, Microsoft allows other researchers and organizations to build upon their work, accelerating advancements in the field.

Detailed scoring (Table 8):

Gartner Group: Human-Centric Security and Privacy	Total Score: 2.64
AI TRiSM	0.58
Cybersecurity Mesh Architecture	0.51
Digital Immune System	0.80
Disinformation Security	0.50
Federated Machine Learning	0.10
Homomorphic Encryption	0.15

5.2.5. Macro & Micro analysis conclusion

Even at this stage of the analysis, before digging into the broader implications of Microsoft's Galletti Index score, analysing whether it leans toward long vs short-term investments or high vs low-impact technologies, the above analysis at both Macro (Gartner categories) and Micro (individual technologies) level already provides valuable insights. In fact, looking at Gartner's four highlighted categories, it becomes clear where Microsoft's investments are focused. For instance, Autonomous AI technologies, such as Machine Customers & Generative AI, showcase Microsoft's efforts towards enterprise autonomous systems, for which human intervention would become almost obsolete. Meanwhile, the focus on Developer Productivity with AI-Augmented Software Engineering as major investment priority reflects Microsoft's commitment to enabling faster and more efficient software development through the augmentation of AI-powered tools like Copilot – continuing a trend the American firm has been pursuing for years. Furthermore, regarding the Total Experience category, Digital Twin of a Customer and Spatial Computing emphasize Microsoft's focus on real-world system simulations (segment for which the firm has already developed some digital solutions), while Superapps suggest a focus on integrating multiple services for a unified user experience. Lastly, in the Human-Centric Security and Privacy category, mid-range scores for AI TRiSM, Cybersecurity Mesh Architecture, and Disinformation Security point to Microsoft's growing focus to secure its AI-driven platforms and protect user data – an ongoing effort to integrate into its already existing cloud offering.

On the other hand, looking at the most significant insights derived from individual technologies, the high investment intensity in AI-Augmented Software Engineering is no surprise, given the significant success and wide adoption of Copilot, Microsoft's flagship tool for boosting developer productivity. As a matter of fact, the latter provides AI assistance in every stage of the software development lifecycle - from code generation to debugging, testing and documenting – making Microsoft's further investments in the field focused on augmenting its capabilities as well as the firm's leading position in the industry, contributing to Microsoft's new revenue streams.
Similarly, Digital Twin of a Customer shows significant investment intensity, reflecting Microsoft's commitment to revolutionizing customer experience through Azure's Digital Twins solutions. This technology is already used in managing complex systems like smart cities, factories or supply chains, perfectly aligning with Microsoft's cloud-first strategy. By digitizing real-world systems, Microsoft ensures optimal operational efficiency for its enterprise clients, reinforcing its leadership in cloud computing.

In addition, Machine Customers stands out as a hidden gem in the model. Although there have been no major public announcements about Microsoft's work in this area, the high score in investment intensity indicates strong internal development. Machine Customers, which autonomously manage transactions, negotiations and supplier relationships – acting on behalf of humans in buying, selling or requesting services – is a technology field Microsoft is extremely interested in and actively exploring. Given the firm's history of integrating AI into supplier-client interactions, for instance with Dynamics 365 – which leverage AI to automate sales and customer services - this could represent a promising opportunity and future direction for the company. This model output is a particularly valuable insight as it uncovers a potential strategy that has not yet been publicly disclosed – an information investors can leverage to beat the market (*Dynamics 365 AI | Microsoft AI*, n.d.).

Furthermore, the high score for Superapps reflects more than just Microsoft's ongoing investment in tools like Power Platform. While the Power Platform has been crucial in building modular, integrated solutions for enterprises, the significant investment intensity score reveals a deeper strategic focus. As a matter of fact, there have been persistent rumours that Microsoft is working on a consumer-facing Superapp similar to WeChat, which would combine its various services like Bing, Teams, and Azure into a single, unified platform. Such high score could be another valuable insight for investors, who can carefully evaluate such opportunity for Microsoft. In fact, patent data further supports the idea that Microsoft is quietly preparing for a larger, more comprehensive push into the Superapp space, which would dramatically expand its service ecosystem and user base.

In sum, this analysis highlights Microsoft's investment strategy across multiple technologies and categories, providing valuable insights for investors both at Macro and Micro level – being able to anticipate public announcements, trust or disregard rumours while evaluating the potential impact initially hidden investments could have on the company's market positioning, providing a unique glimpse into the American firm's priorities beyond industry speculations.

5.3. Visualization – Gartner's Hype Cycle with Investment Intensities

The next step in the thesis project is a compelling, visual representation of the Investment Intensities over the 2024 Gartner Hype Cycle - which leverages an already existing, widely understood framework and augments it with the results of the model at technology (Micro) level, as shown in Table 1.

As a matter of fact, this step is of significant importance not only because it leverages the existing analysis of the investment scores and combines it with the Gartner Hype Cycle, but also because it democratizes the insights, making them accessible and interpretable for a diverse audience of stakeholders - the combination of the well-established framework of the Hype Cycle with a color-coded gradient, facilitates a quick understanding of complex data, previously discussed in technical terms, by decision-makers as well as non-technical stakeholders. The use of a colour gradient, from deep blue (indicating minimal investment) to bright yellow (highlighting the most intense investment), allows even those unfamiliar with the underlying technical aspects of the model to immediately grasp Microsoft's strategic priorities. Given the complexity of this thesis and its supporting model, as well as considering that the audience utilizing this model may not necessarily have technical expertise, this step is crucial in bridging the gap between data analysis and business decision-making. For instance, Venture Capitals, Private Funds or Investors, who might lack technical expertise to explore and analyse the individual scores, can simply leverage this visualization to quickly understand which technologies are drawing the most resources and attention - the high investment levels in areas like AI-Augmented Software Engineering (1.00) and Spatial Computing (0.89), for example, become immediately apparent, conveying a clear message about where Microsoft is placing its future bets. On the other hand, the dark blue and purple dots associated with 6G (0.00) and Humanoid Working Robots (0.10) clearly visualize low Investment Intensities. This visual representation complements the previous analytical work by democratizing the results and their interpretations.

However, looking at the complemented Gartner Hype Cycle image, it is quite difficult to recognize investment trends reflecting Microsoft's medium-to-long term strategy. While we can observe the individual Investment Intensities over different stages, identifying clear patterns is challenging. For instance, understanding whether Microsoft focuses more on short or long-term investments, or prioritizes high vs low-impact technologies, can be extremely difficult. This, on the other hand, is fundamental in enabling investors to align their own investment strategy to the firm's analysed by the model, allowing for an informed investment decision making. This issue arises mainly due to the fact that certain underlying variables used to map the Emerging technologies by Gartner analysts are not visible in the Hype Cycle image, but are instead contained in the documents and data used to create it (i.e. Benefit Rating, Market Penetration, Maturity or Years to Mainstream Adoption variables). As a result, no definitive trends can be spotted by the eye. To address this, we need an additional metric to make these patterns more explicit: the Galletti Index.

5.4. The Galletti Index

Finally, the Galletti Index (G.I.) – which provides a clear, quantitative score of Microsoft's investment strategy in Emerging Technologies, weighting three different variables (i.e. Investment Intensity, Time Horizon and Technology Benefit) – was created in order to identify underlying patent investment trends from just one, single score:

Microsoft's Galletti Index: 5.63

In particular, the American firm's G.I. totalled 5.63, a score in the Mid-Range. On the surface, this might reveal a balanced investment strategy, but with unclear mix between shorter vs longer term investments, and high vs low impact technologies. Consequently, when a mid-range score is estimated, the Galletti Index needs to be accompanied by two closely-related scores, the Time Horizon and the Technology Benefit Scores, which reveal Microsoft's strategic trade-off and improve the GI's interpretability. In particular, both composite scores represent half the Overall Technology Index (the nominator of the Galletti Index) calculation, equally weighted to complement the Investment Intensity:

Overall Time Horizon Score: 0.41

Overall Technology Benefit Score: 0.86

Taking a closer look at the Time Horizon Score, the 0.41 indicates a tendency from the American firm to focus more towards longer-term innovations, compared to others in the Gartner Hype Cycle. The score in the relatively low-range indicates investments are directed toward innovations still in the early stages of maturity, which will take time to develop and penetrate the market. This approach balances the underlying risk of a still undeveloped, embryonic technology (which may fail to meet expectations or evolve more slowly than anticipated) with the high, potential returns given by its early adoption stage. While the score is calculated on the basis of Gartner's Hype Cycle – which by definition includes emerging technologies (thus in early stage), a 0.41 score is indicative of its time horizon compared across all Gartner-selected innovations. Furthermore, breaking down even further the Time Horizon Score, we can derive insights regarding the three different inputs that make it up:

- **Average Years to Mainstream Adoption Score**: 0.531
- **Average Maturity Level Score**: 0.28
- **Average Market Penetration Score**: 0.316

Starting with the Years to Mainstream Adoption Score, which falls in the 5-10 years category, indicates that Microsoft is primarily focusing on technologies that are expected to reach mainstream adoption in the medium term. These innovations, such as Spatial Computing and AI-Augmented Software Engineering, are not immediate game-changers but are anticipated to mature – on average - over the next five years, suggesting that Microsoft's portfolio is oriented toward medium-term growth. In addition, the Maturity Level Score of 0.28 places most of these technologies in the Emerging phase. In this stage, technologies are highly prices, MVPs starting to being commercialized by vendors, still requiring a high level of customization. This category includes technologies like Machine Customers and Prompt Engineering, which are in their first-generation development: although they show considerable promise, they still require significant research and technological advancements, combined with market readiness, before being ready to scale. Finally, the Market Penetration Score of 0.316, which falls between the 1% to 5% market penetration range, indicates that many of these technologies are in their early stages of market adoption – signalling that, on average, the American firm's technologies are currently being adopted by a small percentage of early adopters, not yet achieved widespread use. Technologies like Digital Twin of a Customer and Autonomous Agents fall into this category, being primarily in experimental phases or early commercial deployments. The low market penetration score aligns with Microsoft's focus on long-term opportunities, as it seeks to position itself with a first mover advantage, capturing significant market share once these technologies gain broader acceptance.

In summary, Microsoft's investments are heavily weighted toward emerging technologies that have a relatively long runway to mainstream adoption. These technologies are, on average, expected to take 5-10 years to reach maturity (mostly on the lower range), are still early in their commercialization efforts, and have yet to see broad market penetration (1 to 5%). This strategy suggests that Microsoft is positioning itself for long-term leadership in key areas of innovation, with the expectation that these early investments will pay off as the technologies mature and adoption increases.

On the other hand, while looking at the Technology Benefit Score – indicating whether the firm is investing in innovations with higher or lower market impact & benefit (it is important to keep in mind that all technologies in the 2024 edition of the Gartner Hype Cycle for Emerging Technologies fall into either the “Transformational” or “High” benefit rating category) – we notice Microsoft’s score totalling 0.86. This is a clear hint that the firm’s focus is balanced towards disruptive and transformational technologies (a low score would have not meant low-impact technologies are preferred, but rather “lower” compared to the others). Nevertheless, scoring this high in the Technology Benefit dimension suggests that Microsoft’s investments are mostly directed toward technologies with the potential to reshape industries and enabling new business models. For instance, some of the technologies with the highest Investment Intensity, such as AI-Augmented Software Engineering, Machine Customers or Spatial Computing, belong to the transformational category as defined by Gartner.

5.4.1. The Galletti Index – summary of conclusions

To sum up, Microsoft’s investment strategy becomes evident when considering the Galletti Index, Time Horizon Score and Technology Benefit Score altogether. The 0.41 Time Horizon Score suggests that Microsoft is focused on technologies maturing within an average of five years, which is relatively close to mainstream adoption when considering a maximum time horizon of 20 years. These investments, primarily in Emerging technologies with low market penetration (1% to 5%), indicate that these innovations are on the verge of rapid development and will likely see significant growth over the next few years. In addition, the high Technology Benefit Score of 0.86 shows that Microsoft is prioritizing transformational technologies, which have the ability to redesign and completely transform entire industries and sectors: these technologies are not just incremental improvements, but rather innovations that can fundamentally shift business models and drive disruption. The fact that these investments are expected to reach productivity and widespread adoption within the next five years makes them especially promising for Microsoft, positioning the company to take advantage of major technological breakthroughs.

Consequently, from an investor’s standpoint, the overall result is extremely promising: a combination of very high-impact, fast-maturing technologies makes Microsoft’s strategy particularly attractive. With many of these innovations likely to gain traction in the near term, Microsoft is setting itself up for significant growth and competitive advantage. The blend of high-reward investments in technologies that are both transformational and close to mainstream adoption signals strong potential returns, offering a compelling case for those looking to invest in cutting-edge firms poised to shape future markets.

5.5. Strengths, Limitations and Future Research

This thesis' methodology presented both significant strengths and limitations. First of all, in terms of strengths, the combination of multiple scoring models allowed for a robust and comprehensive patent mapping, capturing the semantic meaning of words within a context, outside of their strict definition. Moreover, equally weighting the different scores reduced the risk of biased or incomplete mappings, acting as a form of *"cross-validation"*. Furthermore, the visualization of the results improved this study's interpretability and accessibility, allowing people not familiar with the leveraged NLP techniques to easily understand and interpret its output. The Hype Cycle, with its investment-based colour coding, provided a clear and intuitive way for stakeholders to understand where investments were being concentrated.

However, the methodology also presented some limitations, namely the reliance on manual adjustments for both the technology definition and variables (i.e. Benefit Rating, Years to Mainstream Adoption etc…) extraction from the Gartner Hype Cycle, as well as for the visualization. Regarding the latter, while the Python interactive tool used to find and select the axis coordinates of technologies on the Hype Cycle mitigated some of the challenges, the need to manually input those coordinates in the model introduced the potential for human error. Alongside the need to manually map technologies with their definitions for the ontology creation, this further manual intervention also limits the scalability and reproducibility of the model, requiring time and effort to ensure both the correct functioning and visualization of the study. Lastly, certain parameters were also chosen to maintain a high model accuracy and flexibility such as the number of top keywords selected from the TF-IDF analysis (i.e. 10 terms), the timeframe of data collection (i.e. past 5 years) or the patent fields selected to create the ontology. In fact, only the title and abstract were utilized, as a more detailed description would have increased the computational complexity. Modifying these parameters could change the output of the model, and thus need to be carefully selected by the users.

This research aimed at providing a comprehensive analysis of patent data, creating an in-depth view of a company's technological trends and investment concentrations. By mapping patents semantically to technology categories, this study offers a robust framework for understanding where companies are focusing their innovation efforts. However, this analysis still leaves out some critical dimensions that investors may consider, such as stock price. The absence of financial metrics limits its direct applicability for investors looking to derive immediate, actionable insights about market value from technological investments. Thus, it could be beneficial for future research to explore methods that integrate the Galletti Index with traditional financial analysis, potentially quantifying a *"forward-looking"* stock price based on patent data. This combined approach would offer a more holistic view, bridging the gap between technological innovation and financial market predictions.

Conclusion

This research project was designed to provide a comprehensive framework for leveraging patent data to identify the investment strategy of any firm chosen to feed the model, mapping corporate technological investments onto the Gartner Hype Cycle. It offers an innovative, powerful tool private investors, venture funds or other investment bodies can leverage to evaluate a firm's medium-to-long term R&D focus, allowing for a quick and intuitive selection of a firm aligning with one owns investment principles. By applying this model to Microsoft's patent portfolio of the past 5 years, this thesis also reveals how the American firm is strategically positioning itself within the disruptive markets and business models of the future. In fact, the combination of NLP techniques and a triple-modelled semantic analysis with the Gartner Hype Cycle introduces a new approach for quantifying investment strategies leveraging patent data. This results in both a quantitative and qualitative analysis, allowing for a comprehensive view on where the American firm is placing its future bets. As a matter of fact, this project's findings reveal Microsoft's focus on transformational technologies - those that have the potential to not only improve existing processes but fundamentally re-invent entire industries - including innovations like AI-Augmented Software Engineering, Machine Customers or Digital Twin of a Customer, all expected to disrupt their respective industry within the next five to ten years. Also, this research shows that - while most of these technologies are still in their emerging stage with a low level of market penetration – Microsoft is positioning itself as one of the first movers for several technologies, taking advantage of their rapid development and future market potential.

While this already provides sufficient information to fill part of the underlying information asymmetry of the market – given the medium-to-long term investments of a Tech firm are often kept secret and under the radar, in order to maintain potential competitive advantages – it might not be informative enough to fully and confidently advice external investors in placing their market bets. As a matter of fact, evaluating the American firm's investment sensitivity towards the 25 technologies present in the 2024th edition of the Gartner Hype Cycle – considered by both Gartner analysts (and consequently by the entire market) the most disruptive and potentially impacting innovations of today – might not be enough to inform a substantial investment decision. Even visualizing such results directly on the Gartner Hype Cycle, highlighting the different investment intensities in each technology – which provides an immediate and intuitive understanding of Microsoft's investment focus – does not fully capture the true, underlying trends the American firm's executives are investing Billions of dollars in to lead in the upcoming years. Thus, this research project proposed a unique, innovative and quantifiable approach to this problem: the Galletti Index.

The introduction of the Galletti Index - an aggregate measure combining Investment Intensity, Technology Benefit as well as Time Horizon – bridges the gap to complete the firm's investment analysis, adding a valuable dimension to understanding how companies allocate resources across various stages of technological development. In fact, this index combines a granular analysis at the technology level with a broader, more comprehensive trend evaluation, revealing Microsoft's investment approach. The Galletti Index, scoring in the mid-range, alongside the two composite Time Horizon and Technology Benefit Score, clearly reveal how the firm is extremely well positioned to lead the technological development in the upcoming years, as Microsoft's next-generation portfolio is mostly composed by transformative, high impact technologies – with the potential of redesigning entire industries and their business models – maturing in the next 5 years, likely yielding substantial gains in the short-to-medium term. The Galletti Index (and its composite scores) provide a useful tool to assess the investment strategy of a firm's technological investments, allowing investors to fill the information asymmetry gap and potentially anticipate corporate strategy shifts, capitalizing on deriving insights from public (but hidden) information. Most importantly, investors can make an investment decision that aligns with their own strategic priorities in the medium-to-long term: placing their bets on companies offering the preferred combination of long vs short(er) term investment and high vs low(er) market impact – being now able to evaluate not only a company financials, but also the latter's unspoken future strategic direction. Consequently, in order to allow for an almost immediate and easy comparison of investments opportunities, the methodology developed in this thesis offers a scalable framework that can be applied beyond Microsoft, providing a versatile and democratic tool for assessing technological investments across various sectors and companies.

In conclusion, this thesis unveils the immense potential of patent examination and technology mapping in decoding the hidden layers of a company's strategic investments. Patents, while publicly available, often remain buried in complexity, making it nearly impossible to analyse each one individually or understand how they fit into the company's broader vision. However, by processing this information through the lens of the Galletti Index (and its composite scores) and aligning it with the most disruptive future trends, as outlined in the Gartner Hype Cycle, this study exposes patterns and insights that a firm may not explicitly reveal. The Galletti Index bridges the gap in market information asymmetry, enabling stakeholders to piece together a company's strategic roadmap - almost as if they were part of the executive Board, being able to access hidden narratives and future strategic investment information otherwise kept secret during internal discussions. Thus, it allows investors to anticipate a company's future business model shifts and strategies, enabling informed and confident decisions based on a firm's long-term ambitions - turning publicly available but hidden insights into a competitive advantage, navigating today's investment opportunities to increase confidence in tomorrow's financial gains.

6.1. Additional remarks

In the code accompanying this thesis (shared separately with the thesis supervisor), I utilized ChatGPT to enhance both the readability and ensure the smooth and correct functionality of my code. In particular, ChatGPT was leveraged in providing clear, explanatory comments for each step, improving the interpretability of the code. Additionally, I leveraged ChatGPT to identify and resolve bugs and errors that were preventing the code from running smoothly, ensuring that the final implementation is robust and efficient.

References:

1. Our World in Data. (2024). *Data page: Annual patent applications*. [Data adapted from World Intellectual Property Organization (via World Bank)]. Retrieved from https://ourworldindata.org/grapher/annual-patent-applications [online resource]
2. World Intellectual Property Organization. (n.d.). *Patents: Frequently asked questions*. WIPO. https://www.wipo.int/web/patents/faq_patents#:~:text=A%20patent%20is%20an%20exclusive,technical%20solution%20to%20a%20problem
3. What is Intellectual Property (IP)? (n.d.). https://www.wipo.int/about-ip/en/#:~:text=%3A%20clipart.com)-,Patents,can%20be%20used%20by%20others.
4. Huawei Technologies Co., Ltd. (2017). Digital Spillover: Measuring the true impact of the digital economy. Extracted from: https://www.huawei.com/minisite/gci/en/digital-spillover/files/gci_digital_spillover.pdf.
5. Millen (2022, March 21). U.S. Patent Grants Fell 7% Last Year, but 'Software-Related' Grants Remained at 63%. (2022, March 23). IPWatchdog.com | Patents & Intellectual Property Law. Extracted from: https://ipwatchdog.com/2022/03/21/us-patent-grants-fell-7-last-year-software-related-grants-remained-63/id=147745/#
6. World Intellectual Property Organization. (n.d.). *Interactive charts: Intellectual property facts and figures*. Extracted from: https://www.wipo.int/en/ipfactsandfigures/patents
7. Argente, D., Baslandze, S., Hanley, D., & Moreira, S. (2020). *Patents to products: Product innovation and firm dynamics*. Federal Reserve Bank of Atlanta. Extracted from: https://www.atlantafed.org/-/media/documents/research/publications/wp/2020/04/17/patents-to-products--product-innovation-and-firm-dynamics.pdf
8. Cohen, L., Gurun, U. G., & Kominers, S. D. (2014). Patent trolls. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2464303
9. Ziedonis, R. H. (2003). *Don't fence me in: Fragmented markets for technology and the patent acquisition strategies of firms*. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.475601

10. GENELÖV, D., & YUN, L. (2018). Using Patent Data to Position Firms in Technology-intensive Environments: How managers can use the world's largest repository of technological information to select technologies. Chalmers University of Technology. https://odr.chalmers.se/server/api/core/bitstreams/3016e5e6-ce9f-4514-bb72-bd999a3633c2/content
11. Wściubiak, Ł. (2016). *Patent protection as a strategic tool for modern enterprises*. *Deleted Journal*, *22*(3), 237–247. https://doi.org/10.58246/sjeconomics.v22i3.325
12. Fenn, J., & Raskino, M. (2008). *How to choose the right innovation at the right time: Mastering the hype cycle*. Harvard Business Press. https://www.gartner.com/imagesrv/books/pdf/mastering_hype_cycle_excerpt.pdf
13. Gartner. (2018, August). *Understanding Gartner's Hype Cycles*. Extracted from: https://www.gartner.com/en/documents/3887767
14. Trappey, C. V., Wu, H., Taghaboni-Dutta, F., & Trappey, A. J. C. (2011). Using patent data for technology forecasting: China RFID patent analysis. *Advanced Engineering Informatics*, *25*(1), 53–64. https://doi.org/10.1016/j.aei.2010.05.007
15. Kwon, K., Jun, S., Lee, Y., Choi, S., & Lee, C. (2022). *Logistics Technology Forecasting Framework using Patent Analysis for Technology Roadmap*. *Sustainability*, *14*(9), 5430. https://doi.org/10.3390/su14095430
16. Jun, S., & Park, S. S. (2013). *Examining technological innovation of Apple using patent analysis. Industrial Management & Data Systems*, *113*(6), 890–907. https://doi.org/10.1108/imds-01-2013-0032
17. Bamakan, S. M. H., Bondarti, A. B., Bondarti, P. B., & Qu, Q. (2021). Blockchain technology forecasting by patent analytics and text mining. *Blockchain Research and Applications*, *2*(2), 100019. https://doi.org/10.1016/j.bcra.2021.100019
18. Yang, C. B. (2012). *Role of patent analysis in corporate R&D. Pharmaceutical Patent Analyst*, *1*(1), 5–7. https://doi.org/10.4155/ppa.12.1
19. Yoon, B., & Lee, S. (2012). *Applicability of patent information in technological forecasting: A sector-specific approach*. *Journal of Intellectual Property Rights*, *17*, 37–45. Estratto da: https://docs.manupatra.in/newsline/articles/Upload/43B00144-9096-46DA-8C4E-50ED7069E06B.pdf
20. Singh, A., Triulzi, G., & Magee, C. L. (2021). *Technological improvement rate predictions for all technologies: Use of patent data and an extended domain description*. *Research Policy*, *50*(9), 104294. https://doi.org/10.1016/j.respol.2021.104294
21. Sharma, A. (2012). *A survey on different text clustering techniques for patent analysis. In International Journal of Engineering Research & Technology (IJERT)* (Vols. 1–1, Issue 9) [Journal-article]. https://www.ijert.org/research/a-survey-on-different-text-clustering-techniques-for-patent-analysis-IJERTV1IS9098.pdf

22. Grundmeyer, T. (2013). *Adopting technology: using student qualitative data and Gartner's hype cycle. Journal of Education and Training Studies*, *2*(1). https://doi.org/10.11114/jets.v2i1.228
23. Stikov, N., & Karakuzu, A. (2023). *The relaxometry hype cycle. Frontiers in Physiology*, *14*. https://doi.org/10.3389/fphys.2023.1281147
24. Collins, F. (2024, February 20). *A genome story: 10th anniversary commentary by Francis Collins. Scientific American. htt*ps://www.scientificamerican.com/blog/guest-blog/a-genome-story-10th-anniversary-commentary-by-francis-collins/
25. Weinraub Lajoie, E., & Bridges, L. (2014). *Innovation Decisions: Using the Gartner Hype Cycle. Library Leadership & Management*, *28*(4). https://doi.org/10.5860/llm.v28i4.7083
26. Oosterhoff, J. H., & Doornberg, J. N. (2020). *Artificial intelligence in orthopaedics: false hope or not? A narrative review along the line of Gartner's hype cycle. EFORT Open Reviews*, *5*(10), 593–603. https://doi.org/10.1302/2058-5241.5.190092
27. Dedehayir, O., & Steinert, M. (2016). *The hype cycle model: A review and future directions. Technological Forecasting and Social Change*, *108*, 28–41. https://doi.org/10.1016/j.techfore.2016.04.005
28. Vern Burkhardt, & Jackie Fenn. (n.d.). *Why the Hype?* Ideaconnection. Extracted from: https://www.ideaconnection.com/interviews/00114-Why-the-Hype.html
29. Adamuthe, A. C., Tomke, J. V., Thampi, G. T., & IJARCCE. (2015). *An Empirical Analysis of Hype-cycle: A case study of cloud computing technologies. In International Journal of Advanced Research in Computer and Communication Engineering* (Vol. 4, Issue 10, pp. 316–317) [Journal-article]. https://www.ijarcce.com/upload/2015/october-15/IJARCCE%2068.pdf
30. Dessaint, O., Foucault, T., & Frésard, L. (2022). *The horizon of investors' information and corporate investment. SSRN Electronic Journal.* https://doi.org/10.2139/ssrn.4276832
31. Syracuse University Libraries. (n.d.). Research guides: Lens guide: How to use Lens.org. Extracted from: https://researchguides.library.syr.edu/lensguide/use#:~:text=The%20Lens%20is%20a%20free,and%20export%20all%20Lens%20data.
32. Kwilliam. (n.d.). *Lens About – Search, analyze and manage patent and scholarly data. Lens About.* Extracted from: https://about.lens.org/
33. Manchekar, A., & Jain, A. (2023). *The role of data visualization in business decision-making: A review of best practices*. *International Research Journal of Engineering and Technology (IRJET)*, *10*(9), 638–639. Extracted from: https://www.irjet.net/archives/V10/i9/IRJET-V10I993.pdf

34. Mozilla Developer Network. (25 luglio 2024). *Regular expressions - JavaScript. MDN Web Docs*. Extracted from: https://developer.mozilla.org/en-US/docs/Web/JavaScript/Guide/Regular_expressions
35. Simha, A. (n.d.). *Understanding TF-IDF for machine learning. Capital One*. Extracted from: https://www.capitalone.com/tech/machine-learning/understanding-tf-idf/
36. What is spaCy? | Domino Data Lab. (n.d.). Domino Data Lab. https://domino.ai/data-science-dictionary/spacy
37. Word embeddings in spaCy. (n.d.). https://applied-language-technology.mooc.fi/html/notebooks/part_iii/05_embeddings_continued.html
38. Silva, E. (2023, July 12). What is Fuzzy Matching? - Redis. Redis. https://redis.io/blog/what-is-fuzzy-matching/#:~:text=The%20Levenshtein%20Distance%20(LD)%20is,terms%20are%20from%20being%20identical.
39. Walker, S. M., II. (n.d.). (n.d.). *What is the Levenshtein distance? Klu.ai.* Exracted from: https://klu.ai/glossary/levenshtein-distance
40. Cloudflare. (n.d.). *What are embeddings?*. Extracted from: https://www.cloudflare.com/en-gb/learning/ai/what-are-embeddings/
41. OpenAI. (n.d.). *Embeddings*. Extracted from: https://platform.openai.com/docs/guides/embeddings
42. Gartner, Chandrasekaran, A. Wong, & A. Khilare. (2024). *Hype Cycle for Emerging Technologies, 2024.*
43. Gartner. (2024). *Gartner 2024 hype cycle for emerging technologies highlights developer productivity, total experience, AI and security*. Extracted from: https://www.gartner.com/en/newsroom/press-releases/2024-08-21-gartner-2024-hype-cycle-for-emerging-technologies-highlights-developer-productivity-total-experience-ai-and-security
44. O'Donnell, J. (2022, November 18). *Microsoft eyes new market with supply chain platform. ERP.* https://www.techtarget.com/searcherp/news/252527483/Microsoft-eyes-new-market-with-supply-chain-platform
45. Research talks: *AI for software development - Microsoft Research*. (2021, November 16). Microsoft Research. https://www.microsoft.com/en-us/research/video/research-talks-ai-for-software-development/
46. Devdiscourse News Desk. (31 luglio 2024). *Microsoft celebrates GitHub Copilot's explosive growth amid AI innovation. Devdiscourse*. Extracted from: https://www.devdiscourse.com/article/business/3036761-microsoft-celebrates-github-copilots-explosive-growth-amid-ai-innovation

47. Ramel, D. (2024, July 31). *Nadella's copilot stats reveal changing Microsoft in earnings call -- Virtualization review.* Virtualization Review. Extracted from: https://virtualizationreview.com/Articles/2024/07/31/copilot-numbers.aspx
48. Digital Twins – Modeling and Simulations | Microsoft Azure. (n.d.). Exctracted from: https://azure.microsoft.com/en-gb/products/digital-twins/
49. *Microsoft Azure. (n.d.). Digital twins – Modeling and simulations.* Estratto da https://azure.microsoft.com/en-gb/products/digital-twins/
50. Lolambean. (n.d.). *Windows Mixed Reality VR - Enthusiast Guide. Microsoft Learn.* Exracted from: https://learn.microsoft.com/en-us/windows/mixed-reality/enthusiast-guide/
51. Ngoma, L. (2022, December 22). *Microsoft is working on a Super App that will dethrone everything.* M2050. Extracted from: https://m2050.media/en/microsoft-is-working-on-a-super-app-that-will-dethrone-everything/
52. AI TRISM: What it is & why it's important | Splunk. (n.d.). Splunk. Extracted from: https://www.splunk.com/en_us/blog/learn/ai-trism-ai-trust-risk-security-management.html
53. AI Trust, Risk, and Security Management (AI TRISM) - Navikenz. (2024, March 13). Navikenz.Extracted from: https://navikenz.com/whitepaper/ai-trust-risk-and-security-management-ai-trism/
54. Sawers, P. (2024, June 6). *Microsoft and OpenAI launch $2M fund to counter election deepfakes.* TechCrunch. Extracted from: https://techcrunch.com/2024/05/07/microsoft-and-openai-launch-2m-fund-to-counter-election-deepfakes/
55. MicrosoftGuyJFlo. (2024, August 13). *Microsoft Cybersecurity Reference Architectures* (MCRA). Microsoft Learn. Extracted from: https://learn.microsoft.com/en-us/security/adoption/mcra
56. MicrosoftGuyJFlo. (2024, August 13). *Microsoft Cybersecurity Reference Architectures* (MCRA). Microsoft Learn. Extracted from: https://learn.microsoft.com/en-us/security/adoption/mcra
57. Federated Learning with Azure Machine Learning: Powering Privacy-Preserving Innovation in AI. (n.d.). *TECHCOMMUNITY.MICROSOFT.COM*. Extracted from: https://techcommunity.microsoft.com/t5/ai-machine-learning-blog/federated-learning-with-azure-machine-learning-powering-privacy/ba-p/3824720
58. Jocontr. (n.d.). *Homomorphic encryption with SEAL - Azure Example Scenarios. Microsoft Learn.* Extracted from: https://learn.microsoft.com/en-us/azure/architecture/solution-ideas/articles/homomorphic-encryption-seal
59. Dynamics 365 AI | Microsoft AI. (n.d.). https://www.microsoft.com/en-us/dynamics-365/solutions/ai
60. Jin, B., & Driebusch, C. (2024). *The $14 Billion Question Dividing OpenAI and Microsoft. The Wall Street Journal.* Extracted from: https://www.wsj.com/tech/ai/the-14-billion-question-dividing-openai-and-microsoft-71cf7d37

61. Linden, A., Fenn, J., & Gartner. (2003). *Understanding Gartner's hype cycles. In Strategic Analysis Report.* Extracted from: http://ask-force.org/web/Discourse/Linden-HypeCycle-2003.pdf
62. From tech investment to impact: Strategies for allocating capital and articulating value. (2024, March 25). Deloitte Insights. https://www2.deloitte.com/us/en/insights/topics/leadership/maximizing-value-of-tech-investments.html
63. Boston Consulting Group. (2021). Meeting the Challenges of Deep Tech Investing. https://www.bcg.com/publications/2021/overcoming-challenges-investing-in-digital-technology
64. Tamplin, T. (2023, May 2). Investment Analysis | Definition, Importance, & Types. Finance Strategists. https://www.financestrategists.com/wealth-management/fundamental-vs-technical-analysis/the-methods-of-investment-analysis/
65. Twin, A. (2023, November 27). Investment Analysis: Definition, types, and importance. Investopedia. https://www.investopedia.com/terms/i/investment-analysis.asp

Appendix

8.1. Patent Example

Kind	Display Key	Lens ID	Publication Date	Application Date	Title	Abstract
A1	WO 2024/191708 A1	151-725-291-072-745	2024-09-19	2024-03-07	ELIDABLE TEXT FOR PROMPT CRAFTING	An elidable text is constructed that prioritizes the content included in a prompt to a large language model having a fixed-size context window. The elidable text is generated from developer-generated instructions or automatically for source code within a source code editor. A source code editor may include a feature that selects certain lines of code as important or focused which are assigned a high-priority value. A changed line, a line of source code at a current

						cursor position, lines of source code at the beginning of a file and those that output data are considered focused lines. Non-focused lines are assigned a priority based on a distance from a focused line. The elidable text constrains the data included in a prompt to the context window size by replacing the lowest-valued lines of text and source code with a replacement string.
A1	WO 2024/191667 A1	136-372-950-658-906	2024-09-19	2024-03-06	CIRCUIT BOARD COOLING CONFIGURATIONS	The discussion relates to thermal management. One example can include a circuit board including inner, intermediate, and outer generally concentric zones and a cryogenically cooled chip located in the inner zone as well as non-cryogenic electronic components positioned in the outer zone. In this example, the intermediate zone can have a skeletonized configuration that slows thermal energy movement from the outer zone to the inner zone.
A1	US 2024/0314128 A1	095-230-476-140-35X	2024-09-19	2024-05-22	SECURING AUTHENTICATION FLOWS USING A DECENTRALIZED IDENTIFIER	A digital wallet generates an identification value associated with a DID of a DID owner. The digital wallet generates a first request including the identification value for an authentication token from an identification provider. The first request is provided to the

						identification provider. The digital wallet receives, in response to the identification provider validating the first request, the authentication token that authenticates the digital wallet with a verifiable claim issuer including the identification value from the identification provider. The digital wallet generates a second request for one or more verifiable claims from the verifiable claim issuer. The second request includes the DID and authentication token including the identification value. In response to the verifiable claim issuer validating the authentication token and the identification value, one or more verifiable claims from the verifiable claim issuer are received by the digital wallet.

ESCP
BUSINESS SCHOOL

Mastères Spécialisés® & MSc

MSc in Big Data & Business Analytics

ESCP benefits from the best international accreditations.
Our European campuses enjoy national recognition.

AACSB ACCREDITED | EQUIS ACCREDITED | EFMD ACCREDITED MBA | EFMD ACCREDITED EMBA | 5 European Higher Ed accreditations

BERLIN
Heubnerweg 8 – 10
14059 Berlin,
Deutschland

LONDON
527 Finchley Road
London NW3 7BG,
United Kingdom

TURIN
Corso Unione Sovietica,
218 bis
10134 Torino, Italia

MADRID

ARROYOFRESNO
Arroyofresno 1
28035 Madrid, España

MARIA DE MOLINA
Maria de Molina
28006 Madrid, España

NAVALMANZANO
Navalmanzano 6
28035 Madrid, España

PARIS

CHAMPERRET
6-8, av. de la Porte de
Champerret
75017 Paris, France

MONTPARNASSE
3, rue Armand Moisant
75015 Paris, France

WARSAW
c/o Kozminski University,
International Relation
Office,
57/59 Jagiellońska St.
03-301 Warsaw, Poland

BRANCH CAMPUS

DUBAÏ
DIFC Academy
Dubaï International
Financial Center
Dubaï, UAE