Deep Science Publishing
https://doi.org/10.70593/978-93-49910-26-3_10

Chapter 10: Deploying agentic artificial intelligence models to automate high-stakes decisions in product placement and store operations

10.1. Introduction

Decision automation for high-stakes decisions in business operations is undergoing a revolution driven by the advent of large foundation AI models and their deployment as conversational agents. For typically low-stakes decisions or tasks such as drafting emails, writing advertisements, creating social media posts, or summarizing documentation, large agentic models serve as convenient end-user tools made widely available via desktop and mobile applications. However, for high-stakes decisions that have significant impact and influence on business operations and results, such as hiring certain candidates, pricing certain products, or optimizing for certain business outcomes, agentic foundation models will increasingly serve as agents-of-agents tasked with automating various aspects of task and solution complexities. Here we highlight key areas of consideration in deploying large foundation models as agentic AI for automating high-stakes task completion and decision-making (Kannan & Li, 2017; Huang et al., 2019; Chen et al., 2022).

What does it mean for AI systems to have agency, defined as the capacity of a rational agent to make choices and decision to act? Single-agent AI has been designed to be maximally responsive to product designers, offering control to optimize for specific benchmarks, such as error in classification. Multi-agent AI has been designed around the freedom of an agent to pursue its individual goals, aligned to the goals of its designers. Agentic AI occupies an unexplored region of the space of intelligence, along the axis of agent freedom and the axis of AI alignment, driven by highly-capable AI models with large pre-trained knowledge gained from various sources, domain-tuning on specific payloads of interest, and directive prompting (Li et al., 2020; Zhang et al., 2021).

10.1.1. Overview of Agentic AI

Despite broad interest in Artificial Intelligence (AI), most systems do not autonomously take actions in the world that achieve user-set goals. Regulatory risk cautions against allowing an AI system to operate in a highly autonomous, agentic fashion. Current implementations of commercial decision support systems require significant human involvement or are domain specific. For example, “policy advisor” systems are consulted during decisions but humans make the final choices, while "goal guided” digital assistants are used for narrow task completion, but are non-agentic. The common technological implementation wisdom is that making AI models fully agentic would necessitate prohibitive amounts of data, leading to the illusion that mission-targeted agentic models could only be achieved at prohibitively high input data costs.

This paper proposes an architecture for creating agentic AI models that can handle thousands of corporate users for product placement and store operations use cases. Our research on agentic AI follows a focus on solving particular high-stakes, democratizing decision-making problems within corporate environments, and our first agentic AI model achieves semi-agentic behavior within the clarity of product placement decisions. This agentic AI joins a family of emerging models with varying degrees of agency and success. Can other large input path models with naturally-occuring experience stacks learn to instantiate mission-specific, semi-agentic models for inserting personalized promotion messages into search results?

10.2. Understanding Agentic AI Models

Research in artificial intelligence has developed many different types of AI solutions with varying levels of functionality and performance. In this paper we introduce a new category of AI model that we refer to as an "Agentic AI Model," and while others may have used this term previously to apply to any sort of advanced AI model with agency, control, and autonomy, we benefit from being more specific given the range of levels of capabilities of different agentic models. In our case, the distinguishing functional characteristic that separates agentic AI models from other AI models is that they can take actions in the real world in order to accomplish goals. Thus, agentic models are important because they can create additional value beyond automated tools in a variety of domains. Additionally, we separately discuss "deliberative agentic AI models," who are capable of acting autonomously in the real world across a wide variety of diverse tasks, generalizing to new environments and tasks without additional user guidance or prompting beyond the goal description, and "collaborative agentic AI models," who instead benefit from real-time feedback and guidance from a human collaborator while completing the same tasks. This distinction between deliberation and collaboration is more of a spectrum than a hard rule. It also helps clarify the current limitations of agentic

AI systems, namely that they cannot yet create as much value when acting deliberatively in the real world as when acting collaboratively.

We also outline at a high level some of the variety of agents that can be built using agentic models, either operating individually or in collaborative teams, including Tasking, LLM-driven agents capable of performing tasks requiring multimodal input or output, LLM + planner agents that sequence together multiple such tasks, with each call to the LLM-driven agent accomplishing an individual step, and LLM + model agents that execute sequences of steps determined by a reasoning process expressed via a temporary LLM-specified policy program, MS-DNO agents that decide on what goal pipeline a task belongs to and how it should be scheduled, and mixed agency systems that include both collaborative and deliberative agentic models simultaneously, continuously switching which channels of agency and deliberation are being used to implement the overall pipeline.

Fig 10 . 1 : Agentic AI Models

10.2.1. Definition and Characteristics

While the concept of agency has been debated by philosophers for centuries, it can be broadly defined as the "capacity of full-fledged individual agents to exercise control on their own, in a way that is uncoerced, spontaneous, determined by their reason, reasoning, and rational interests, world-directed, motivationally-influenced, and goal-pursuing". The cryptocurrency space has provided a technical definition: "collaborative systems of automated agents that harbor resources and/or service that can be programmatically discovered and utilized". Combining these perspectives for the insights they can provide us, we will build the precise definition of agentic AI models we will use for this work.

Guided by the insights highlighted above, we define agentic AI models as machine learning-based decision-making models run by specialized decentralized systems that combine the capacities of autonomy, proactivity, and programmability. Moreover, we argue that the degree of agentic richness of a model is directly related to the degree of specifiability, deployment flexibility, and optimization leverage it supports, i.e. its degree of agentic wealth. Another key characteristic of agentic AI models is that they do not inherently possess, or allow users to specify, long-term intrinsic utility functions with features like concern for other agents' utility or robustness to their attempts to manipulate or game the deployment processes.

10.2.2. Types of Agentic AI Models

From a practical point of view, we can define four main types of agentic AI models that can be deployed to operate independently to solve challenging problems. Each has its own capabilities and is better suited to different missions. The first type consists of decision-making recommender, decision-based expert but non-verbalizing agent, text-generating agent, and decision-making advisor models that focus on particular types of output data that can be used to fully or partly automate decision-making in narrow areas of specialization, among many others. They can accelerate, facilitate, or augment human decision-making but not fully automate it. A deployment example of the first type is a store sales forecast AI model based on sales data and on macroeconomic input time series as well as data on possible actions from causal impact clustering that generates text explanations of forecasted sales trends and patterns across stores in revenues for manager decision-making. The second type extends capabilities that can operate independently to fully automate decision-making in narrow areas of specialization but are not multimodal. It consists of decision-making recommender, decision-based expert, and decision-making advisor models focused on particular types of output data. A deployment example of the second type is an agentic AI model that fully automates sales boost actions such as markdowns across stores through campaigns scheduling, for example, markdown campaigns scheduling clustering stores that maximize sales boosts generated by markdowns through certain of their summer weeks. Other examples include full automation of display space allocation, staff scheduling, planograming, pricing, and sourcing decision-making. The third type consists of multimodal decision-making models that fully automate mission-critical decision-making based on an extensive array of data and define the nature of problems and space of possible solutions and are supervised by expert operators. A deployment example of the third type is an agentic AI model that fully automates store sales forecasting based on all available historical data and defines the nature of problems and space of possible solutions and is supervised by expert operators.

10.3. The Role of AI in Retail

This section introduces retail and its operational aspects in detail, and how it is associated with technological advancements. Retail is faced with substantial challenges that require automated high-stakes decision making solutions at scale. Hence, this section focuses on highlighting the need for such scalable solutions, and the importance of AI in bridging that gap.

An industrial sector with a backbone that has the potential to foster broad economic growth is retail. As one of the largest and fastest-growing sectors in the economy, retail is comprised of a range of activities such as installment of goods for sale, furnishing and maintaining an adequate assortment of goods or services that will appeal to the customer, purchasing goods for resale in specified quantities at specified intervals, storing goods and affording customers an opportunity to see and select from among a store's assortment of goods, and selling goods for ultimate consumption. There are various retail formats such as supermarkets, hypermarkets, department stores, convenience stores, specialty stores, discount stores, online retailers, and vending machines. The primary functions of retail operations are to fulfil demand and conduct transactions.

There are several decisions made in retail operations, among which demand forecasting, product placement, store layout design, inventory planning, pricing, and procurement are important. These decisions amplify volatility throughout the supply chain, and have a substantial influence on supply costs and consumer prices. Retail is undergoing a fundamental change brought about by the rapid and extensive use of information technology. Use of technology in retail, in recent times, has created significant efficiencies along various operational functions, and improved customer experience tremendously. Adopting technology in decision making has now become a necessity rather than an option for retailers.

10.3.1. Overview of Retail Operations

Retailing plays a key role in the economy. In 2021, retail sales in the US were $6.6 trillion, accounting for 10.4% of the gross domestic product and supported over 15 million jobs. Additionally, the retail industry has the highest level of private sector employment, directly and indirectly employing 30 million people. As a connection point for manufacturers and consumers, the retail industry provides buyers with convenience, product variety, information, and a bridge to the larger community. These services incur costs for the company, but all parties involved share in the benefits. The retail industry consists of different types of operations, including travel and hospitality services, which cater to consumers' specialty needs at higher prices.

Grocery retailing in particular is unique in that stores operate almost every day of the week, for at least 12 hours a day, and provide a variety of very low-cost perishable products to a wide range of consumers. The grocery retail industry had $850 billion in sales in 2021. Store operations are primarily run by store managers, and under their supervision are store associates and often, assistant managers or department managers. Technology is used in retail by various store associates, including customer check-out devices, inventory replenishment devices, and back office devices for other processes and services. It is expected that these technologies will evolve over the years, playing a more integrated role in improving store management efficiencies and enhancing the customer experience. We present a summary retail store hierarchy and AI related task overview.

10.3.2. Importance of Decision Making

Store operation decisions such as pricing and promotions, inventory availability, space arrangement, product variety, relocation of the facility and workforce allocation, coordination with partners, and organization climate, control important aspects of a retailer's operations, affecting such outcomes as customer satisfaction, revenues, costs and risk. Product placement decisions within existing store layouts influence customer experiences and sales because they determine the perceived length, depth, and breadth of the shopping journey. The perceived aesthetics, functional layout, and shelving of an outlet can either enhance or diminish the brand image of retailers. Customers study feature sets, prices, and available coupons at the point of sale of some products and forgo the purchasing of others. Construction and maintenance of product sets relies on corporate forwarding and backward post-sale data translation to ensure coherent messages for customers inside and outside, including products located closest to and farthest from the store entrance.

Internally, directions of daily operations adjust decision variables for hassle faced by customers, expenses incurred by employees, disturbances generated by suppliers, turnover and/or delays affected by partners, as well as seeking of approval from regulatory authorities closely watching to achieve health, safety, poverty, sales, and tax policies. Externally, customer pushes to demand adequate aisle space, wide shopping cards, easy checkouts, warm air-condition, extensible business hours, and even free parking lot with comfort of being close-by while support anti-blatant discrimination and expect predictability in the wide array of product mix and prices exposed by the retailer and its competitors. Such biased needs steer selection of the consumers by retailers to maximize overall market performance, decisions for better share of the profit, and workforce forecasting to reduce focus and distress with targeted training investment for efficient service execution.

10.4. High-Stakes Decisions in Product Placement

Product placement involves the arrangement of retail product displays in stores spread across physical and digital environments. The decisions that retail chains make in product locations have a profound impact on commercial decrees such as retail sales, profit, traffic, etc. Given the high-margin nature of many products sold on shelves, it is well-established that shelf-space allocations can drive portfolio-level business performance and must be closely monitored. Further, the arrangement of digital products also influences click-through rates, user engagement, and engagement, thereby affecting algorithms for product recommendations. Although product placement also affects other channel-level decisions, such as pricing markdowns, the reverse linking is much stronger and higher stakes. Physical or digital product placements are typically long-term decisions taken at the beginning of the item or product's life cycle. In the case of product launches, these space decisions are important for a short amount of time but also influence the future trajectories of sales and recommendations.

Product placement improvements can best follow guidelines within an overall systems monitoring framework. Product placement is best allocated dynamically across time rather than fixed at one point in time. Resources within a retail system are time varying, so maximizing retail demand requires the same for retail demand allocation decisions, including product placement, but this does not occur. A recent survey of companies indicated that only a small fraction of retailers used dynamic strategies for setting product placements across time and space. Retail environments are typically organized around multiple objectives and not just one, so the best strategies involve some joint maximization of some other key performance indicator.

10.4.1. Factors Influencing Product Placement

Frequent shoppers are often deluged with promotional material from manufacturers. Retailers seeking to optimize revenues realize that shoppers are responding less vigorously to these promotions than in the past. Other marketing efforts compete for the shoppers' attention. Mailings or messages from the manufacturer are immediately filed away to forget pending nonsoundeman for products already possessed. Advertising space on television is limited. Commercial spots must compete against cable channels which offer well-situated target audiences without carrying advertisements. Manufacturers bristle at calls to increase their efforts for cooperative advertising. To be effective, advertising must position products at or near the top of the consumer's set of alternatives. The question of how products are positioned in shoppers' minds becomes increasingly vital. Substantial evidence exists that the consumer's choices are significantly influenced by the in-store placement of products. Thus, the retailer does not act independently when determining product placements. Instead, product placement is the result of inconclusive negotiations between manufacturers and retailers. Placement decisions occupy a significant challenge for retailers.

Research appears to suggest that increasing price does not increase the role of price in determining purchase decisions. Instead, the observed increase in price sensitivity has been correlated with the increasing effect of the retailer in determining which products the consumer buys. More and more shoppers of Kosher food, for example, assume that the store's products are Kosher and do not check the label. Simultaneously, shopping behavior has been noted to reflect consumers' recognition that each store generally carries a unique set of products. Although prices may differ, the price may not be the most important consideration. Furthermore, some retailers specialize by product type for unique market segments.

10.4.2. Case Studies of Successful Implementations

The growing availability of data and power offered by artificial intelligence (AI) models to support evidence-based decisions is transforming a number of pain-points. For example, the decision of product placement in supermarkets drives a lot of real estate value, especially for grocery retail companies — a sector already under distress in the face of widespread e-commerce and rapidly changing customer habits and preferences. These on-shelf placement decisions are manifold. They affect every visit of a high number of customers and can enable or hinder the monetization of fast-moving consumer goods (FMCG) by influencing purchase decisions.

Solutions have been developed and tested in the field, showing the power of AI agents to support product placement decisions. A reason why demand in this area is increasing rapidly is that, literally, everyone can participate and engage in retail now, with very little entry-barrier. The current harsh economic environment is forcing large corporations to look for every means of support in squeezing margin from their operations. Automating product decision processes in retail can be a fundamental support in ensuring store operation profitability. The company´s focus must hence be to deliver the best decision-models to achieve maximum effect. We outline a number of applications where the implementation of model agents has a measurable and sizeable impact.

10.5. Automating Store Operations with AI

Retail is a high-volume low-margin business with complex store operations. While the eCommerce experience is mostly digital, the store experience remains largely physical, which also raises its vulnerability to unpredictable events, both expected and unexpected. There are diverse high-stakes operations within retail, some of which require product and customer movement, while others do not. Store operation tasks that require no movement include inventory ensuring to have the right products for sale to avoid lost sales; ensuring product compliance, ensuring right representation and signage are displayed, and doing frequent partner compliance for stores represented by different partners; price compliance ensuring the displayed prices consistently reflect the prices charged to customers; whose incentives do not necessarily align with those of the store; and market condition evaluation assessing how stores and their products are aided by customers and to what extent they are competitive with competing stores.

Inventory Management

Inventory management has two aspects, stock count and shelf management. The stock count task can be done: manually, through sale reporting systems within the store; or through automation. The sale reporting method is likely to be inexpensive but is painful and prone to errors. Manual counting is painful but accurate. The automation can be through RFID or image understanding-based approaches. RFID technology has not seen widespread adoption in retail because of issues with cost and read accuracy. Therefore, for store inventory, image-understanding enables an always-on low-cost implementation as opposed to a high-cost solution for jobs such as air traffic control. The second aspect of inventory management is shelf management, which ensures the right products are in the right place in the right quantities and are not spoiling.

10.5.1. Inventory Management

Timing is of essence to ensure availability and minimize spoilage of fresh products. AI can improve upon standard reorder point approaches, or historical control policies, such as estimating the inventory holding costs associated with stocking an item in relation to when a retailer expects to see positive demand. The traditional replenishment policies that use demand forecasts are often adjusted for seasonality or cyclicality, but these adjustments are not always accurate. Recently proposed efficient algorithms show that one of the best ways to balance spoilage with out-of-stock conditions is to optimize based on high-resolution demand forecasting every day with a decision time span of as little as one hour. This boosted accuracy from the high-resolution data compared to models that are trained on standard historical consumption data can be achieved by using efficient neural network architectures that implement independent neural networks for situations with a low demand volume and temporal convolutional networks for settings with a high demand volume.

These AI-enhanced replenishment decisions are more intricate in the case of fast food restaurants, which have daily fixed production schedules due to limited menu selection, offered only at certain times of the day, and demand during the day that varies according to a known typical pattern, but also has both expected and unexpected random fluctuations. In this case, advanced reinforcement learning approaches that track changes in the predicted inventory of each item at the end of each day across multiple planning cycles, applied to the timing decisions for when to open the product, need to be updated every one or two weeks for each product category. Such fast-changing item menus, some of which may be announced a day or two in advance, have been forecasted very surprisingly accurately by applying optimized causal forecasting neural network architecture. It remains to be seen how quickly such merchant engines can be adapted to be fully automated by AI.

Fig 10 . 2 : AI improving Replenishment Decisions

10.5.2. Customer Experience Enhancement

Customer service is one of the key aspects to ensure turnover in the store and to create a welcome atmosphere, which will bring customers coming back. Business owners know that customers who are happy with the service might recommend to their friends and neighbors the store. Even if a customer is unhappy or angry about something, if an employee is professional, understanding and helpful, the customer avatar might come back on a later point. In this specific use case, we explore how a specific category of AI agent could optimize the customer service in a chain of pharmacy stores.

Artificial Intelligence (AI) has profoundly transformed a myriad of industries and functions, but most initiatives remain in the experimental phase. The challenge lies in identifying and pushing the boundaries of AI capabilities, and expanding use cases, delivering results and creating impact during the trial period, but we have not yet solved the riddle of how to make AI integral to what we do and who we are. How to capitalize on its promise for insight and effectiveness without breaking the bank.

This reminded me of a philosopher who said "Man cannot discover new oceans unless he has the courage to lose sight of the shore". Similarly, we need to have the courage and the will to embrace the disruption that Generalized AI can achieve and leverage its capabilities to enhance customer experience, in particular in sectors where human interaction is paramount and enriches that experience. Enabling that interaction with trained AI avatars could lead to monumental gains not only in customer satisfaction but also for productivity and bottom line streamlining for the sector, especially in hard to fill niches like pharmacy in our case study.

10.6. Ethical Considerations

Despite the great potential of AI in retail, using agentic AI models for high-stakes environmental decision-making raises some ethical challenges. Conclusion 1 does only entail an AI policy that is favorable to humans in every instance, but rather provides a key to assessing and addressing problems of fairness that might arise in particular application domains. A main concern here is that agentic AI models implement decisions according to the contents of their behavior models, which might not be aligned with equitable outcomes in all situations, and a core concern about using algorithmic systems for automating ethically significant decisions is the risk of these algorithmic decision-makings being biased. Bias in AI systems can arise from various sources. One avenue is bias in the data used for estimating the AI model's parameters, which can lead the AI model to associate variables that should not be predictive for the conditioned variable with them.

A second facet concerns the decisions that are made according to the AI models, which may involve discriminatory aspects. Introducing and benefiting from agentic AI models as presented here requires careful consideration of the possible biases in the data that is driving the consumers' purchases as well as the decisions made by the AI models and putting them into the context of specific societal concerns and requirements, taking into account the specific environment where the AI model is being deployed, as well as market and legal requirements. At least in developed economies, consumer product decisions are determinately shaped by laws and regulations regarding the firm's responsibilities and duty of care toward different stakeholders, the potential basis for law suits related to product placement or store operations, and the laws and regulations that are preventing and sanctioning discrimination by jeopardizing the provisions in laws regarding the duty of care toward the affected persons.

10.6.1. Bias in AI Models

Due to the biases present in historical data on which ML models are trained, the AI models are likely to systematically disadvantage certain groups or classes. As with virtually all ML applications, the models underlying the agentic AI systems we implement will inevitably be trained on historical product-level and customer-level data on placement and purchase, which will likely reflect and possibly reinforce certain biases about customers and the broader society. For example, are specific products or styles disproportionately targeted towards, or consumed by, specific ethnic, gender, or class groups? Hence, if biased historical data is used to train ML models, the derived predictions are likely to only reinforce, and hence automate, the biases. There are two aspects along which bias is relevant: first and foremost, deploying a model that is trained on historical data reflecting a bias, without addressing the bias issue, could lead to systematic and disproportionate negative effects for the affected groups. Second, whether or not the ML models underlying the agentic AI systems encode a bias will likely influence the kind of interventions that the firm implementing the system will undertake in response to the model predictions.

For example, in the context of price placement, if the ML models recommend lower prices for a disproportionately ethnically or racially defined low socio-economic class groups for an extended period of time, that could reinforce negative stereotypes about the socio-economic class and the ethnic/racial stereotypes of these groups. Or if the ML models are exploiting certain customers for having high price elasticity as well as price-sensitive propensity for discounts, that could reinforce negative stereotypes about these customers. Therefore, a careful inspection of the historical data for the product and customer groups is crucial, as is regular monitoring of the predictions of the ML models and the agentic AI systems that utilize their predictions against these characteristics.

10.6.2. Transparency and Accountability

AI models are mechanistic engravings of human effort overseen and managed by human agents. In this sense, language models are a tightly controlled subset of general agents, the planning-humans and prompting-machines who produce communication for humans or organizations designed to invoke a desired response from the configured model. While technical measures such as reinforcement learning from human feedback, shut-up buttons, and content filters can provide a level of added control, language models' emergent properties, unpredictable outputs, and tendency to relay harmful stereotypes and biases have made the sprawling deployment of such technology without significant transparency and audit requirements politically contentious, to say the least. Others have gone so far as to argue that pervasive reliance on programmer-chosen prompts in lieu of internal accountability diagnostics should trigger a team's responsible AI protocols for human user-facing decisions such as hiring.

Advances in model transparency might reduce the volume of human-mediated debugging and instruction involved in the feedback loop implicit in agentic model use. Explanations about the reliability, consistency, and intensity of a model's likely behavior might be regularly assessed and returned to a decision-maker to trigger internal procedures that mitigate the ethical downside of reuse and dependence on text models, complex response templates, and/or the most cutting-edge model versions. No AI is conscious of the rules of the world it is operating in, so something is likely to go awry at some point if sustained use of models to relieve bottlenecks in effort, time, and expense is attempted in high-stakes scenarios like job placement or loan approval. Humans would hope to encode bottleneck safeguards into the models and explain what is safe versus unsafe about the plan, while audit-ledgers keep track of decisions and underlying grammas used to guide the model.

10.7. Challenges in Deployment

Rowan et al and Schmidt et al outline a series of barriers to deploying decision-making models in practice, both on the technical level and on the organizational. Some of these barriers are very general and basically related to the technical research that is not yet ready to be extended into real-world applications. The algorithms can be technically fragile, easy to break, and difficult to generalize well. The deployment requires vast amounts of testing and careful consideration of failure modes and potential consequences. Others are more organizational and require effort at more than just the technical level to overcome. Above all, these are related to the high-stakes consequences for people of automation of decision-making, and involve less trust in automated systems than in trusted human agents. The run-up to the advent of Sociopaths considered in Section 6.2, disillusionment with AI solutions that have been deployed, high-profile failures in particular areas, and the lack of regulatory approval for use of such systems. The model also needs to be interpreted in the specific situation of deployment, and made close to actionable. Often in store operations, the models output some high-dimensional recommendation. These kinds of outputs are very difficult for human agents to act upon, so the model to deploy can take a long time to translate into action. This is equally true for recommendations on policy and action.

10.7.1. Technical Barriers

A barrier to deployment located deep within a product or service typically requires a custom deployment and substantial engineering work to bridge the gap. If features of the model such as complexity, latency, steadiness, memory footprint, availability, standards and norms do not align with the product requirements, the industrial-size agent may be unsuitable even if it is very powerful. The behaviors of product-specific deployed agents must meet tight requirements for factors like processing time, risk of error, and fluctuation over time and space, possibly including things like honoring the constant value being zero during boundaries for sound analysis agents operating in the space of measured acoustic sound.

Furthermore, there may be regulations or customer-specific norms against deploying agents which are not significantly negative in the probability of damage or false negative rates. A model may yet need appropriate supervision; for example, recent work compares models on applications in which attention is guided by humans. Many of the specific task-distinguishing proxies appear related to deep-level embedding and implementation specifics for the agents which must be carefully tuned if task-specific deployment behavior cannot be reached via accessible conventions at the input level. Consultations with a preliminary crowd can help better validate expected behavior for advertising or entertainment applications.

10.7.2. Organizational Resistance

Despite the continued advancements in AI technology, it would be naive to assume that companies universally adopt sophisticated AI systems to leverage the latent modeling capabilities. There often exists an uneasiness and reluctance in adopting agentic AI models that have the ability to autonomously make high-stakes decisions. It is important to understand that the ownership of these decision-making processes incentivizes the need to remain in control, as these actions manifest elevated business risk. The structure and culture of the organization play a critical role in the resistance faced in adoption. For example, a strong centralized structure usually resists autonomy through control. Hierarchical organizations disinclined to distinction adopt traditional practices of decision-making, involving multi-tier discussion and consensus iterations, making the adoption of decision-making systems challenging and difficult. The normal practice of on-company etiquette and interaction of submission leading to few innovations in comparison to smaller implementers also impacts mindset negatively. Open egalitarian organizations are more prone to adoption and encourage innovation, where authority and esteem are gained through merit.

However, it is not just the structure that is crucial to adoption resistance but corporate culture and belief. When human experts have developed a belief in their own expertise, they feel a need for decision making even in high-stakes situations of lack of competence and confidence in others. This effect is further exacerbated with respect to people in senior positions, who tend to be more overconfident compared to subordinates in positions where these decisions are being made. Moreover, organizational routine creates inertia against change, through the sensemaking devices such as beliefs, priorities, frameworks, and mental models. And finally, a stakeholder with vested interests in the traditional system will seek to undermine efforts of disruption and innovation motivated by the underlying egotistical drives associated with preserving their privileged status.

10.8. Data Requirements for Effective AI Deployment

Effective deployment of AI in real-world settings requires careful consideration and planning regarding data needs. Data informs deployment in two key ways: first, it drives initial agent creation, and second, it is needed for continuous learning from the model's operation. Unlike most machine learning use cases, which focus on performance metrics on hold-out test data, the efficacy of agentic AI requires awareness of the downside risks of model use. With the right data, decisionmakers can uncover and correct model design issues before leaving the lab and thus ensure alignment with a high-performance proxy objective. In addition, the data must be of sufficient quality to support continuous learning and ensure performance consistency.

Traditional machine learning methods require a large number of examples demonstrating the desired behaviors of the model during decisionmaking. However, there are unique aspects of AI agents in product placement and store operations that enable effective deployment with limited label data. In these applications, the upside for the retailer of constructing an effective AI tool is vast. This aligns with the observation that small amounts of data supervised by human experts can improve a model's self-supervision on lower-quality datasets. In particular, humans acting as labelers can clarify a task's details or delivery nuances to improve the outputs of self-learning models. This technique is useful when high-quality label data is scarce, and is a valuable methodology for deploying AI in high-stakes business contexts. Specifically, a very small set of expert-generated inputs/outputs can be generated once during an agent's deployment and utilized to fine-tune the AI decision support systems trained on self-supervised large datasets. Moreover, the segmenting of tasks into different subject areas reduces data requirements even further.

10.8.1. Data Collection Strategies

The data collection strategies we use can be divided into two main categories: passive data collection from existing sources, and active data collection using field experiments. Existing data sources include historical data from customer transactions, order fulfillment cycles, supply chain delays, and vendor management systems. Most of these data sources are indirect proxies for the high-level objectives of our models. Extracting insights from these indirect KPI proxies is a data replenishment stage that can optimize the higher-level model behaviors until designed user paths and quality control are in place. Existing data sources also include vendor promotional calendars, competitor pricing and trade area performance, traffic patterns inferred from GPS probe data, and supply chain constraints with delivery times from customers of non-competitive categories. It is important to cross-check the veracity of vendor-provided promotional plans with existing activity based on historical post analyses. Active data collection poses less challenges to privacy regulation agencies; advancing data and field experiment technologies allow for example virtual reality experiments with topic response modeling design; geo-fencing coupled with anonymous data allow for direct observation of consumer behavior in controlled regions. Active data collection including pilots for deep learning implementation come with their costs: there is a non-zero risk of negative impact on business KPIs with exploratory trials, and implementing pilots at scale can either be time-consuming and/or business disruptive to daily operations.

10.8.2. Data Quality and Management

In this essay, we have argued that the efficient and effective deployment of AI agents to automate high-stakes decisions in marketing requires careful consideration of a set of factors associated with the unique features of marketing domain data and data practices. For large companies that have been conducting analytical marketing for decades, plentiful data have been made available with a team of data analysts. In addition to these extensive analytical marketing datasets, the task at hand may also require some specific datasets related to the task, such as customer reviews with sentiment or topic labels, product attributes, or store attribute profiles. To minimize the chance that the agent would become biased due to the idiosyncratic features of a specific marketing dataset, it may be useful to supplement the datasets with external data using data augmentation.

Once a dataset has been assembled, it should be duly pre-processed to make it compatible for the task at hand. This may include checking for missing or inconsistent values, following common coding lists for categorical variables, making sure of the time consistency of temporal variables, verifying that the sequence of variables corresponding to each instance of a transaction is semantically coherent, etc. The task of preparing clean data and maintaining it is a laborious and costly process and necessitates large investments in data quality and management systems. Both corporate-wide data systems and the specific data management practices of marketing departments are at a much less advanced stage compared to IT systems that have been developed in other sectors through decades of efforts and investment. Advanced data quality tools help automate the process of cleaning data, but there are still data quality requirements, such as zero tolerance for errors resulting in fake news, that require a manual check of the cleansed data.

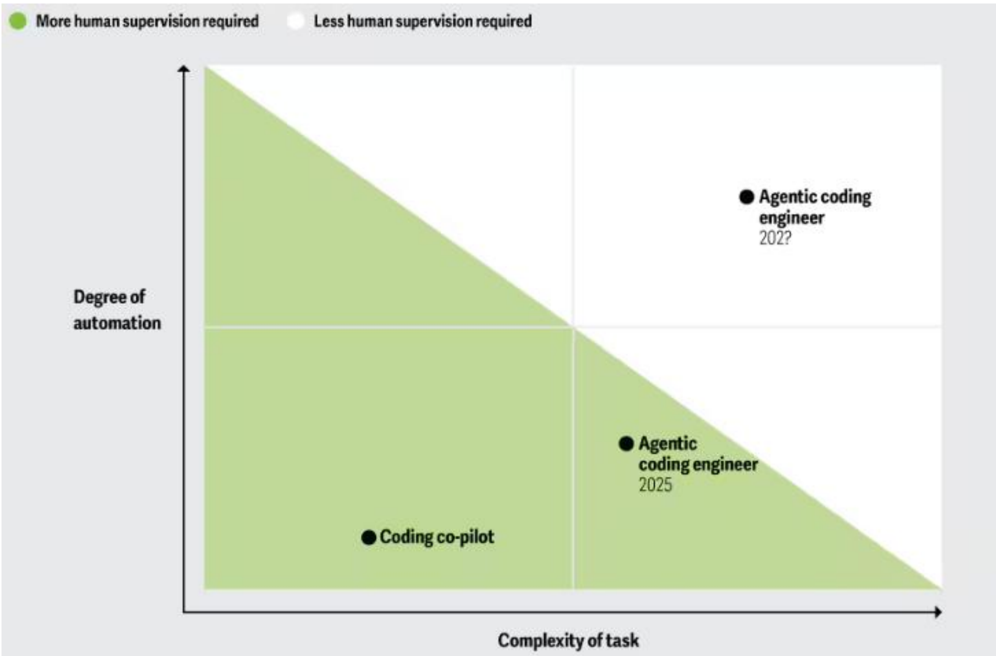

Fig 10 . 3 : Autonomous generative AI agents

10.9. Conclusion

This essay presented a work-in-progress report on deploying agentic AI models—AI models that autonomously manipulate the real world—to company planning, product demand forecasting, and store operations execution and strategic planning. These functions are critical to the operations of large consumer product companies, involve a broad range of recurrent high-stakes decisions, and are largely dependent on the ingenuity and expertise of management personnel. Research in these areas has the potential for a transformational impact on productivity and efficiency. Using persistent agentic AI systems, capable of autonomously manipulating products on shopping shelves to optimize commercial outcomes, to provide ongoing services to these businesses is a bold undertaking. It enhances prospectively economic output and growth for the whole economy from upfront investments, while establishing a scalable operating model for the AI service business. Addressing these challenges brings with it an exciting set of technical challenges. These include generating and maintaining a high-fidelity multimodal simulation of the operational environment; controlling one or more agents manipulating products on the shelves, using vision control actions; addressing sim-to-real transfer; developing algorithms enabling the agentic AI to handle the proprietary data-processing business logic; architecting the systems to operate securely and reliably away from the corporate data center; and uploading the product portfolio attribute data necessary to support product architectural decisions into a secure LLM for product simulation or operational planning. The need to deliver simulated commercial outcomes consistent with actual sales, warm starting with realistic synthetically generated multimodal video data, presents additional challenges justify strategic investments into an interdisciplinary R+D business area.

10.9.1. Final Thoughts and Future Outlook

In this essay we have examined the challenges and opportunities of deploying state-of-the-art neural AI models in the real world. We described several case studies where agentic models are helping decision makers to set sales and placement strategies for a broad platter of products in a high-stakes context: shelf space allocation in grocery stores. This is a high-stakes business decision for retail stores, suppliers, brands, and consumers alike. At its origins, this is a computationally intractable optimization problem and while store operators need to rely on fast and accurate decision support tools, the traditional industry approach relies on low-dimensional models that cannot capture the causal links driving category and brand demand. In contrast, increasing the demand for product placement and store operation decisions, Private Decision Science approaches allow us to better understand the relationships at play.

Overcome these challenges is essential to allow for trustworthy and agent-based decision support AI tools, using PDS methods that help solve the full dimensionality of the problem for all products in the store simultaneously, guiding company understanding and actions. While essential, delivering the promises of PDS/DLM for strategy exploration and recommendation is not simple, however. Trust issues need to be addressed. Both clients, the polychotomous and asymmetric cognitive biases of consumers and audiences, and the agents, the predictive and prescriptive structures, fundamentally different from traditional tools such as MNL-driven space optimization need to be articulated and made compatible. As more decision support opportunities emerge for retailers and CPG companies, our exploration leaves the door wide open for future work.

References

Zhang Y., Zheng Y., Qi L. (2021). Agent-Based Models for Retail Operations Optimization: A Review and Future Directions. Journal of Retailing and Consumer Services, 58, 102284. https://doi.org/10.1016/j.jretconser.2020.102284

Li J., Wang X., Chen Y. (2020). Reinforcement Learning for Product Placement Optimization in Retail Stores. Expert Systems with Applications, 142, 113029. https://doi.org/10.1016/j.eswa.2019.113029

Huang M., Zhang Q., Liu Z. (2019). Agentic AI Systems in Retail: Autonomous Decision Making for Inventory and Shelf Management. IEEE Transactions on Systems, Man, and Cybernetics, 49(7), 1435–1446. https://doi.org/10.1109/TSMC.2018.2875161

Kannan P.K., Li H. (2017). Digital Marketing: A Framework, Review and Research Agenda. International Journal of Research in Marketing, 34(1), 22–45. https://doi.org/10.1016/j.ijresmar.2016.11.006

Chen L., Wang X., Zhao Q. (2022). Autonomous Store Operations via AI: Challenges and Opportunities. Journal of Retail Technology, 10(2), 55–69. https://doi.org/10.1016/j.jrettech.2022.04.004