Artificial Intelligence + Machine Learning:
Current Applications in Real Estate

MIT | Center for Real Estate

Many real estate investors, bloggers, and technology firms are asking: Are machine learning (ML) and artificial intelligence (AI) transforming our industry, or is it just hype? This question served as the basis of an exploration of the emerging ML/AI technology firms in the real estate space. The result of interviews, research, and culling a list of a few thousand firms is a collection of 80 real estate technology companies that use AI and ML, and a dive into what it means for the industry today. My conclusion: machine learning and artificial intelligence can and will be used to facilitate real estate investment in myriad ways, spanning all aspects of the real estate profession-- from property management, to investment decisions, to development processes-- transforming real estate into a more efficient and data-driven industry.

ABOUT THE AUTHOR
Believer in the future of real estate technology and the impacts it will have on industry. Investment and development professional and life-long learner looking to make a difference through the built world.

j-conway.com

jennifer conway

The 80 firms identified in this research are rare finds at the forefront of a well-hyped trend towards advanced analytics, automation, and "deep learning," and span a remarkably wide array of real estate applications and machine learning and AI techniques. We are in the early years of these developments and there is a long way to go before we see widespread changes in the real estate industry stemming from ML and AI.

As new technologies continue to quickly emerge it will be important for technology firms to be aware of the pitfalls of machine learning, the competitive set in the market, and the areas where they can have the most impact. It also behooves real estate professionals to be keenly aware of the opportunities for efficiencies and the potential to look for in these technologies.

The following summary of the emerging impacts of these early real estate tech ML/AI adopters covers a portion of the ground explored in a much larger MIT thesis paper. For the purposes of this article I will present the overall results of this research and a deeper dive into a few of the exciting ways in which machine learning and AI are changing the real estate technology landscape.
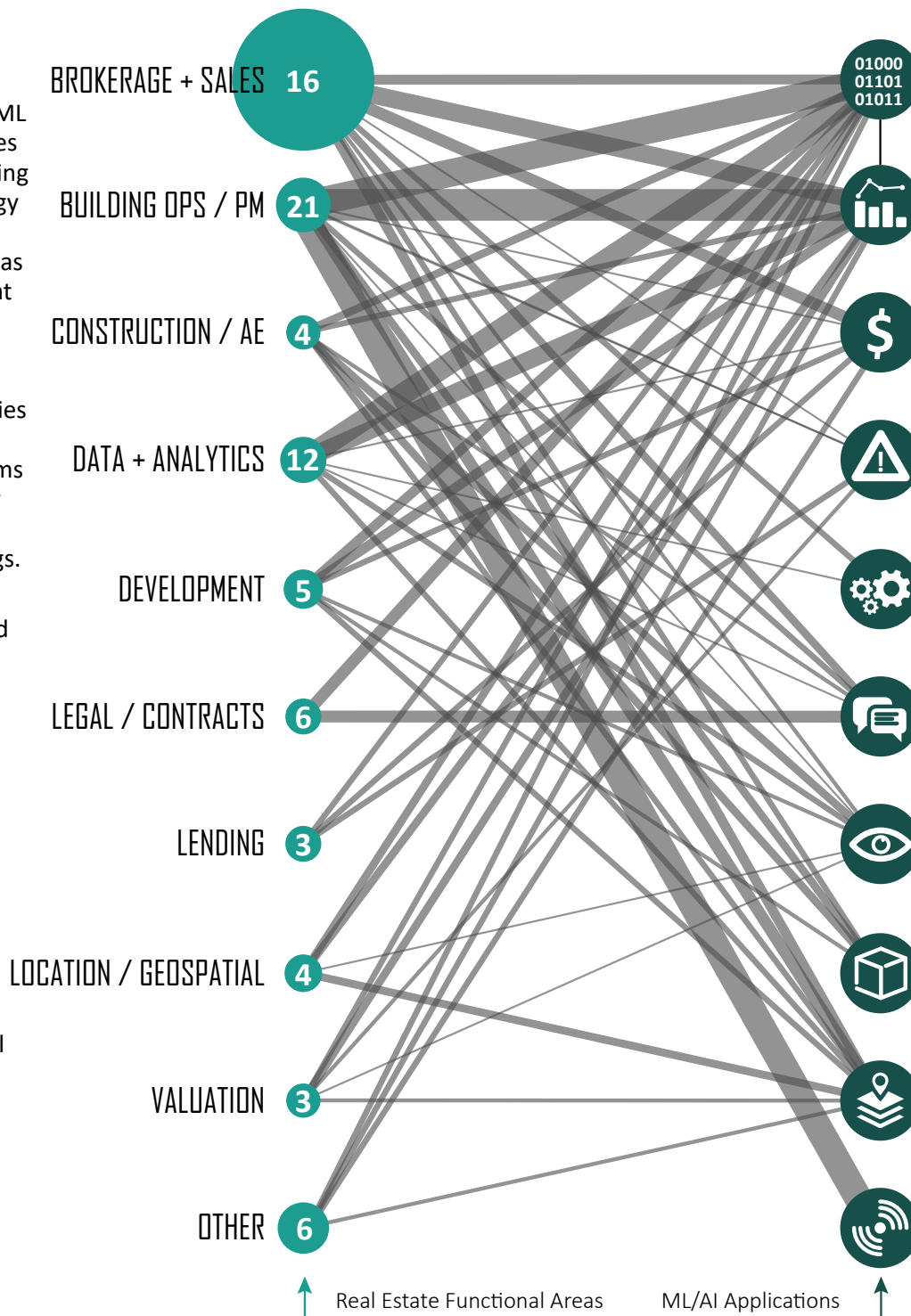
DEFINING THE FUNCTIONS

It is helpful to start with definitions of these often over-used terms. Artificial intelligence is a general term for machines performing tasks that typically require human intelligence. A wide variety of applications could fall under this umbrella, but as technology advances, our expectations for what computers should be able to automate and augment and what really qualifies as AI keeps shifting. Machine learning allows systems to learn from the past to predict the future by leveraging "big data" through data aggregation, tracking, analytics, and more, and is one piece of the AI puzzle.

VISUALIZING THE DATA

The chart on the next page breaks down the 10 real estate functional areas along with the number of companies in each category. These 80 companies are shown on the following page. The size of the associated circles represents the relative funding amount going to the companies in those categories. The 10 machine learning and AI applications for real estate are linked to the functional areas with lines sized to reflect the number of firms using a particular technology.

As demonstrated here, many functional areas are impacted by ML and AI today, with several examples for each category. The area receiving the most attention from technology investors is brokerage and sales, where AI and ML are now serving as the basis for real estate investment decisions. It should be noted that some of this funding is directed into investments in physical real estate, and not just the technologies themselves. Building operations is another major area with many firms entering the space, particularly by making use of IoT technologies in commercial and residential settings.

Data and analytics is a functional area for real estate, while data and analytics are also key applications where machine learning can be used to help any business, including real estate. Data is an important area of focus as it is the first step to advancing our understanding of real estate and using ML effectively. Machine learning can be a helpful tool in structuring data and aggregating various data sources. We are also seeing exciting developments in valuation, applications of image processing, and locational data, all of which are promising areas for growing this trend in the industry. These developments are still in early phases, but the potential for improved investment data and greater efficiencies mean ample opportunities for widespread adoption in the future.

Real Estate Functional Areas

- BROKERAGE + SALES — 16
- BUILDING OPS / PM — 21
- CONSTRUCTION / AE — 4
- DATA + ANALYTICS — 12
- DEVELOPMENT — 5
- LEGAL / CONTRACTS — 6
- LENDING — 3
- LOCATION / GEOSPATIAL — 4
- VALUATION — 3
- OTHER — 6

ML/AI Applications

DATA — **Data** gathering and distribution can be accomplished with new ML tools, which have the power to capture every software interaction, sensor output, and message, turning day-to-day elements of the real estate business into rich sources of data.

ANALYTICS — **Analytics** using machine learning has great potential to enhance our understanding of the built world and investment. As datasets grow, it is important to track which companies begin to develop truly sophisticated data analysis methods to leverage that data.

VALUATION — V**aluation** is key to the business in areas such as sales, portfolio management, REITs, tax assessment, and lending. Automated valuation models using ML tools improve predictive abilities, and also streamline appraisal and assessment processes.

RISK — **Risk** assessment using machine learning can produce a better understanding of thousands of potential risk factors. Lenders can optimize borrowing levels and rates and insurance companies can balance customer demand and potential costs.

BUSINESS PROCESSES — **Business processes** can be enhanced through new software with structured, purpose taylored environments. These environments become data sources in themselves, recording and learning from every customer interaction.

NLP / NLG — **Natural language processing (NLP)** and **generation (NLG)** give computers to ability to read text and communicate back to us in "human" terms. These tools are used in chatbots, contract review and data extraction, data gathering, and document writing.

COMPUTER VISION — A picture is worth a thousand words, especially with powerful tools using **computer vision**, which allows computers to interpret images and video. It can be used for value assessments, people tracking, and reading documents.

3D ANALYSIS — AI/ML tech is enabling a wide range of new approaches to mapping, designing, and constructing the built world. With machine learning, **3D analysis** for drawing or image augmentation and space planning can be performed quickly and efficiently.

GEOSPATIAL ANALYSIS — **Geospatial analysis** is the use of geographic coordinates in data display and analysis. The importance of location is not lost on the machine learning scientists who have developed interesting locational anlaysis tools.

IOT — The massive datasets coming from **Internet of things (IoT)** technology, where physical objects in our environment are internet connected, will enable us to see patterns in human behavior, including interactions with the physical environment.

REAL ESTATE TECH COMPANIES USING ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING
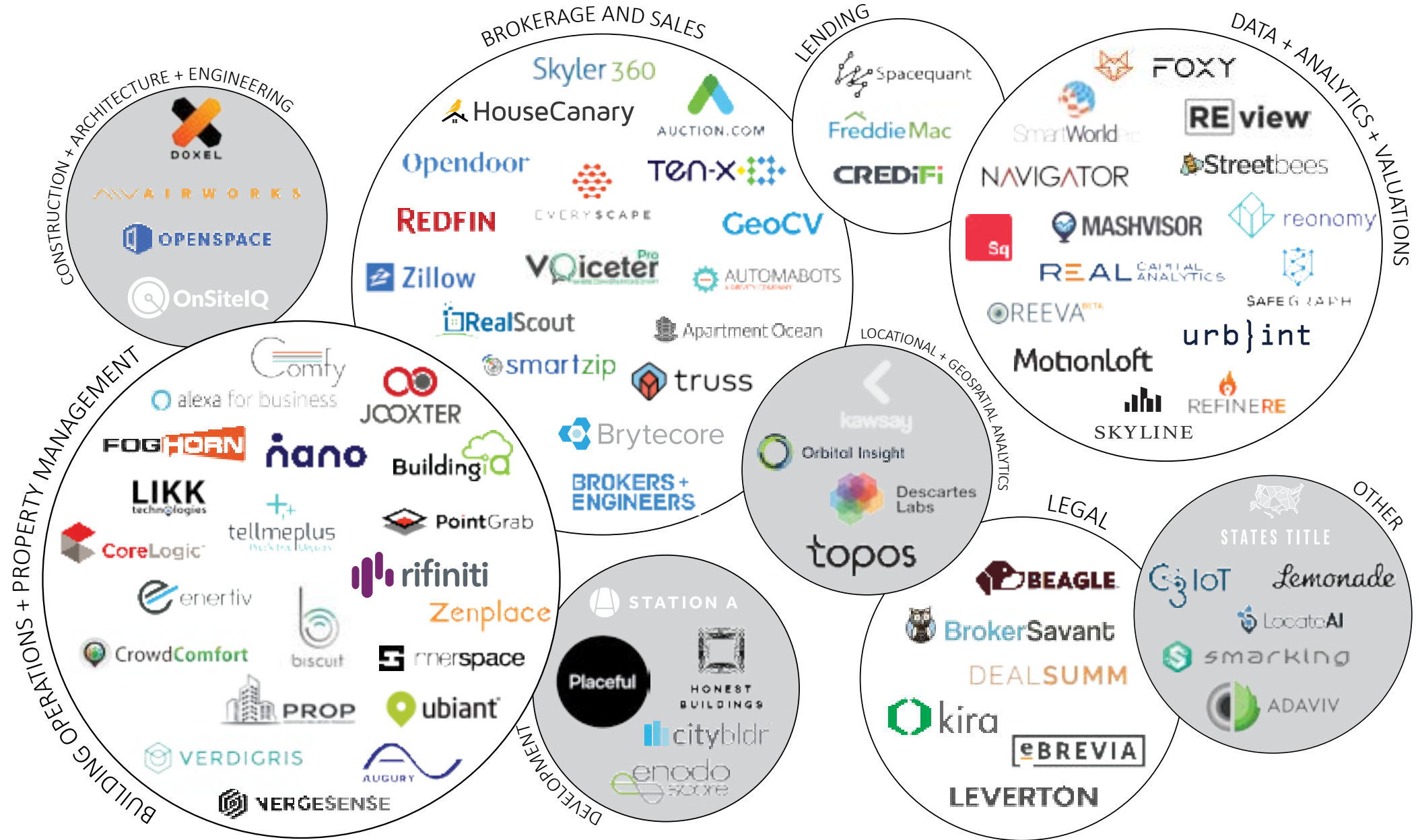
LEARNING DEEPER

The functional areas of real estate identified here should be fairly familiar to the real estate audience, but let's dig a little deeper into a few of the technology applications identified and the ways that each of them can be transformative to our industry.

01000 01101 01011 Data Gathering and Distribution

Machine learning can't happen without good data – garbage in, garbage out is a commonly heard aphorism in investment analytics, but also in the world of data science. This is one area where we have seen significant progress in the real estate tech realm through the use of machine learning. Nearly three quarters of the RE tech firms identified use advanced processing techniques to aggregate disparate data sources, clean and structure data, or generate new data through their machine learning processes, creating bigger, and wider, sets of structured data.

The concept of wide data comes from researchers at the MIT Real Estate Innovation Lab. This concept tackles an issue inherent to real estate data: there are often not enough buildings, sales, and other types of comparable data points to make an effective analysis. Through the "NYC Wide Data Project" the lab is gathering enormous amounts of detail on New York real estate assets and factors that influence real estate. The database is not truly "big," but it is "wide." In other words, there aren't that many rows, but there are thousands of columns that represent a wide array of attributes.

Some firms that are using AI to generate the new data for real estate include Street Bees, Foxy AI, MotionLoft, Enertiv, and Urbint. Leverton, a contract data extraction firm, turns language extracted from contracts into a structured output of the data. REview, DealPath, and Truss have identified the interactions with their users as important data points in themselves, a la Google. There are many more companies that are taking existing data sources, aggregating them using ML (with techniques such as string matching and record linkage), and then performing analysis, such as Skyline AI, CoreLogic, Reonomy, Enodo, CrediFi, LocateAI, Real Capital Analytics (RCA), and SpaceQuant. There are a variety of data sources that can provide insights into real estate, such as cellular device tracking data, mapping data, transactions data, demographics and government census data, and more. When combined with real estate portfolio data, new insights might be found.

Data security and rules around ownership will be paramount to enabling a data based real estate market. Ideally, proprietary data can be protected, anonymized, and combined to get a better picture of the market. However, the incentives for data sharing will impact what data points become part of a shared resource. For example, leasing information would be bad for landlords to share because transparency might bring down rents. On the other hand, open construction cost data would be beneficial to developers, owners, and investors as it would push down costs and make development pricing more predictable.

Data is the basis for any kind of analytics, including those that fall under the machine learning umbrella. Real estate firms of all kinds have long suffered from the issues inherent to disparate spreadsheets, unstructured data, lack of transparency, and an inability to dig into the details. As powerful tools for data aggregation and management become widespread, the impacts of machine learning and AI in real estate will become evident.

$ Valuation

Real estate valuation is a key part of the real estate business; professionals including real estate investors, appraisers, assessors, and lenders are all concerned with the value of real estate assets and the factors that drive that value. Automated valuation models (AVMs) have emerged as analytics tools to derive valuations more quickly, and in some cases more accurately, than traditional valuation methods. Those that apply machine learning analytics to improve valuations are able to process more variables and find new insights into variables that have an impact on value. Companies that use machine learning for their models include Zillow, Redfin, HouseCanary, Opendoor and Enodo. Sites like Zillow and Redfin have popularized single family and condo AVMs with home searches accompanied by a "Zestimate" or "Redfin Estimate" that gives home buyers a best guess of the selling price based on the latest analytics. Lenders in the residential market also are turning to AVMs, including Fannie Mae, which also provides its own value estimating service. Opendoor has a model that puts its own investors' money on the line with automated offers on homes based on its proprietary algorithms. Enodo provides a service to multifamily developers with large amounts of data driving predictions for the value of certain amenities and apartment features.

There are similar efforts beginning in the commercial sector as well. Research has shown that using machine learning methods it is possible to quickly and accurately predict value, even with missing information about a property. A goal of the MIT REIL NYC Wide Data Project is to achieve a better understanding of the drivers of commercial asset valuation, for example, the value that can be achieved by bringing co-working space into an office building.

Some sources stated that machine learning AVMs today are directionally accurate and good enough to use when valuing large numbers of properties. The real test is whether real estate professionals, particularly in the commercial sector, will trust these assessments. It may take a long time before human verification is no longer deemed necessary.

Risk Assessment

Real estate lending is a data-driven business where banks and other lenders do careful analytics on property values, deal terms, and risk. Insurance is also an area that requires risk assessments. For both applications machine learning is a useful tool to process large data sources and get a better understanding of risk factors that may not have been considered when using traditional evaluation models. Lenders can optimize borrowing levels and rates, and insurance companies can balance customer demand and potential costs.

One example of risk assessment in action is Freddie Mac's Loan Advisor Suite which includes a feature for automated risk assessments for potential borrowers. This feature allows them to assess borrower risk even when they do not have a credit score, opening up access to loans to a broader pool of potential home buyers, and providing instant feedback to customers.

Similarly, insurance risk and, therefore, optimal premiums can be calculated using machine learning techniques. One example is Lemonade, which automates the renter's insurance process. Customers simply enter data online and purchase a policy through the AI powered system.

As ML and AI continue to expand there are further opportunities for risk assessment in real estate. In a world where the "risk adjusted return" is the key metric for success, a true measure of the risks involved in any deal would give investors a whole different mode of operating. However, there are many risk factors in any deal, such as cost risk, weather and other "acts of God," partner risk, and capital market risk, to name just a few. It will take a long time for computers to get a good handle on any one of those factors.

Natural Language Processing/Natural Language Generation

Natural language processing (NLP) and natural language generation (NLG) are becoming commonplace technologies in the form of chatbots. Natural language processing provides the ability for computers to "read" text and extract pertinent data points, while generation gives computers the ability to "talk" back using written or spoken words. Applications in real estate include data scraping from text sources, contract data extraction, chatbots, and document production, thus allowing large amounts of previously "unstructured" data to be gathered and sorted, and making computer systems more accessible to a wide range of people.

One quickly developing area is contract analysis and abstracting. Legal documents contain important business terms buried in legalese. Computers can now read and sort this information to achieve quick and accurate lease abstraction, verification of proper contract drafting, and data analytics on the terms seen throughout a portfolio or even in the market broadly. There is still a lot of work to be done to achieve an acceptable level of accuracy where human verification is not required, but taking the human out of the first step creates meaningful efficiencies.

Computer Vision/Image Processing

Computer vision refers to the interpretation of images and videos, sometimes in ways that mimic human visual perception and sometimes to pick up on information and gather data in novel ways, like translating images into symbolic data. As a physical asset, real estate offers exciting opportunities to apply these quickly developing technologies in a variety of ways.

One interesting example is Foxy AI, a company that uses deep learning to generate qualitative assessments of real estate assets based on photos of the property. Another is MotionLoft, which uses advanced processing power to analyze video imagery for retailers to better understand foot traffic and how people flow through spaces. Others are using visual data for urban planning and economics studies. MIT Professor Sarah Williams has directed projects exploring how people use moveable furniture in public spaces. She found that ML for image video processing performed better than the sensors that were tested in tracking motion and people.

ASSESSING THE RESULTS

This research provides some hints of things to come in ML and AI for real estate tech. The overall impression from discussions with real estate technology firms and industry professionals is that we have a long way to go before the machine learning tools reach their full potential and we enter the realm of true artificial intelligence in real estate. Those companies that are using machine learning effectively stand to gain a lot, but will have to contend with consumer buy-in in an industry that has proven slow to adapt in the past and that may be reluctant to trust a machine, especially a black box result.

The insights gleaned through this research show a high level of rigor that goes into producing new models, and should provide some comfort and an appreciation for the non-hype promise of these tools. Machine learning platforms are already opening some doors and providing insights from the quickly growing farms / warehouses / lakes of data. But these new technologies won't change absolutely everything; there may forever be a time and a place for a simple hedonic regression analysis, a market cap rate comparison, or a human element, no matter how far the technology comes.