

Project Title: AI-Driven Agents for Fair and Efficient Property Valuation

Student Name: Fahad Abubaker Bahashwan

Advisor: Dr. Ashraf Elsayed

Institution: MidOcean University

Abstract

This project aims to build an AI-powered system that automatically analyzes maps and estimates property values. The motivation is to make property valuation faster, more accurate, and less reliant on manual work. The approach combines deep learning for land use classification, AI agents for collecting data, machine learning for predicting values, and natural language generation for creating reports. Key datasets include satellite images, real estate listings, and public records. The expected outcome is a tool that helps investors, planners, and homeowners understand property values and the factors influencing them.

1. Introduction

Property valuation is essential for real estate, urban planning, and investment. Traditional methods are slow and require experts to gather and analyze lots of data. With advances in AI, it's possible to automate much of this process. However, challenges remain in combining different data sources, making accurate predictions, and explaining the results. This report presents a structured plan to build an AI system that addresses these challenges, with sections covering the problem, motivations, objectives, related work, methodology, datasets, timeline, and references.

2. Problem Specification

This research addresses the problem of unfair and inconsistent property valuations that happen when humans estimate prices manually. Since different evaluators use different methods and personal judgments, the same property can get very different price estimates. To fix this, we're building an AI system that automatically analyzes property values in a fair and consistent way. The system examines satellite images to identify what type of land or buildings exist in an area, collects

Project – AI- Driven Agents for Fair and Efficient Property Evaluation

real estate information from reliable sources, and uses smart algorithms to calculate accurate prices without human bias. It also explains how it reached each valuation in simple terms, making the process transparent and trustworthy. By replacing outdated manual methods with this automated approach, we can make property valuations faster, more accurate, and fair for everyone involved.

3. Motivations

This research is driven by the urgent need to improve how property values are determined. Currently, the manual valuation process creates frustrations for homeowners, buyers, and urban planners alike - it's slow, often inconsistent, and can be influenced by human biases. By developing an AI-powered solution, we can provide faster, fairer, and more reliable property assessments that benefit everyone involved in real estate transactions. The project is particularly exciting because it brings together cutting-edge AI technologies in new ways, teaching computers to analyze property features from satellite images and combine this with other real estate data to make intelligent valuations. Beyond the practical benefits, this work pushes forward important research in artificial intelligence, showing how different AI techniques can work together to solve complex real-world problems. With property valuation being such a crucial factor in personal finances and urban development, creating a system that's both more accurate and more transparent could have significant positive impacts on how we manage and understand property markets.

4. Objectives

The primary objective of this research is to develop an artificial intelligence system capable of automatically analyzing maps and predicting property values. This system will utilize deep learning techniques to classify land use from satellite imagery, enabling the identification of various land types. In addition, AI agents will be developed to autonomously collect and organize relevant data from multiple sources, ensuring a comprehensive dataset for accurate predictions. Machine learning models will then be created to predict property values, with an emphasis on explaining the key factors that influence these predictions. Furthermore, the system will generate user-friendly, understandable reports that summarize the predicted values and the factors impacting

them. Ultimately, the goal is to make property valuation more transparent, fair, and accessible, thereby improving the decision-making process for various stakeholders.

5. Literature Review

In ^[1] The study presents Autonomous GIS, an AI-driven system that leverages Large Language Models (LLMs) like GPT-4 to automate spatial data analysis, aiming to minimize human intervention in data collection, processing, and visualization. The prototype, LLM-Geo, demonstrated strong performance across case studies, producing accurate maps and analytics while significantly reducing manual work. While the system highlights the potential of integrating LLMs into GIS for automating routine tasks, it remains limited in handling complex spatial reasoning, real-time data updates, and contextual decision-making. My research could address these gaps by developing more adaptive, dynamic GIS-AI systems capable of deeper spatial understanding and real-time responsiveness.

In ^[2] the research introduces a domain-based self-supervised learning approach for real estate appraisal, particularly suited for low-resource environments. By combining intra-sample geographic prediction with inter-sample contrastive learning, DoRA effectively processes unlabeled real estate data and enhances model generalization. Tested on Taiwan's real estate transactions from 2015–2021, DoRA significantly outperformed existing methods, and its practical value is evidenced by its adoption by E.SUN Bank for mortgage appraisals. However, while DoRA improves valuation accuracy in limited data settings, its reliance on pre-existing geographic and transactional data may limit adaptability in rapidly changing or highly diverse markets. My research could build on this by integrating real-time dynamic data, enhancing cross-region generalization, and developing frameworks that adjust to market shifts without heavy dependence on historical datasets.

In ^[3] the research is closely related to my focus on AI-driven property evaluation. It emphasizes how artificial intelligence (AI) and geospatial technologies can be combined to improve the accuracy and objectivity of property valuations. The study highlights the use of neural networks, machine learning, and GIS-based spatial analysis—important tools for automating the valuation process—which align well with my interest in AI agents. One of the key points made is that AI can

help reduce human bias often found in traditional property appraisal methods, which is also a major goal of my research. However, the study has some limitations. It mainly relies on static spatial data and does not fully address the use of real-time dynamic factors, such as market changes. It also faces challenges when trying to apply the models across different regions. Furthermore, while the study mentions the importance of collaboration across disciplines, it does not explore how to practically implement such collaboration. My research aims to address these gaps by developing adaptive AI frameworks that can use real-time data, work effectively across different regions, and include clear ethical guidelines for data use. This would contribute to creating stronger, more context-aware property valuation systems.

In ^[4] The study explores how AI is transforming real estate through automated valuation models (AVMs), image recognition, predictive analytics, and customer interaction tools. AI enhances efficiency, accuracy, and market forecasting, enabling faster and more data-driven decisions. However, key limitations remain, including biases in AI models, difficulties in adapting to sudden market changes, high implementation costs, and data privacy challenges. My research aims to address these gaps by developing bias-resistant models, integrating real-time dynamic data, creating more affordable AI frameworks, and embedding stronger privacy and regulatory compliance measures.

In ^[5] The study presents a real estate AI agent built using LangChain and Pinecone, capable of assisting users with mortgage and finance decisions by processing unstructured data in both Danish and English. The agent demonstrates superior performance compared to models like ChatGPT 4, offering professional, concise outputs that help reduce information asymmetry and consulting costs for investors. However, the model's limitations include its dependence on static, unstructured datasets, challenges in adapting to rapid market changes, and weaker performance in minority languages. My research could address these gaps by incorporating real-time dynamic financial data, enhancing multilingual optimization, and building more adaptive, self-updating AI agents.

6. Proposed Methodology

This research proposes a multi-stage methodology for the automated evaluation of properties utilizing artificial intelligence (AI) agents, integrating deep learning, natural language processing, and machine learning techniques. The methodology is designed to systematically address the challenges associated with data collection, land classification, valuation prediction, model interpretability, and result presentation.

The first stage involves **land use classification**, a critical component for contextualizing property values. Deep learning techniques, particularly convolutional neural networks (CNNs), will be employed to segment high-resolution satellite imagery and classify land into various categories such as residential, commercial, agricultural, and undeveloped areas. Accurate land classification provides essential spatial context that significantly impacts property valuation, enabling the system to capture localized land-use patterns that influence real estate prices.

The second stage focuses on **data collection and structuring**, leveraging autonomous AI agents. These agents will be developed using a combination of reinforcement learning strategies and natural language processing (NLP) models, with a particular focus on fine-tuning transformer architectures such as BERT. The agents will be tasked with scraping and extracting relevant property information from online real estate portals, governmental databases, and third-party APIs. The collected data will include property characteristics (e.g., size, age, amenities), spatial attributes (e.g., location coordinates, proximity to points of interest), and economic indicators (e.g., market trends, local economic statistics). To ensure high-quality datasets, the agents will also perform data validation, cleansing, and normalization.

The third stage addresses **property valuation prediction** through the application of advanced machine learning models. In particular, XGBoost will be utilized for its high predictive performance on structured tabular data, while Graph Neural Networks (GNNs) will be employed to capture the complex relational and spatial dependencies between properties within a geographic area. Features incorporated into the valuation models will include spatial features, property-specific attributes, and broader economic variables. A rigorous feature engineering process will be

conducted to maximize model performance, including spatial encoding, economic index normalization, and interaction feature creation.

Recognizing the importance of transparency and trust in AI-driven decision-making, the methodology integrates a dedicated model interpretability step. Specifically, Shapley Additive Explanations (SHAP) will be applied to the predictive models. SHAP values provide a theoretically grounded method for attributing the contribution of each input feature to a given prediction, thereby allowing stakeholders to understand and validate the factors influencing property valuations. This interpretability is critical for fostering user trust and ensuring regulatory compliance in real-world applications.

The final stage involves **automated report generation** to present the results in a human-readable and accessible format. A hybrid approach combining template-based methods and transformer-based text generation models will be implemented. The template-based component ensures consistency and reliability in report structures, while the transformer models enable dynamic generation of narrative explanations tailored to individual property evaluations. These reports will summarize the predicted property value, influential factors, spatial and economic context, and recommendations where applicable.

The proposed methodology will be implemented using a set of modern AI development tools. PyTorch will serve as the primary framework for building and training deep learning and GNN models. Scikit-learn will be employed for traditional machine learning tasks and feature engineering. Hugging Face libraries will support NLP tasks and transformer fine-tuning. FastAPI will be utilized to develop an efficient, scalable API infrastructure to integrate the various components into a cohesive, operational system.

In summary, the proposed methodology offers a holistic, explainable, and automated approach to property evaluation, harnessing the full potential of AI agents across data collection, modeling, and reporting phases. It is designed to contribute both practically to real estate valuation practices and academically to the fields of applied AI and geospatial machine learning.

7. Datasets Description

The research will utilize multiple datasets to support the development and evaluation of the proposed system. Satellite imagery will be sourced from publicly available repositories such as **OpenStreetMap**⁶ and **Bing Maps open building footprints**⁷, providing detailed information necessary for land use classification tasks. Additionally, **real estate listings**⁸ will be collected by scraping data from various property websites. These listings will include critical attributes such as property price, area, number of rooms, amenities, and other relevant features. To complement this information, **public records**⁹ from governmental sources will be incorporated, offering verified data on property sales histories, issued permits, demographic distributions, and economic indicators.

8. Timeline of Second Semester Project (10 Weeks)

Week	Planned Activity
1	Collect initial datasets, and study similar AI property valuation research.
2	Preprocess satellite images, start CNN model for land use classification.
3	Train and evaluate CNN model for land classification.
4	Build AI agents for scraping property listings and public data.
5	Clean and organize the collected data.
6	Train XGBoost and GNN models for property valuation.
7	Evaluate models, apply SHAP for model explainability.
8	Build report generation system (template + transformer).
9	Connect all parts (FastAPI backend integration / frontend view), full system test.
10	Write final thesis report, prepare for presentation and demo.

9. References

1. Li, Z., & Ning, H. (2023). Autonomous GIS: the next-generation AI-powered GIS. arXiv preprint arXiv:2305.06453. <https://arxiv.org/abs/2305.06453>
2. Du, W.-W. (2023). DoRA: Domain-Based Self-Supervised Learning Framework for Low-Resource Real Estate Appraisal. arXiv preprint arXiv:2309.00855. <https://arxiv.org/abs/2309.00855>
3. Droj, G.; Kwartnik-Pruc, A.; Droj, L. A (2024) Comprehensive Overview Regarding the Impact of GIS on Property Valuation. ISPRS Int. J. Geo-Inf. 2024, 13, 175. <https://doi.org/10.3390/ijgi13060175>
4. Alizadeh, B.; Gholizadeh, H.; Alizadeh, M. (2024) AI-Driven Innovations in Real Estate: From Property Analysis to Market Predictions. ResearchGate. <https://www.researchgate.net/publication/387672248>
5. Haurum, K. R.; Ma, R.; Long, W. (2024) Real Estate with AI: An Agent Based on LangChain. Procedia Computer Science, 242, 1082–1088. <https://doi.org/10.1016/j.procs.2024.08.199>
6. OpenStreetMap has an editing API for fetching and saving raw geodata from/to the OpenStreetMap database - <https://www.openstreetmap.org/>
7. Microsoft Bing - open building footprints around the world 1.4B buildings
<https://github.com/microsoft/GlobalMLBuildingFootprints>
8. Property listing data was collected through web scraping of major real estate platforms comprising approximately greater than 200,000 records with attributes including price, location, property features.
9. Open Data Platform platforms comprising transaction of approximately greater than 300,000 records with attributes including price, location, property features. <https://open.data.gov.sa/>