Human-Machine Synergy in Real Estate Similarity Concept

Małgorzata Renigier-Biłozor & Artur Janowski

Faculty of Geoengeering
University of Warmia and Mazury in POLAND

INTERNATIONAL RESEARCH SYMPOSIUM

Amsterdam, The Netherlands · December 4 - 5, 2024

INSPIRATION

Faculty of Geoengeering
University of Warmia and Mazury in POLAND

dr hab. inż. **Małgorzata Renigier-Biłozor**, University Professor
&
dr hab. inż. **Artur Janowski**, University Professor

malgorzata.renigier@uwm.edu.pl
artur.janowski@uwm.edu.pl

IAAO Rich Almy Research Grant

The original methodology for homogeneous area determination (HAD) for the purpose of property taxation procedures' fairness and equity increase
GRANT 2023-2024

CHALLENGES AND INNOVATIONS IN NEW ERA OF PROPERTY VALUATION

Human-Machine Synergy in Real Estate Similarity Concept:
Will New Trends Break Stereotypes?

THE INTERNATIONAL PROPERTY TAX INSTITUTE (IPTI) AND THE INTERNATIONAL ASSOCIATION OF ASSESSING OFFICERS (IAAO) PRESENT A (VIRTUAL)

MASS APPRAISAL VALUATION SYMPOSIUM
JUNE 26–27, 2024
Challenges & Opportunities

NATIONAL SCIENCE CENTER
Methodology of consensus-oriented roadmaps development for property tax transition.
GRANT 2024-2027

IFI IPTI
International Property Tax Institute
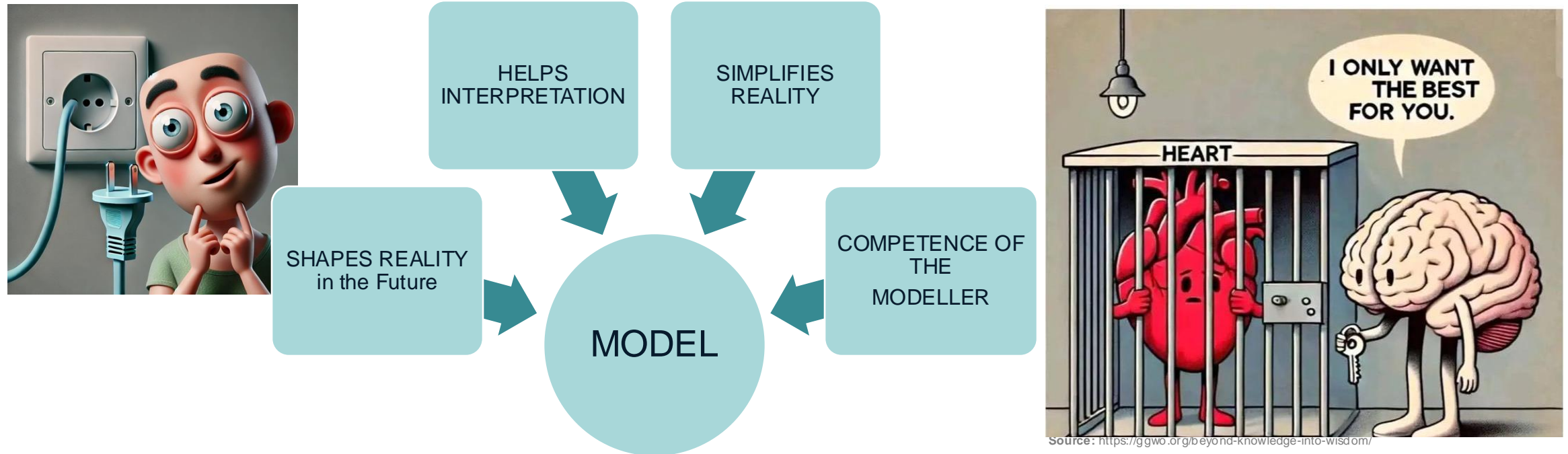Marriott Hotel, The Hague, Netherlands
December 6th and 7th, 2023
PROPERTY TAX CONFERENCE
Policy and Practice

Utilization of AI in the Most Challenging Aspects of Real Estate Market Analyses

According to classical decision theory, an economic entity (**homo economicus**) has **access to complete and perfect information,** which they process optimally to achieve maximum benefits.

Stiglitz pointed out that in **markets with imperfect information**, **specific decision-makers' behaviors contain the "missing" information**, making it impossible to obtain full and perfect information (in the classical sense).

HELPS INTERPRETATION

SIMPLIFIES REALITY

SHAPES REALITY in the Future

COMPETENCE OF THE MODELLER

MODEL

HEART

I ONLY WANT THE BEST FOR YOU.

The application of **hybrid solutions** based on cutting-edge technologies, including artificial intelligence, aims to demonstrate that **behavioral factors** are not just a part of the **real estate market** but **constitute and define it**, and that it is possible to understand their implications.
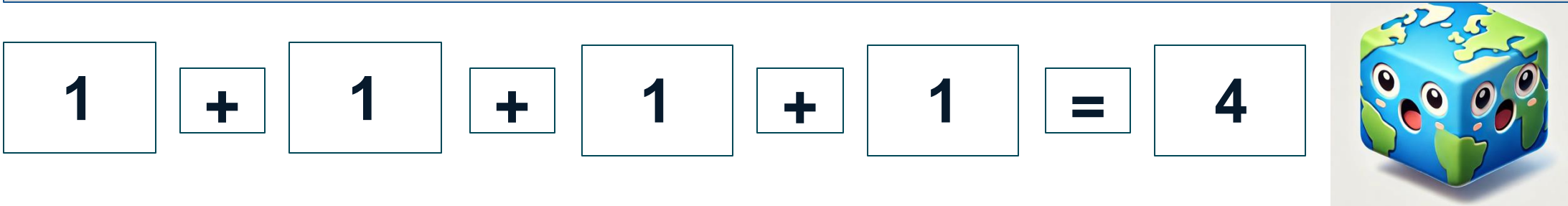
The **greatest PARADOX** lies in the fact that we use **extremely simple tools** and models to interpret some of the **most complex phenomena** and the **most important assets** in every person's life: PROPERTY
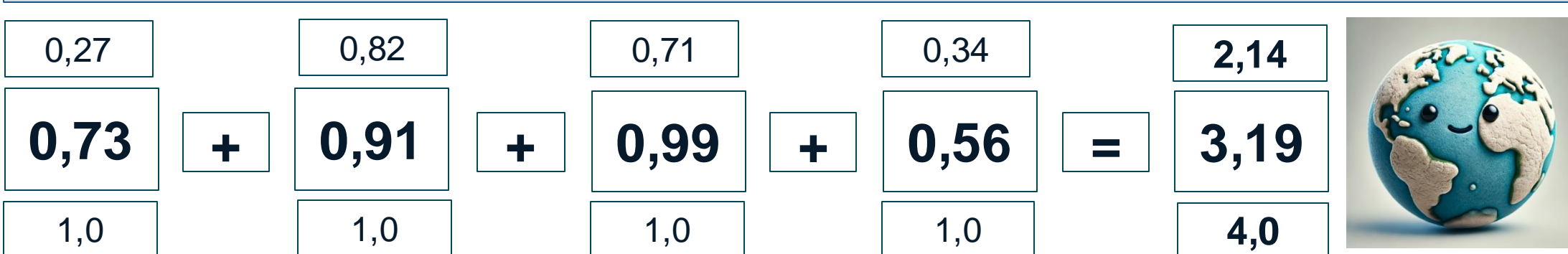
Models' vs Real Estate World

REALITY

Real estate features	+	Economy/ Politics	+	Fashion/ Trends	+	Human Behaviour	=	Genuine Value

Less valuable / **Genuine Value** / More valuable

CLASSIC MODEL/METHODS

1	+	1	+	1	+	1	=	4

AI & COGNITIVE BASED MODEL

0,27		0,82		0,71		0,34		2,14
0,73	+	**0,91**	+	**0,99**	+	**0,56**	=	**3,19**
1,0		1,0		1,0		1,0		**4,0**

What does it mean to be "similar"?

Two objects are more similar the less they differ from each other and significantly differ from the rest.

SIMILAR= HOMOGENIUS= INDISCERNIBLE ?

„**Similar**" means that two or more things have certain common features, enabling a comparison between them.

"**Homogeneous**" indicates objects that are practically identical, with little difference, often implying that they have the same specificity, characteristics, making it difficult (or even impossible) to distinguish significant differences between them.

It is imperative to emphasize that the concept of market homogeneity is intricately linked to property similarity, and essentially, they should not be subjected to separate analyses. In general, this interdependence can be expressed through a mathematical function, denoted as:

$$F(H,S) = k \cdot H(\alpha) \cdot S(\beta)$$

One thing we know for sure is the vague and fuzzy nature of the assessment of 'similarity'.

HUMAN-MACHINE FUSION

Current artificial intelligence - **ANI (Artificial Narrow Intelligence)**
Development of superintelligence - **AGI (Artificial General Intelligence)**

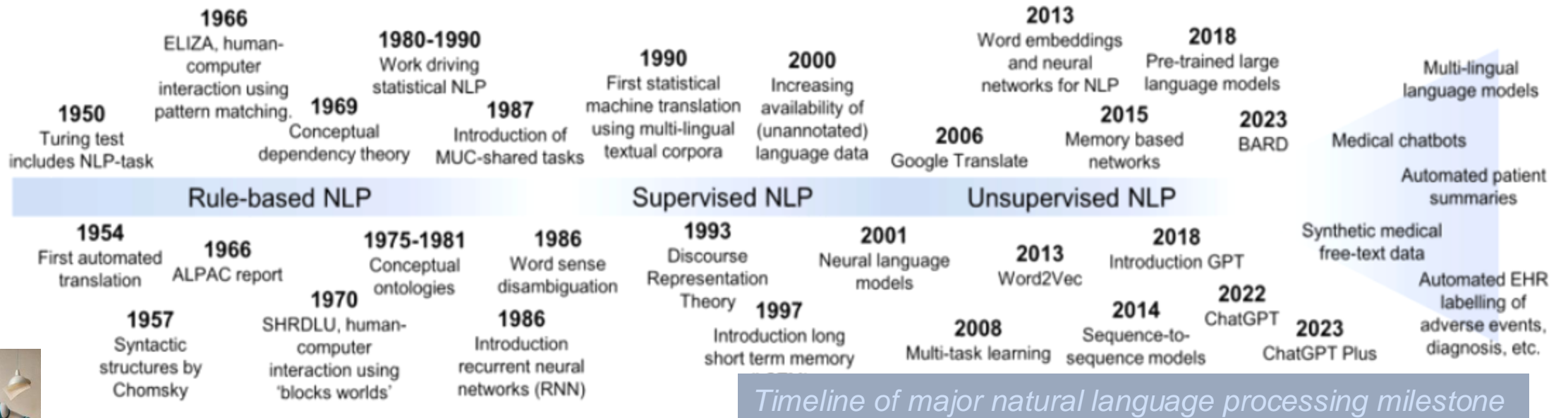**TECHNOLOGY IN THE SERVICE OF THE REAL ESTATE MARKET**

Cognitive systems and algorithms aim to faithfully **mimic** certain functions of the **human mind**, sometimes even surpassing it by **eliminating** some **human imperfections**, such as making errors, limited access to information, or subjectivity.

HOW CAN AI HELP IN SIMILARITY AND HOMOGENEITY ANALYSES?

Machine Learning Technologies

Large **L**anguage **M**odels
&
Cognitive systems

✓ synthesize information from text/language
✓ recognize context
✓ adapt responses
✓ uncovers hidden patterns,
✓ uncovers market sentiment and consumer preferences

1950
Turing test includes NLP-task

1966
ELIZA, human-computer interaction using pattern matching.

1969
Conceptual dependency theory

1980-1990
Work driving statistical NLP

1987
Introduction of MUC-shared tasks

1990
First statistical machine translation using multi-lingual textual corpora

2000
Increasing availability of (unannotated) language data

2006
Google Translate

2013
Word embeddings and neural networks for NLP

2015
Memory based networks

2018
Pre-trained large language models

2023
BARD

Multi-lingual language models

Medical chatbots

Automated patient summaries

Rule-based NLP | Supervised NLP | Unsupervised NLP

1954
First automated translation

1957
Syntactic structures by Chomsky

1966
ALPAC report

1970
SHRDLU, human-computer interaction using 'blocks worlds'

1975-1981
Conceptual ontologies

1986
Word sense disambiguation

1986
Introduction recurrent neural networks (RNN)

1993
Discourse Representation Theory

1997
Introduction long short term memory

2001
Neural language models

2008
Multi-task learning

2013
Word2Vec

2014
Sequence-to-sequence models

2018
Introduction GPT

2022
ChatGPT

2023
ChatGPT Plus

Synthetic medical free-text data

Automated EHR labelling of adverse events, diagnosis, etc.

Timeline of major natural language processing milestone

Cognitive systems within **Large-Language-Model (LLM)** frameworks operate through **language processing**, logical reasoning, and memory functions.

# Homogeneity Estate Linguistic Intelligence Omniscient Support (HELIOS)

2/2024 vol. 32

EDITOR'S CHOICE

Human-machine synergy in real estate similarity concept

Małgorzata Renigier-Biłozor [1], Artur Janowski [2]

This exploration **aims** to contribute to the definition of market homogeneity and the establishment of a **set of similar properties fingerprint** utilizing Large Language Models (LLM).

Real Estate Fingerprint - *utilizing LLM technology, represents a comprehensive approach involving the synergistic combination of standardized **tokens to capture and analyze unique characteristics and features** within the real estate domain.*
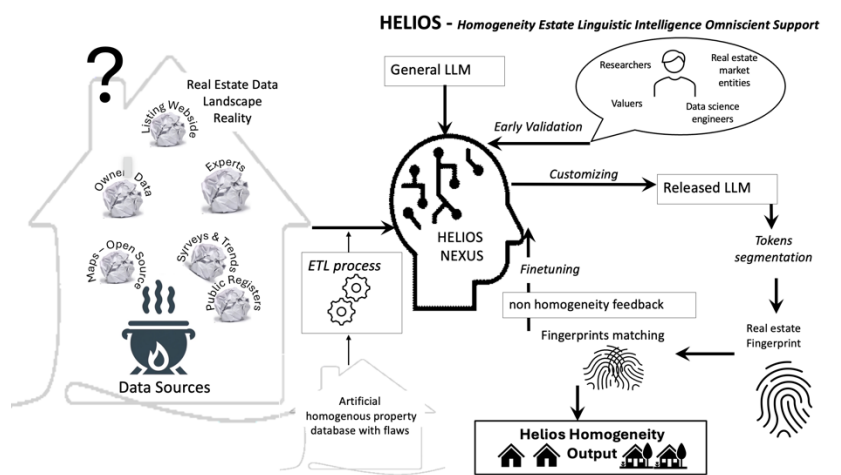
HELIOS: AN EFFECTIVE ALTERNATIVE TO CETERIS PARIBUS?

DATA
TRANSACTION DATABASE

4372 residential properties from 2018-2022y

28 features

Transaction ID, Type of Record, Document ID, Transaction Date, Price per square meter, Premises area in square meter, Market type, Seller, Buyer, Type of property, Gross property type, Type of object, Precinct, Registered area, Premises area in square meter, Share, Share in joint area, Type of rights, Address, Function, Storey, Building Construction, Year of Construction, Associated premises area, Elevator, other, Longitude, Latitude.

FRACTIONS

MISSING INFORMATION

ABBREVIATIONS

NUMBERS

NO DATA

TEXT

REPETITIONS

O.N.	Transaction ID	Type	Document	Transaction Date/Valuation	Date Preliminary Agreement	Type of Estimated Value	Gross Transaction Price	Net Transaction Price	VAT Rate (%)	Value	Price per m2 of Land	Price per m2 of Building	Price per m2 of Premises	Building area in m²	Building area m²	Premises area m²
1	65296	T	Akt not. 16/18 FAL z dnia 2018-01-02 wydany przez: NOT.FAL I.	2018-01-02			199 000,00		-				12 772,79	7		15.58
2	65296	T	Akt not. 16/18 FAL z dnia 2018-01-02 wydany przez: NOT.FAL I.	2018-01-02			199 000,00		-				12 772,79	7		15.58
3	65296	T	Akt not. 16/18 FAL z dnia 2018-01-02 wydany przez: NOT.FAL I.	2018-01-02			199 000,00		-				12 772,79	7		15.58
4	65413	T	Akt not. 17/18 JUR z dnia 2018-01-02 wydany przez: NOT.JUREWICZ-SZULC ELŻBIETA	2018-01-02			315 000,00		-				6 176,47	73		51.00
5	65413	T	Akt not. 17/18 JUR z dnia 2018-01-02 wydany przez: NOT.JUREWICZ-SZULC ELŻBIETA	2018-01-02			315 000,00		-				6 176,47	73		51.00
6	65413	T	Akt not. 17/18 JUR z dnia 2018-01-02 wydany przez: NOT.JUREWICZ-SZULC ELŻBIETA	2018-01-02			315 000,00		-				6 176,47	73		51.00
7	65413	T	Akt not. 17/18 JUR z dnia 2018-01-02 wydany przez: NOT.JUREWICZ-SZULC ELŻBIETA	2018-01-02			315 000,00		-				6 176,47	73		51.00
8	65414	T	Akt not. 24/17 JUR z dnia 2018-01-03 wydany przez: NOT.JUREWICZ-SZULC ELŻBIETA	2018-01-03			271 922,80		-		0,04		5 652,08	49	47.79	48.11
9	65414	T	Akt not. 24/17 JUR z dnia 2018-01-03 wydany przez: NOT.JUREWICZ-SZULC ELŻBIETA	2018-01-03			271 922,80		-		0,04		5 652,08	49	47.79	48.11
10	65414	T	Akt not. 24/17 JUR z dnia 2018-01-03 wydany przez: NOT.JUREWICZ-SZULC ELŻBIETA	2018-01-03			271 922,80		-		0,04		5 652,08	49	47.79	48.11
11	65414	T	Akt not. 24/17 JUR z dnia 2018-01-03 wydany przez: NOT.JUREWICZ-SZULC ELŻBIETA	2018-01-03			271 922,80		-		0,04		5 652,08	49	47.79	48.11
12	65931	T	2/18 MAC3 z dnia 2018-01-12 wydany przez: NOT.MACIĄG E., GOWKIELEWICZ-KOTLARZ I	2018-01-12			365 000,00		-				4 850,50	177		75.25
13	65931	T	2/18 MAC3 z dnia 2018-01-12 wydany przez: NOT.MACIĄG E., GOWKIELEWICZ-KOTLARZ I	2018-01-12			365 000,00		-				4 850,50	177		75.25
14	65931	T	2/18 MAC3 z dnia 2018-01-12 wydany przez: NOT.MACIĄG E., GOWKIELEWICZ-KOTLARZ I	2018-01-12			365 000,00		-				4 850,50	177		75.25
15	65931	T	2/18 MAC3 z dnia 2018-01-12 wydany przez: NOT.MACIĄG E., GOWKIELEWICZ-KOTLARZ I	2018-01-12			365 000,00		-				4 850,50	177		75.25
16	65931	T	2/18 MAC3 z dnia 2018-01-12 wydany przez: NOT.MACIĄG E., GOWKIELEWICZ-KOTLARZ I	2018-01-12			365 000,00		-				4 850,50	177		75.25
17	65931	T	2/18 MAC3 z dnia 2018-01-12 wydany przez: NOT.MACIĄG E., GOWKIELEWICZ-KOTLARZ I	2018-01-12			365 000,00		-				4 850,50	177		75.25

THE DATASET MODEL WAS TAILORED MADE TO LLM TECHNOLOGY REQUIREMENTS BASED ON ETL SOLUTION (EXTRACT, TRANSFORM, LOAD).

most usefulness features extraction

transformation in the text mode for LLM requirements

public property transaction register

Transaction Date: Seventeen of september, two thousand twenty year. Price per square meter: eight thousand and eighty-four. Premises area in square meter: forty-one. Market type: primary. Seller: legal entity. Type of property: premises. Type of object: UE. Precinct: Jasień. Share: one hundred and one out of ten thousand . Unknow Share in joint area. Type of rights: co-ownership. Address: ul.Krzysztofa Komedy fourteen Function: three-unit building.Storey: fourth. Building Construction: brick.Year of Construction: two thousand and twenty. Associated premises area: no data. Elevator: no elevator. Longitude: five hundred and thirty-one. Latitude: three hundred and forty-five

LOADING to

In order to obtaine results we employ a solution based on a broader **understanding of phrases rather than individual words**.

BERT Model Utilization

BERT transforms input text into numerical vectors representing the text's context, processing each word within its immediate context for semantic placement within the BERT domain.

Transformation to BERT Space

The property description database is transformed into a matrix of vectors (4372 x 769), where rows correspond to transactions and columns to BERT model features.

Vector Clustering:

Using the Self-Organizing Map (SOM) method for clustering vectors in this matrix, with cosine similarity for comparing vectors in the BERT domain.

Agile Method for Optimization

Agile method helped identify optimal parameters for the SOM grid size, iteration number, learning rate, and sigma size, achieving optimal results.
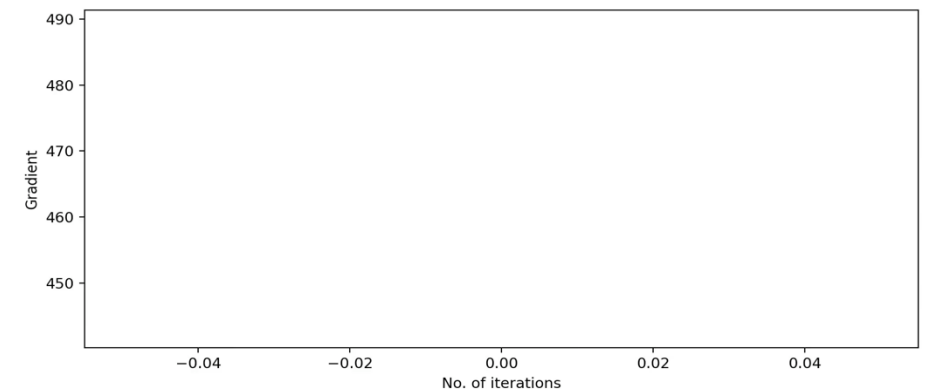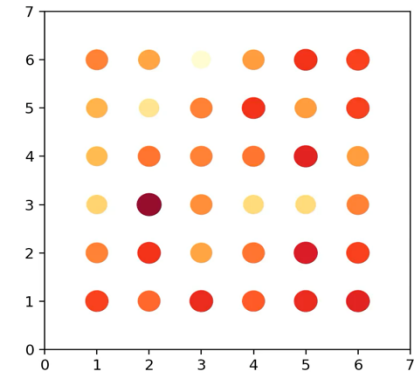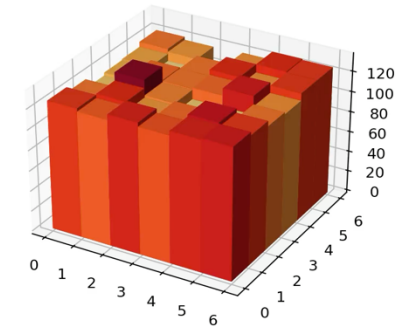
Number of Transactions	Node X	Node Y	Real Estate Fingerprint Characteristic
388 (the largest group)	2	0	transaction date: fourteen of November, two thousand twenty years, price per square meter: six thousand, two hundred and seventy-one, associated premises area: no data, building construction: other than brick, type of rights: co-ownership, seller: legal entity, market type: primary, precinct: Osowa
350 (the second largest group)	4	0	price per square meter: five thousand, eight hundred and eighty associated premises area: no data, precinct: Jasien, share in joint area: unknow, type of rights: ownership, type of object: ud, seller: individual, market type: secondary,
99 (the average group)			price per square meter: five thousand, three hundred and twenty-six, transaction date: ten of April, two thousand nineteen year.
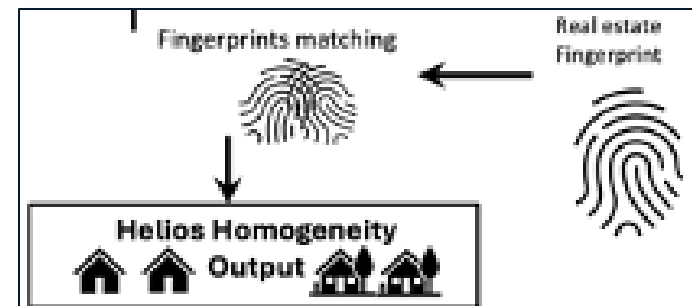
Node X	Node Y	Mean Distance	Quantity Vectors
0	0	7047	142
0	1	1362	95
0	2	3691	139
0	3	10008	155
0	4	12754	179
0	5	15113	50
1	0	6076	99
1	1	9219	195
1	2	3447	81
1	3	13874	57
1	4	10171	208
1	5	13292	131
2	0	12400	388
2	1	6089	178
2	2	14112	85
2	3	14801	61
2	4	12870	40
2	5	14899	62
3	0	5907	85
3	1	12342	118
3	2	14164	109
3	3	14780	162
3	4	13627	128
3	5	12202	129
4	0	15658	350
4	1	13937	37
4	2	13263	105
4	3	1225	52
4	4	13186	98
4	5	14883	54
5	0	8542	113
5	1	11567	216
5	2	3087	25
5	3	14855	58
5	4	13784	134
5	5	13690	54
Average	10887		
			Sum 4372

Fingerprints matching — Real estate Fingerprint — Helios Homogeneity Output

The table presents a homogeneous group of properties, each with a unique "**fingerprint**" identified using the TF-IDF (Term Frequency-Inverse Document Frequency) method. This technique highlights the most important features by comparing term frequency within a document to its frequency across all documents, enabling focused analysis and prioritization.

Parameters	The lower value	The highest value	Optimal value
Iterations	1000	10000	4000
Sigma	3.1	0.1	1.5
Learning rate	3.1	0.1	1.1
Average of mean distance	31841	8653	10887
Grid resolution	2x2	10x10	6x6

NATIONAL SCIENCE CENTER

Methodology of consensus-oriented roadmaps development for property tax transition.
GRANT 2024-2027

dr hab. inż. Małgorzata Beata Renigier-Biłozor
prof. dr hab. inż. Sabina Źróbek
dr Alina Źróbek-Różanska
dr inż. Elżbieta Zysk
dr hab. inż. Artur Janowski

„It is the duty of a good Shepherd to shear his flock, not to slaughter it" Tiberius Caesar (42 BC – 37 AD)

The project aims to increase social inclusion in transforming the property tax system, transitioning from area-based to value-based taxation.

The key challenge is resistance to change due to fears of worsening personal well-being, hindering the implementation of necessary reforms

„The art of taxation consists in so plucking the goose as to obtain the most feathers with the least hissing." Jean Baptiste Colbert (1619 -1683)

ANALYSIS OF SURVEY RESPONSES

Taxpayers Resilience Analysis

Stressor Reduction
Stressor Compensation
Stress Resilience

493 respondents
447 features in one hot encoding form (0,1)
220371 cells for analysis
3.57 x 10^{134} number of response combinations

Polish Language???

Corpus Viewer

Tokens: 1933
Types: 864
Matching documents: 479/479
Matches: n/a

Search features
Filter…
S level_4
S level_5
S level_6
C id;Wynik

Display features
Filter…
S level_0
S level_1
S level_2
S level_3

☑ Show Tokens & Tags
☑ Auto send is on

441 Document 441
442 Document 442
443 Document 443
444 Document 444
445 Document 445
446 Document 446
447 Document 447
448 Document 448
449 Document 449
450 Document 450
451 Document 451
452 Document 452
453 Document 453
454 Document 454

LLM MODEL FOR SURVEYS ANALYSES

Model MS MARCO MiniLM - strictly dedicated for semantic similarity tasks (cross-encoder) – multilanguage, lightweight and efficient

Features	BERT	MiniLM (ms-marco-MiniLM-L-v2)
Architecture	Full-sized Transformer model (24 layers)	Lightweight MiniLM model (12 layers)
Model size	110M – 340M parameters	33M parameters
Training	Masked LM + Next Sentence	Fine-tuning on MS MARCO
Text comparison	Bi-encoder	Cross-encoder
Speed	Slower	Faster
Specialization	General NLP tasks	Text comparison, retrieval
Hardware requirements	Higher	Lower

SEMANTIC SIMILARITY OF TEXTS IN LLM's

Statitics =

Semantic/Meaning ≠

The dog eats the frog.

The frog eats the dog.

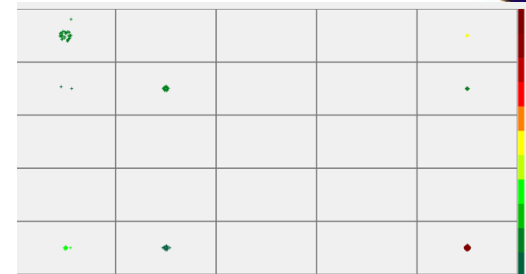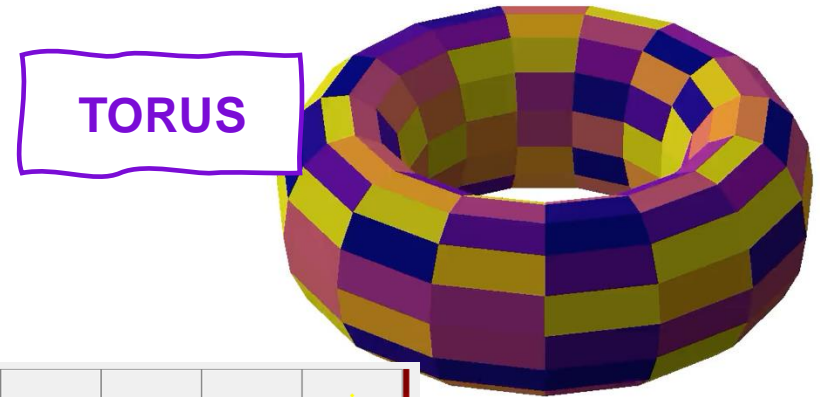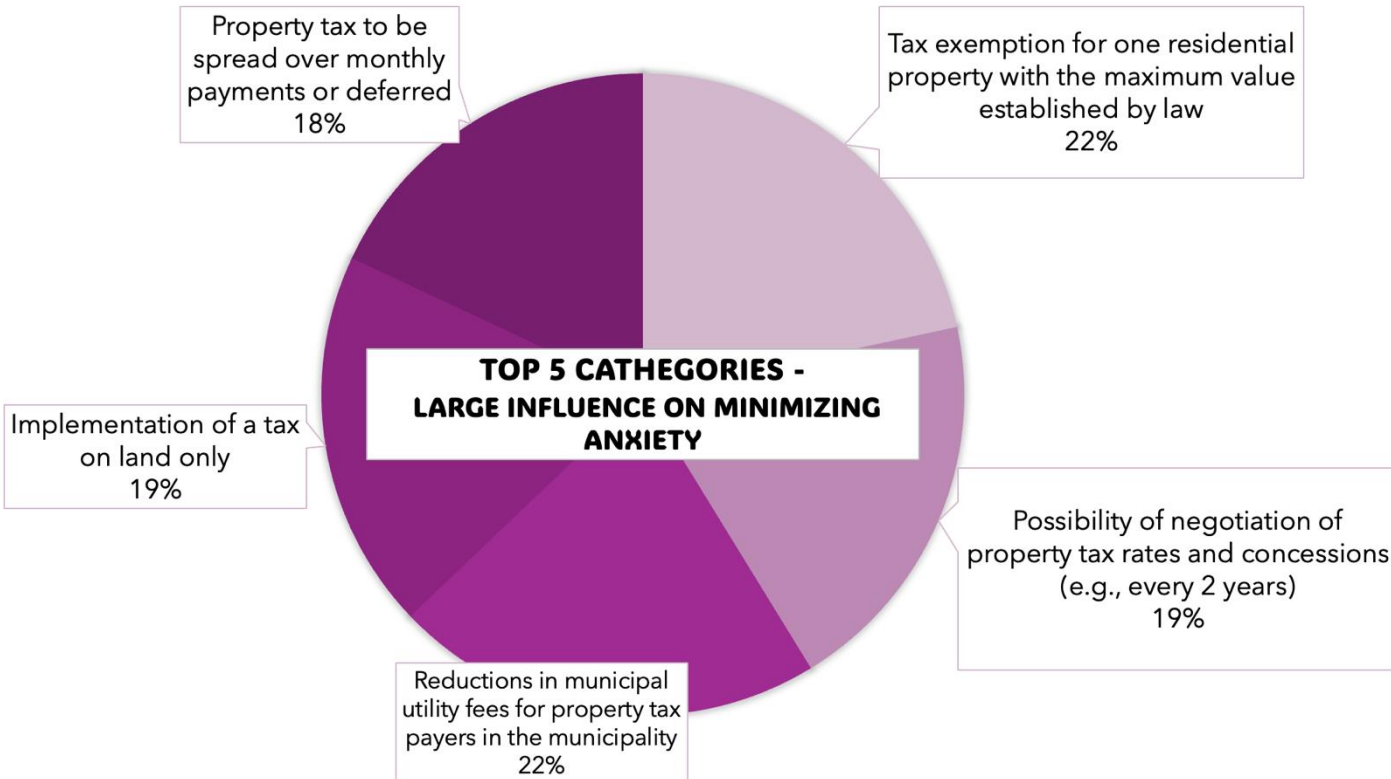# The efficiency of the HELIOS (for surveys) **verify** based on **numerical** and **substantial** approach

No of respondents	Node x	Node y	Fingerprint of a Homogeneous Respondents' Profile
1	11	4	1. Transparency: Access to clear information about taxes and their allocation. 2. Engagement: Participation in consultations and decision-making. 3. Flexibility: Options for installments, deferrals, and tax reliefs. 4. Social Sensitivity: Support for local initiatives funded by taxes. 5. Negotiability: Preference for negotiating rates and reliefs.
147	11	3	1. Prefer active participation in municipal decisions, e.g., online voting. 2. Expect flexibility in payments, such as installments, deferrals, and personalized consultations. 3. Appreciate income-based reliefs, family (3+ children) benefits, and income-dependent tax thresholds. 4. Have moderate trust in digital tools, such as tax calculators. 5. Low impact of municipal spending information on reducing their concerns.
150	12	4	1. Perception of information: Moderate impact of detailed information about municipal spending and tax calculation on reducing concerns. 2. Preferences for reliefs and negotiations: Interest in tax reliefs and negotiation options, though less emphasized compared to other groups. 3. Payment flexibility: Clear payment terms and the ability to adapt them are important. 4. Skepticism towards digital tools: Relatively low significance of online tools, such as municipal websites, in alleviating concerns.
6	11	1	1. High need for consultations: Taxpayers strongly value personalized consultations with officials to address tax-related concerns. 2. Moderate interest in municipal decision-making: Limited preference for active participation in decision-making processes. 3. High flexibility in payment options: Clear terms with options for installments or deferrals are essential for comfort 4. Low trust in digital tools: Digital tools like tax calculators are considered less significant. 5. Moderate interest in tax reliefs: Tax reliefs are moderately valued, with a focus on practical solutions. 6. Balanced view on project visualizations: Moderate appreciation for visualizations of tax-funded projects to enhance transparency.
			1. Information and transparency: Moderate impact of detailed information about municipal spending and tax calculation changes on reducing concerns. 2. Tax reliefs and negotiation options: High importance of tax reliefs and the ability to negotiate rates, particularly every two years, as a key factor in alleviating concerns. 3. Flexible payment conditions: Preference for clear payment terms with options for installments or deferrals. 4. Significance of tax-funded investments: Emphasis on the importance of visualizing projects funded by taxes to build trust.

Parameters	Optimal results
Iteration	1000
Sigma	5
Topological Error	0.22
Quantization Error	1.82
N-Gram Token Frequency Extraction	25
Grid resolution	15x15

TORUS

What reduces taxpayers' stress and anxiety during a property tax system transformation?

TOP 5 CATHEGORIES - LARGE INFLUENCE ON MINIMIZING ANXIETY

Pie chart segments:
- Property tax to be spread over monthly payments or deferred — 18%
- Tax exemption for one residential property with the maximum value established by law — 22%
- Implementation of a tax on land only — 19%
- Possibility of negotiation of property tax rates and concessions (e.g., every 2 years) — 19%
- Reductions in municipal utility fees for property tax payers in the municipality — 22%

	11.3	12.3	12.4	11.1
11.3	Same group	11.3 prefers more participation; 12.3 values tax reliefs more and trusts digital tools slightly more.	11.3 focuses on participation; 12.4 emphasizes pro-environmental projects and visualizations.	11.1 values consultations more; 11.3 emphasizes participation.
12.3	12.3 values tax reliefs and digital tools more; 11.3 prefers active participation.	Same group	12.3 prioritizes participation and payment flexibility; 12.4 highlights pro-environmental focus.	11.1 trusts digital tools less; 12.3 values reliefs more.
12.4	12.4 emphasizes pro-environmental focus and visualizations; 11.3 prefers active participation.	12.4 focuses on visualizations and environmental aspects; 12.3 prioritizes participation and payment flexibility.	Same group	11.1 prioritizes consultations; 12.4 focuses on visualizations.
11.1	11.1 values consultations more; 11.3 emphasizes participation.	11.1 trusts digital tools less; 12.3 values reliefs more.	11.1 prioritizes consultations; 12.4 focuses on visualizations.	Same group

Strategic roadmap indicators

Conditions of stress reduction	11.1	11.3	12.3	12.4
High need for consultations with officials	✓			
Active participation in municipal decision-making		✓		
Flexibility in payment options (installments, deferrals)	✓	✓	✓	
Trust in digital tools (e.g., calculators, visualizations)				✓
Focus on pro-environmental projects				✓
High importance of tax reliefs (income-based, family benefits)			✓	
Significance of visualizations of tax-funded projects			✓	✓
Emphasis on negotiation opportunities (e.g., bi-annual reviews)		✓		

The EU Artificial Intelligence Act, 2024

The **European Union** became **the first global organization** to regulate the use of **AI through** the implementation of the AI Act. It was signed on March 13, 2024, published on July 12, 2024, and came into force on August 2, 2024. The regulations will be implemented in stages, with full application by August 2026. **All 27 member states unanimously approved the AI Act.**

AI Risk Classification:
• **Low Risk**: Light transparency obligations (move recommendation, e-commerce)
• **High Risk**: Strict requirements (risk management, medical treatments, documentation).
• **Unacceptable Risk**: Banned AI (e.g., manipulative behavior techniques, crime prediction).

Exceptions apply to systems used exclusively for:
- Military and defence purposes
- Research purposes
- Special provisions for AI use by law enforcement in extraordinary situations

These regulatory initiatives are poised to significantly impact various sectors, **including real estate valuation**, by providing clarity regarding legal responsibilities and fostering a safer and more accountable environment for AI deployment in the EU.

| FORMAL SIGNING: MARCH 13, 2024 | PUBLICATION: JULY 12, 2024 | IN FORCE: AUGUST 2, 2024 | FULL APPLICATION: AUGUST 2026 |