Fachbereich
Geoinformatik

GeoAI für die Datenvalidierung

Key Concepts um Daten für AI „fit“ zu machen

Paris-Lodron-University Salzburg
Department of Geoinformatics – Z_GIS

Johannes Scholz

Department of Geoinformatics – Z_GIS
Paris-Lodron-University Salzburg

johannes.scholz@plus.ac.at
www.zgis.at || www.johannesscholz.net
https://linkedin.com/in/johannes-scholz-gisc
@Joe_GISc
@Joe_GISc@mastodon.online
@joegisc.bsky.social

Z_GIS

Table of Contents

Geospatial Artificial Intelligence

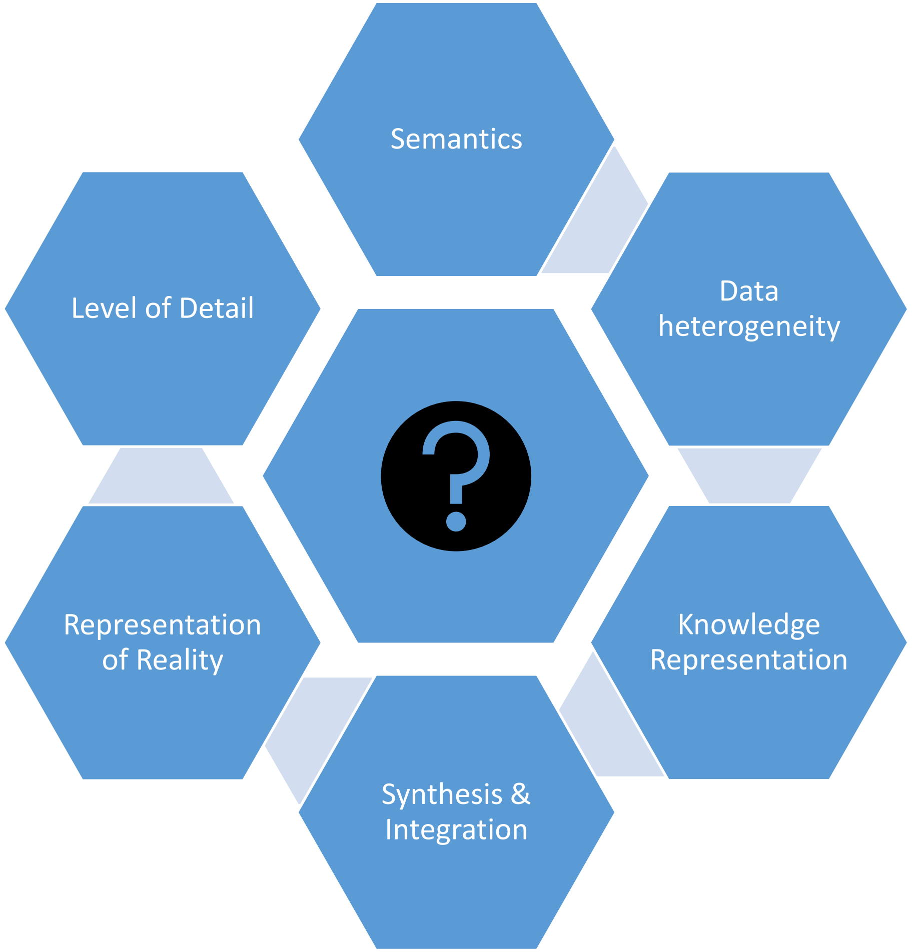
- "**GeoAI** can be regarded as a study subject to **develop intelligent computer programs** to mimic the processes of **human perception, spatial reasoning**, and **discovery** about **geographical phenomena** and **dynamics**
  - to advance our knowledge,
  - to solve problems in human environmental systems and their interactions,
  - with a focus on spatial contexts and roots in geography or GIScience." (Gao, 2021)

- **Spatially explicit models incorporating spatial contexts** (Yan et al., 2018) can **outperform traditional nonspatial AI** models in many tasks:
  - image classification,
  - geographic knowledge graph summarization (Yan et al., 2019),
  - and geographic question-answering problems (Mai et al., 2019).

GeoAI – Additional Questions?

- Questions that may surface when:
  - Representing
  - Manipulating
  - Storing
  - Analyzing, and
  - Visualizing

Geographic Data …!

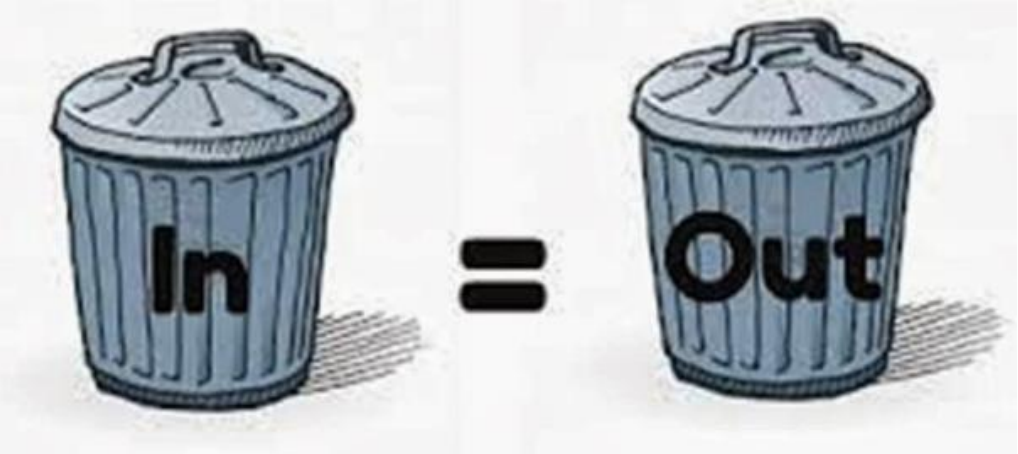


Motivation

„Garbage in, Garbage out.“

AI Models are dependent on training and verification data

- Heterogeneity
- Incompleteness
- Bias
- Flaws



Data Quality as Critical Success Factor

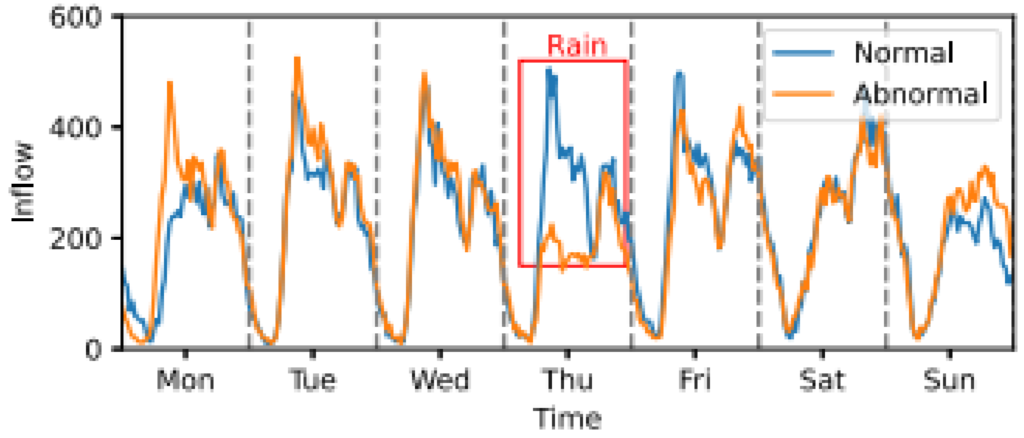
- **Completeness:**
  - Missing values can distort patterns or hide important relationships.
- **Consistency:**
  - Inconsistent formats or conflicting entries complicate modeling and interpretation.
- **Representativeness:**
  - Skewed samples result in unfair or non-generalizable models
- **Timeliness:**
  - Outdated data fails to reflect current realities—especially critical in dynamic domains like mobility or energy. ng and interpretation.

GeoAI for Anomaly Detection

PARIS LODRON UNIVERSITÄT SALZBURG
Fachbereich Geoinformatik

- **GeoAI leverages spatial and temporal patterns** to identify outliers
  - **Spatial context** (e.g. neighborhood similarity, geographic clustering)
  - **Temporal dynamics** (e.g. recurring patterns, seasonality)

- Key Methods:
  - **DBSCAN:** Density-based clustering to detect spatial outliers without needing labeled data.
  - **Autoencoders:** Neural networks trained to reconstruct input—large reconstruction errors signal anomalies.
  - **ST-GCN** (Spatio-Temporal Graph Convolutional Networks): Models spatial dependencies and temporal evolution simultaneously, ideal for dynamic sensor networks.



Deng, L., Lian, D., Huang, Z., & Chen, E. (2022). Graph convolutional adversarial networks for spatiotemporal anomaly detection. *IEEE Transactions on Neural Networks and Learning Systems*, *33*(6), 2416-2428.

Bias Correction with GeoAI (1)

PARIS LODRON UNIVERSITÄT SALZBURG | Fachbereich Geoinformatik

- GeoAI integrates **spatial context to detect and mitigate bias** that arises from geographic (and/or demographic) imbalances.
- Spatial fairness-aware learning ensures that models treat regions and populations equitably, even when data availability varies.

- Key Methods
  - **Reweighting**: Adjusts the influence of overrepresented regions or groups to balance model learning.
  - **Fairness-aware algorithms**: Incorporate fairness constraints during training to reduce disparate impact.
  - **Spatial stratification**: Ensures that training samples are geographically diverse and representative.

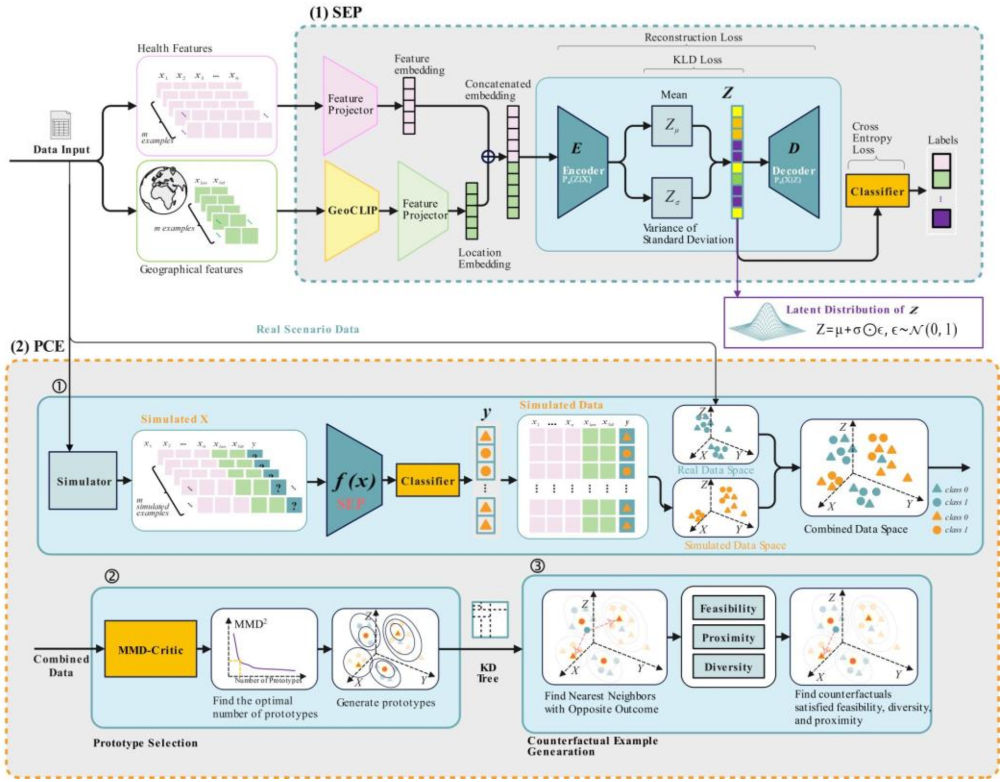
Wang, Z., Wu, N., Cao, Q., Xia, J., Liu, Z., Xie, Y., ... & Mai, G. (2025). GeoBS: Information-Theoretic Quantification of Geographic Bias in AI Models. *arXiv preprint arXiv:2509.23482*.
Xie, Y., Jia, X., Chen, W., & He, E. (2023). Heterogeneity-aware deep learning in space: Performance and fairness. In Handbook of Geospatial Artificial Intelligence (pp. 151-176). CRC Press.

Bias Correction with GeoAI (2)

Fachbereich Geoinformatik

- Some issues we should pay attention to:
  - **Spatial Autocorrelation**: Nearby locations often share similar characteristics. Fairness-aware models must account for this to avoid overfitting to dense urban clusters.
  - **Sensitive Attributes**: In geospatial contexts, attributes like income level, ethnicity, or infrastructure access may correlate with location—raising fairness concerns.
  - **Representation Bias**: Volunteered geographic information (VGI) and sensor data often reflect the interests of more connected or affluent communities.

- **Counterfactual Fairness in Spatial Models:**
  - Ensure that predictions remain consistent if a **location's demographic profile** were different.



Ma, J., Guo, R., Zhang, A., & Li, J. (2023, August). Learning for counterfactual fairness from observational data. In *Proceedings of the 29th ACM SIGKDD conference on knowledge discovery and data mining* (pp. 1620-1630).
Ma, J., Guo, R., Wan, M., Yang, L., Zhang, A., & Li, J. (2022, February). Learning fair node representations with graph counterfactual fairness. In Proceedings of the fifteenth ACM international conference on web search and data mining (pp. 695-703).
Zhang, J., Mu, L., Zhang, D., Chen, Z., Rajbhandari-Thapa, J., Pagan, J. A., ... & Zhou, Z. (2025). SpaCE: a spatial counterfactual explainable deep learning model for predicting out-of-hospital cardiac arrest survival outcome. International Journal of Geographical Information Science, 1-32.

GeoAI for Semantic Enrichment

Fachbereich Geoinformatik

- **Semantic enrichment** refers to enhancing **raw geospatial data with meaningful, contextual information** – improving **interpretability** & **usability** for **downstream tasks**

Ontology + (Geo)Data = (Geo)Knowledge Graph

- GeoAI Capabilities:
  - **Text + Location Fusion:** Combine textual metadata (e.g., descriptions, tags) with geolocation to classify or cluster features.
  - **Embedding Techniques:** with NLP, LLMs or BERT generating semantic vectors for objects, places and/or events.
  - **Ontology Integration:** Link data to structured vocabularies (e.g., INSPIRE themes, GeoNames) for interoperability.

Qiu, Q., Zheng, S., Tian, M., Li, J., Ma, K., Tao, L., & Xie, Z. (2024). A deep neural network model for Chinese toponym matching with geographic pre-training model. *International Journal of Digital Earth*, *17*(1), 2353111.

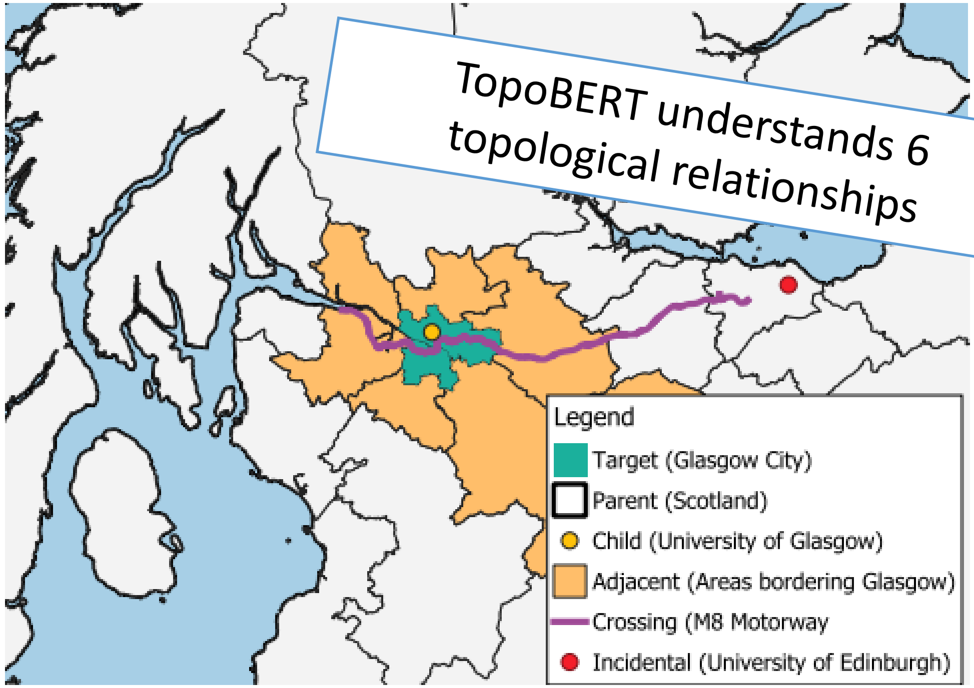
GeoAI for Semantic Enrichment

GeoBERT

"London is a city in Ontario, Canada. Its station, situated on York Street, has rail links to the neighbouring towns of Woodstock, and onward to Toronto."

- **GeoBERT**
  - built around a Bidirectional Encoder Representations from Transformers (BERT) model (Devlin et al., 2019)
  - BERT can extract syntactical and semantic information from sentences, including:
    - subject/object/verb relationships (Nastase and Merlo, 2023)
    - and capturing structural linguistic information (Jawahar et al., 2019).
  - BERT has also been shown to be effective in identifying spatial relationships within natural language (Shin et al., 2020),
    - "Tom is on the box"
    - "The cat is in the house"

TopoBERT understands 6 topological relationships



Shingleton, J., & Basiri, A. (2024). Enhancing toponym identification: Leveraging Topo-BERT and open-source data to differentiate between toponyms and extract spatial relationships. *AGILE: GIScience Series*, *5*, 1-10.
Shingleton, J., & Basiri, A. (2025). How close is" close"? An analysis of the spatial characteristics of perceived proximity using Large Language Models. *AGILE: GIScience Series*, *6*, 11.
Qiu, Q., Zheng, S., Tian, M., Li, J., Ma, K., Tao, L., & Xie, Z. (2024). A deep neural network model for Chinese toponym matching with geographic pre-training model. *International Journal of Digital Earth*, *17*(1), 2353111.

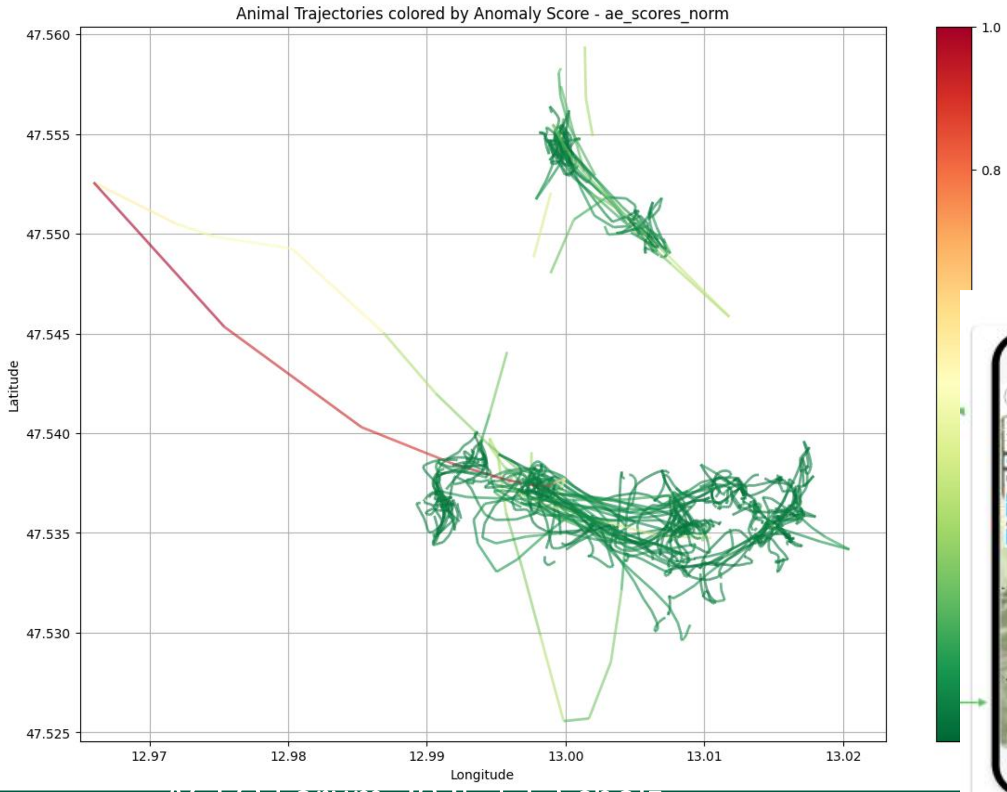
Applications in Selected Projects

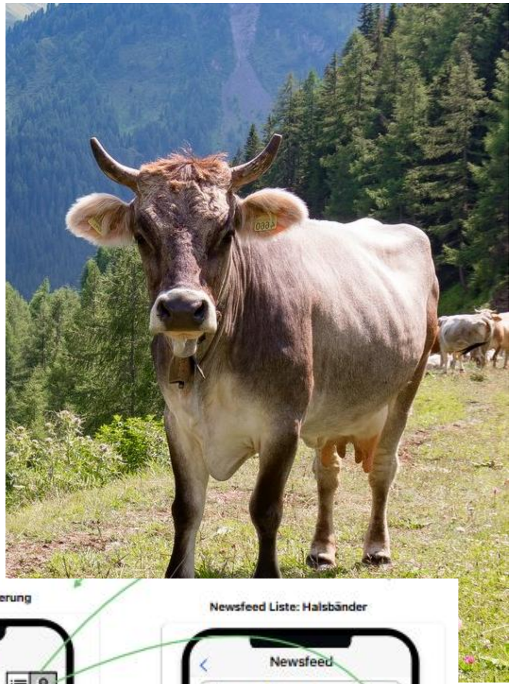
Virtual Shepherd

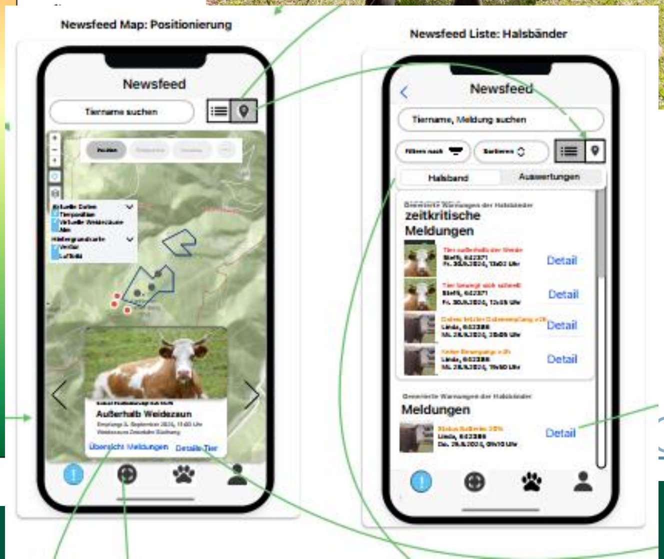
Paris Lodron Universität Salzburg | Fachbereich Geoinformatik

TU Wien Informatics · Moving Layers · Viehfinder

- Virtual Fencing and Virtual Shepherd that supervises lifestock in Alpine regions
- Detection of anomalies in cow movement
  - Trajectories
  - Additional sensors (accelerometer, temperature, …)
- GeoAI:
  - Autoencoder-based anomaly detection
  - Edge Devices(!)





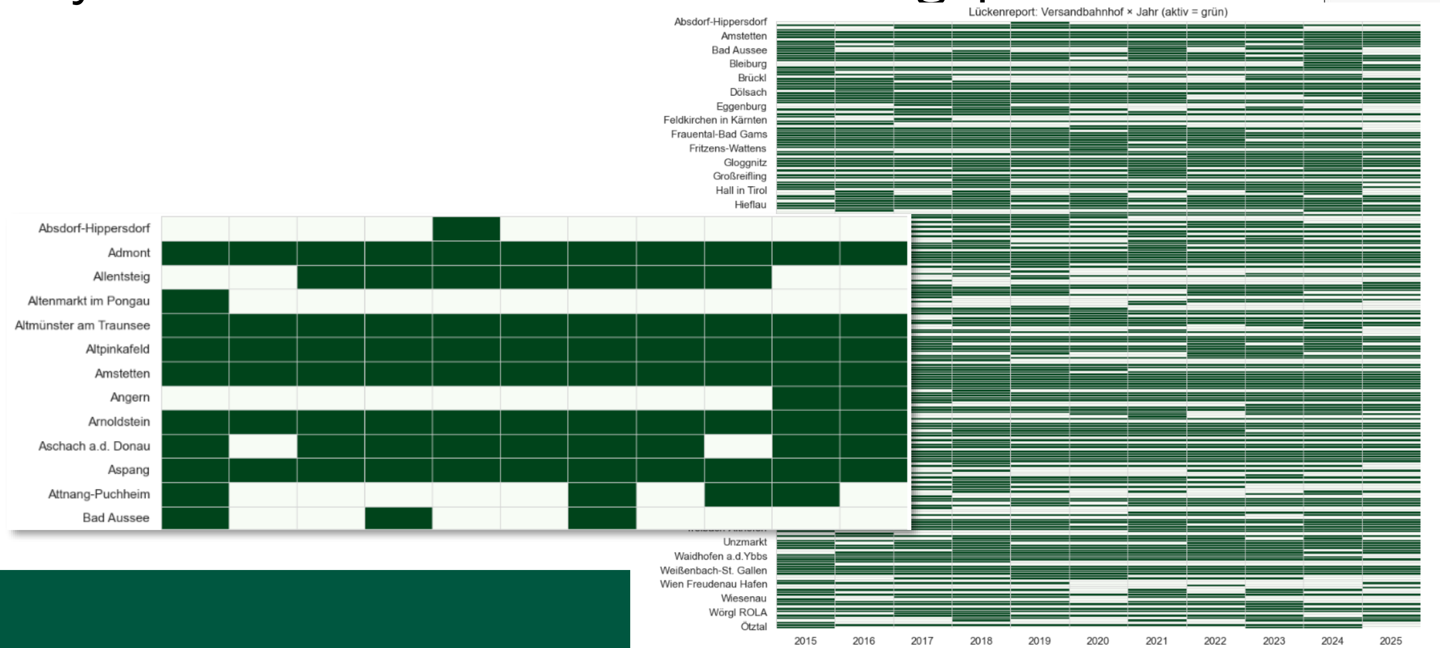


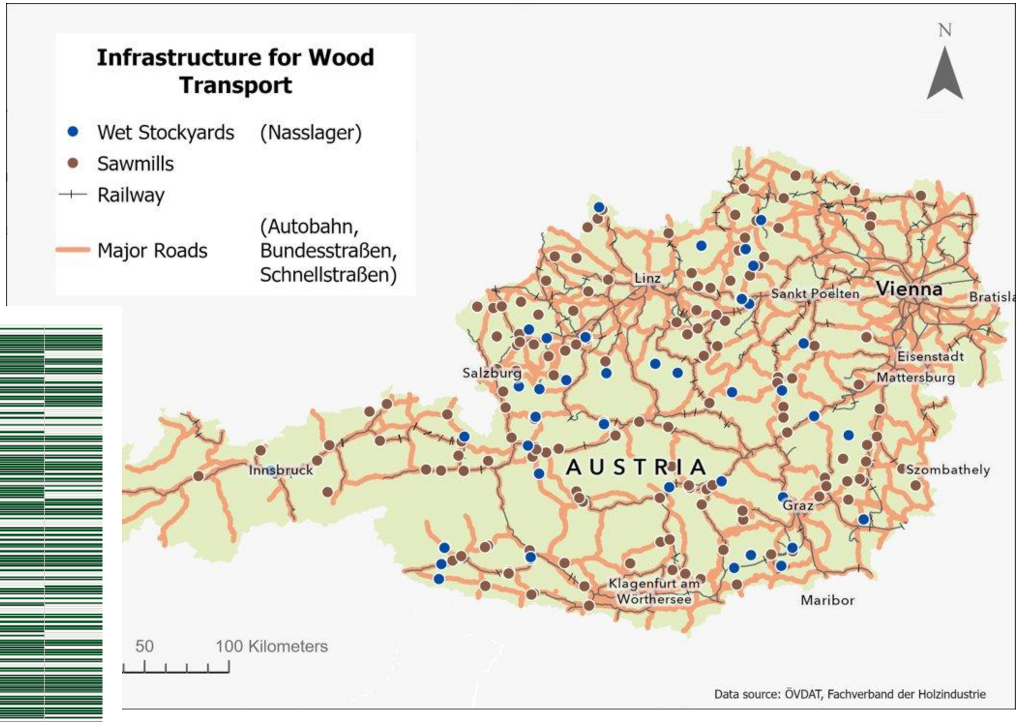
GIS

RegioWoodTrain

Fachbereich Geoinformatik

- The project develops cooperative, AI-driven strategies to shift wood transport toward climate-friendly rail logistics, improving resilience and sustainability in Austria's regional supply chains.
- ST-GNNs based on a graph-based representation of the data
- Identification of missing data and usage of synthetic data to close some "data gaps"



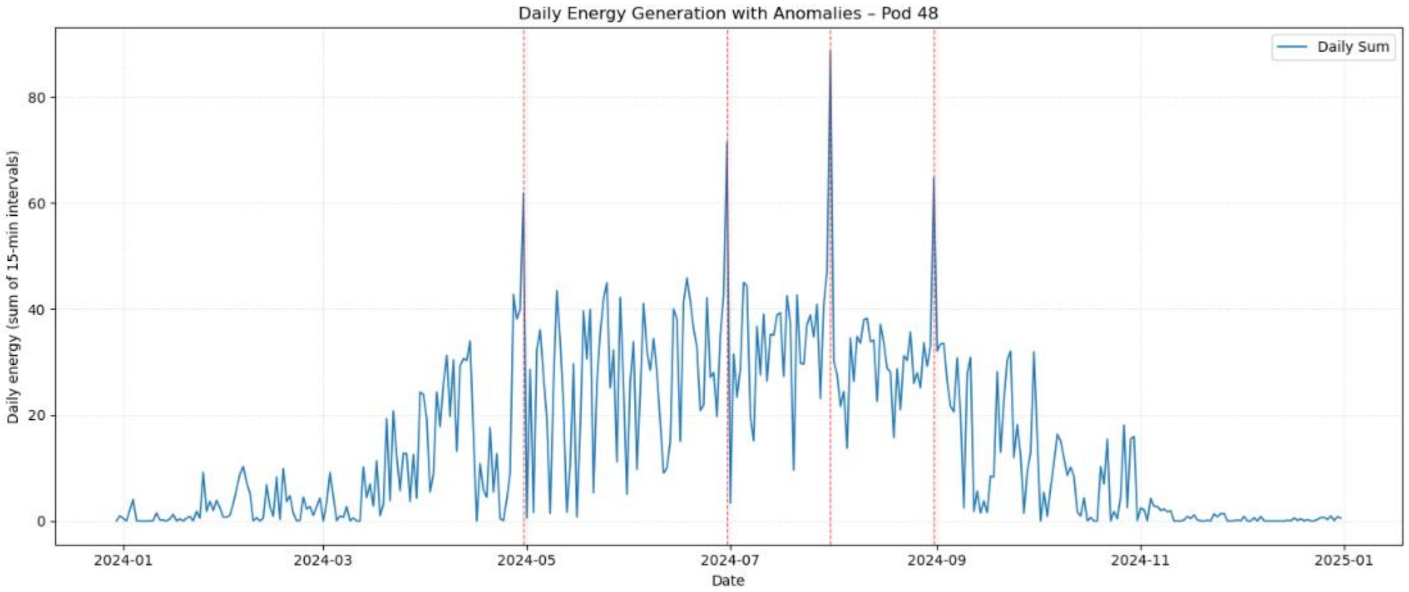


ZGIS

iKlimEt

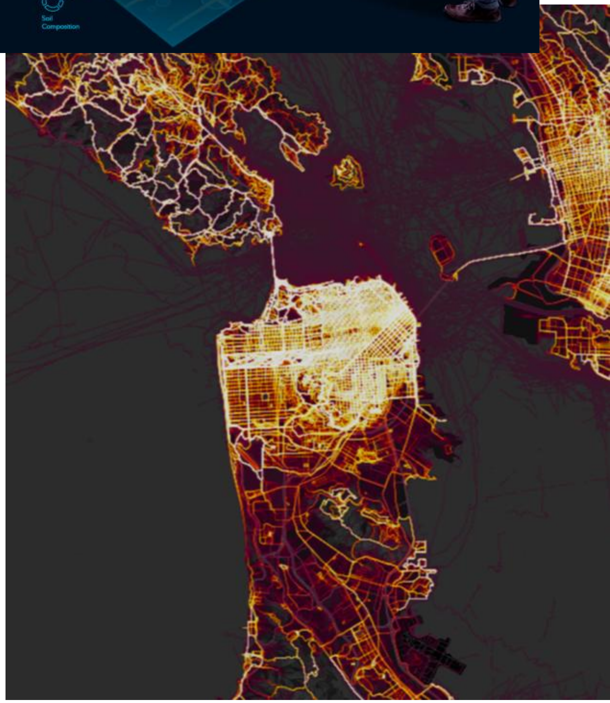
- The iKlimEt project develops AI-driven simulation tools for integrated, climate-resilient energy system planning, focusing on sector-coupled optimization and validating them through a case study in Styria to support Austria’s 2040 climate neutrality goals.

- LSTM autoencoder detects anomalies in smart meter time series data (energy generation).
  - Divided into daily sliding windows (96 values of 15 minutes each) with a 12-hour overlap and then normalised so that all pods are comparable.
- Each sliding window therefore represents one day of energy generation → training data.
- Reconstruction errors were calculated, and windows with the highest errors top 3% were marked as anomalies.



Summary

- **Why GeoAI?**
  - Enhances AI models with spatial reasoning & geographic context
  - Tackles challenges in data quality, bias, and semantic richness
- **Core Use Cases**
  - **Anomaly Detection:** DBSCAN, Autoencoders, ST-GCN
  - **Bias Correction:** Reweighting, fairness-aware learning, spatial stratification
  - **Semantic Enrichment:** GeoBERT, TopoBERT, Ontology integration



GeoAI bridges spatial intelligence and machine learning to improve data quality, fairness, and interpretability—critical for robust, ethical AI systems

Fachbereich
Geoinformatik

Paris Lodron Universität Salzburg
Fachbereich Geoinformatik

GeoAI für die Datenvalidierung

Key Concepts um Daten für AI „fit“ zu machen

Paris-Lodron-University Salzburg
Department of Geoinformatics – Z_GIS

Johannes Scholz

Department of Geoinformatics – Z_GIS
Paris-Lodron-University Salzburg

johannes.scholz@plus.ac.at
www.zgis.at || www.johannesscholz.net
https://linkedin.com/in/johannes-scholz-gisc
@Joe_GISc
@Joe_GISc@mastodon.online
@joegisc.bsky.social

Z_GIS