

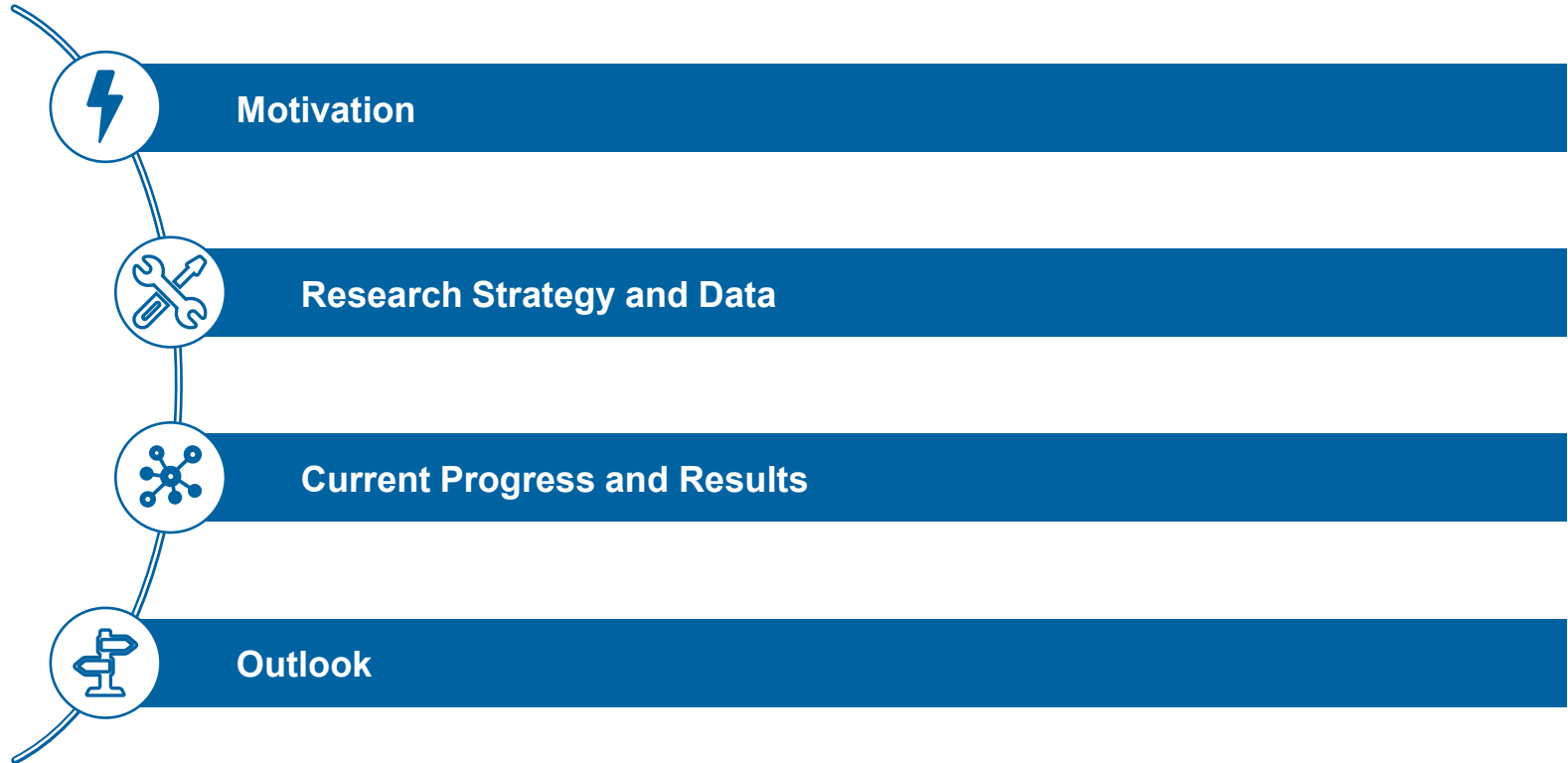
Geospatial Data and Multimodal LLM for Validating Company Data

Gabriela Alves Werb, Patrick Felka, Lisa Reichenbach, Gabriel Thiem, Susanne Walter, Ece Yalcin-Roder

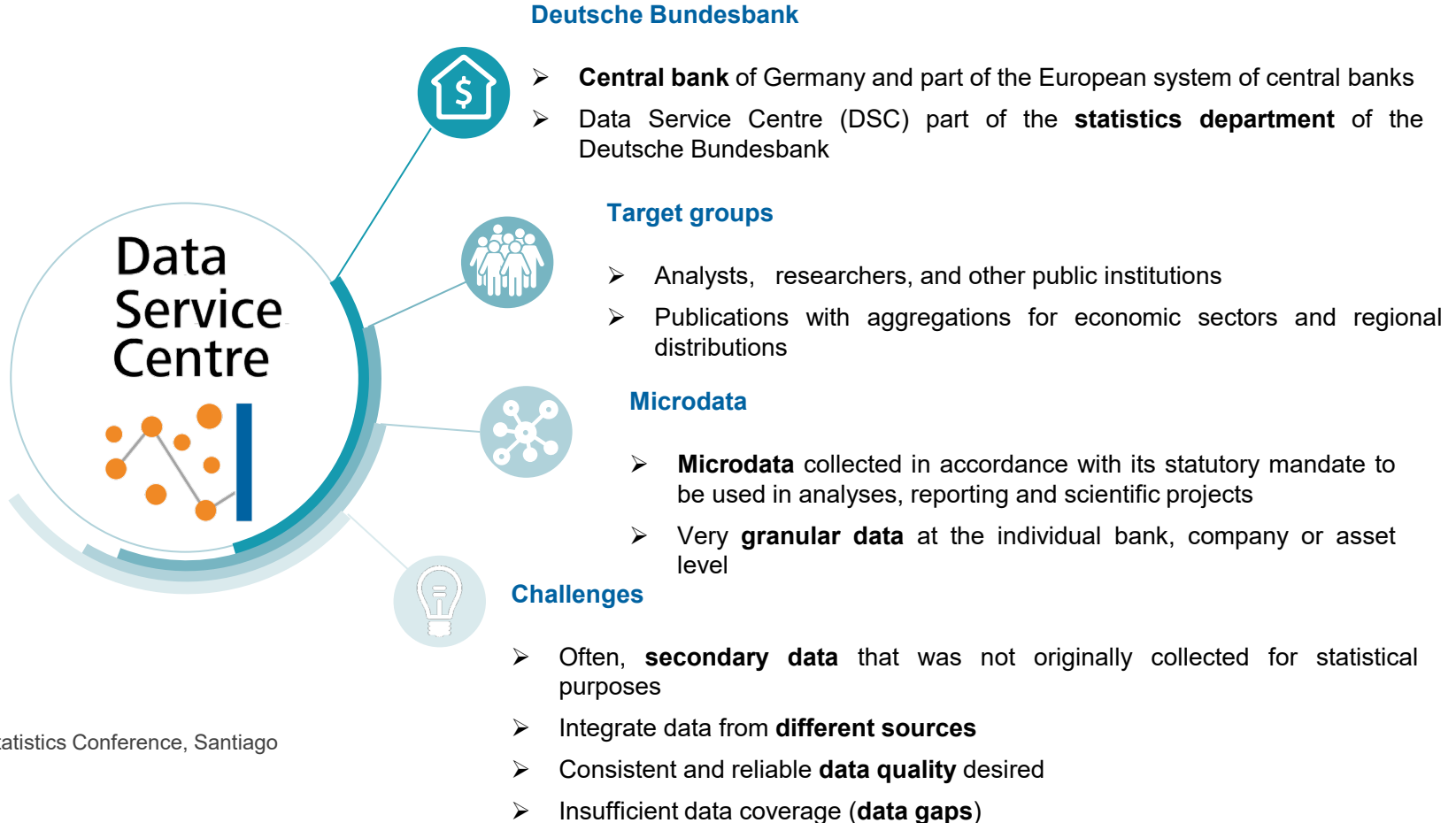
Dr. Susanne Walter, susanne.walter@Bundesbank.de

Disclaimer: The views expressed in this presentation are personal views of the authors and do not necessarily reflect the views of Deutsche Bundesbank or the Eurosystem.

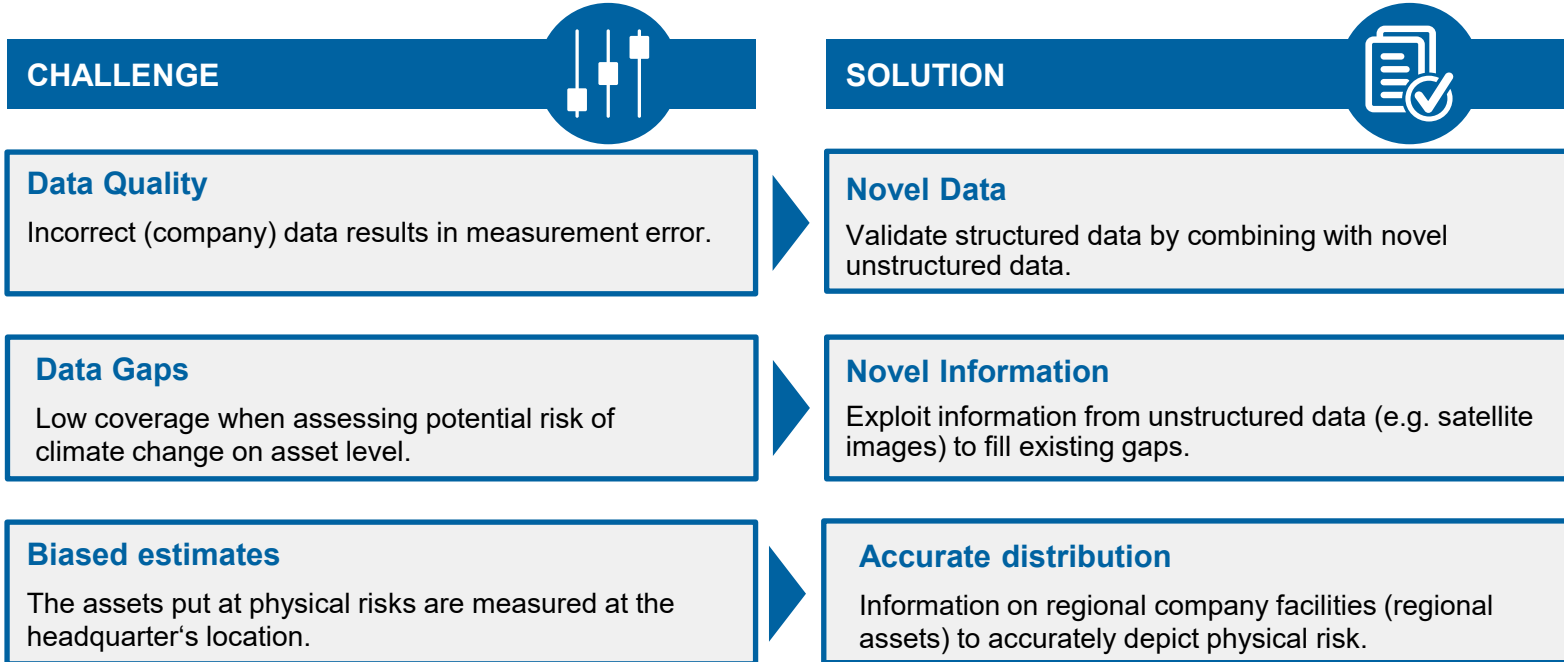
Agenda



Deutsche Bundesbank as Data Provider



Recent challenges of providing granular microdata



Project Goal – Exploit novel data sources with AI



Can we use **visual information** about the **company facilities** and **textual information** about its **activities** to automate some of the validation checks and enrich (company) master data?

Target

Validate structured company data

ID	Company	Street	City	Postal Code	Economic Sector	Employees	Parent Company
1	Company 1	Street X	City A	1234	Car manufacturing	500	X
2	Company 2	Street Y	City B	5678	Car manufacturing	400	X

Unstructured data

- High resolution overflight data (BKG)
- LOD-3D-building models (BKG)
- Websites
- Social Media
- Financial reports

Related research

- Using **satellite images** to detect surface characteristics
 - Training and testing of **urban building** detection models on **high-resolution satellite imagery** (Sirko et al., 2021)
 - Detection of **location of water surface** and prediction of changes from a data stream of **satellite images** (Nguyen et al., 2022)
- **Fact-checking** with transfer learning and generative models
 - Transfer learning from **text-based fact-checking** (Zlatkova et al., 2019)
 - Explore generating **explainable fact verification** (Atanasova et al., 2020)
- Using **multiple modalities of data**, such as text and images, to improve the results of classification models
 - **Multimodal sentiment analysis** approach using a dataset of Tumblr posts containing images and text (Hu and Flaxman., 2018)
 - **Multimodal bitransformer** models that effectively combines text and image data for classification tasks (Kiela et al., 2019)
 - Fact-checking with **multimodal large language models** that understand both text and images (Geng et al., 2024)

Geospatial Data - Satellite and building data from BKG

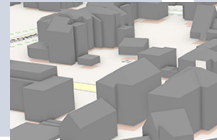
Sat4Bund

- Commercial satellite data
- 3,1 TB
- Geotiff
- Airbus / Satellite imagery
- Raster data
- [BKG - Servicestelle Fernerkundung \(bund.de\)](https://www.bkg.bund.de/Service/Fernerkundung)



3D Buildings (LoD1/LoD2)

- Level of Detail 1/2, Digital Twin
- 18 GB
- CityGML
- Building function (e.g. factory, residential building, building label)
- Height (Number of floors, volume)
- Roof types
- Geocoordinates (multipolygons)
- [3D-Gebäudemodelle LoD2 Deutschland \(bund.de\)](https://www.bkg.bund.de/Service/3D-Geb%C3%A4udemodelle)



DOP20

- Digital Orthophotos
- 6 TB
- Geotiff
- Aerial photos
- 20cm x 20cm resolution
- Raster data
- [WMS Digitale Orthophotos Bodenauflösung 20 cm \(bund.de\)](https://www.bkg.bund.de/Service/WMS-Digitale-Orthophotos)



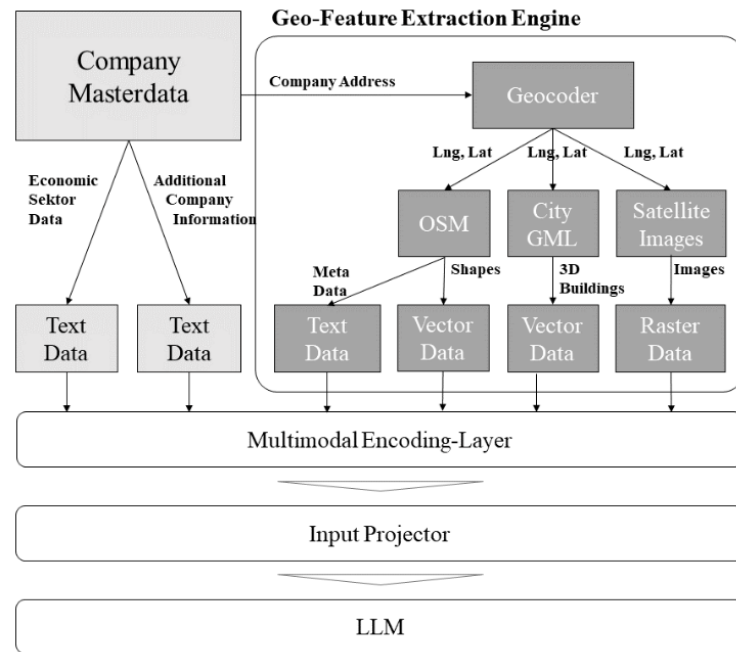
Storage

Processing

Linkage

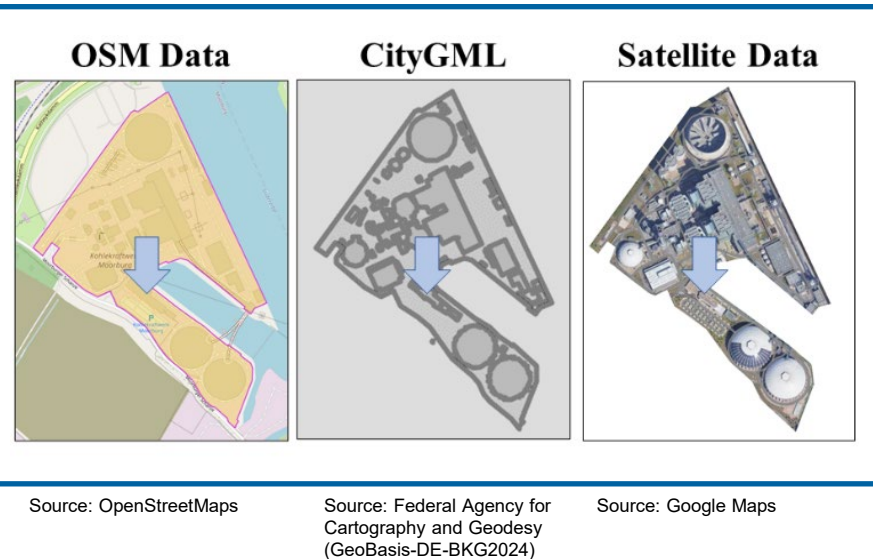
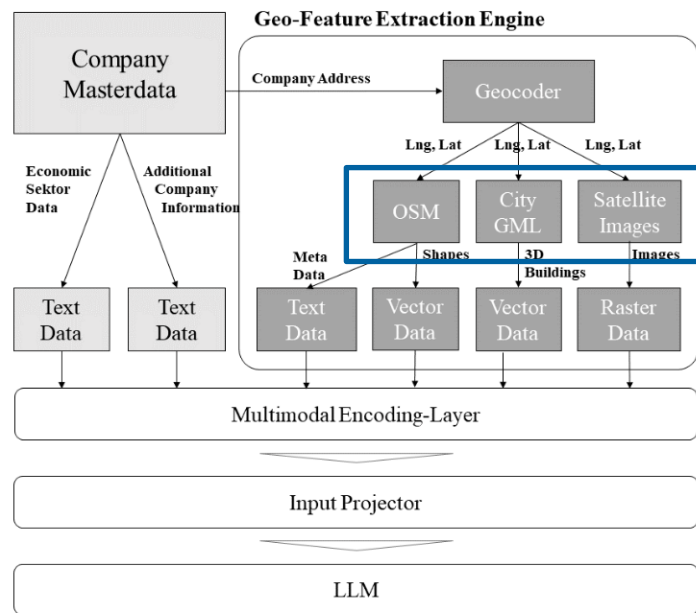
System Architecture

- We developed a potential system model architecture to encompass and link all relevant data sources.
- Preparation of the different modalities to adapt them to the specifications of the input layer using a so called Geo-Feature-Extraction Engine.

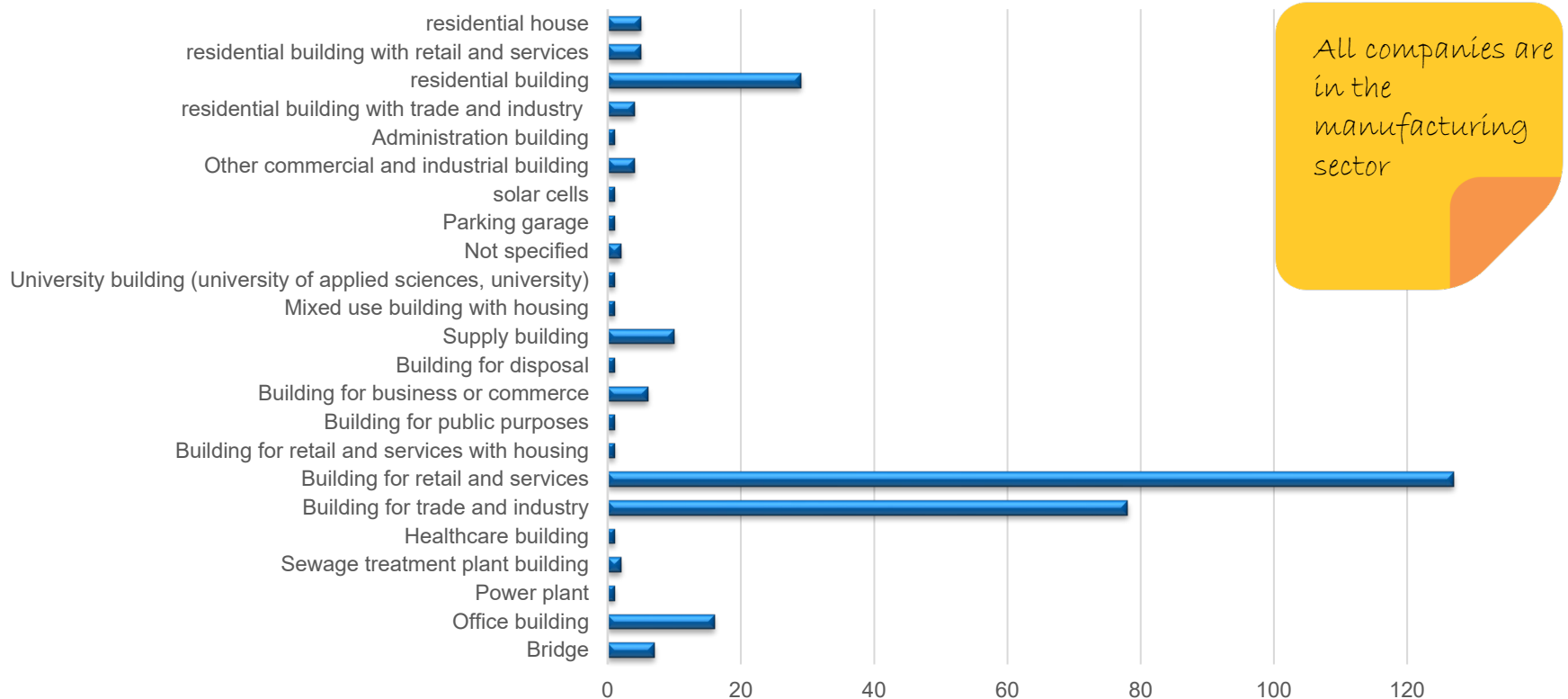


Incorporating Geospatial Data

Geo-Feature-Extraction Engine



Linkage company data and building function (LoD1 from BKG)



Merging company data and overflight images (DOP20 from BKG)



Outlook for multimodal learning

➤ Example: Proof-of-concept with sample data

Company Registration Nr	Installation Name	Contact Address	Contact City	Contact PCode	Activity Type	Account holder name
FN 87132 k	Baumit Baustoffe Bad Ischl	Linzer Straße 8	Bad Ischl	4820	Production of lime, or calcination of dolomite/magnesite	Calmit GmbH

➤ Satellite image from Google:



Google street views:



Reviews:

"Delivering raw material is simple and straight forward. Place looks busy by the way. I've been ones till now but hope get back soon."

"I came to them for loading, they greeted me pleasantly in the office and told me in detail where to go. The forklift driver loaded me very quickly. Then we go to the office again, pick up the documents and go where we need to go"

Example 1: Image analysis with LLMs



Description of the Building in the Left Image

The building in the left image appears to be part of an **industrial complex**. It consists of several structures, including what seem to be large warehouses or production facilities, silos, and outdoor storage areas. The layout suggests it is a facility designed for large-scale material handling and processing. The presence of open areas with large piles of materials and various vehicles indicates active **industrial operations, likely involving heavy machinery**.

(...)

Conclusion

On a scale from 1 (very unlikely) to 10 (completely certain), the likelihood that the activity type of the company (production of lime, or calcination of dolomite/magnesite) is correct can be assessed as follows:

Likelihood Score: 9

Reasoning:

- The **industrial setup, presence of heavy machinery, large storage areas, and the specific types of customer reviews** all strongly suggest that the facility is involved in processing and production activities consistent with lime production or the calcination of dolomite/magnesite.
- The specific activity type mentioned aligns well with the **observable infrastructure and operations at the site**.

Company	Production	Location Address	Contact Info	Customer Details	Product Type	Customer Reviews	Customer Feedback
Industriale	Production of Lime, or calcination of dolomite/magnesite	Industriale	Industriale	Industriale	Industriale	Industriale	Industriale

> Satellite image from google:

Google Street Views:

Reviews:

"Delivering our material is simple and so right forward. Place looks clean for the size. I've been down to see how they get their work done."

"I came to them for loading. They greeted me promptly in the office and had me in about 10 minutes. The facility shows loaded very quickly. They will go to the office again, and up the documents and generate me everything to go."



Please describe the building on the left image and what function it is likely to have.

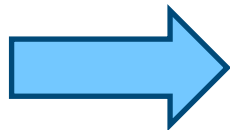
Then, looking at the table above, please decide on a scale from 1 (very unlikely) to 10 (completely certain), how likely it is that the activity type of the company is correct, given the satellite image of the site and the additional information from street view and customer reviews.

Example 2: Baseline for data validation

- Prompting through API (on premise), internal data can be processed warranting data security
- **Input:** Prompts, Satellite Image and information from structured company data
- **System prompt** = *"You are a helpful assistant that can precisely verify the location and presence of company buildings. You will see a satellite image and assess if it shows the given installation from a specific company or if the address is incorrect. Here is your task:"*
- **Prompt template** = *"Do you think the following image shows part of an installation called {installation_name} from the company {company_name}? The activity listed for this company is {activity_type}. ALWAYS start your answer with one of these two options: "Yes." or "No.". After that, ALWAYS provide your reasoning step by step."*
- **Example response** *"No. The image does not show part of an installation called RUHR OEL GmbH - Werk Scholven - CO2-Glocke from the company Ruhr Oel GmbH. Reasoning step by step: 1. The activity listed for Ruhr Oel GmbH is the refining of mineral oil, which typically involves large industrial facilities with extensive infrastructure such as storage tanks, pipelines, and processing units. 2. The satellite image provided shows a location with a large parking area, several buildings, and a significant amount of green space, which is more characteristic of an office complex, residential area, or a park rather than an industrial refinery. 3. There are no visible signs of industrial activity, such as large storage tanks, chimneys, or heavy machinery, which are commonly associated with oil refining facilities. 4. The layout and structures in the image do not match the typical appearance of a refinery installation. Therefore, the image does not correspond to the described installation of RUHR OEL GmbH"*
- Calculate a confidence score from the logprobs (for yes or no)

Example 3: Multimodality to predict economic activity

P: The picture shows a company's plant. Based on the 2008 Classification of Economic Activities (WZ), under which WZ 2008 code would you classify the company depicted?



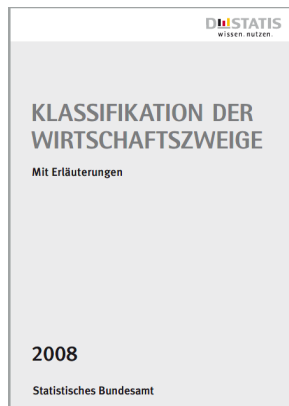
R: The plant shown in the picture could indicate a waste incineration plant, as it includes typical elements such as waste bunkers and chimneys that are characteristic of such plants. According to the Classification of Economic Activities 2008 (WZ 2008), such plants fall under:

- Section E: ****Water supply; sewerage, waste management and remediation activities**
- ***** Subgroup 38.21.0: **Treatment and disposal of non-hazardous waste****, which includes energy recovery by incineration.

If it is an energy production facility instead, it could be classified under:

- Section D: ****Energy supply**
- ***** Group 35.11: **Electricity production****, in particular if it produces electricity by burning fossil or renewable fuels.

The exact classification depends on whether the main activity is the disposal of waste or the generation of energy.



Results

- Current **multimodal LLMs** perform really **well** on the task of identifying the **presence of specific building types** from high resolution **satellite images**.
- Current LLMs are able to identify **links between company names** even if there is some modification/noise such as added prefixes, suffixes or abbreviations.
- **Multimodality** is an inherent part of the company data validation task, as images or structured text alone don't provide enough information for a complete analysis.
- Creating a solid **ground truth** from multiple data sources is crucial and not trivial
- The comparison between smaller open and larger closed-source models shows **big differences in performance** on the company linking task. While smaller models are able to find more matches, larger models focus on correct matches - finding fewer links while being less prone to hallucinations and false positives.
- Added **reasoning** to LLM predictions provides some form of interpretability, which can in turn be used for validation. The trade-off between the higher number of output tokens and the performance improvement from the added reasoning is a valid point of discussion.

Next steps and challenges ahead

Evaluate the results of LLMs

- Process CityGML completely (including building parts)
- Analyse the Logprobs for the model evaluation
- Compare LLMs results to traditional approaches
- Run classification task systematically on the full set of companies and compare results to the validated classifications
- Test parameters (temperature e.g. for data linkage) to find optimal values



Thank you!

| Backup

- Atanasova, P., Simonsen, J. G., Lioma, C., & Augenstein, I. (2020). Generating Fact Checking Explanations. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 7352–7364. Available at: <https://arxiv.org/abs/2004.05773>
- Gebru, T., Krause, J., Wang, Y., Chen, D., Deng, J., Aiden, E. L., & Fei-Fei, L. (2017). Using Deep Learning and Google Street View to Estimate the Demographic Makeup of the US. Available at arXiv: <https://arxiv.org/abs/1702.06683v2>
- Geng, J., Kementchedjhieva, Y., Nakov, P., & Gurevych, I. (2024). Multimodal Large Language Models to Support Real-World Fact-Checking. Available at arXiv: <https://arxiv.org/abs/2403.03627>
- Hu, A., & Flaxman, S. (2018). Multimodal Sentiment Analysis To Explore the Structure of Emotions. *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 350–358. Available at arXiv: <https://arxiv.org/abs/1805.10205>
- Kiela, D., Bhooshan, S., Firooz, H., Perez, E., & Testuggine, D. (2019). Supervised Multimodal Bitransformers for Classifying Images and Text. arXiv:1909.02950. Available at arXiv: <https://arxiv.org/abs/1909.02950>

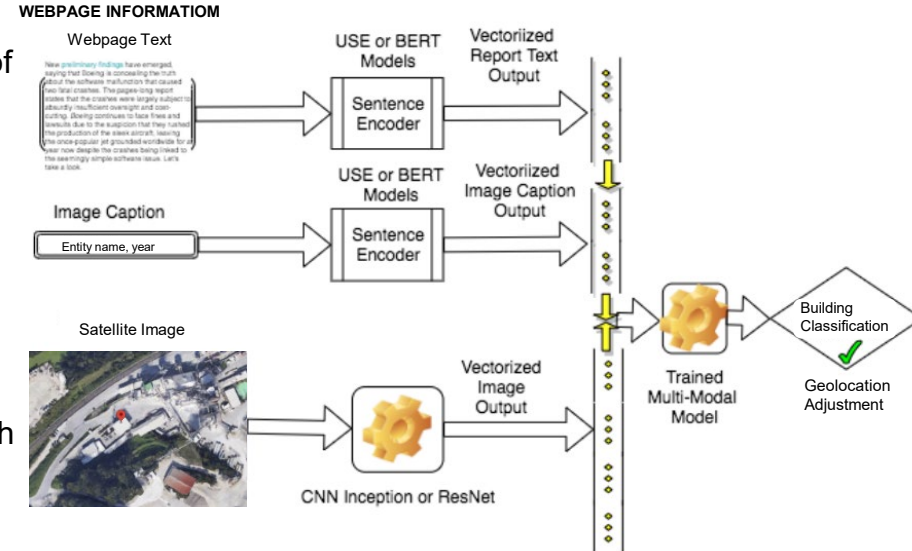
- Miller, S. J., Howard, J., Adams, P., Schwan, M., & Slater, R. (2020). Multi-Modal Classification Using Images and Text. *SMU Data Science Review*, 3(3), 6. Available at: <https://scholar.smu.edu/datasciencereview/vol3/iss3/6/>
- Nguyen, T., Nguyen Thanh, T., Phan, T., & Hung, N. (2022). Mapping and Monitoring Water Areas with Satellite Images and Deep Learning. 10.13140/RG.2.2.22869.70886. Available at: [\(PDF\)](#)
[Mapping and Monitoring Water Areas with Satellite Images and Deep Learning \(researchgate.net\)](#)
- Pradeep, R., Ma, X., Nogueira, R., & Lin, J. (2020). Scientific Claim Verification with VERT5ERINI. *Proceedings of the 12th International Workshop on Health Text Mining and Information Analysis*, 94–103. Available at: <https://arxiv.org/abs/2010.11930>
- Sirko, W., Kashubin, S., Ritter, M., Annkah, A., Salah Eddine Bouchareb, Y., Dauphin, Y., . . . Quinn, J. (2021). Continental-Scale Building Detection from High Resolution Satellite Imagery. Available at arXiv: <https://arxiv.org/abs/2107.12283>
- Zlatkova, D., Nakov, P., & Koychev, I. (2019). Fact-Checking Meets Fauxtography: Verifying Claims About Images. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 2099–2108. Available at: <https://arxiv.org/abs/1908.11722>

Further methods that will be explored for the use-case:

- Compare existing facilities with other economic core indicators (e.g. legal requirements about the minimum legal working space to approximate the size of the facility given a number of employees)
- Transfer learning from text-based fact-checking (Zlatkova et al., 2019)
- Large pre-trained generative transformers like T5 (Pradeep et al., 2020)
- Explore generating explainable fact verification (Atanasova et al., 2020)

Solution: Multimodal Fact-Checking and GeoAI

- **Multimodal:** Processing of different data types such as images and text or audio (language models such as ChatGPT mainly trained with text data)
- **Fact-Checking:** Verification of information and identification of misinformation
- **GeoAI:** Use of artificial intelligence to analyze geodata
- Combine **natural language processing** with **georeferenced image data** (visual inspection and visual segmentation)
- Advantages (Miller et al., 2020):
 - Improve classification accuracy
 - Robust against missing values
 - Reduced training time and resource consumption through parallel processing
 - Cooperation with the Ubiquitous Knowledge Processing Lab der TU Darmstadt



Source: (Miller et al., 2020), adjusted