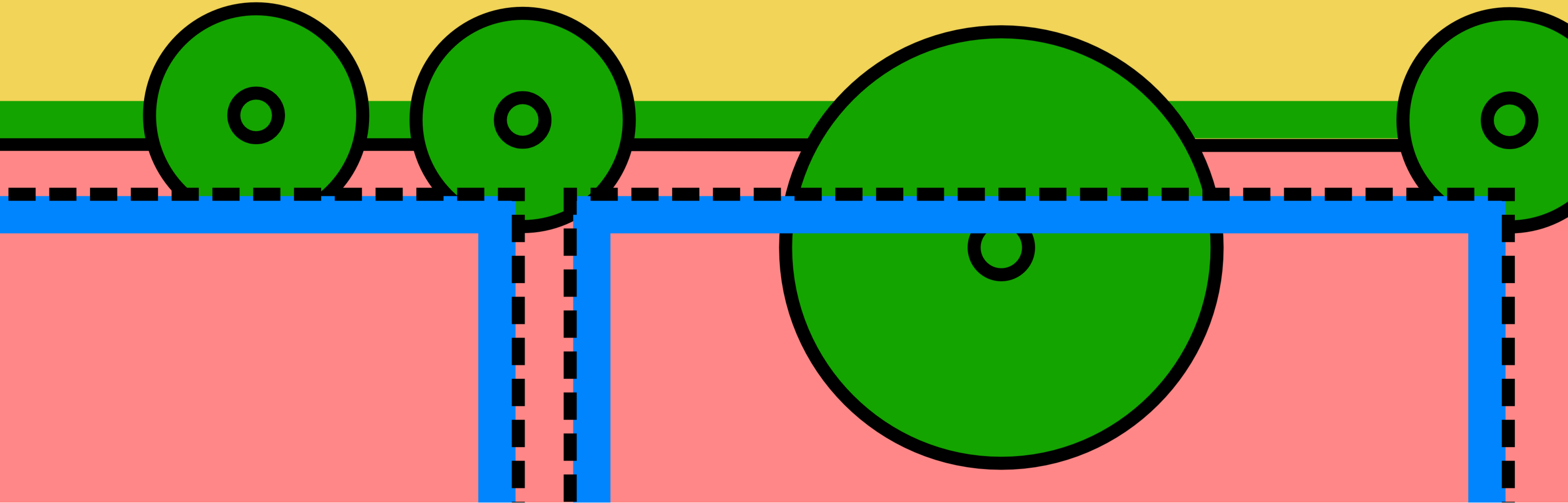
Digital Urban Land-Use Planning –

How AI Systems Understand Development Plans

October 30th / Michael Schwarz



Moin! 👋

Michael Schwarz **(He/Him)**

Studied **Computer Science M.Sc.** with specialization in **Data Science** at the TUAS Augsburg

Digitale Visitenkarte: https://zeeg.me/schwamic

Master Thesis

The aim of my master's thesis was to **analyse the extent to which *Multimodal Large Language Models* are able to understand *Development Plans***. The investigations for the design and evaluation of an ai system explicitly relate to the ***text comprehension of dev-plans***.

🎉 *15.05.2024 – 06.09.2024*

Process Steps

✓

Domain Knowledge

Multimodal Large Language Models

Experiments & Evaluation

Would you like a short recap?

URBAN LAND-USE PLANNING / DEVELOPMENT PLAN

Development Plan

Legally binding!

The development plan defines the **details of land use** and the **authorisation of urban development activities**, i.e. it regulates whether and how a property may be built on.

Art der baulichen Nutzung	Zahl der Vollgeschosse
Grundflächenzahl (GRZ)	Geschossflächenzahl (GFZ)
Bauweise	Dachform/Dachneigung

Figure 1

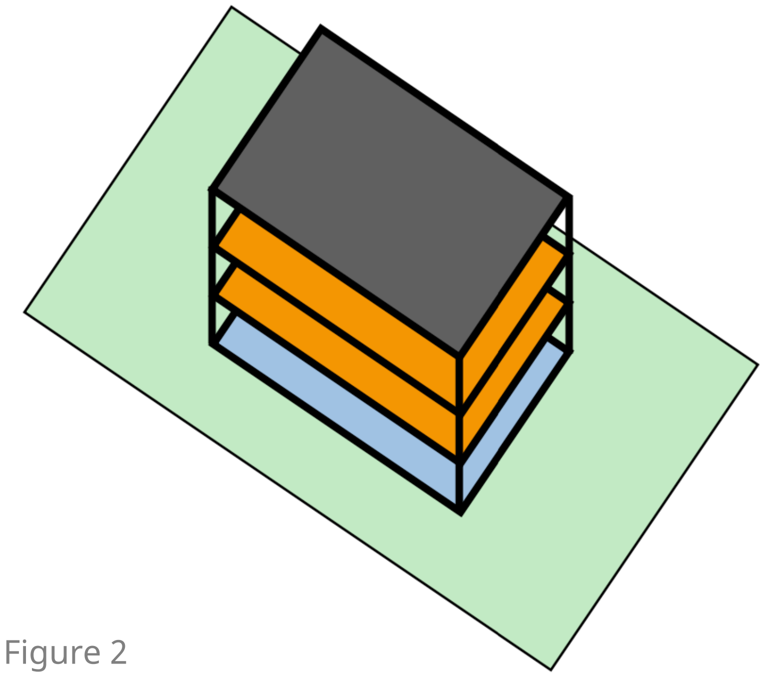


Figure 2

GRZ indicates how many qm of (building) floor area (GR) can be built on per square metre of plot area.

GFZ indicates how many square metres of floor area (GF) are permitted per square metre of plot area.

$$\textbf{GRZ} = \frac{\text{Grundfläche}}{\text{Grundstücksfläche}}$$

$$\textbf{GFZ} = \frac{\text{Geschossfläche}}{\text{Grundstücksfläche}}$$

Beispiel:

Gebäude mit drei Vollgeschossen*

Grundstücksfläche = 1.200 m²

Grundfläche = 300 m²

Geschossfläche = 900 m²

GRZ = 0,25 **GFZ = 0,75**

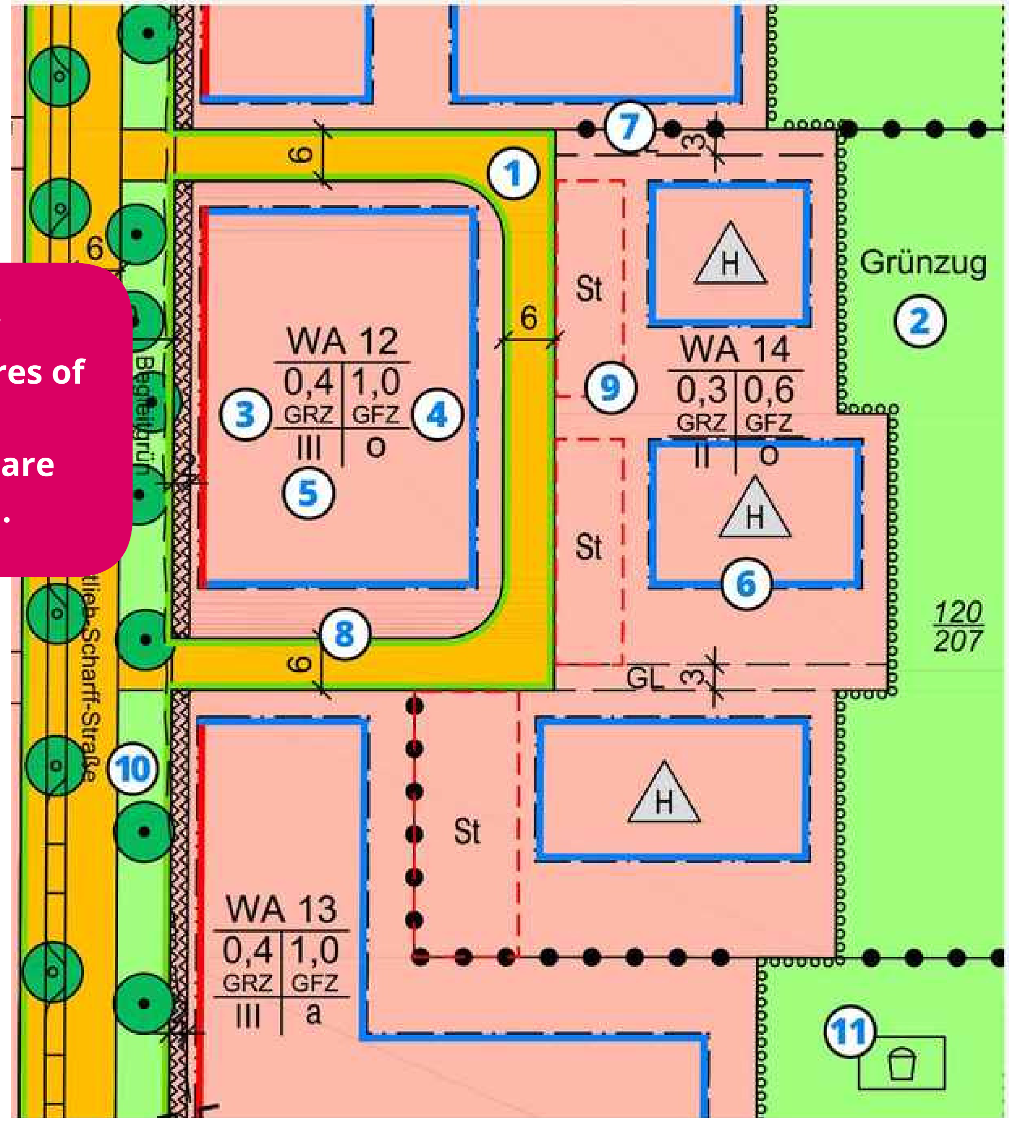


Figure 3

Documents of a Development Plan

	Plan Drawing	Explanation of Symbols	Text Part	Geo Referenzing
Data Types	**PDF, Image**	**PDF, Image**	**PDF, Image**	**TIF, TFW**
	DWG, DXF, XPlanGML	**Docx, XPlanGML**	**Docx, XPlanGML (partly)**	

Documents of a Development Plan (2024)

- high resolution
- high information density (need to zoom in)
- no coordinates/geo-ref
(...)

Stadt Augsburg

Aufstellung

Bebauungsplan Nr. 872 A
„Zwischen Waldstraße und Döllgaststraße“

mit integriertem Grünordnungsplan

Text_Part.pdf

Stadt Augsburg

Aufstellung

Bebauungsplan Nr. 872 A
„Zwischen Waldstraße und Döllgaststraße“

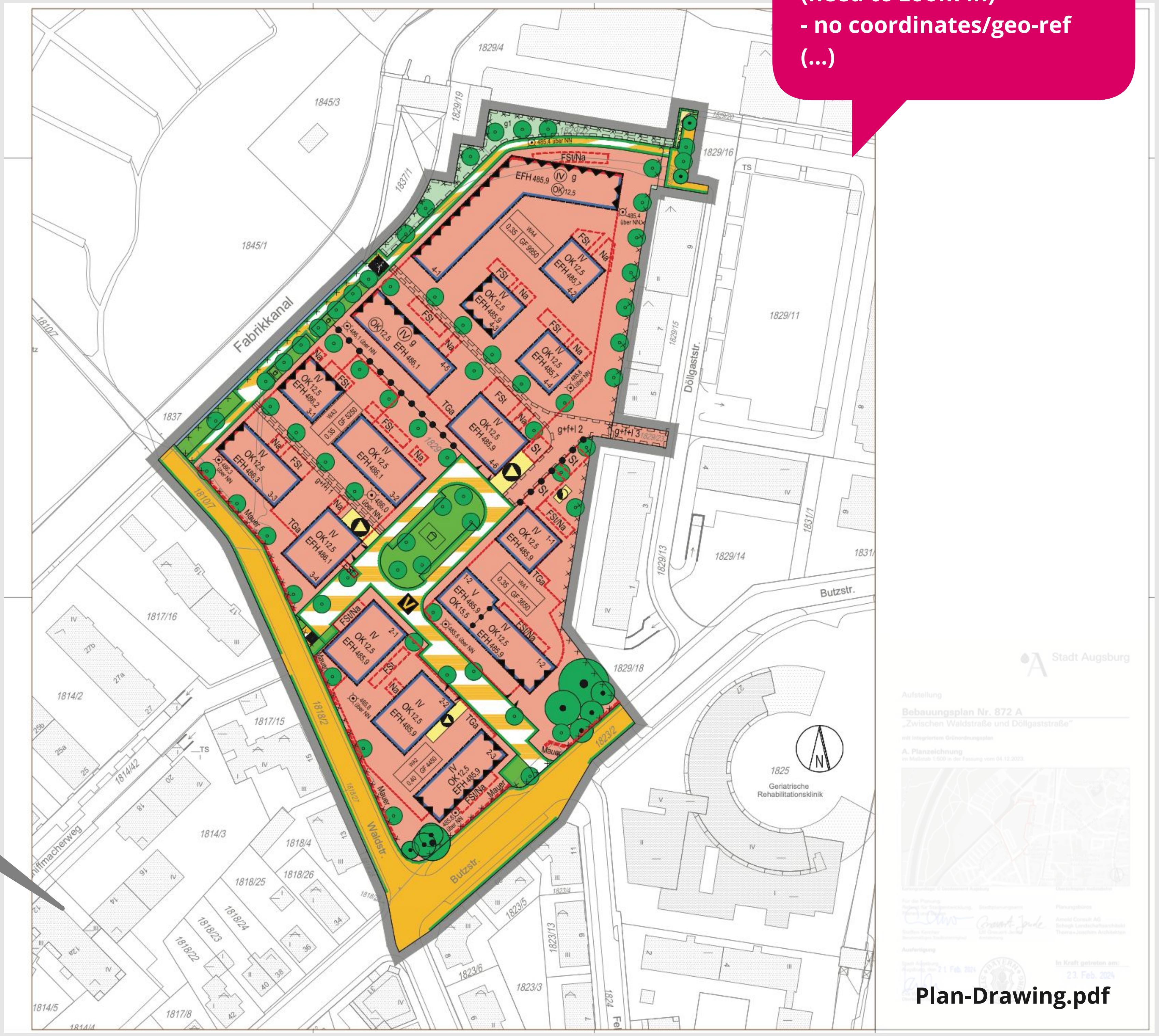
mit integriertem Grünordnungsplan

Explanation_Symbols.pdf

Geo-Referencing.tfw

Development-Plan.gml

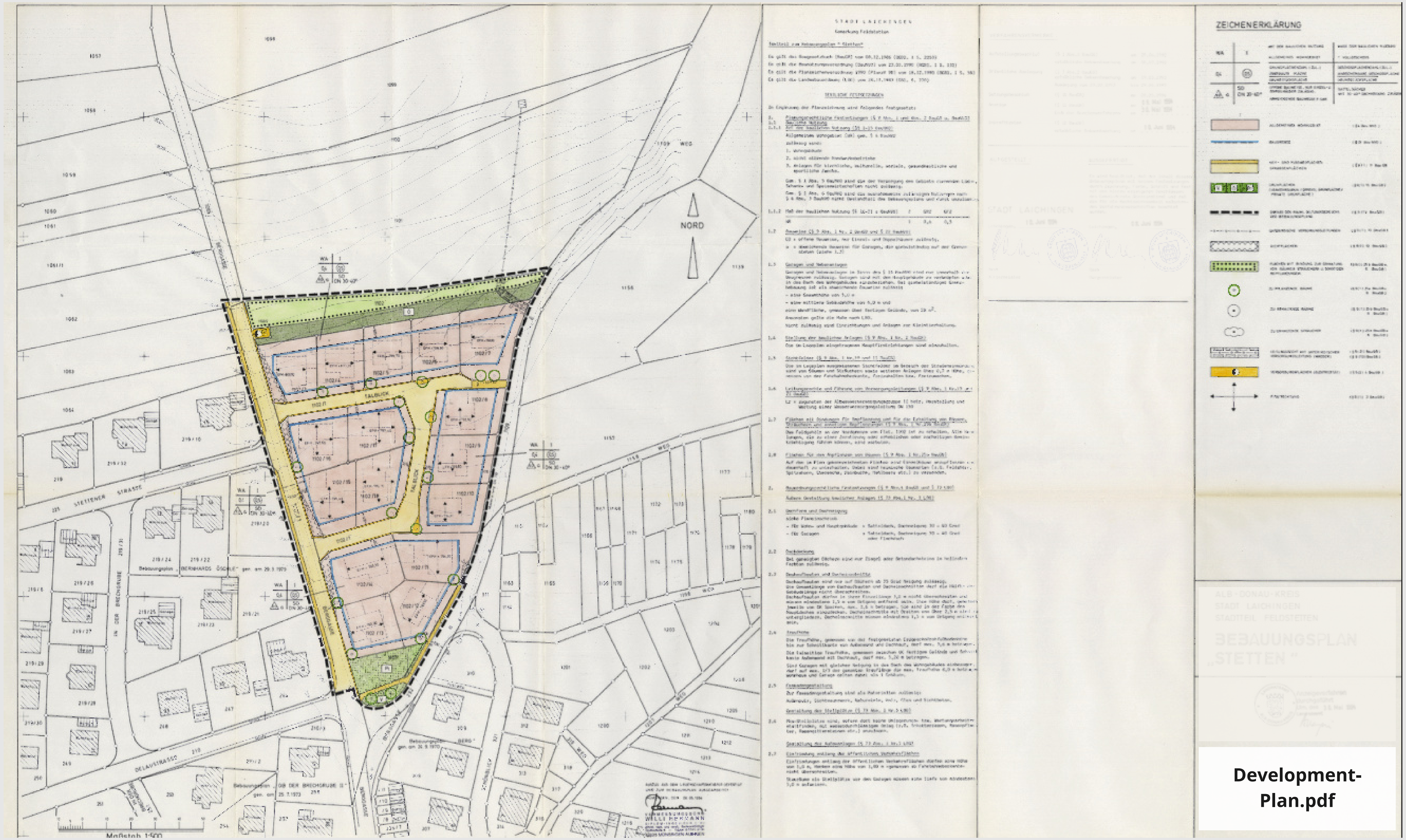
Geo-Referencing.tif



Plan-Drawing.pdf

Figure 4

Documents of a Development Plan (1994)



Plan-Drawing.DWG/.DXF

?

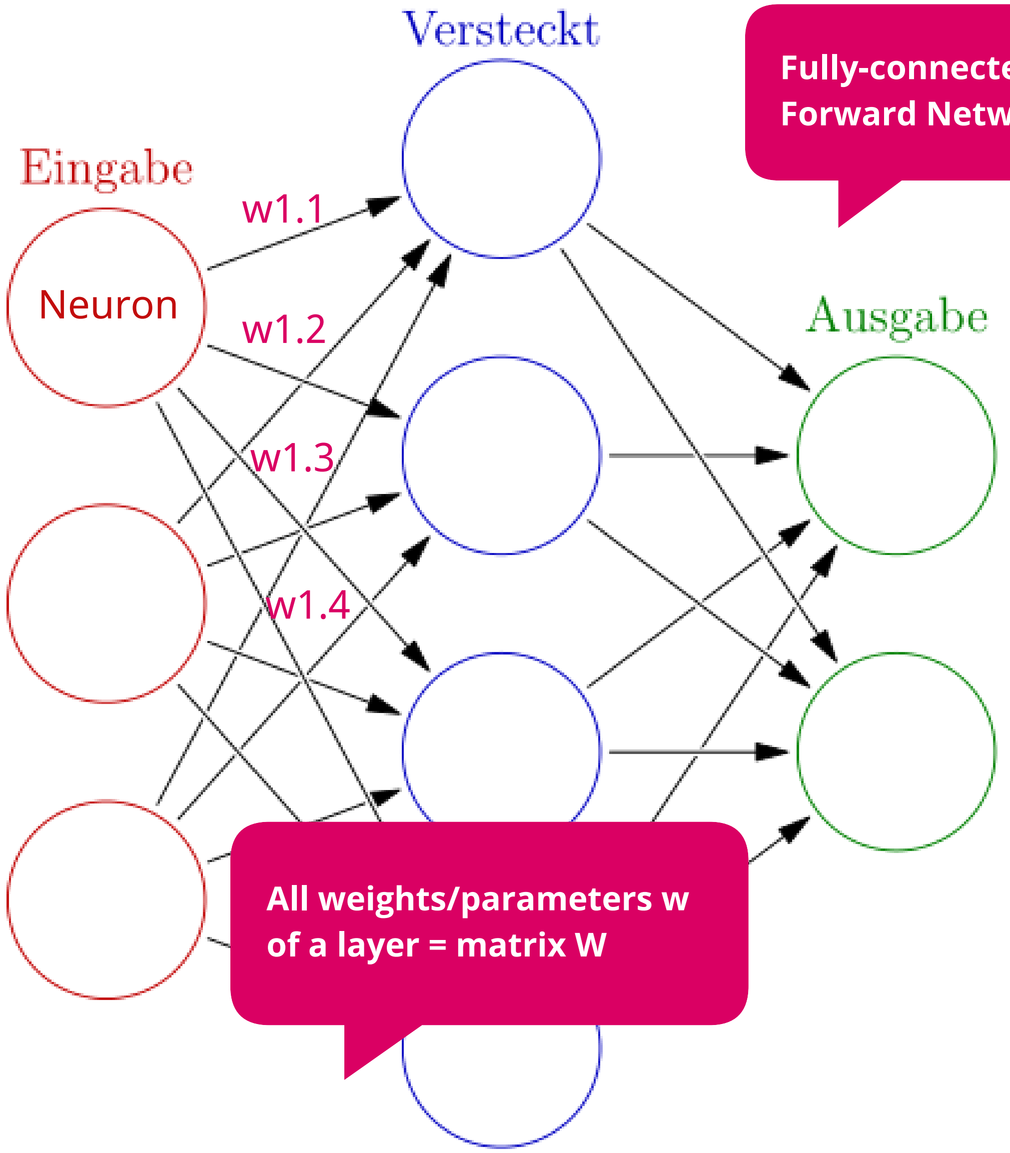
Geo-Referencing.TIF/.TFW

Development-Plan.pdf

MULTIMODAL LARGE LANGUAGE MODELS

Neural Networks

Fully-connected Feed Forward Network



All weights/parameters w of a layer = matrix W

Figure 5

TRANSFORMER-Architecture

"Attention Is All You Need" (Vaswani et.al. 2017)

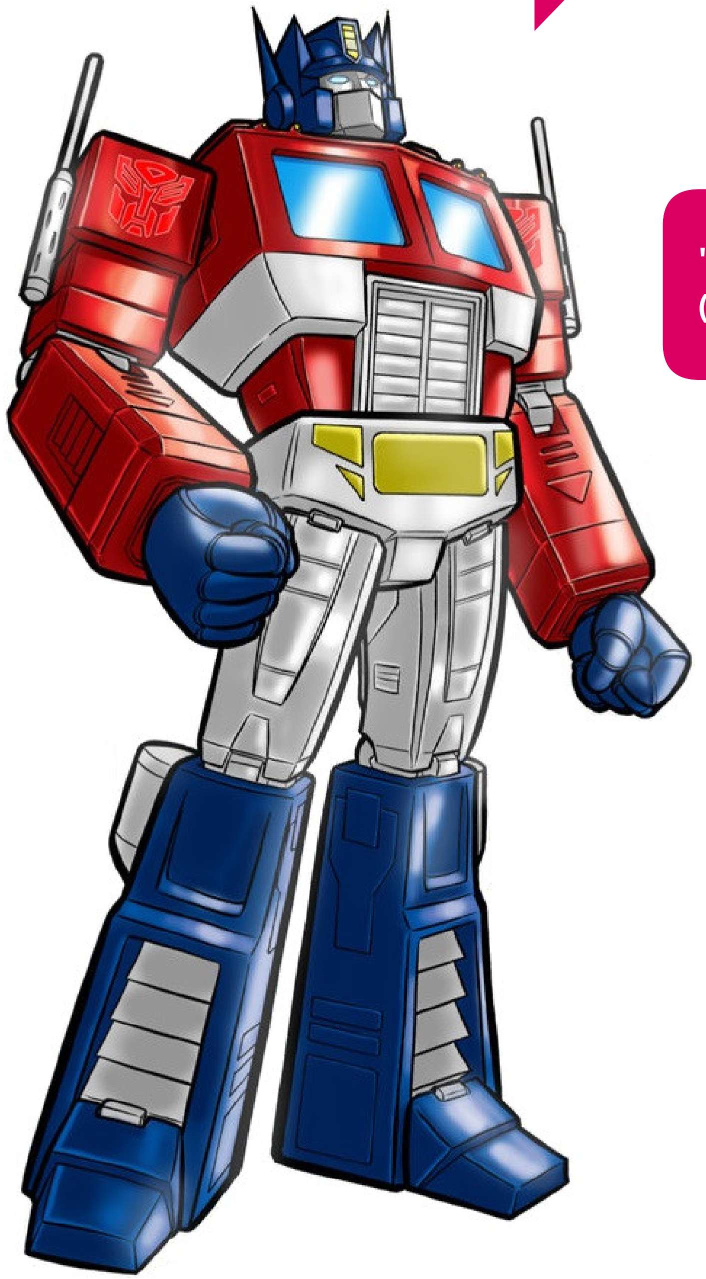


Figure 6

Machine Processing Of Language – Vector Representations

King + Woman - Man ≈ Queen (Analogies) 😈

"Efficient Estimation of Word Representations in Vector Space"
(Mikolov et.al. 2013)

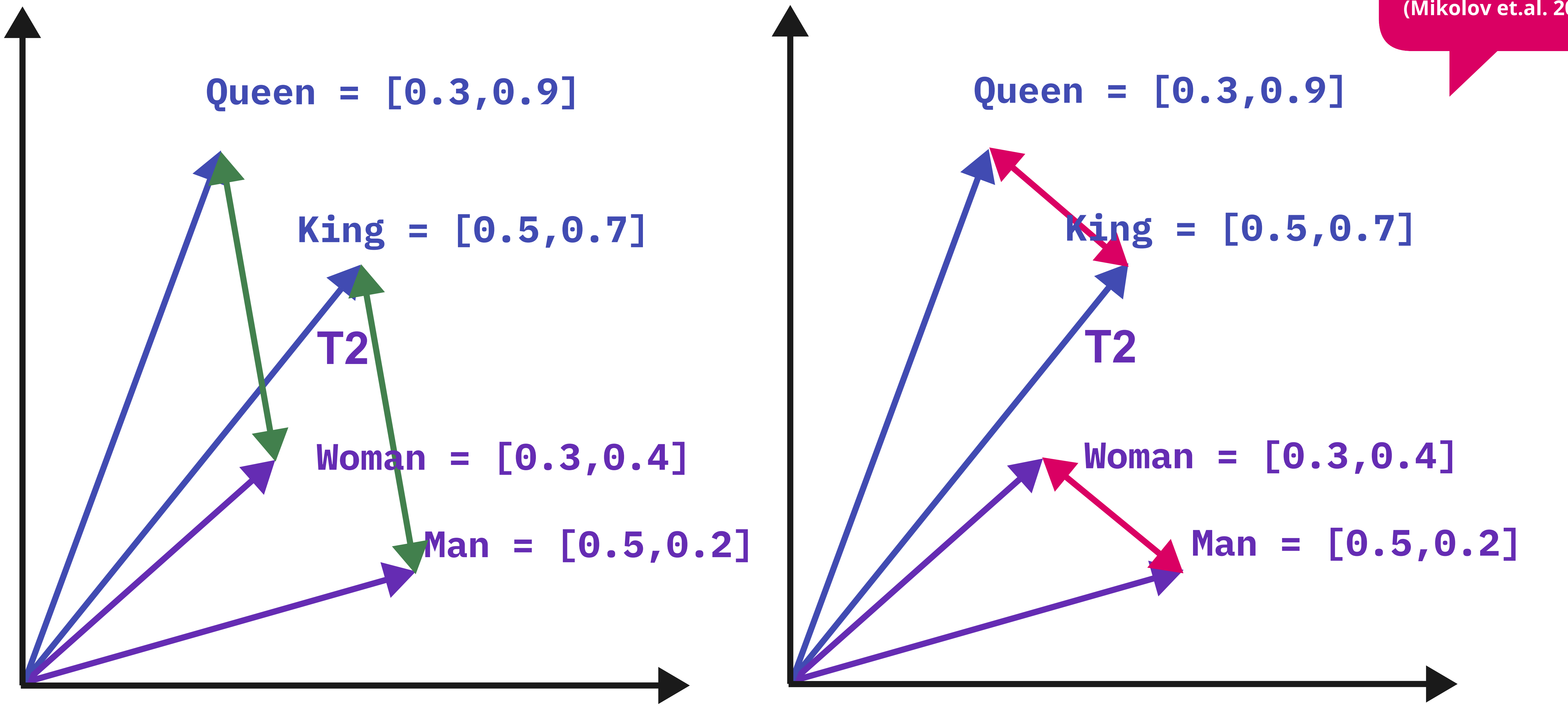
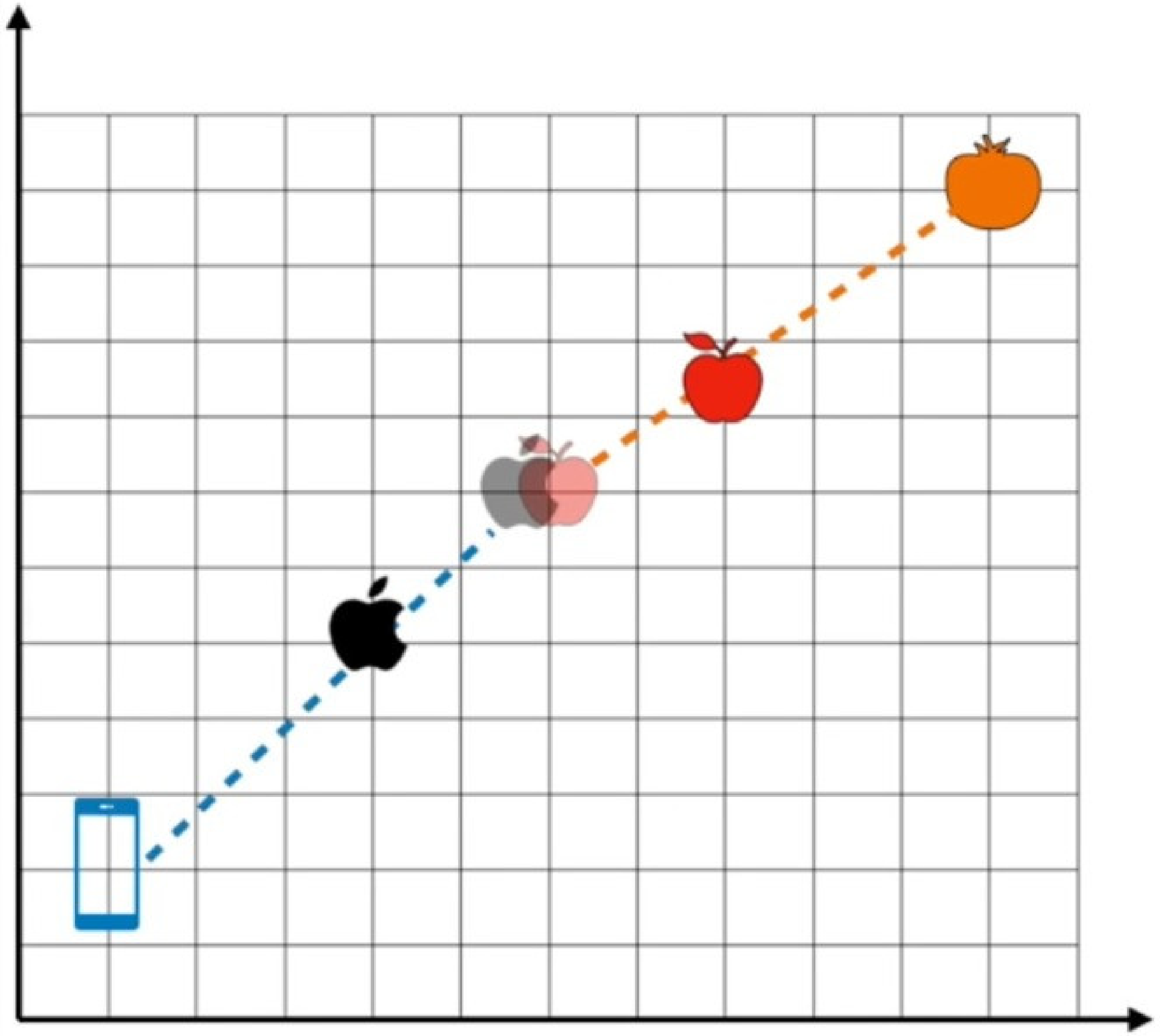


Figure 7

Machine Processing Of Language – Attention



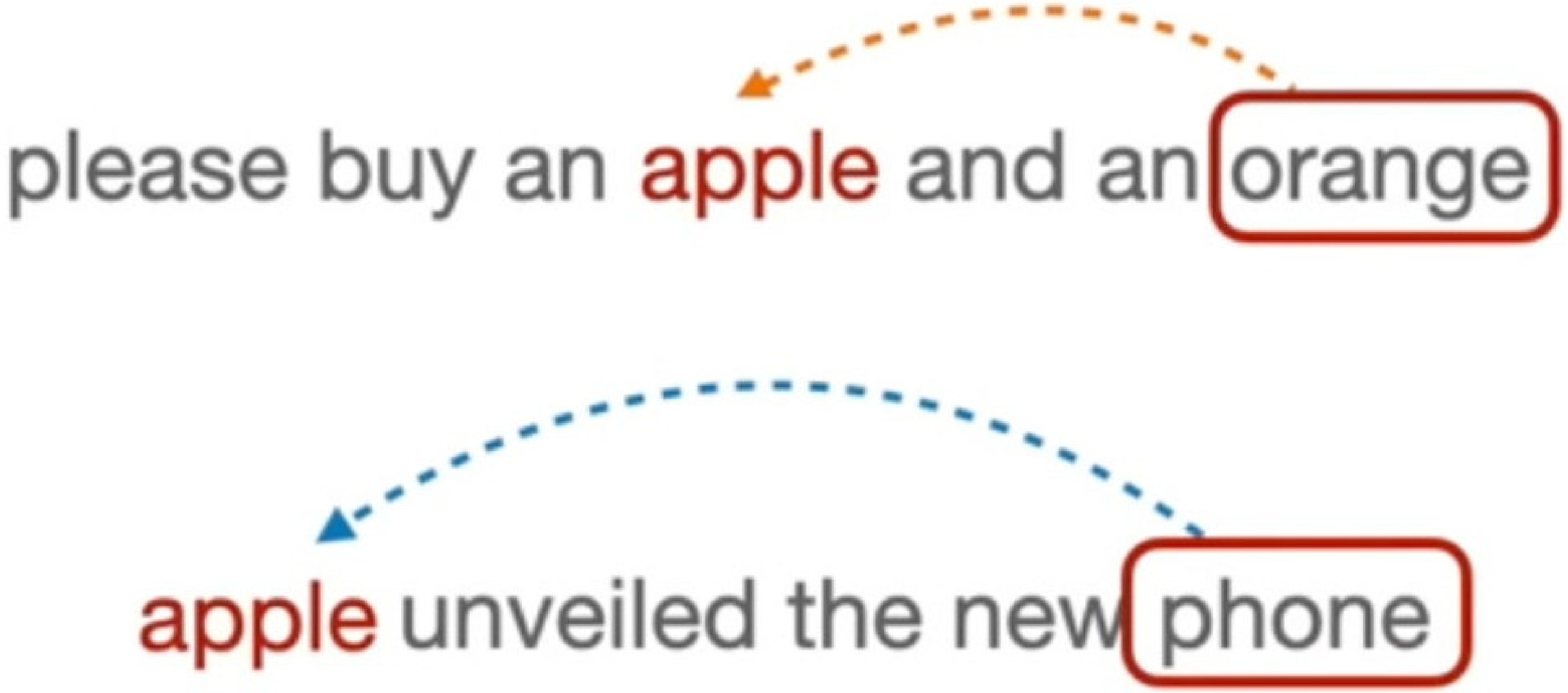
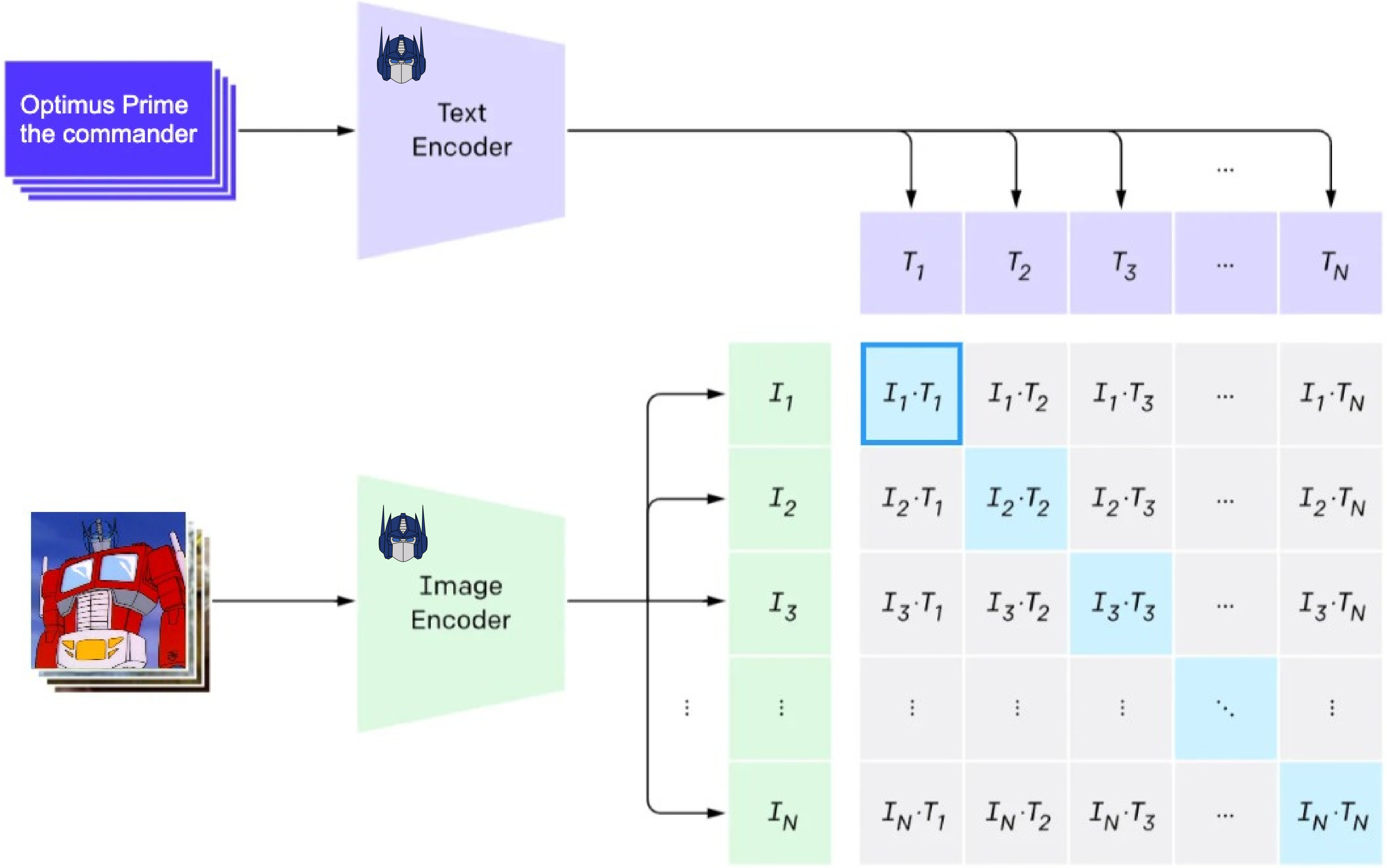


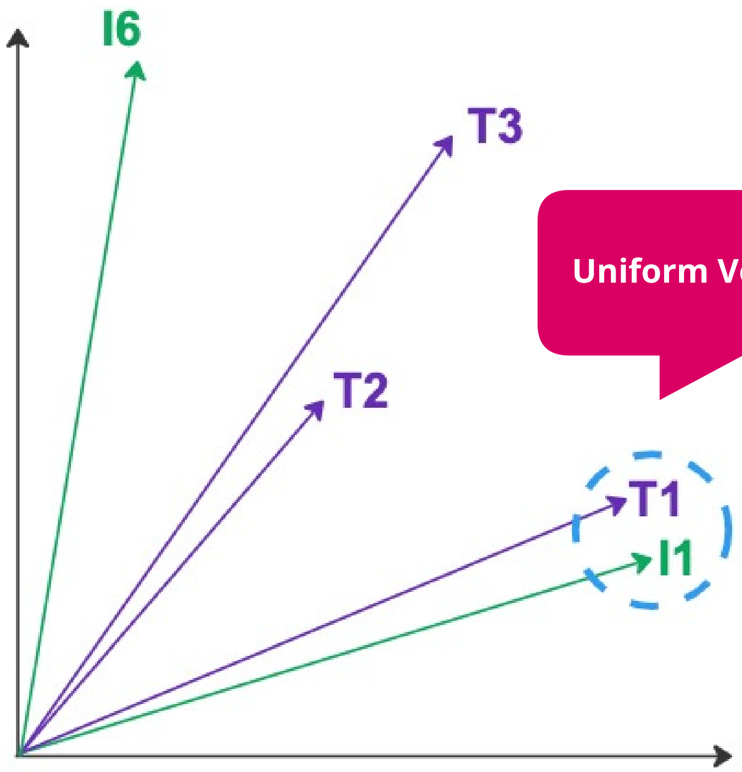
Figure 8

Machine Processing Of Language + Vision via CLIP

"Learning Transferable Visual Models From Natural Language Supervision" (Radford et.al. 2021)

OpenAI





Uniform Vector Language

Figure 9

Large Language and Vision Assistant – LLaVA 1.0

"Visual Instruction Tuning"
(Liu et.al. 2023)

Generative Language Models (Decoder)
Vicuna Claude GPT-4
Mistral PaLM ...

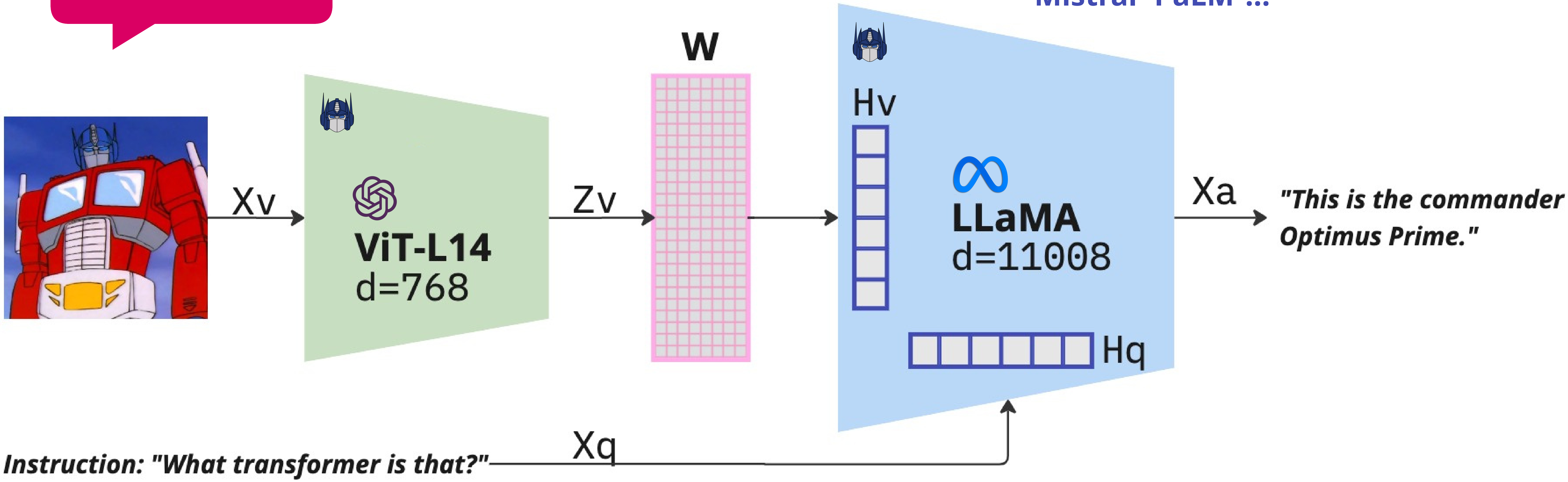
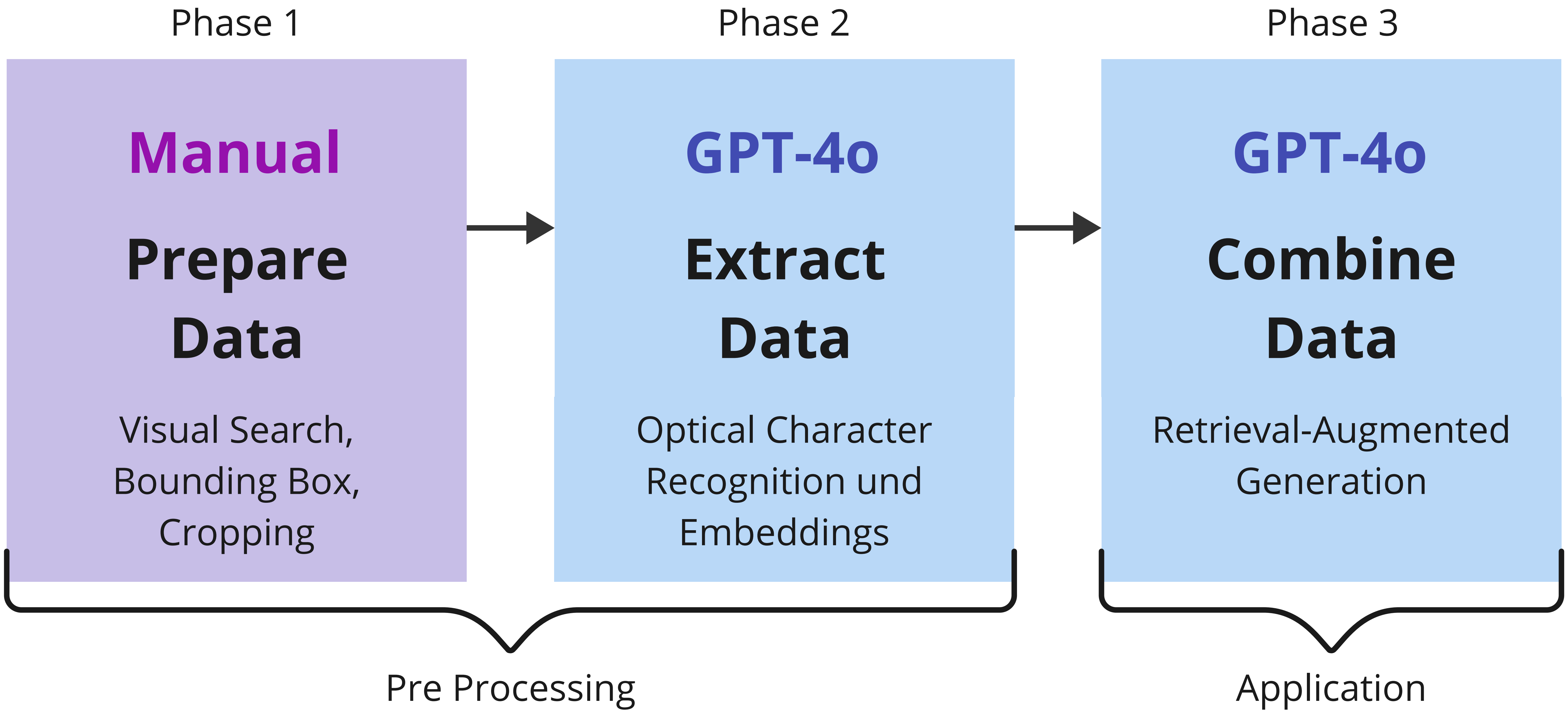


Figure 10

EXPERIMENTS & EVALUATION

Architecture – Prompt-Chain



Planning Permission Process – Final Result of GPT-4o [Phase 3]

Festsetzungen	Bauantrag	GPT-4o	widerspricht / entspricht	Korrekt	Konsistent
Art der baulichen Nutzungs	• Wohnhaus	• Erlaubt ist ein Gebäude im **Allgemeinen Wohngebiet (WA)** (§ 4 BauNVO). **(Kapitel 1.1.1)**	✅	3/3	3/3
Maß der baulichen Nutzungs	• Lageplan (zeichnerischer und schriftlicher Teil) GRZ 0,17 / GFZ 0,35 • Zahl der Vollgeschosse II	• Die festgesetzte Grundflächenzahl **(GRZ) ist 0,3** und die Geschossflächenzahl **(GFZ) ist 0,5**. **(Kapitel 1.1.1)** • Es ist **ein Vollgeschoss** erlaubt. **(Kapitel 1.1.3)**	✅ ❌	3/3	3/3
Bauweise	• Lageplan (zeichnerischer Teil) mit seitlichen Abstandsflächen	• Die festgesetzte Bauweise ist die **offene Bauweise** (0). **(Kapitel 1.2)**	**Das ist mir nicht bekannt.**	3/3	3/3

Prompt Setup – Jupyter Notebooks

```
from utils.openai import OpenAI
from utils.runner import Runner

# Initialisierung
instructions = "Du bist ein Assistent zur getreuen Wiedergabe von Informationen aus einem Bebauungsplan. Achte auf Vollständigkeit."
chatGPT = OpenAI(instructions)
runner = Runner()
```

```
# Datenquelle
img_path = "../data/processed/bpläne/1_alles_in_einem_dokument/L04-ZE-TT-crop.png"
pdf_path = "../data/raw/bpläne/2_zeichnung_textteil_getrennt/F11-PZ-ZE.pdf"

# Aufgabe durchführen
instruction = 'In welchem Jahr ist der vorliegende Bebauungsplan in Kraft getreten? Ausgabe im JSON-Format: {"Gültig seit": <YYYY>}'
async def run():
    msg1 = await chatGPT.extractTextFromImage(instruction, img_path)
    msg2 = await chatGPT.extractTextFromImage(instruction, pdf_path, "pdf")
    return [msg1, msg2]
results = await runner.async_consistency_check(run)

# Ergebnisse speichern
msg1_l04_year, msg1_f11_year = results[0]
%store msg1_l04_year msg1_f11_year
```

„To what extent are MLLMs able to understand dev-plans?“ *[RQ1]*

„How good is the quality of the extracted data?“ *[SQ1.1] – Phase2*

- *SL-1A*: Plan drawing
  SL-1B: Symbol declaration document
- *SL-1C*: Text part document

„To what extent can an understanding of the extracted data be created?“ *[SQ1.2] – Phase3*

- *SL-2A*: Extract information from collected *SL-1* data
- *SL-2B*: QA capabilities with regard to the known data from *SL-1*

How good is the quality of the extracted data?

[RQ1][SQ1.1] – Phase2

- GPT-4o was able to extract text information in *adequate quality BUT depends very much on the pre-processing and visual structure* of the dev-plan documents.
- Accordingly, it can be assumed that GPT-4o extracts content mostly *based on visual characteristics*, so that *domain specific connections can be overlooked or misinterpreted.*
- Limitations also became apparent, such as the *lack of geo referencing*.

➡️ In order *to achieve reliable and high quality, the MLLM would have to be trained for specific tasks and learn domain knowledge*.

To what extent can an understanding of the extracted data be created? *[RQ1][SQ1.2] – Phase3*

- MLLMs such as GPT-4o have a *sufficient understanding of language to correctly link domain-specific content* with each other and to deal with user instructions.
- However, the understanding of MLLMs is *strongly dependent on the expertise provided*,which was also illustrated in the step "Bauweise".
- Likewise, *GPT-4o did not question the context information provided*, which indicates that *GPT-4o has little domain specific understanding*.

➡️ *For a robust understanding, the MLLM would need to learn basic domain knowledge and retrival evaluation skills*.

„Can MLLMs understand the contents of machine-readable development plans at least as well as humans can?“ [RQ2]

➡️ *GPT-4o can not understand dev-plans as well as humans can*. However, the results also show that a *human and MLLM as a team, may be able to understand dev-plans better* than a human could without the support of AI.

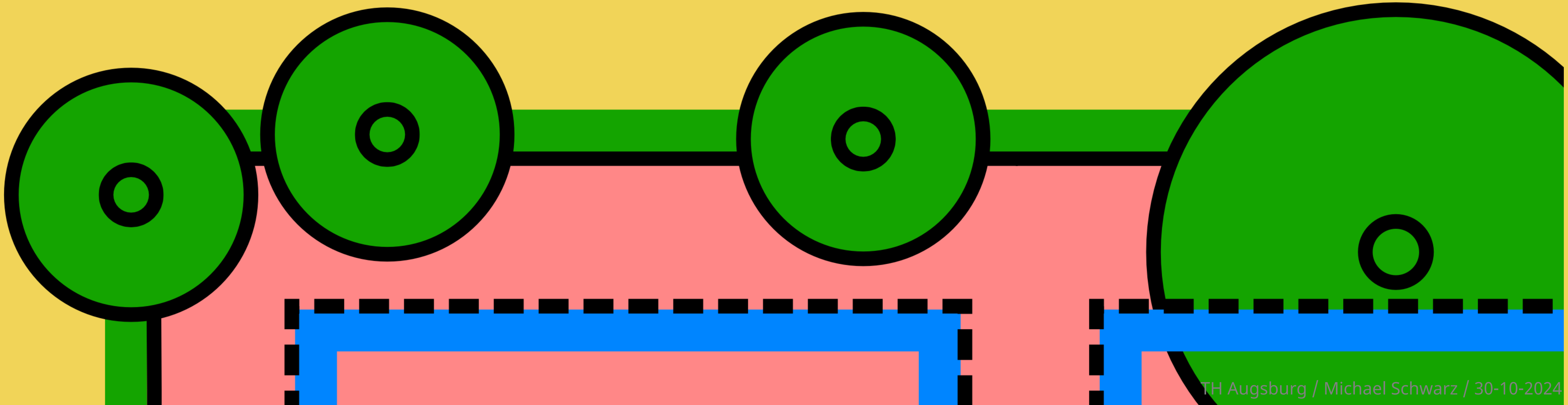
- *GPT-4o did not have to figure out which test steps are relevant for the permission process of the construction project*.
- GPT-4o is *currently not able to carry out geo-referencing or a precise comparison between a dev-plan and a construction plan*, which severely *limits the number of feasible test steps* within an permission process.

- *The strengths of GPT-4o are its ability to recognise associations and relationships* between text content.
- This allows information to be *extracted and searched efficiently*. Accordingly, GPT-4o was able to *achieve very good results in the minimal RAG application* on the basis of the contextual knowledge provided.

Further Research Opportunities

- Possible applications of ***MLLMs*** for ***RAG scenarios*** in the field of urban land use planning.

- Pre-train ***CNN*** models to extract ***graphical unstructured information*** from dev-plans reliably.



MANY THANKS! ✨

for the great projects we did together during my Master's programme!

- *Study Project:* **Mobile App for Managing Building and Room Data for AI-Supported Analyses**

- *Master Thesis:* **Digital Urban Land-Use Planning – How AI Systems Understand Development Plans**

Links & Sources

- Master Thesis (GitHub)
- Huggingface (Transformer) / LangChain / LLamaIndex

Figure 1 und 2: THA Prof. Fina, Skript Umfeldplanung 2023
Figure 3: https://www.musterhaus.net/ratgeber/bebauungsplan-lesen-und-verstehen
Figure 4: Bebauungsplan der Stadt Augsburg
Figure 5: https://wikimedia.org/wikipedia/neural_network_de.svg.png
Figure 6: http://img3.wikia.nocookie.net/__cb20140629204205
Figure 7: https://www.researchgate.net/figure/The-classical-king-woman-man-queen-example
Figure 8: https://www.youtube.com/watch?v=UPtG_38Oq8o
Figure 9: Darstellung angelehnt an https://arxiv.org/abs/2103.00020
Figure 10: Darstellung angelehnt an https://arxiv.org/abs/2304.08485