

...









## The topics

#### Opening

- Ronald Maier
- Ulf Busch





#### Artificial Intelligence, powered by Azure

- Fabian Jusufi: Introducing Azure Services at the University of Vienna
- Oliver Wieder, Christian Binder: Decoding Odor Perception with Attention-Based Graph Neural Networks through Self-Supervised Learning
- Wolfgang Klas, Mara Sophie Aichinger: Using Azure Cloud Services in the FactCheck Project

#### **ZID Services and AI**



 Janós Békési: Data Enrichment by AI: UNIDAM as an Example

#### News and Feedback

• Michaela Bociurko



#### Summary and Outlook

Ronald Maier





#### **Organisational notes**

• Please ask your questions as a chat comment.



• If you want to ask questions with microphone, please click the "Raise Hand" button.









# Opening

- Ronald Maier
- Ulf Busch

© Images by storyset on Freepik







## Artificial Intelligence, powered by Azure

- Introducing Azure Services at the University of Vienna
- Decoding Odor Perception with Attention-Based Graph Neural Networks through Self-Supervised Learning
- Using Azure Cloud Services in the FactCheck Project

© Images by storyset on Freepik







## Introducing Azure Services at the University of Vienna

Presentation by Fabian Jusufi

© Image by storyset on Freepik





#### What is Microsoft Azure?

- Public cloud platform from Microsoft
- Offers over 200 different services
  - Infrastructure Service: Classic hardware such as computing, storage & network resources
  - **Platform Services** (SaaS & PaaS): Ready to use software/applications managed by Microsoft
- Service of the ZID since February 2023
  - Employees eligible for usage
  - Pay-as-you-go with extra 15 % discount
  - <u>zid.univie.ac.at/en/azure/</u>







#### **Azure AI Services**



https://azure.microsoft.com/en-us/products/ai-services





## **Funding for research with Azure services**

- ZID funding of 22,000 euros in the summer semester 2024
  - Goal: Supporting research activities in Azure
  - Criteria: Usage of Azure services for which the ZID does not offer alternative IT services (e.g. **AI services**) or working with hybrid approaches
- Our strategy: Enabling research projects that currently cannot be realised on-premises by offering a secure environment in the cloud
  - Possibility for researchers to continue in Azure after funding period
  - <u>zid.univie.ac.at/en/azure/funding/</u>





Project name	Submitter	Funding amount
FactCheck	Wolfgang Klas	4,000€
Revolutionizing Olfactory Perception Mapping: A Contrastive Learning Graph Neural Network Approach	Oliver Wieder	4,000€
FLEXWEB	Marina Dütsch	4,000€
Very Large-scale Distributed Micromagnetic Research Tools	Abert Claas	2,000€
selscape: Automated and Distributed Pipelines for Investigating the Landscape of Natural Selection from Large-scale Genomic Datasets	Xin Huang	2,000€
MULTIREP – Multidimensional Representation: Enabling an Alternative Research Agenda on the Citizen- Politician Relationship	Dylan Paltra	2,000€
Controlled Machine Translation with Large Language Models for the Technical Domain	Miguel Angel Rios Gaona	2,000€
Determining Scientific Uncertainty in Academic Publications	Petro Tolochko	2,000€





#### Questions about Introducing Azure Services at the University of Vienna

• Please ask your questions as a chat comment.



• If you want to ask questions with microphone, please click the "Raise Hand" button.





© Image by storyset on Freepik







## Decoding Odor Perception with Attention-Based Graph Neural Networks through Self-Supervised Learning

Presentation by Christian Binder and Oliver Wieder

© Image by pana on Freepik





#### Main objective

# Link and highlight molecular features to odor perception



© Image by storyset on Freepik





## **Odor perception importance**

- Fragrance and flavour development
- Environmental monitoring
- Food and safety monitoring
- Healthcare and medicine



© Image by storyset on Freepik



## **Odor perception challenges**

- 400 odor receptors (GPCRs)
- Multi-sensory data
- Subjectivity of perception







## **Odor perception challenges**

- Complex interaction
  - Relationship between structure and odor
- Data scarcity
  - Proteins and odor structures







#### **Odor perception challenges**

# Highly complex relationship meets data scarcity



© Images by storyset on Freepik





#### Considerations for developing the algorithm

- Not enough protein information available
- Which molecule inherent features (3D, bond, atom, and molecule info) are necessary?
- Explainable architecture



![](_page_18_Picture_0.jpeg)

![](_page_18_Picture_1.jpeg)

![](_page_18_Figure_2.jpeg)

https://www.biorxiv.org/content/10.1101/2022.09.01.504602v4.full.pdf

![](_page_19_Picture_0.jpeg)

![](_page_19_Picture_1.jpeg)

#### **Considerations – Embeddings**

![](_page_19_Figure_3.jpeg)

![](_page_19_Figure_4.jpeg)

https://www.biorxiv.org/content/10.1101/2022.09.01.504602v4.full.pdf

![](_page_20_Picture_0.jpeg)

![](_page_20_Picture_1.jpeg)

#### **Considerations – Data scarcity**

Self-supervised learning

Self-supervised learning

- 1. Pre-train on millions of molecules
- 2. Fine-tune on labeled data

![](_page_20_Figure_7.jpeg)

![](_page_20_Figure_8.jpeg)

![](_page_21_Picture_0.jpeg)

![](_page_21_Picture_1.jpeg)

![](_page_21_Figure_2.jpeg)

![](_page_22_Picture_0.jpeg)

## Our approach

#### **Pre-training**

- Predicted with POM (Principal Odor Map) model
- "Low quality" labels
- Focus on contrastive loss

![](_page_22_Picture_6.jpeg)

![](_page_22_Figure_7.jpeg)

![](_page_23_Picture_0.jpeg)

![](_page_23_Picture_1.jpeg)

#### **Considerations – 2 training phases**

![](_page_23_Figure_3.jpeg)

GNN = Graph Neural Networks SSL = self-supervised learning

![](_page_24_Picture_0.jpeg)

![](_page_24_Picture_1.jpeg)

![](_page_24_Figure_2.jpeg)

![](_page_25_Picture_0.jpeg)

![](_page_25_Picture_1.jpeg)

## **Odor perception challenges – Summary**

- Self-supervised learning with Classifier
  - Pre-training
  - Fine-tuning
- Attention for explainability

![](_page_25_Picture_7.jpeg)

© Image by storyset on Freepik

![](_page_26_Picture_0.jpeg)

![](_page_26_Picture_1.jpeg)

#### Azure

#### POM prediction and training

- Azure cloud infrastructure
- GPU-enabled VMs (NVIDIA)

File Edit View Search Terminal	Help				
odormap-003-admin@odormap-vm-00 Mon Apr 1 12:33:00 2024 +	03:∼\$ nvidia-smi				
NVIDIA-SMI 535.161.07	Driver Version: 535.161.07	CUDA Version: 12.2			
GPU Name Per  Fan Temp Perf Pwr	rsistence-M   Bus-Id Disp.A r:Usage/Cap   Memory-Usage	Volatile Uncorr. ECC     GPU-Util Compute M.     MIG M.			
0 NVIDIA A100 80GB PCIe N/A 32C P0	Off   00000001:00:00.0 Off 42W / 300W   0MiB / 81920MiB	0    0% Default    Disabled			
+   Processes:   GPU GI CI PID Type Process name GPU Memory     ID ID Usage					
No running processes found					
dormap-003-admin@odormap-vm-00	03:~\$	+			

NVIDIA A100	24 Cores	4:12 min Training	4.13€/h
NVIDIA Tesla T4	4 Cores	6:19 min	0.55 €/h

![](_page_27_Picture_0.jpeg)

![](_page_27_Picture_1.jpeg)

## Special thanks to:

#### **Cheminformatics Research Group**

- Prof. Thierry Langer
- Daniel Rose, M.Sc. M.Sc.
- Christian Binder B.Sc
- Dr. Thomas Seidel

ZID

Christian Doppler Laboratory for Molecular Informatics in the Biosciences (CD-Lab MIB)

![](_page_27_Picture_9.jpeg)

![](_page_28_Picture_0.jpeg)

![](_page_28_Picture_1.jpeg)

#### Questions about Decoding Odor Perception with Attention-Based Graph Neural Networks through Self-Supervised Learning

• Please ask your questions as a chat comment.

![](_page_28_Picture_4.jpeg)

• If you want to ask questions with microphone, please click the "Raise Hand" button.

![](_page_28_Figure_6.jpeg)

![](_page_28_Picture_7.jpeg)

© Image by storyset on Freepik

![](_page_29_Picture_0.jpeg)

![](_page_29_Picture_1.jpeg)

![](_page_29_Picture_2.jpeg)

## Using Azure Cloud Services in the FactCheck Project

Presentation by Wolfgang Klas and Mara Sophie Aichinger

![](_page_30_Picture_0.jpeg)

![](_page_30_Picture_1.jpeg)

#### **Speakers**

![](_page_30_Picture_3.jpeg)

Wolfgang Klas

![](_page_30_Picture_5.jpeg)

Mara Sophie Aichinger

#### Further team

- Daniel Berger
- Adrian Hofer
- Peter Kalchgruber

## Agenda

- The FactCheck Framework
- FactCheck & Azure

![](_page_31_Picture_0.jpeg)

![](_page_31_Picture_1.jpeg)

![](_page_31_Picture_2.jpeg)

#### **The FactCheck Framework**

Introduction to the Research Context

© Image by storyset on Freepik

![](_page_32_Picture_0.jpeg)

![](_page_32_Picture_1.jpeg)

#### **Conflicting Information on the Web**

![](_page_32_Figure_3.jpeg)

www.univie.ac.at/fileadmin/user\_upload/startseite/Dokumente/2023\_UW\_in\_Zahlen\_EN\_\_1\_.pdf

![](_page_33_Picture_0.jpeg)

![](_page_33_Picture_1.jpeg)

#### **Conflicting Information on the Web**

Google	university of vienna	a how many students	×	<u>ତ</u> . ଦ			
All Images Ne	ws Videos Books	: More		Tools			
About 44.300.000 re	sults (0,43 seconds)						
University of Vienna	7 Total enrollment						
<b>88,900</b>							
With about 90,000 students, 178 degree programmes as well as more than 60 university continuing education and training programmes, the University of Vienna is the largest and most diverse educational institution in Austria.							

![](_page_34_Picture_0.jpeg)

![](_page_34_Picture_1.jpeg)

## Information on the Web: Challenges

- Finding information and **facts**
- Time dimension: how **"up to date"** are the data?
- Heterogeneity and abstraction level of information
  - Rounding (e.g. "about 90,000 students")
  - Time/number formats
- Trust aspect

![](_page_34_Picture_9.jpeg)

Image by storyset on Freepik

![](_page_35_Picture_0.jpeg)

![](_page_35_Picture_1.jpeg)

#### **FactCheck: Introduction**

![](_page_35_Picture_3.jpeg)

- Toolkit for **exploring and detecting inconsistent** related data on the web
- Ongoing research topic at Multimedia Information Systems (MIS) group
- **Prototype** currently hosted mostly on-premises, partially on Azure

Image by storyset on Freepik

![](_page_36_Picture_0.jpeg)

![](_page_36_Picture_1.jpeg)

#### **FactCheck: Research Aspects**

- Extraction of information/facts on the web
  - Entity resolution/topic extraction
  - Extracting facts from text or multimedia content (e.g. videos)
- Context-aware **comparison** of information/facts
- Crowd-sourced conflict detection

![](_page_36_Figure_8.jpeg)

Image by storyset <u>https://storyset.com</u>

![](_page_37_Picture_0.jpeg)

![](_page_37_Picture_1.jpeg)

![](_page_37_Picture_2.jpeg)

#### FactCheck & Azure

Hybrid Cloud Computing in Research

![](_page_38_Picture_0.jpeg)

![](_page_38_Picture_1.jpeg)

## **Azure / Cloud Computing for Research**

- **Scalability:** prototype can easily scale horizontally (more instances) and vertically (better hardware)
- Elasticity: scale up and down dynamically to meet current demands
- Availability: high availability backed by Service-Level Agreements (99.9%+ uptime)
- Fault tolerance and disaster recovery: suitable replication and backup in case of emergency
- Cost-efficiency: economies of scale and consumption-based billing
- Use of cloud-based services such as, e.g.
  - Artificial Intelligence (AI)
  - Natural Language Processing (NLP)

![](_page_38_Picture_11.jpeg)

Hybrid Approach

- Joint use of **on-premises and Azure infrastructure**
- **Combines benefits** of cloud computing and on-premises deployment
- Allows for research data to be stored and processed on-premises
- Allows for use of **local infrastructure** (e.g. databases, VMs) if more cost-efficient
- Needed also in the **development** phase of the system

© Image by storyset on Freepik

![](_page_39_Picture_8.jpeg)

![](_page_39_Picture_9.jpeg)

![](_page_39_Picture_10.jpeg)

#### Current Work Packages: Cluster 1 – Hybrid Architecture

- **1.1** Use of Azure API Management as central point to expose and test our API
- **1.2** Implement secure connection between on-premises and Azure services
- 1.3 Reimplement parts of on-premises prototype as cloud-native microservices (Azure Functions)
- 1.4 Migrate parts of locally hosted data storages to Azure CosmosDB

![](_page_40_Picture_6.jpeg)

![](_page_40_Picture_7.jpeg)

![](_page_40_Picture_8.jpeg)

![](_page_41_Picture_0.jpeg)

![](_page_41_Picture_1.jpeg)

#### **Current Work Packages: Cluster 2 – Fact Extraction**

- **2.1** Extraction of information from video material using Azure AI Video Indexer
- **2.2** Dynamic and scalable collection of information from the web ("crawling")

![](_page_41_Picture_5.jpeg)

![](_page_42_Picture_0.jpeg)

![](_page_42_Picture_1.jpeg)

#### WP 2.1 – Fact Extraction from Videos

#### • Main service:

Azure Al Video Indexer

- Extraction of information from unstructured video material
- **Leverages** Azure's mature Al services

![](_page_42_Picture_7.jpeg)

![](_page_42_Picture_8.jpeg)

![](_page_43_Picture_0.jpeg)

![](_page_43_Picture_1.jpeg)

#### WP 2.2 – Scalable Crawling of Web content

- Main services: Azure VMs, Azure Service Bus
- Scalable **extraction** of information on the web
- Dynamically (re-)crawl websites for information
- Process and store information at scale
  - e.g. using a data lake built upon Azure Data Lake Storage (ADLS) Generation 2

#### All About Bernie All About Bernie This is an example page! Not an actual website. Bernie Sanders was born in Brooklyn, 1941. He is an US politician, and has been the senator from Vermont since 200 Why d natura "@context": "https://schema.org", him w "@type": "Person", The h "givenName": "Bernie", "familyName": "Sanders", "jobTitle": "Politician", "birthDate": "1941-09-08", "age": "80"

![](_page_43_Picture_9.jpeg)

![](_page_44_Picture_0.jpeg)

![](_page_44_Picture_1.jpeg)

#### Questions about Using Azure Cloud Services in the FactCheck Project

• Please ask your questions as a chat comment.

![](_page_44_Picture_4.jpeg)

• If you want to ask questions with microphone, please click the "Raise Hand" button.

![](_page_44_Figure_6.jpeg)

![](_page_44_Picture_7.jpeg)

© Image by storyset on Freepik

![](_page_45_Picture_0.jpeg)

![](_page_45_Picture_1.jpeg)

![](_page_45_Picture_2.jpeg)

#### Q & A Session Materials

Additional information, scrapped slides, more material

© Image by storyset on Freepik

#### WP 1.1 – Azure API Management for API Development

- Main service: Azure API Management (APIM)
- **Decouple** API definition and implementation
- Hide internal endpoints behind public URL
- Cloud-based API documentation and testing

![](_page_46_Figure_6.jpeg)

![](_page_46_Picture_7.jpeg)

![](_page_46_Picture_8.jpeg)

![](_page_47_Picture_0.jpeg)

![](_page_47_Picture_1.jpeg)

#### WP 1.2 – Connect On-Premises with Azure

- Main service: Azure VPN Gateway
- Securely **connect** on-premises infrastructure with the cloud
- Uses IPSec IKE P2P (industry standard)
- Close collaboration of MIS and the ZID

![](_page_47_Picture_7.jpeg)

![](_page_47_Figure_8.jpeg)

https://learn.microsoft.com/en-us/azure/vpn-gateway/about-site-to-site-tunneling

![](_page_48_Picture_0.jpeg)

![](_page_48_Picture_1.jpeg)

#### WP 1.3 – Cloud-native reimplementation on Azure

- Main service: Azure Functions
- Split up monolithic business logic, and (partially) reimplement it as **microservices**
- Eases independent scaling and deployment

![](_page_48_Figure_6.jpeg)

![](_page_48_Picture_7.jpeg)

#### WP 1.4 – Data Store Migration to Azure CosmosDB

**D** ~

NOSQL API

ß

**D** ~ **O** 

🖪 🗗 v 🛱 v

 $\bigcirc$  <

- Main service: Azure CosmosDB
- **Hosting** (non-sensitive) research data on Azure CosmosDB (multi-model document database)
- Advanced **indexing** and **querying** functionality
- Distributed database with high availability

▼ DATA	SELECT * EDON	Call Calls (	llan	
DebugDB     Comparison-result-digests	id	(mid		
Items		/1110		1 {
Settings  Stored Procedures	2022-01-30 2022-01-30	kg://msa/be kg://msa/wu.		<pre>2 "mid": "kg://msa/bernie", 3 "property": "givenName", 4 "value": "Bernard", 5 "reasons_count": {</pre>
<ul><li>User Defined Functions</li><li>Triggers</li></ul>	2022-01-30 2022-02-01	kg://msa/wu. kg://msa/be		6 "outdated": 1, 7 "not-equal": 2 8 }, 9 "reasons freetext": {
▼ NOTEBOOKS	2022-02-01	kg://msa/be.	·	10 "outdated": [ 11 "He used to be called Bernard when he starts
Notebooks is currently not available. We are working on it.	2022-01-31 2022-02-09	kg://msa/be		12 ], 13 "not-equal": [ 14 "His name is Bernie, and not Bernard!".
	2022-02-09	kg://msa/be.	·	15 "Who even is Bernard Sanders?" 16 ]
	2022-02-03	kg://msa/be		"day": "2022-01-30", 18 "day": "2022-01-30", 19 "id": "2022-01-30-kg:++msa+bernie-givenName-Bernard"
	2022-02-24	kg://msa/be		<pre>20rla": '3KUFAPVFU2/mEAAAAAAAAAa=-'' 21self": 'dbs/9kUFAA==/colls/9kUFAPvF02M=/docs/9kUFA 22etag": '\"8300d9af-0000-0d00-0000-61d361480000\""</pre>
	2022-02-24 Load	kg://msa/be	•	23 "_attachments": "attachments/", 24 "_ts": 1641242952

🗋 New Item 🔚 Update 🏷 Discard 🗋 Delete

![](_page_49_Picture_7.jpeg)

Upload Item

![](_page_49_Picture_8.jpeg)

![](_page_49_Picture_9.jpeg)

![](_page_50_Picture_0.jpeg)

![](_page_50_Picture_1.jpeg)

![](_page_50_Picture_2.jpeg)

## **ZID Services and AI**

• Data Enrichment by AI: UNIDAM as an Example

© Images by storyset on Freepik

![](_page_51_Picture_0.jpeg)

![](_page_51_Picture_1.jpeg)

![](_page_51_Picture_2.jpeg)

## Data Enrichment by AI: UNIDAM as an Example

Presentation by János Békési

© Image by storyset on Freepik

![](_page_52_Picture_0.jpeg)

UNIDAM search view

#### UNIDAM: Digital Asset Management System at the University of Vienna

- Mainly picture material for teaching and research purposes
- Closed user base (for copyright reasons)
- About 500,000 assets
- About 150 metadata fields (not used simultaneously)
- URL: <u>unidam1.univie.ac.at</u>

![](_page_52_Picture_7.jpeg)

![](_page_52_Picture_8.jpeg)

![](_page_52_Picture_9.jpeg)

![](_page_53_Picture_0.jpeg)

![](_page_53_Picture_1.jpeg)

#### Metadata Enrichments already attempted

- Named Entity Recognition (NER): Artists' names in UNIDAM are supplemented with Virtual International Authority File (VIAF) and GND (Gemeinsame Normdatei) entries
- **Topic Models** (only for textual assets): not an enrichment as such but a research tool
- Optical Character Recognition (OCR )/ Handwritten Text Recognition (HTR) result improvement (only for textual assets): via open source LLM – not yet feasible
- Azure/OpenAI: ICONCLASS

Ξ	Recherche	🛓 🖉 📤 jbekesi DE 🔗 Karr
+	Schnellzugriff < Mappe	Detail
+ Q Ⅲ ≌ ≯ ☆ 6	Schneitzugriff   Mappe  Fingencen auf.  MAPPEN  Meine Mappen  Meine Mappen  Monie Mappen  Mappen  Monie Mappen  Mo	Detail
	Daten Lebensdaten Luce 25.01.1706 - 04.02.1787 Rom Geschlecht männlich Künstlerart Maler	Info
	Normdaten GND Person Batoni, Pompeo Girolamo & Wikidata http://www.wikidata.org/entity/Q505613 & Getty Batoni, Pompeo &	Fundort / Aufbewahrung / Herstellung Aufbewahrung (Ort) Istalen / Goma Privatbeaitz Technik (en) Italen / Goma Italen / Gom
	Split [1366] GNDJGetty autom. Ja   SND Getty autom. System-10 #12931 UUID 4x86e401-1914-49c-1857-92x4cba3ddd Objekttyp Kinstler Maske GND und Getty Erstellt am 13.02.2018 f2:08 Zuletzt geänder 31.03.221351 Versior	scniagworter Malerei Iconclass fite2(Teres Ugurgieri Spinola) - historische Person (Teresa Ugurgieri Spinola) fite2(Teres Ugurgieri Spinola) - historische Person (Teresa Ugurgieri Spinola) fite2(Teresa Ugurgieri Spinola) - historische Person (Teresa Ugurgieri Spinola) - historische Pe

UNIDAM detail view with artist normdata

![](_page_54_Picture_0.jpeg)

#### IT<sup>4</sup>S IT for Science FORUM

🛣 Foru

#### ICONCLASS

- Is a "comprehensive classification system for the content of images" devised by Henry de Waal in the 1940s
- Has 10 main subdivisions (and then some)

**Working example:** WienMuseum (Iconclass codes implemented manually) <u>sammlung.wienmuseum.at</u>

Browse	Search	2 · Nature
• 0 · Abs	stract, Non-representational Art	25 - earlin, wond as celesital body
▶ 1 · Reli	gion and Magic	25F2 , mammale
2 · Nat	ure	25E22 - hearte of prov. prodatory animale
· 20·'	Natura' (allegorical figure or scene; or as Diana of Ephesus, with many	25F23 beasts of prey, predatory animals
breasts)	; 'Natura' (Ripa)	
▶ 21 · t	he four elements, and ether, the fifth element	25F23(LION) a beasts of prey, predatory animals
▶ 22 · r	natural phenomena	lion
▶ 23·t	ime	lion
▶ 24 · t	he heavens (celestial bodies)	Search with these related keywords:
▶ 25 · e	earth, world as celestial body	animal, earth, lion, mammal, nature, predatory animal, world
▶ 25A	· maps, atlases	Add more detail:
▶ 25B	- continents represented allegorically	25FF23(LION) · beasts of prey, predatory animals: lion - FF - fabulous animals
▶ 250	C · geological phenomena	25F23(LION)(+0) · beasts of prey, predatory animals: lion (+ variant)
▶ 25D	) · rock types; minerals and metals; soil types	25F23(LION)(+1) · beasts of prey, predatory animals: lion (+ animals used symbolically) 25F23(LION)(+2) · beasts of prey, predatory animals: lion (+ sex and age of animals: propagation
▶ 25E	· geological-chronological division; historical geology; geological era	of animals)
▶ 25F	· animals	25F23(LION)(+3) · beasts of prey, predatory animals: lion (+ anatomy of animals)
▶ 25	F1 · groups of animals	25F23(LION) (+4) · beasts of prey, predatory animals: lion (+ animal behaviour)
▶ 25	F2 · mammals	25F23(LION)(+5) · beasts of prey, predatory animals: lion (+ animal(s) in motion; positions, expressions of animals)
▶ 2	25F21 · groups of mammals	$25F23(LION)(+6) \cdot beasts of prey, predatory animals: lion (+ disease and death of animal(s))$
▶ 2	25F22 · monkeys, apes	25F23(LION)(+7) · beasts of prey, predatory animals: lion (+ animals ~ biological investigation)
▶ 2	25F23 · beasts of prey, predatory animals	25F23(LION)(+8) - beasts of prey, predatory animals: lion (+ man and animal)
•	25F23() · beasts of prey, predatory animals (with NAME)	25F23(LION)(+9) - beasts of prey, predatory animals: lion (+ products of animals)
)	<ul> <li>25F23(BADGER) · beasts of prey, predatory animals: badger</li> </ul>	
1	25F23(BEAR) · beasts of prey, predatory animals: bear	For more features, you need to log in
	<ul> <li>25F23(CHEETAH) · beasts of prey, predatory animals: cheetah</li> </ul>	r or more reatures, <u>you need to log m</u> .
)	<ul> <li>25F23(ERMINE) · beasts of prey, predatory animals: ermine</li> </ul>	
)	<ul> <li>25F23(FERRET) · beasts of prey, predatory animals: ferret</li> </ul>	
	<ul> <li>25F23(FOX) · beasts of prey, predatory animals: fox</li> </ul>	
)	<ul> <li>25F23(HYENA) · beasts of prey, predatory animals: hyena</li> </ul>	
)	25F23(ICHNEUMON) · beasts of prey, predatory animals: ichneumon	
•	25F23(LEOPARD) · beasts of prey, predatory animals: leopard	
•	<ul> <li>25F23(LION) · beasts of prey, predatory animals: lion</li> </ul>	
	<ul> <li>25FF23(LION) · beasts of prey, predatory animals: lion - FF - fabulous</li> </ul>	
	animals	
	25F23(LYNX) · beasts of prev. predatory animals: lvnx	

ICONCLASS Browse & Search Documentation News

ICONCLASS website with opened concept "LION"

![](_page_55_Picture_0.jpeg)

![](_page_55_Picture_1.jpeg)

ICONCLASS (cont'd)

#### Iconclass codes can

- work as facets in a faceted search
- thereby collect assets related by their visual (iconographical) content

![](_page_55_Picture_6.jpeg)

UNIDAM detail view with iconclass metadata

UNIDAM search view filtered by iconclass

![](_page_56_Picture_0.jpeg)

![](_page_56_Picture_1.jpeg)

#### **OpenAl/Azure**

- Prompts are being iterated to find a reasonable output for processing
- Cost/benefit has yet to be assessed
- After authentication etc. is in place, everything works smoothly
- Later on, the wealth of query metadata could be analysed

	,	25	p2024-02-02:
/* prompt_history.yml	0	26	content: Reschreibe nach kunstaeschichtlichen Kriterien, stelle Vermutungen an
🗋 test-gpt.log	0	27	aib mänliche Tronclass-Reschreibungen an füge diese als Code für das Feld "subject" in
test_gpt.dudb	0	28	ein Dublin Core-Schema ein und publiziere das im JSON-Format. Danke!"
/* test_gpt.py	•	29	
/* testgpt_results.json		30	
easydb-library			system: Du bist ein Professor der Kunstgeschichte und erklärst das Bild für eine
🖿 110n		32	Studentin.
src			p2024-02-23:
.gitignore			content: 'Beschreibe nach kunstgeschichtlichen ikonographischen und ikonologischen
.gitmodules			Kriterien, stelle Vermutungen an, gib mögliche Iconclass-Beschreibungen an,
/* gpt-pb.yml		36	füge diese als Code für das Feld "subject" in ein Dublin Core-Schema ein und
/* iterate.sh		37	publiziere das im JSUN-Format. Danke!
/* Makefile			system. Du hist ein Professon der Kunstgeschichte und erklänst des Bild für einen
/* manifest.yml		41	Kollegen.
mycsv.csv	0	42	p2024-02-23a:
mycsv_notxt.csv	0		content: 'Für das Bild gibt es folgende Informationen: Wien, Triton- und Najade-Brunnen
/* Pipfile			
/* Pipfile.lock		45	Ort: (Hauptstadt) Wien
pixi.lock	0	46	
/* pixi.toml		47	. Beschreibe nach kunstgeschichtlichen ikonographischen und ikonologischen Kriterien,
🗅 pixi.toml~		48	stelle vermutungen an, gib mogliche iconclass-beschreibungen an, fuge alese
/* plugin-gpt.sublime-proj	iect		als code fur das Feld "subject" in ein Dublin Core-Schema ein und publiziere
⇔ README.md	1998	50	aas im JSON-Format. Danke!
		52	
		53	system: Du bist ein Professor der Kunstgeschichte und erklarst das Bild für einen
		54	Kollegen.

![](_page_57_Picture_0.jpeg)

![](_page_57_Picture_1.jpeg)

#### **Our Workflow (Prototype)**

- 1. Export data (pictures and some metadata) from UNIDAM
- 2. Use prompt to query Azure/OpenAI model with data
- 3. Clean result and write import file for UNIDAM
- 4. Import result into UNIDAM
- 5. Take expert assessment into account and
- 6. Iterate 2-5 until satisfied

![](_page_57_Figure_9.jpeg)

![](_page_58_Picture_0.jpeg)

#### **Our Workflow (cont'd)**

- Prompts are being evaluated after each run lacksquare
- Prompts from the OpenAI Web GUI are being evaluated as well, often used as a starting point
- A multi-step prompt will be attempted to let OpenAI auto-correct the results
- Results (ICONCLASS codes) are checked for validity (to avoid hallucinations)

![](_page_58_Picture_6.jpeg)

. . .

FOLDERS 🛛 📄 plugin-gp ⊧ 💼 .pb

🔻 📄 cli

🕨 📰 data

▶ 📰 data-2

🕒 .env.ir

lin 🖿

⊳ IIII src

.gitignore

/\* gpt-pb.ym

iterate.sh LICENS

\* Makefile /\* manifest.yr

mvcsv.csv

/a Pinfile /\* Pipfile.lock pixi.lock

pixi.tom

est apt

59

**Future perspectives** 

- Obtain half-automated workflow for enrichment
- Implement quality control for results of enrichment runs
- Widen the scope of possible enrichment targets in UNIDAM
- Reduce cost per request (i.e. number of tokens lacksquareused) by smart prompting or other measures
- Evaluate cost/benefit

Normdaten GND Person Batoni, Pompeo Girolamo 🚱 Wikidata http://www.wikidata.org/entity/Q505613 🗗 Getty Batoni, Pompeo 🚱 Split [1366] GND/Getty autom. GND prüfen Getty prüfen! Nein Nein

Ja

UNIDAM: toggle switch for later checks

![](_page_59_Picture_9.jpeg)

![](_page_59_Picture_10.jpeg)

![](_page_60_Picture_0.jpeg)

![](_page_60_Picture_1.jpeg)

#### Questions about Data Enrichment by AI: UNIDAM as an Example

• Please ask your questions as a chat comment.

![](_page_60_Picture_4.jpeg)

• If you want to ask questions with microphone, please click the "Raise Hand" button.

![](_page_60_Figure_6.jpeg)

![](_page_60_Picture_7.jpeg)

If you have any further questions, please contact us at: <u>unidam@univie.ac.at</u>

© Image by storyset on Freepik

![](_page_61_Picture_0.jpeg)

![](_page_61_Picture_1.jpeg)

![](_page_61_Picture_2.jpeg)

## **News and Feedback**

Michaela Bociurko

© Images by storyset on Freepik

![](_page_62_Picture_0.jpeg)

![](_page_62_Picture_1.jpeg)

#### Free online course about digital accessibility

≡	<u>iMᡂ</u>	K			Über	iMooX Partner:inne	en 🗸 All	e Kurse	🐚 Deutsch	(de) 🗸	Login
E	€ DE		<b>Æ</b>	Ø	>	Î	)	>_		¢	Ŋ
	Betreuter Kurs	ationstechnik - Medi	enwissenschaften Digitale Universität Rainer Jant	<ul> <li>Kommunikationswissensch</li> <li>Barrierefreiheit</li> <li>Wien</li> <li>scher</li> <li>Über</li> </ul>	aften	Inhalt		☆☆ Bev	☆ ☆ ☆ rertung	4	
Über	0 U	Dauer 6 Lektionen Verfügbarkeit Unbegrenzt	•	Einheiten 1 Stunde/Lektion Startdatum 1. April 2024	©© €	Lizenz CC BY-SA 4.0 Kosten € 0.00	*	Kursteilnehme 141	r.innen		

![](_page_62_Picture_4.jpeg)

#### Find out more at: <u>imoox.at/course/diba</u>

![](_page_63_Picture_0.jpeg)

![](_page_63_Picture_1.jpeg)

#### We asked for your feedback

![](_page_63_Figure_3.jpeg)

![](_page_64_Picture_0.jpeg)

#### Topics you want to hear about

16 Antworten

Which topics should be covered in upcoming sessions of the IT for Science Forum?

History

Register Data (AMDC) https://rdm.univie.ac.at/services/austrian-micro-data-center/ maybe first reports on applications? Al and Data (Who ownes the data?) IP and Data How do we deal with data being used by large language models without the consent of researchers and affected?

It was great that todays forum was dedicated to digital humanities and I hope to see more about this field of research in the future.

Big data sharing; dealing with sensitive data, i.e. very selective sharing of securely archived material

KI

FAIR data

Perhaps a bit more on working with data sheets and storing those in a collaborative environment. Best practice examples ... etc But so far the inputs were very interesting!

Unicorns in space

Enthaltungen

![](_page_64_Picture_13.jpeg)

ıl.

х б

![](_page_65_Picture_0.jpeg)

## Thank you!

14 Antworten	11.	
General feedback on the IT for Science Forum (optional)?		
Very informativ		
Thank you! It was very interesting so far :)		
Very nice		
Perfect.		
The forum is a great initiative to help connect not only different fields of research but especially - as it's title states - IT and science(/humanities). I'd explicitly like to acknowledge the great support beforehand, which really facilitated my preparation. Regarding language: Obviously, English is practical as an universal language, but maybe the forum language could be decided from event to event, depending on the speakers, thereby allowing conversation to flow a little more freely.		
I've attended all events so far and really like the scope of topics. General organization and time management are always excellent and highly appreciated!		
Good initiative! Keep up the good work.		
Great.		
thank you very much. It is great to have cases and examples from research!		
Enthaltungen	x 4	

![](_page_65_Picture_3.jpeg)

![](_page_66_Picture_0.jpeg)

![](_page_66_Picture_1.jpeg)

![](_page_66_Picture_2.jpeg)

## Outlook

Ronald Maier

© Images by storyset on Freepik

![](_page_67_Picture_0.jpeg)

![](_page_67_Picture_1.jpeg)

# Thank you for your attention! See you at the next IT for Science Forum: 30 October 2024, 9:00–11:00

Contact us: it4s@univie.ac.at