DEMOCRATIZING DATA & AI
FOR NHR

CARSTEN BINNIG
CSG DATA MANAGEMENT, NHR4CES, TU DARMSTADT
VISION: DEMOCRATIZING DATA & AI
TODAY’S DATA & AI SYSTEMS!

- Flink
- mxnet
- Spark
- hadoop
- TensorFlow
- Asterix DB
WHAT ARE THE PAIN POINTS?

Research Agenda: Make Data & AI Systems more useable, efficient, .... for NHR
REVISIT DATA & AI SYSTEM STACK

1. System Efficiency
   (Scalability on Modern Hardware)

2. User Efficiency
   (Visual & NL interfaces)

3. Automation
   (System & User Tasks)

Focus: Automation of User Tasks
(Data & AI Systems)

Manual Data Preparation
HOW ARE DATA & AI PIPELINES BUILT?

Today: Manually-composed “pipelines” → high overheads + expert skills needed
What is the energy consumption if we reduce cooling by 1%?

Idea: Use AI to automate construction pipelines → reduce high manual overheads
DIRECTION 1: DATA EXTRACTION

Patients (table)

<table>
<thead>
<tr>
<th>name</th>
<th>age</th>
<th>gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alice</td>
<td>42</td>
<td>f</td>
</tr>
<tr>
<td>Bob</td>
<td>23</td>
<td>m</td>
</tr>
</tbody>
</table>

Examinations (text)

Alice was diagnosed with fever ...

Scans (images)

What is the age and diagnosis of patient Alice?

Database

GPT-like Model

Use a GPT-like model to "query" a mix of table, text, image data
DIRECTION 2: DATA AUGMENTATION

What is the average revenue of movies with actor „Zoe S“?

<table>
<thead>
<tr>
<th>movies</th>
<th>genre</th>
<th>revenue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avatar</td>
<td>Sci-Fi</td>
<td>2.74B $</td>
</tr>
<tr>
<td>Avatar2</td>
<td>Sci-Fi</td>
<td>2.24B $</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>actors (GPT-table / virtual)</th>
<th>movie</th>
<th>actor</th>
<th>age</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avatar</td>
<td>Zoe S</td>
<td>...</td>
<td></td>
</tr>
<tr>
<td>Avatar</td>
<td>San W</td>
<td>...</td>
<td></td>
</tr>
</tbody>
</table>

Use a GPT-like model to augment data with external “world-knowledge”
DATA & AI

LITERACY
DATA & AI EDUCATION: HIGHLY INDIVIDUALIZED DEMANDS

Python Programming?

Data Engineering?

ML/AI?

HPC?

RDM?
Where to start?
The Data and AI Literacy Hub is a guided learning platform that offers individual learning paths for different backgrounds and skill levels. It enables you to develop the skills necessary to navigate the data-driven world and apply AI techniques in custom projects.
Learning Paths

Our carefully curated learning paths guide you in finding learning resources that fit your needs and level of knowledge.

- Get started with HPC
- Introduction to Programming
- Learn the Basics of Data Science & AI
- The essentials of RDM
- Theoretical Foundations of AI
OPEN EDUCATION: EVERYONE CAN PROPOSE NEW PATHS & COURSES

GitHub
TRY IT YOURSELF!

https://data-ai-literacy.ml
THANK YOU FOR YOUR ATTENTION!