Data Science and Artificial Intelligence

Institute of Interactive Systems and Data Science

Viktoria Pammer-Schindler

Data Science is

The science of using data as key part in the process of creating knowledge.

Artificial Intelligence is…

1) Systems that are (partially) intelligent.

2) The science of engineering technologies that fulfill some criteria of intelligence.

What does a system need to be able to do in order to have a chance at passing as intelligent?

4

Perceive

Senses and sensors – Digital systems mostly: Data received through interactions with other systems, and with humans.

Think

"Brain" - Memory, knowledge representation, reasoning. This lecture focused on multiple paradigms of how to represent knowledge, and how to reason/think given data/knowledge in a particular form

Broadly, two approaches:

- Symbolic (rules, object-oriented KR, graphs)
- Sub-symbolic, data-driven (graphs, vector representations, neural networks – more from statistics would be: Bayesion networks, Hidden Markov models, etc.)

Act

Human body, and actuators – Digital systems mostly: Interactions with other systems, and with humans (interactive systems)

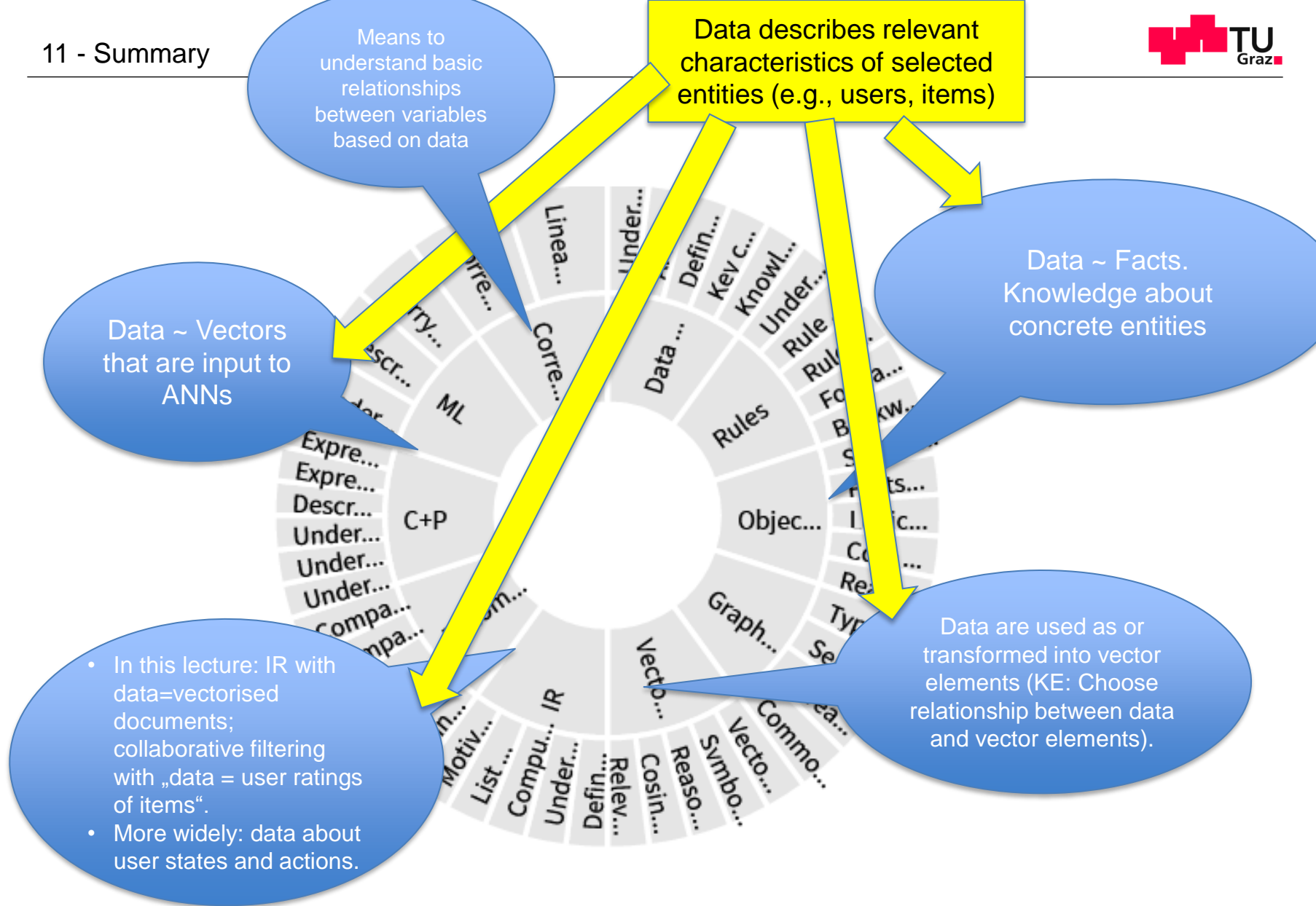## Key capabilities of intelligent systems

5

What is a representation? Knowledge representation and data structures

A **representation** Y conforms in a systematic manner to X , preserving pre-selected characteristics of X.

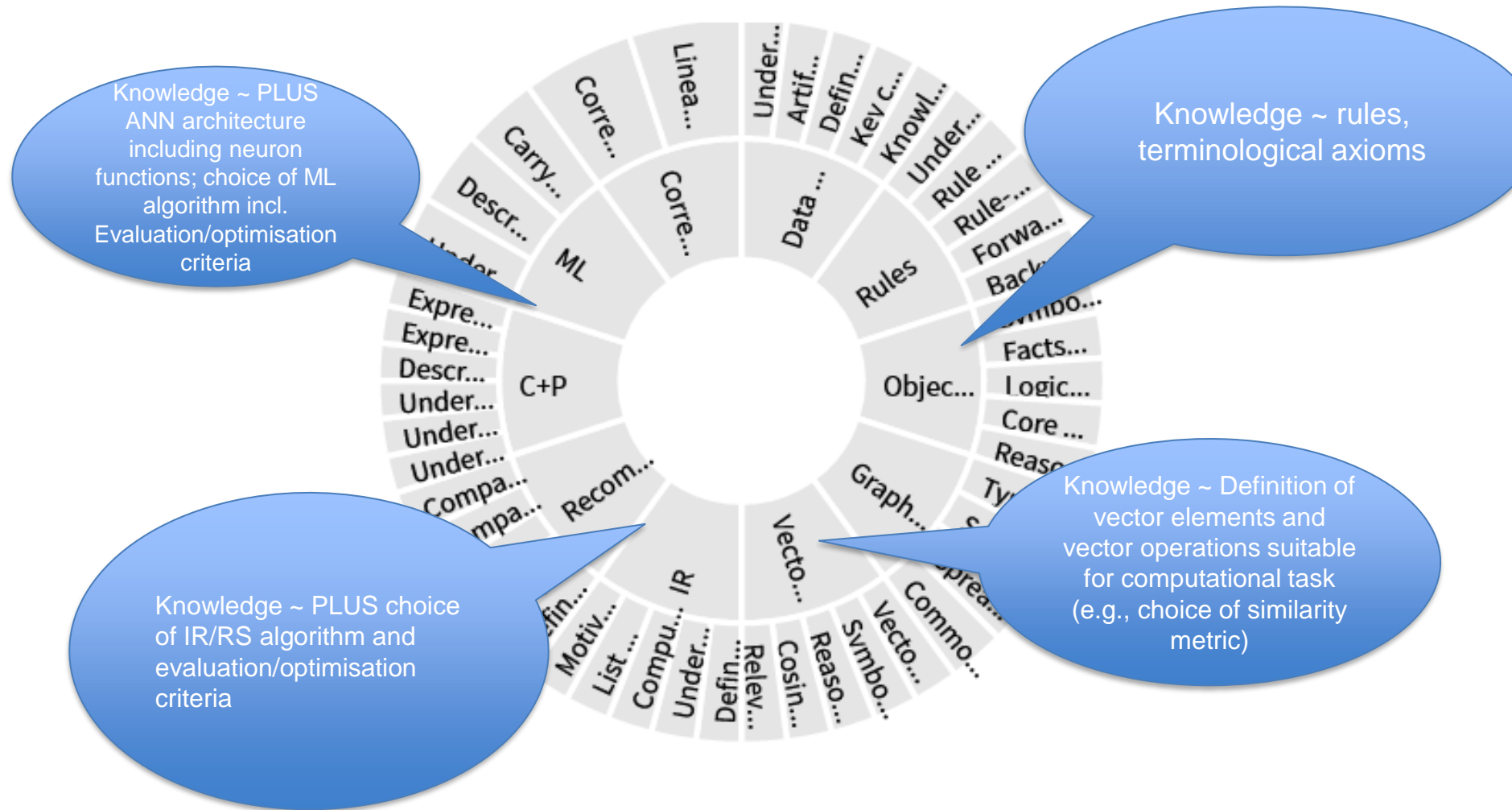**A representation always loses something; it is an approximation.**

- This lecture: A knowledge representation represents (human) knowledge. It serves as structuring this knowledge for computers, as a basis for thinking.
    - A knowledge representation could also have other uses, like communicating knowledge between humans.
- To what extent are vectors a knowledge representation, rather than a data structure? - (some) general knowledge is represented in the choice of which kinds of entities to represent as vectors, and in the choice of features.
    - In symbolic knowledge representation: Terminological axioms vs. assertional axioms (~data)
    - This is, conceptually, a major difference between symbolic knowledge representations and data-driven knowledge representations, which don't formally represent the general knowledge underlying the chosen representation.

6

Where did DATA come up within this lecture?

Means to understand basic relationships between variables based on data

Data describes relevant characteristics of selected entities (e.g., users, items)

Data ~ Facts. Knowledge about concrete entities

Data ~ Vectors that are input to ANNs

Data are used as or transformed into vector elements (KE: Choose relationship between data and vector elements).

- In this lecture: IR with data=vectorised documents; collaborative filtering with „data = user ratings of items".
- More widely: data about user states and actions.

Where did KNOWLEDGE come up within this lecture?

Beyond all this: Knowledge in domain and data science/AI is required in order to decide on suitable problem formulation and implementation

Knowledge ~ PLUS ANN architecture including neuron functions; choice of ML algorithm incl. Evaluation/optimisation criteria

Knowledge ~ rules, terminological axioms

Knowledge ~ PLUS choice of IR/RS algorithm and evaluation/optimisation criteria

Knowledge ~ Definition of vector elements and vector operations suitable for computational task (e.g., choice of similarity metric)

10

Intelligence as…

- 1) Acting humanly? 2) Acting rationally? 3) Thinking humanly? 4) Thinking rationally?

5)

- an entity's capability
- to adapt behavior
    - in response to own interactions with environment
    - to a changing environment
- in order to achieve goals

= an entity's capability to learn from experience.

11

Where did INTELLIGENCE come up in this lecture?

What this lectures covers (very little) and what not (a huge space in data science, AI, and interactive systems design)

13

We covered:

- What is data science, what is AI, data and knowledge, data structures and knowledge representation
- knowledge representation and reasoning paradigms and underlying assumption. This achieves „thinking and acting (in a digital world) rationally)
- A very first ML algorithm, which achieves „learning from experience"

We didn't cover:

- Many many important technical and implementation aspects, e.g., how to set different goals for AI systems (evaluation/optimisation criteria); how to implement scalability, robustness, privacy; how to test or prove that a system is doing what it should, etc.
- Socio-technical embedding: How do AI systems inter-relate with the surrounding environment? What experiences are we „allowing" an AI to have, what impact do AI systems have on the surrounding environment?

Learning Goals

- Understand what data science is
- Understand what artificial intelligence is
- Understand the relationship between data science and artificial intelligence
  - *and the relationships of knowledge representation and reasoning approaches taught in this lecture to both.*

- Remember and understand different definitions of intelligence, and apply them to concrete examples
- Remember and explain building blocks of intellient systems