Deep X: Deep Learning with Deep Knowledge

SAP Leonardo

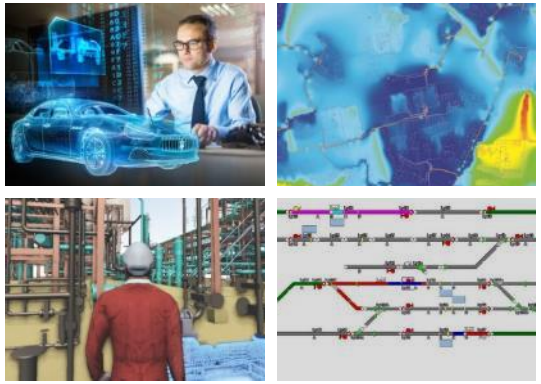
Volker Tresp, Siemens AG and LMU Munich

Digitalization is disrupting entire customer value chains

- Data analytics
- Artificial Intelligence
- Simulation tools
- Cloud and platform technology
- Secure connectivity
- Cyber-Security

Enabling the next level of ...

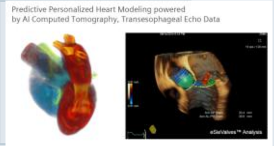
... productivity and time-to-market ...	... flexibility and resilience ...	... availability and efficiency ...
Design and engineering	Automation and operation	Maintenance and services



Artificial Intelligence has a transformative impact on business

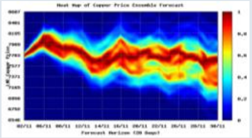
Achievements 1995 – 2017

2015: Autonomous gas turbine optimization using reinforcement learning and RNN is productive



2016: Predictive personalized heart model powered by AI

1995: Online learning neural nets deployed in >30 steel plants



2009: Commodity price forecasting using RNN ensembles

Siemens Corporate Technology - Business Analytics and Monitoring 200 Data Scientists & AI experts at 9 locations globally

Vision for AI Driven Business

AI driven Enterprise

Digital Companion

Autonomous Trains

Smart Grid/City Monitoring

Artificial Intelligence

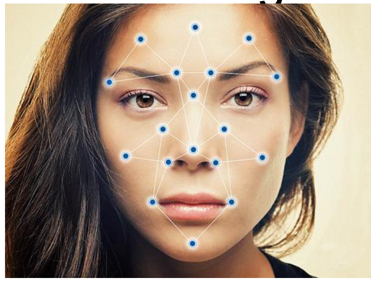
- *Creating machines that perform functions that require intelligence when performed by people (Kurzweil, 1990)*

Games

Translation

Q&A

Face Recognition



Auton. Driving

Speech Recognition

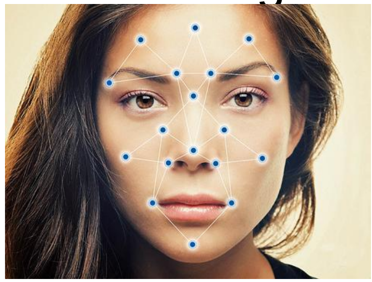
Drones, Robots

?

Deep Learning

- Deep Learning is the reason for the emerging huge interest in AI
  - Convolutional DL
  - Recurrent DL
  - Reinforcement DL
  - Generative Adversarial Networks (GANs)

Face Recognition



Translation

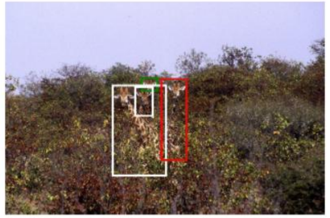
Speech Recognition

CycleGan

Games

Student Magic: Visual Q&A

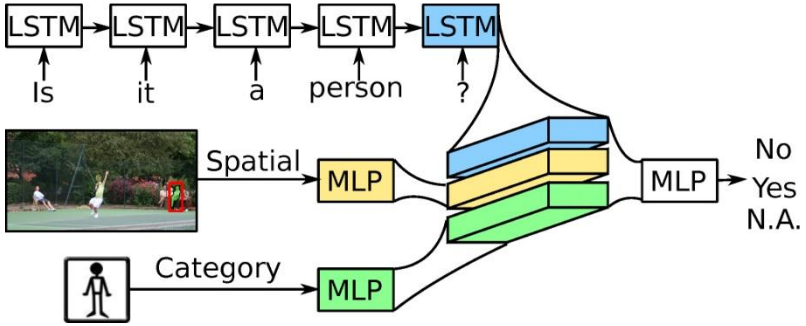
"I spy with my little eye ..."

Image	Policy Gradient		Tempered Policy Gradient	
	Is it in left?	No	Is it a person?	No
	Is it in front?	No	Is it a **vehicle**?	Yes
	Is it in right?	Yes	Is it a **truck**?	Yes
	Is it in middle?	Yes	Is it in front of photo?	No
	Is it person?	No	In the left half?	No
	Is it ball?	No	In the middle of photo?	Yes
	Is it bat?	No	Is it to the right photo?	Yes
	Is it **car**?	Yes	Is it in the middle of photo?	Yes
	Status:	Failure	Status:	Success
	Is it in **left**?	No	Is it a giraffe?	Yes
	Is it in front?	Yes	In front of photo?	Yes
	Is it in right?	No	In the **left half**?	Yes
	Is it in middle?	Yes	Is it in the middle of photo?	Yes
	Is it person?	No	Is it **to the left** of photo?	Yes
	Is it giraffe?	Yes	Is it to the right photo?	No
	Is in middle?	Yes	**In the left** in photo?	No
	Is in middle?	Yes	In the middle of photo?	Yes
	Status:	Failure	Status:	Success

Convolutional DL
+ Recurrent DL
+ Reinforcement DL

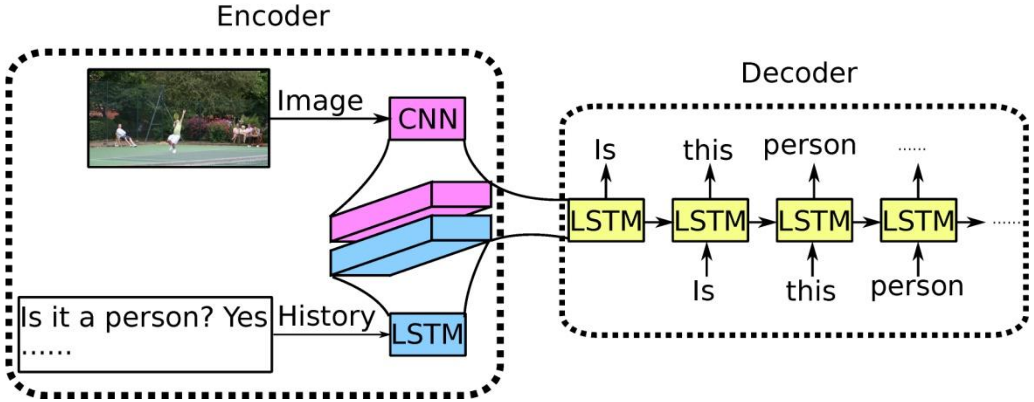
Talents, Talents Talents!

The Oracle Model

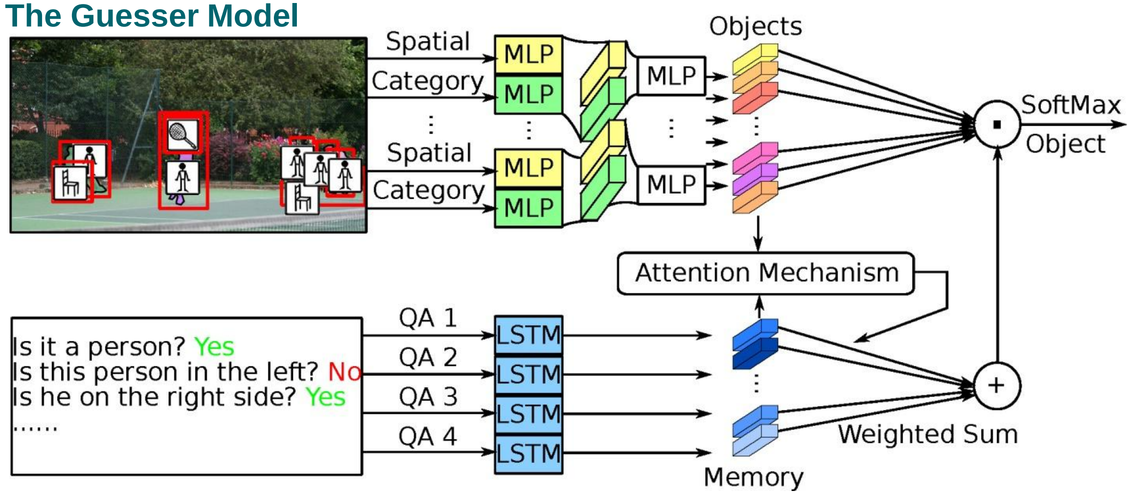


Rui Zhao, 2018

The Question-Generator Model



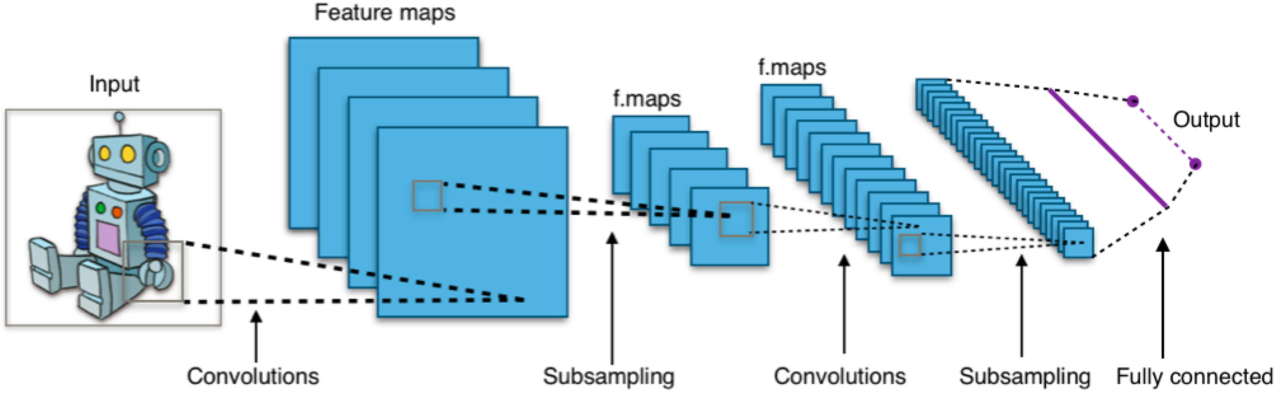
The Guesser Model



Deep X Technologies behind Artificial Intelligence

- **Deep Learning**; Machine Learning; Data Mining; Statistics
  - More (Labeled) Data
  - Deeper Models
  - New Algorithms
  - End-to-End Training; Differentiable-Computing (no Feature Engineering)
  - Computational Power
  - Community

- **Deep Knowledge: Facts and Models**
  - Huge Document Repositories with Rapid IE / QA (IBM Watson)
  - Maps with GPS for Autonomous Driving
  - Ubiquitous IoT and Big Data in Industry
  - Detailed (Patient) Profiles
  - Web Content, Wikipedia for Humans
  - **Knowledge Graphs for Machines**



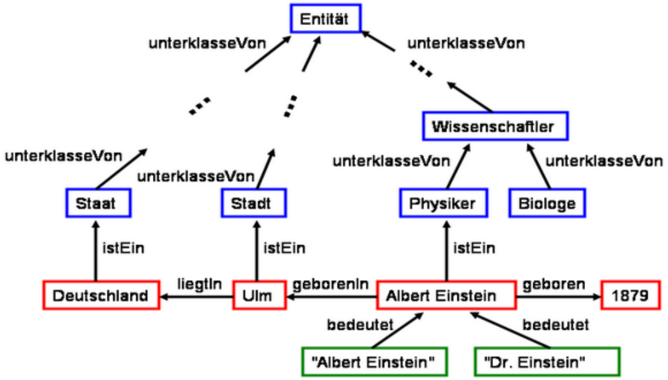
Deep X: Laplace's Demon

- Laplace's demon was the first published articulation of causal or scientific determinism (Pierre-Simon Laplace, 1814)
- According to determinism, if someone (the Demon) knows the precise location and momentum of every atom in the universe ***(Deep Facts),*** their past and future values for any given time are entailed; they can be calculated from the laws of classical mechanics ***(Deep Laws, Deep Learning, Deep Insights and Deep Models)***

Deep Knowledge: Knowledge Graphs

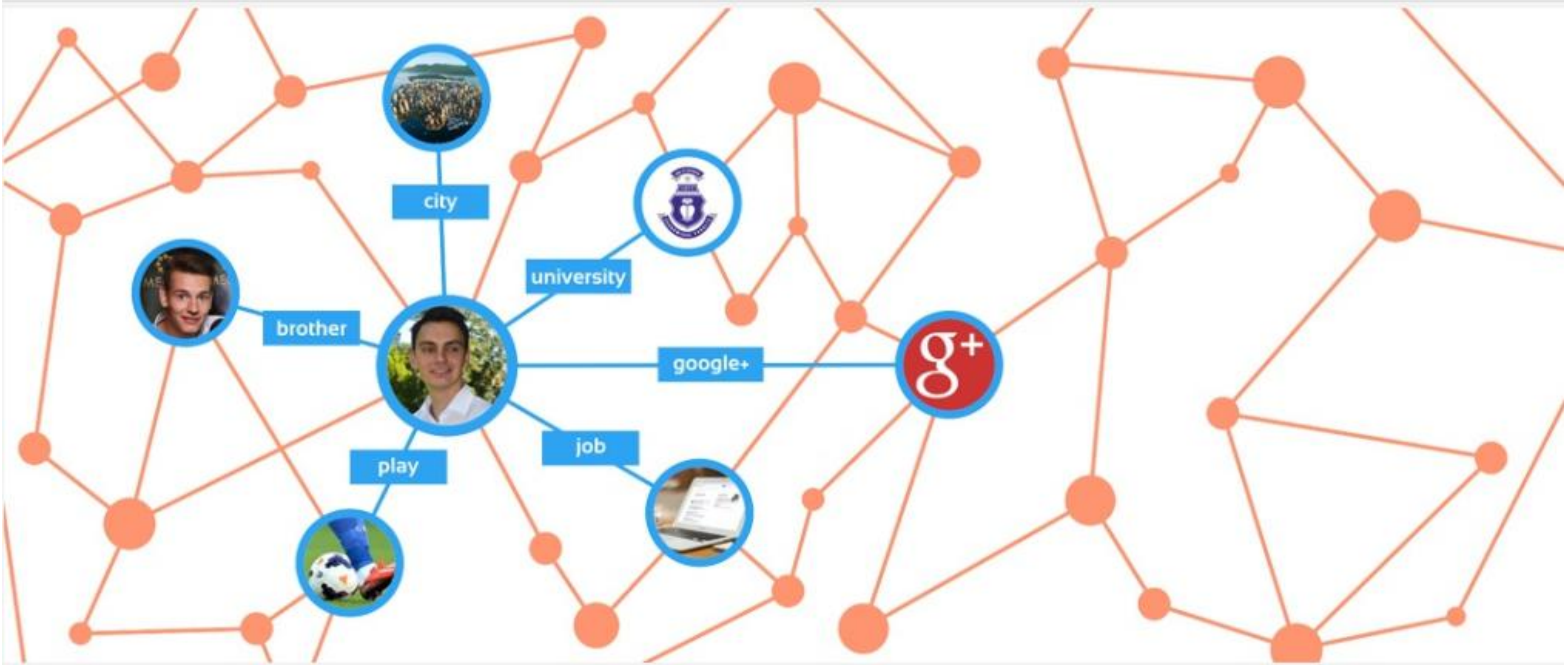
- The Google Knowledge Graph is a major break through in the field of Knowledge Representation
  - Scalability: >100B fact
  - Reliability: >99% fidelity
  - Maintainability
  - Usefulness: Search, Q&A, text understanding
- The basis for the Google Knowledge Graph are facts
- *Growing interest across industries*
- *Well suited for information integration (easier than RDBs)*

>100B Facts



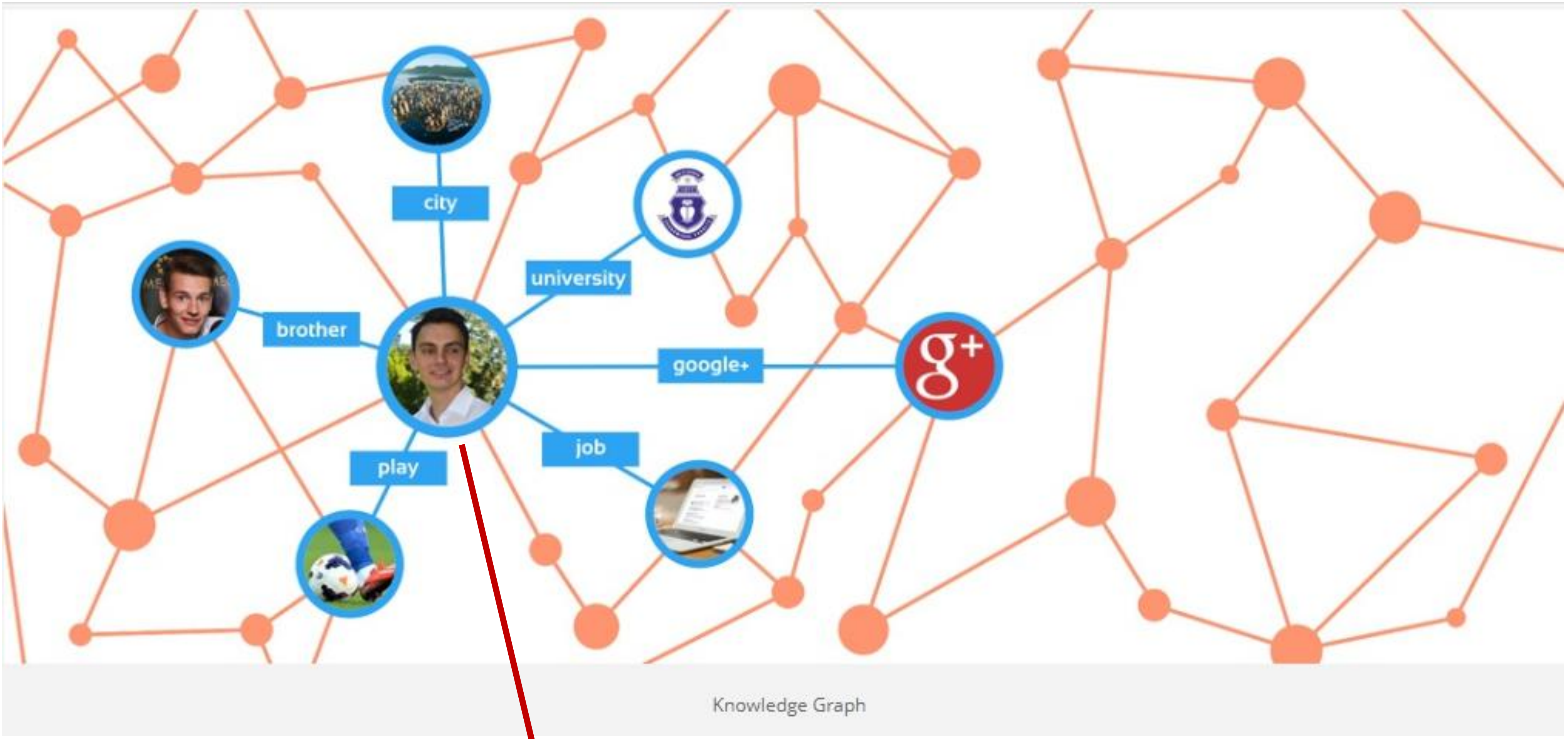
Singhal. Introducing the Knowledge Graph: things, not strings. Official Google Blog, 2012

A Knowledge Graph for Structured Information: Relationships and not Just Attributes!



Knowledge Graph

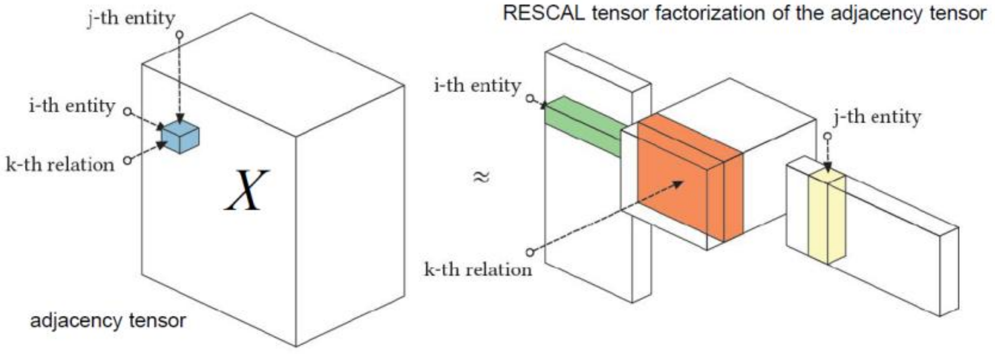
Links to Unstructured Information



web pages, texts, videos, images

Machine Learning with Knowledge Graphs

- The Knowledge Graph *stores* Knowledge: Can we *learn and generalize* from stored knowledge?
- The RESCAL model is based on an approximation of the Knowledge Graph adjacency tensor
- It was the basis for further research in our group, but also other groups



Training Data:

$x_{s,p,o} = 1$ If *(s,p,o)* is known to be true

$x_{s,p,o} = 0$ otherwise

After factorization (RESCAL2: constr. Tucker2):

$$P((s,p,o)) = \mathrm{sig}(\theta_{s,p,o})$$

$$\theta_{s,p,o} = \sum_{r_1} \sum_{r_3} a_{e_s,r_1} a_{e_o,r_3} g(r_1, p, r_3)$$

$$\Theta = G \times_1 A \times_2 A$$

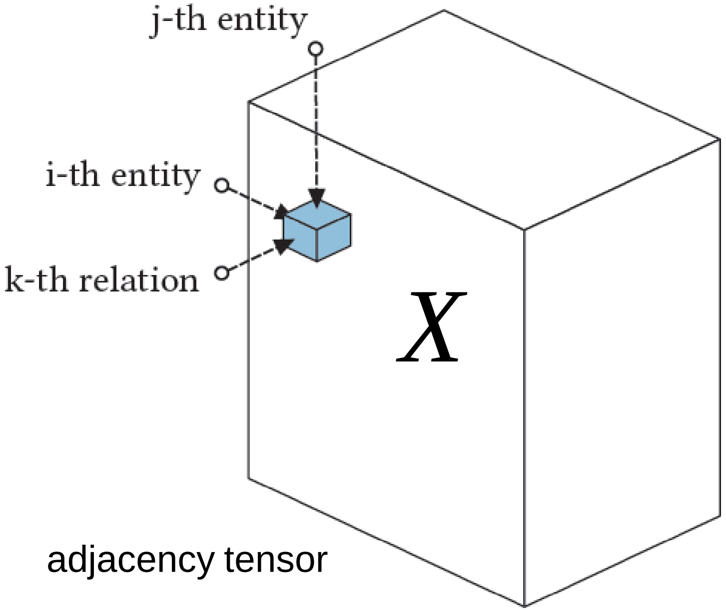
- Inferential queries
  - What disease does Jack likely have?
- Automatic filling of KG
  - Knowledge Vault projects
- KG priors to understand texts and images
- Detection of KG errors
- Learning Database
- Use as background information (compressed as latent factors) that can be used in other applications (predictions, decision support)

Tresp, et al. Materializing and querying learned knowledge. IRMLeS, 2009

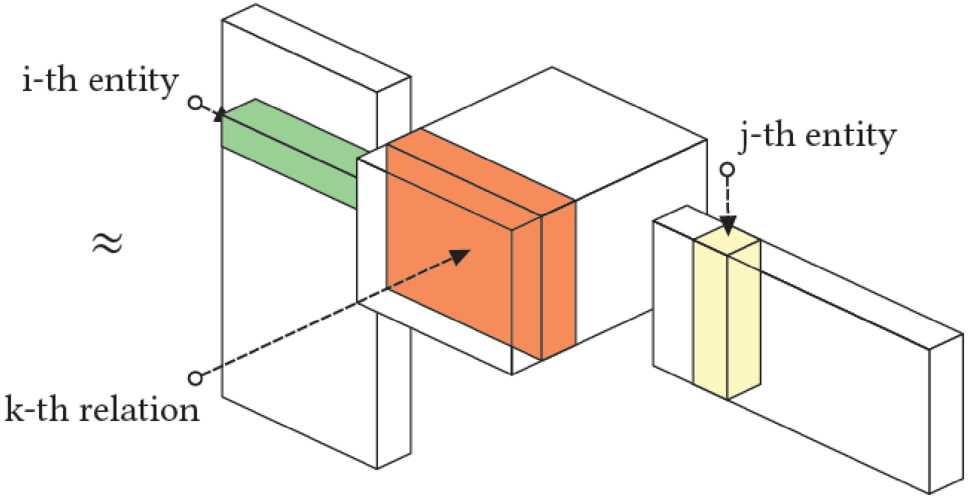
Nickel, et al. A Three-Way Model for Collective Learning. ICML, 2011

Nickel, et al. A review of relational machine learning for knowledge graphs. Proceedings of the IEEE, 2015

Machine Learning: Generalization via Tensor Factorization



RESCAL tensor factorization of the adjacency tensor



Training Data:

$x_{s,p,o} = 1$ If *(s,p,o)* is known to be true

$x_{s,p,o} = 0$ otherwise

After factorization (RESCAL: constr. Tucker2):

$$P((s,p,o)) = \mathrm{sig}(\theta_{s,p,o})$$

$$\theta_{s,p,o} = \sum_{r_1} \sum_{r_3} a_{e_s,r_1}\, a_{e_o,r_3}\, g(r_1, p, r_3)$$

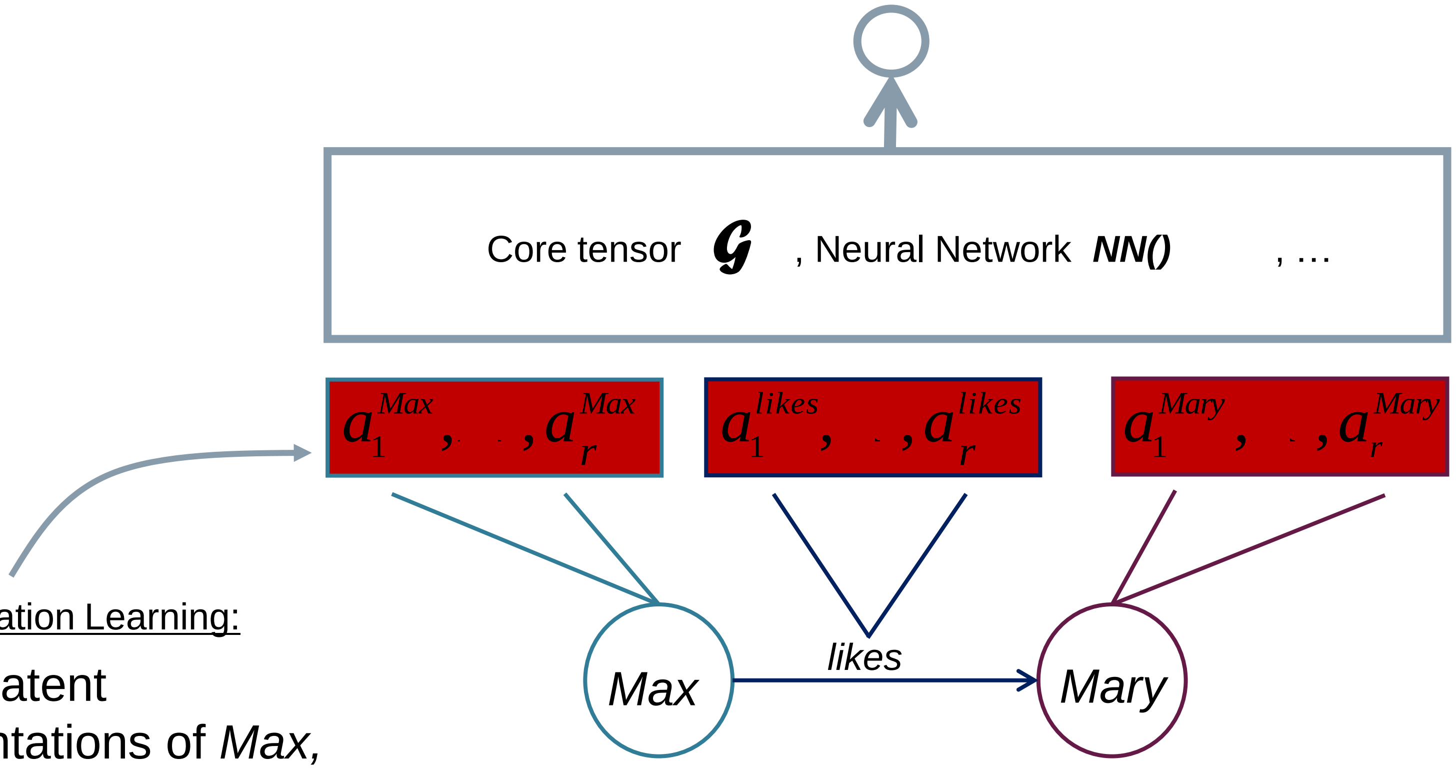
$$\Theta = \mathcal{G} \times_1 A \times_2 A$$

Nickel, Tresp, Kriegel. A Three-Way Model for Collective Learning on Multi-Relational Data. ICML 2011

Tensor Factorization as Representation Learning

- We maintain that an adjacency tensor is the appropriate representation
- Different forms of representation learning

$$P(\quad (Max, likes, Mary) = true \quad)$$



Representation Learning:

Shared latent representations of *Max, likes, and Mary*

We are able to predict all typed links (> 100 types) between several million of nodes!

Families of approaches:

- SUNS, RESCAL, DistMult, CompleEx,
- HolE, TransE, multiway Neural Networks, Poincare Embeddings, Holistic Embedding, …
- Graph Convolutional Network, Graph SAGE, …
- …

Configuration Recommendation System

SIEMENS
Ingenuity for life

Historical data

Contains information about 35,888 previously configured (anonymized) solutions containing 6,865 different items.

Technical features

Contains information about technical features of the items, such as voltage, size, weight, material, etc.

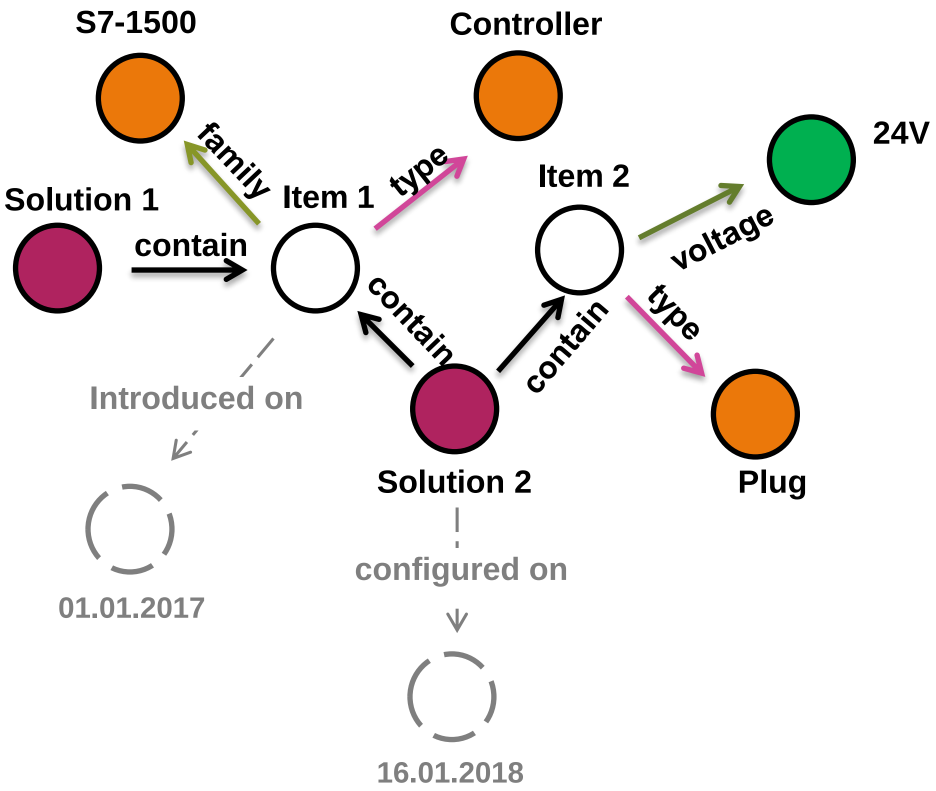
While most of the features are **numerical**, they belong to different scales: nominal, ordinal, interval, ratio.

Catalog data

Contains the information for categorization of the product.

Temporal data

Contains information about when a given solution was configured and when a given item was first introduced to the TIA Portal.



Hildebrandt, Shyam, Mogoreanu, Thon, Tresp, Runkler. Configuration of Industrial Automation Solutions Using Multi-relational Recommender Systems. ECML, 2018.

Deep Knowledge in Healthcare

Detailed Information about each individual patient (more dimensions; over time)		Precision medicine ▪ Information overload! ▪ Need for IT support and automation
Information about many patients (more instances)		Learning healthcare system ▪ Descriptive A. (what has happened?) ▪ Diagnostic A. (why? Insight!) ▪ Predictive A. (what will happen?) ▪ Prescriptive A. (what should be done?)

Tresp, Overhage, Bundschus, Rabizadeh, Fasching, Yu. Going Digital: A Survey on Digitalization and Large Scale Data Analytics in Healthcare, Proceedings of the IEEE, 2016.

Clinical Deep Knowledge in the Research Project: *Klinische Datenintelligenz*

- EHR: structured information
- Annotations of unstructured information

- Terminologies, Ontologies, Semantics
- Guidelines

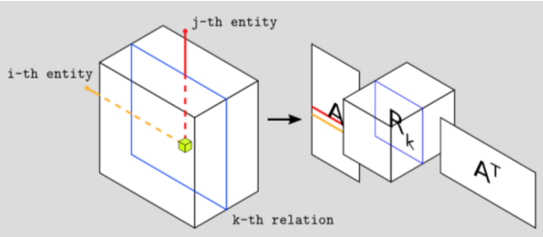
- Medical reports
- Literature

- Molecular Data
- Omics, Genetics
- Expression data

- Radiology
- Pathology

Research Database (I2B2; tranSMART, OMOP)

Statistical model of the research database:
Analysis, Prediction, Prescription



Sonntag, Tresp, et al., The Clinical Data Intelligence Project. Informatik-Spektrum, 2016.

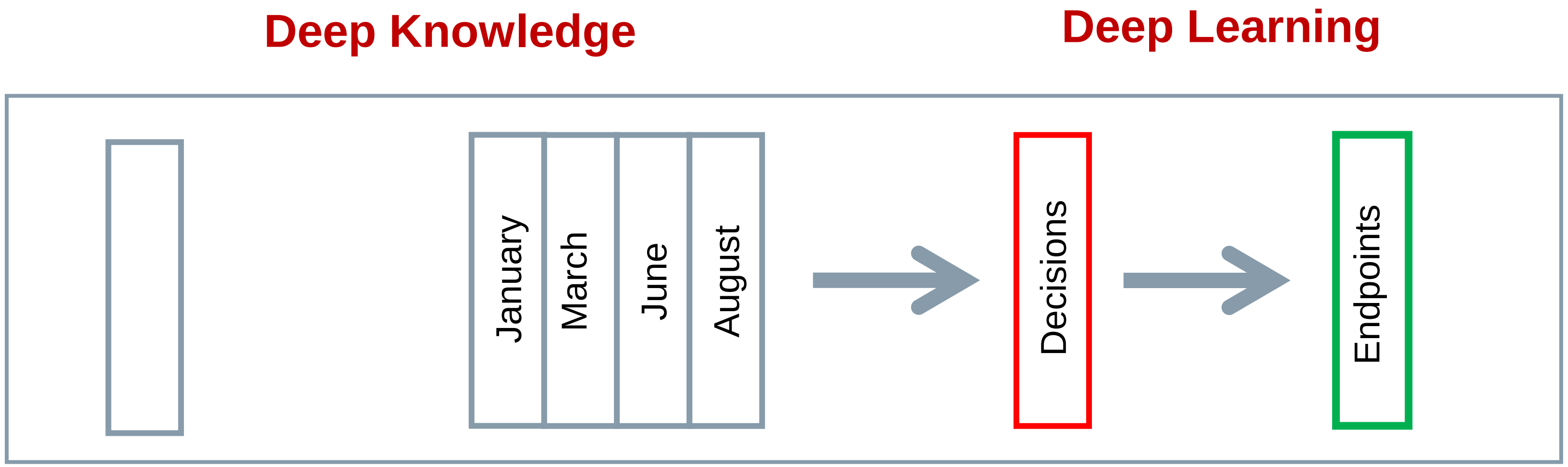
Deep Medical Knowledge and Deep Learning

Principle: Decision Modeling

- „The knowledge of the physicians --- including years of training, experience, publications they read--- is only relevant in as much as it influences medical decisions”
- And it is reflected in their decisions!

Principle: Endpoint Prediction

- Endpoint prediction can serve many purposes
- Decision support: „Propose decisions which are optimal under the predictive model to reach best end points“



Background (sKG)

- Age, gender, preconditions, …, primary tumor, history of metastasis before the study

Sequential (eKG)

- measurements, decisions

Recurrent Deep Learning

Endpoint

- Progression-free survival

Deep Learning and Knowledge Graphs for Decision Modeling

Synergies

Smart Data

BMWi Smart Data Project:
“Clinical Data Intelligence”

Semantic Knowledge Graph	Episodic Knowledge Graph	Unstructured Data
▪What do I know about the patient?	▪What happened to the patient?	▪Medical Reports ▪Radiology ▪Genetics

sKG, eKG,

LSTM, ...

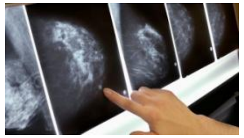
- Decision Modeling
- Endpoint Prediction

By disagreement: Physicians find our proposed decisions better than the ones of their peers

	Acceptable alternative	Don‘t agree	Don‘t agree at all
Re-T-Board	11%	64%	23%
ML	33%	58%	8%

Esteban et al. IEEE ICHI 2015 / Yang et al. IEEE ICHI 2017

Rohm et al., Ophthalmology, 2018



Universitätsklinikum Erlangen

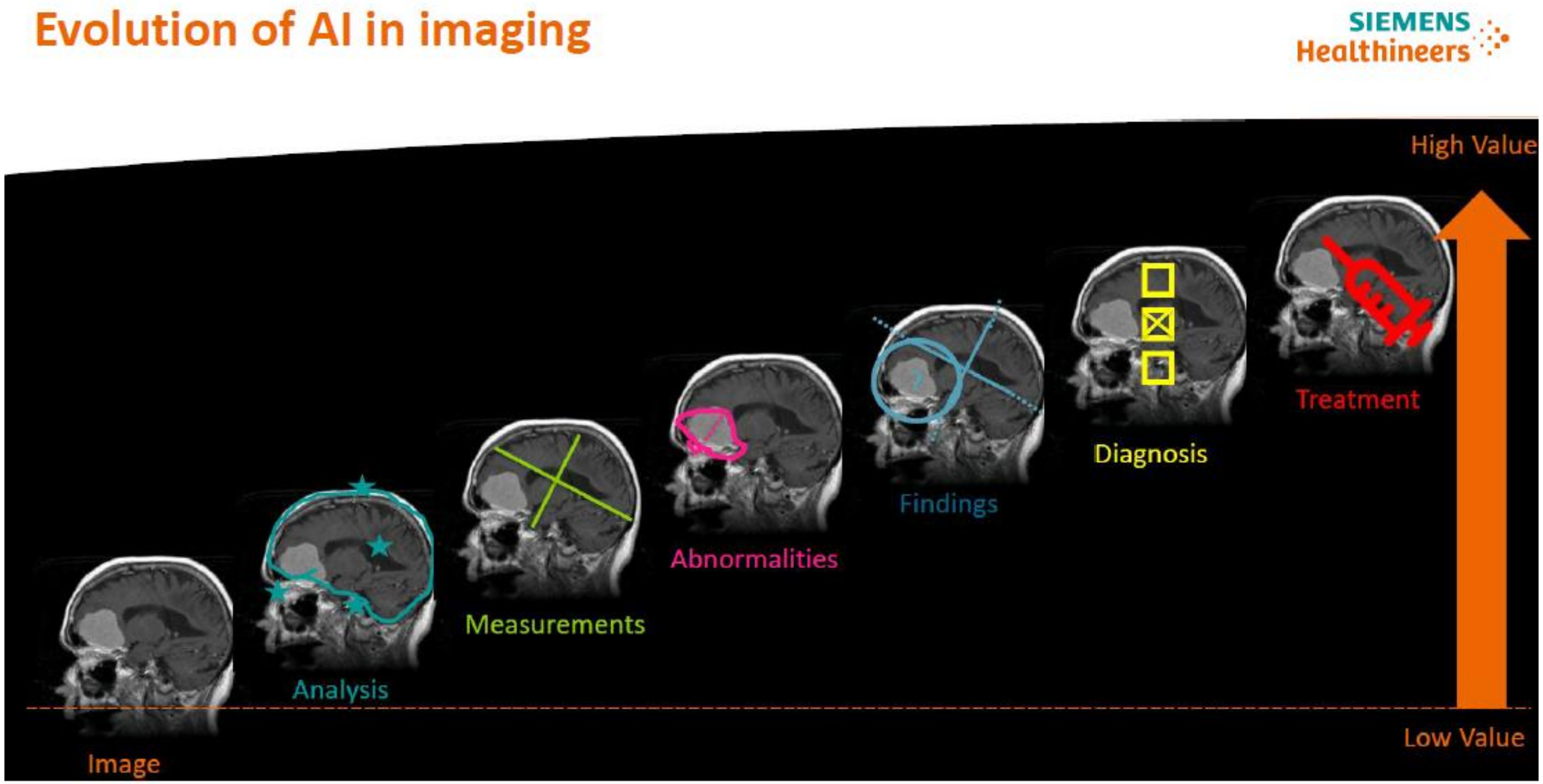
CHARITÉ UNIVERSITÄTSMEDIZIN BERLIN

LMU KLINIKUM DER UNIVERSITÄT MÜNCHEN

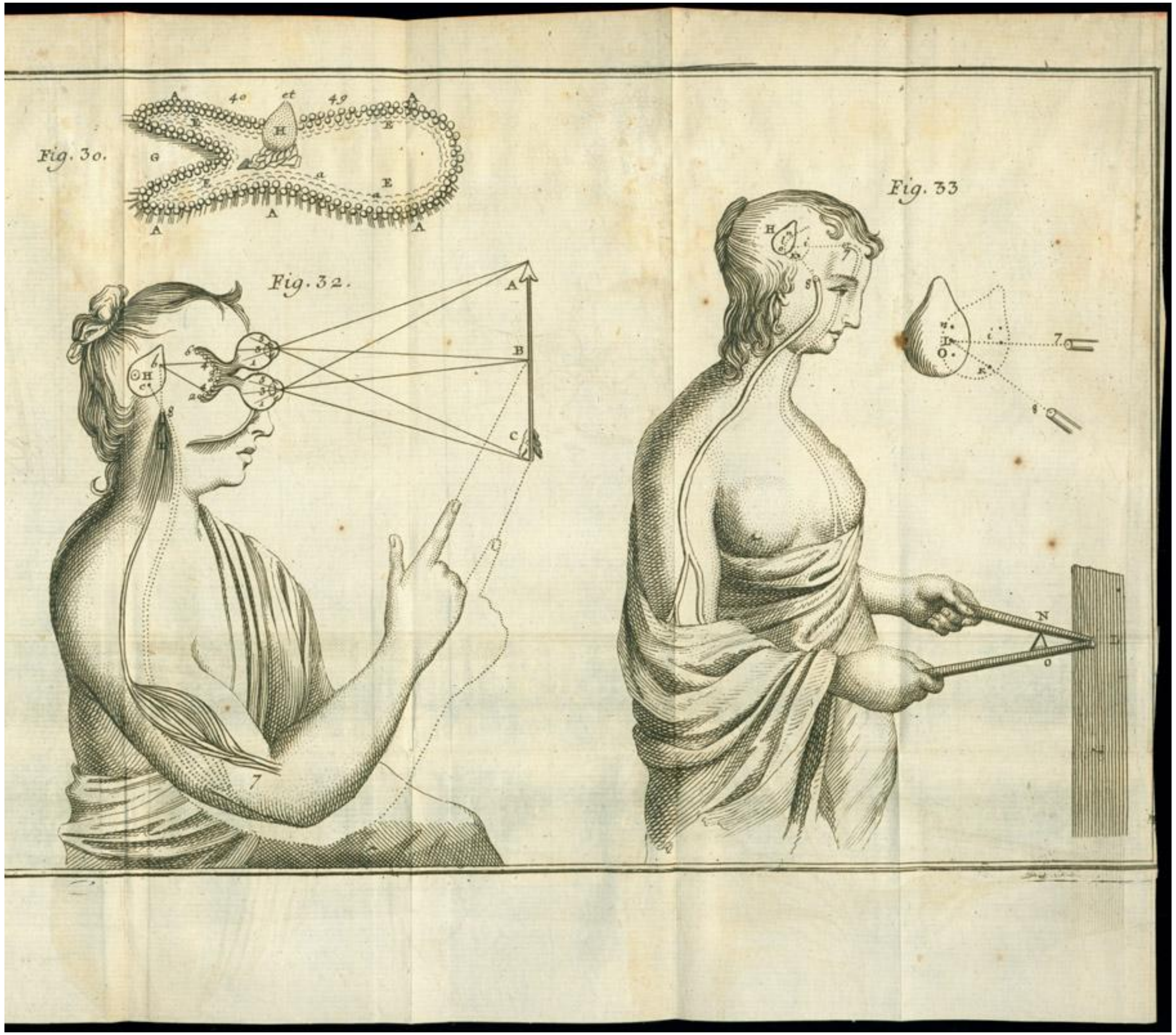
age-related macular degeneration

Another Line of AI Research at Siemens: Radiology

Increasing relevance for Siemens Healthcare



Smart Perception:
Integrating Knowledge Graphs with Deep Learning

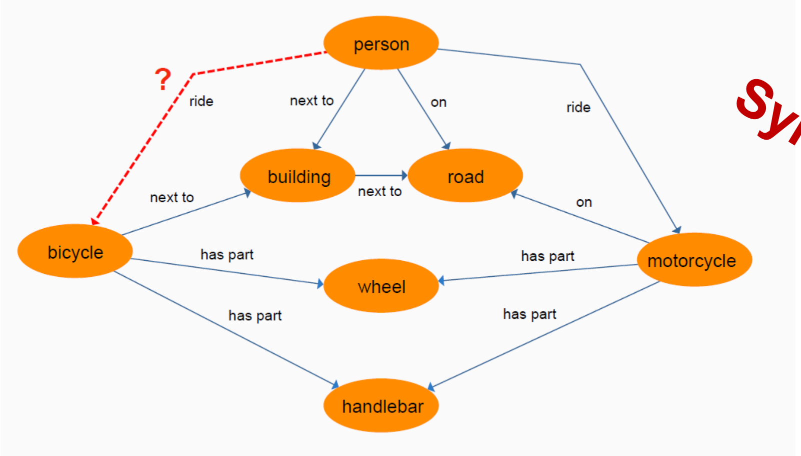


"Man sieht nur, was man weiß“

“You only see what you know"

--Johann Wolfgang von Goethe

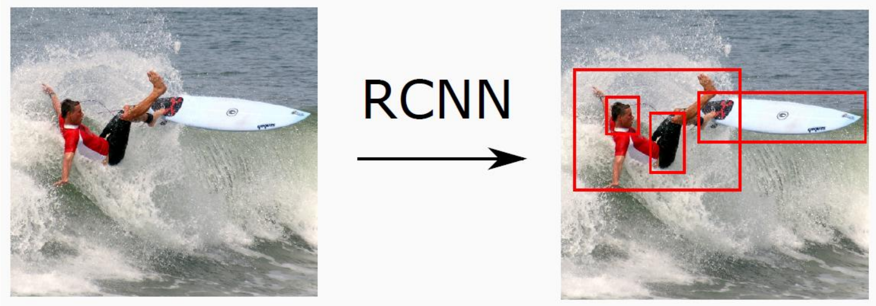
Smart Perception: Integrating Knowledge Graphs with Deep Learning



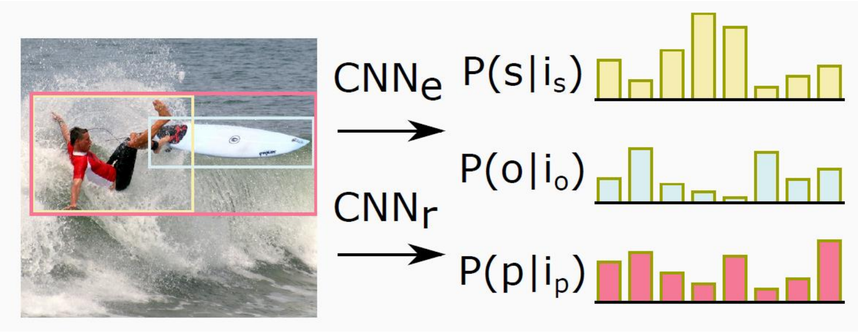
Synergies

Learning Knowledge Graph (with Generalization)

Deep Learning



Deep Learning



S = Person
P = nextTo
O= Surfboard

- By using a KG prior, we obtained better results than the Stanford group: *Lu, Krishna, Bernstein, Fei-Fei, 2016*
- ISWC 2017 best student paper; IJCAI 2018 Best Paper Track

Baier, et al. Improving Visual Relationship Detection using Semantic Modeling of Scene Descriptions, ISWC 2017

☺ person-next to-person

☺ truck-on-road

☺ person-next to-person

☺ truck-on-road

☹ lamp-on-box

☹ motorcycle-has-wheel

Cognition: a Perspective for AI

- Learning from human cognition

Perception and Memory

Sensor Processing

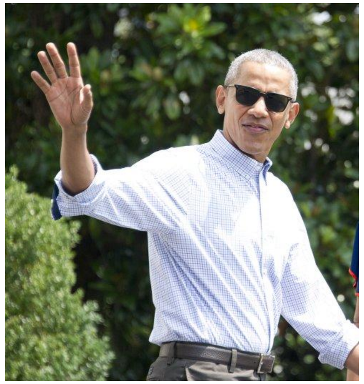
- Fast, skillful reaction
- Human declarative capabilities
  - Deep understanding of sensory inputs; **declarative decoding**; with a link to language

Episodic memory ("events we remember")

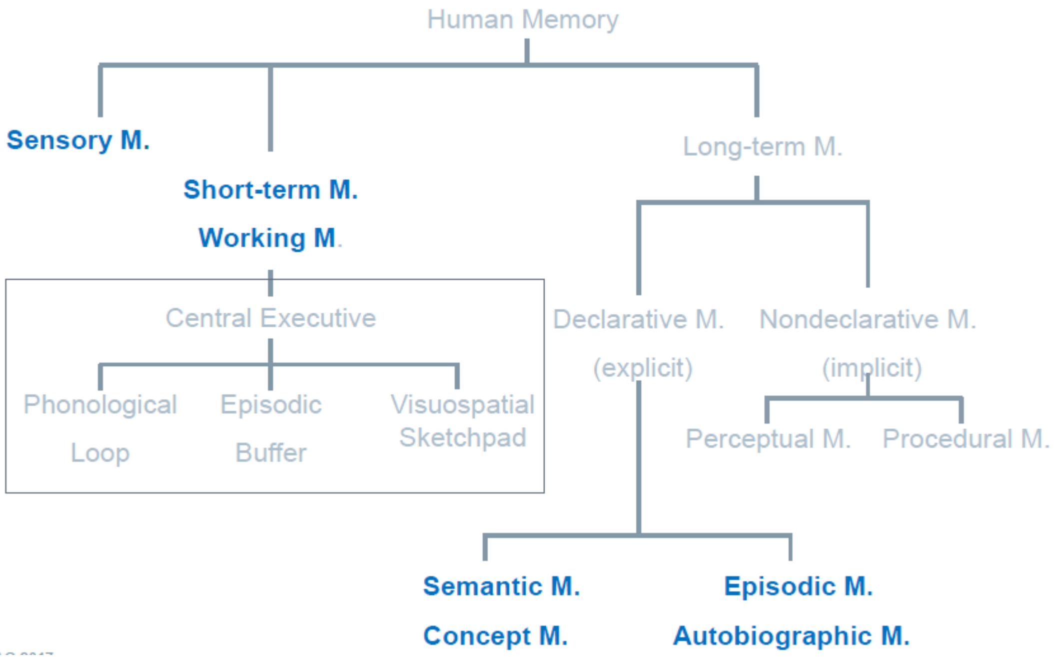
- Recall a sensory impression of past events
- Human **declarative memory**

Semantic memory ("facts we know")

- "Obama is ex-president of the United States"; "Munich is in Bavaria"
- Human **declarative memory**



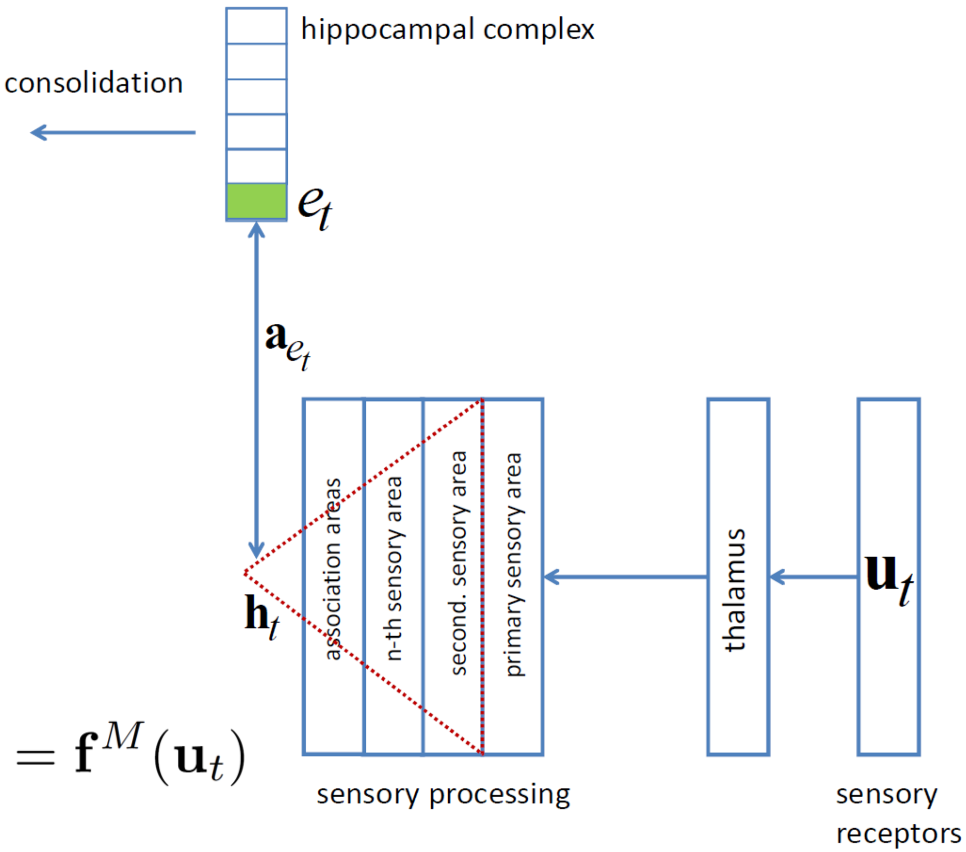
More: decisions; prediction; reasoning; action; learning from episodes, ….



Declarative!

Tresp, et al. Learning with memory embeddings. arXiv 2015 / Tresp, et al. The tensor memory hypothesis. NIPS Workshop 2016

Hippocampal Memory Indexing Theory: Representation Learning for Episodes



$$\mathbf{h}_t = \mathbf{f}^M(\mathbf{u}_t)$$

Teyler & DiScenna, 1986; Teyler & Rudy, 2007

- The latent representation for time is represented in the higher order layers of sensory processing / association cortex
- For meaningful episodes, an index for the time instance is generated in the hippocampal area
- Engram: index & representation $(e_t, \mathbf{a}_{e_t})$
- Recollection (internal stage, subsymbolic) by back projection: $(e_t, \mathbf{a}_{e_t})$ is reactivated, including the bound neocortical traces
- *But: not a model for explicit memory!*

Starting Point: Sensor Hierarchy

Sensory and other representations:

looks, sounds,

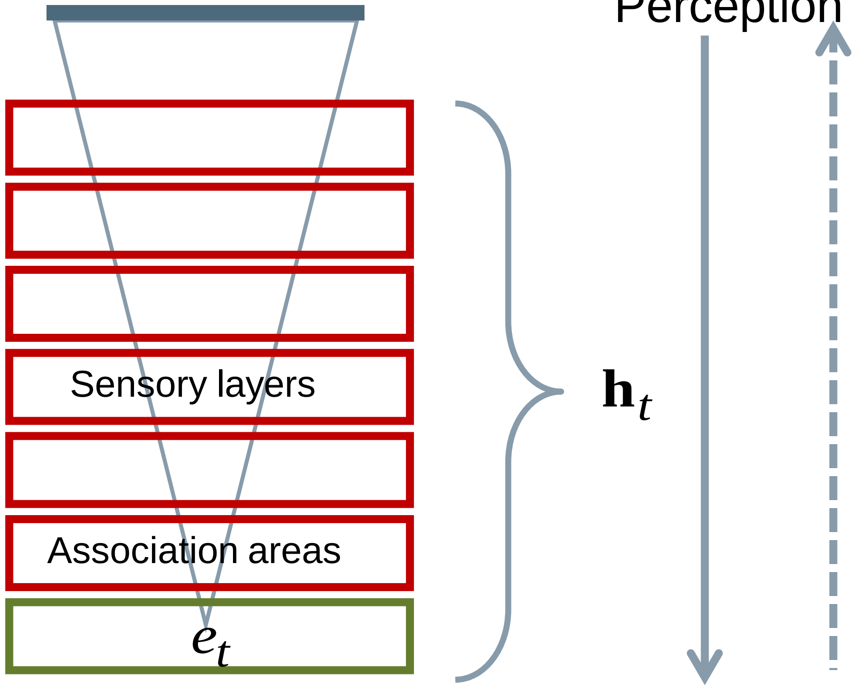
feels, …

Large overlap

(general)

Little overlap

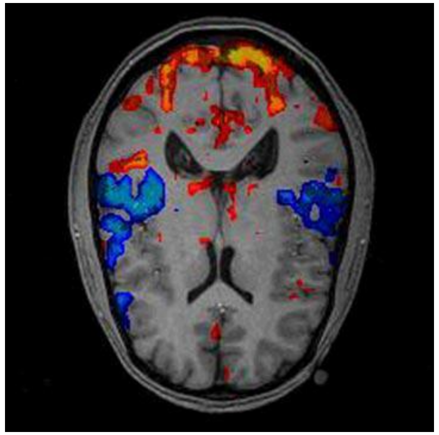
(specific)

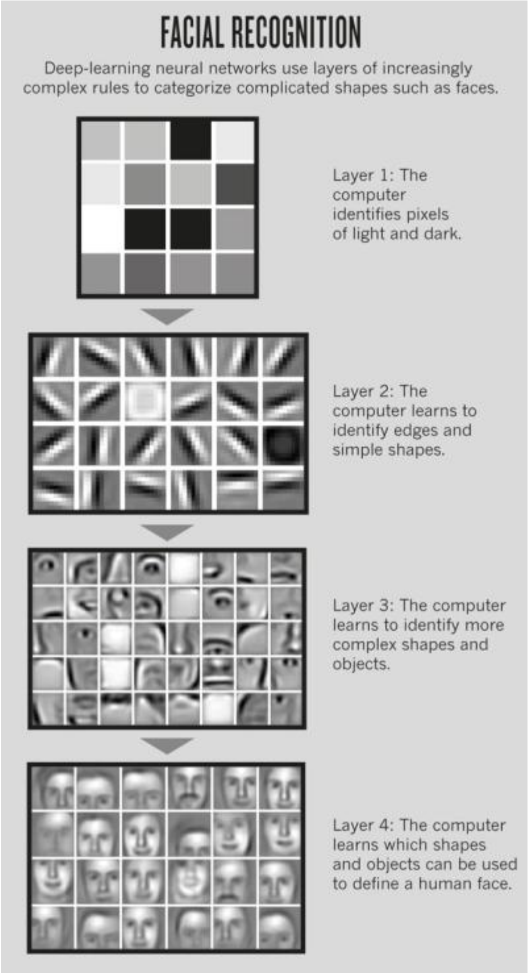


Perception

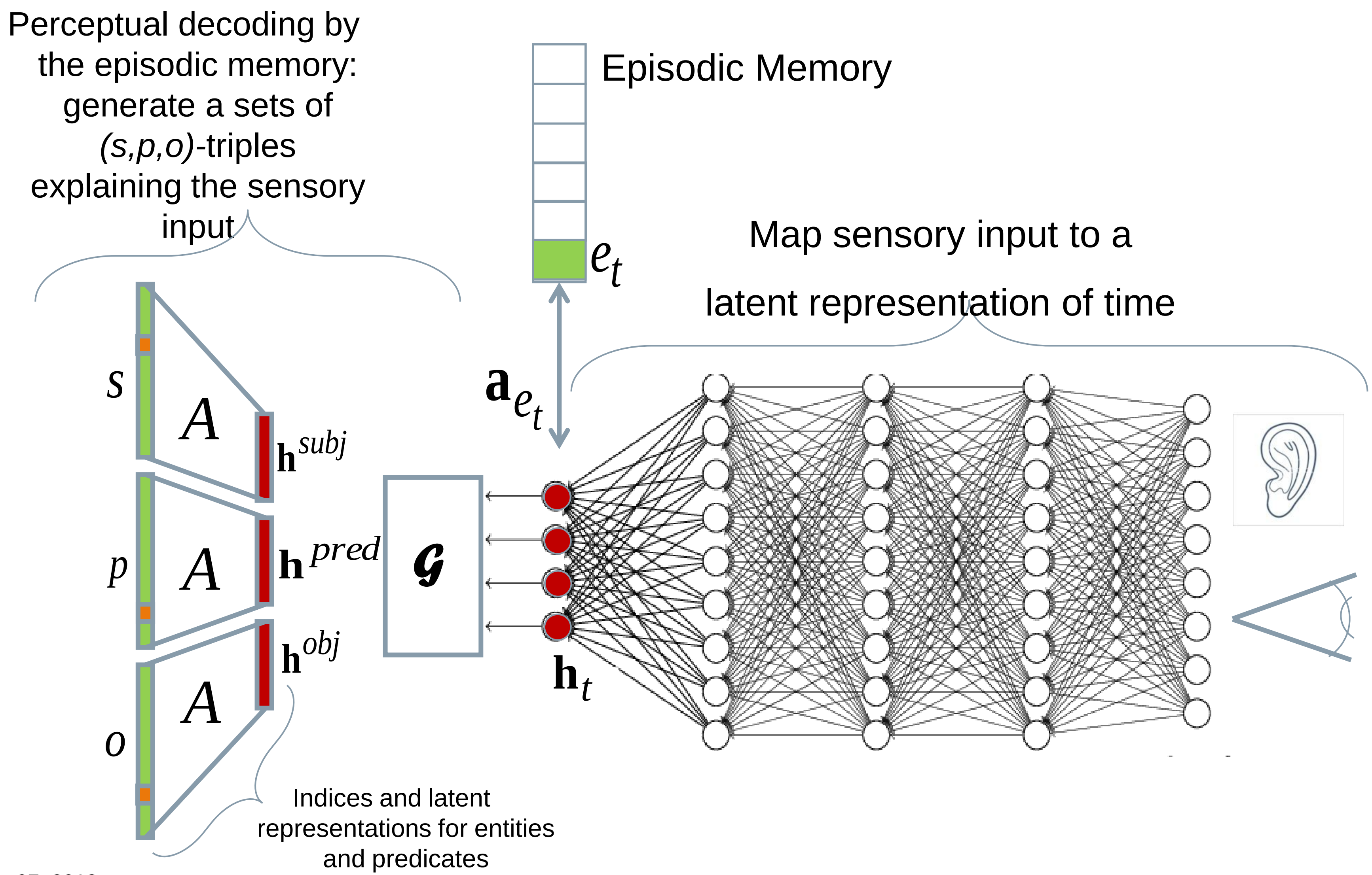
Index, new for each new memory

Recollection





All-In-One Hypothesis: Perception, Episodic Memory and Semantic Memory all Use the Same Functional Brain Modules

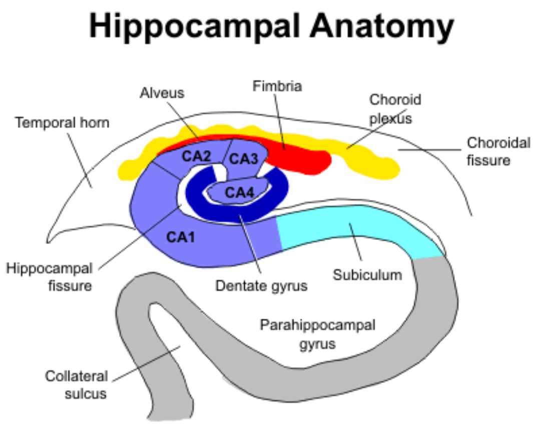


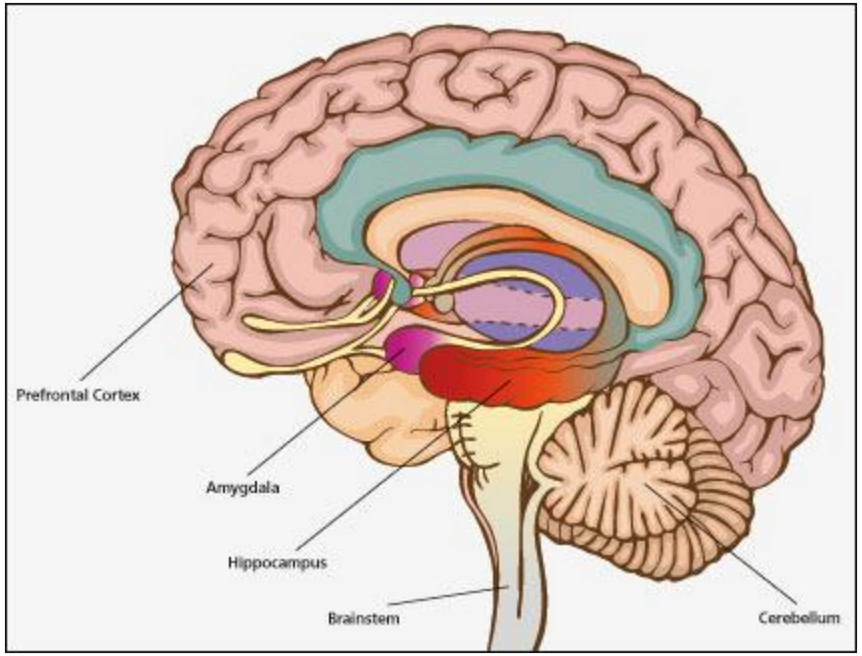
Hippocampus and MTL Play Significant Roles in the Generation of New Declarative Memories

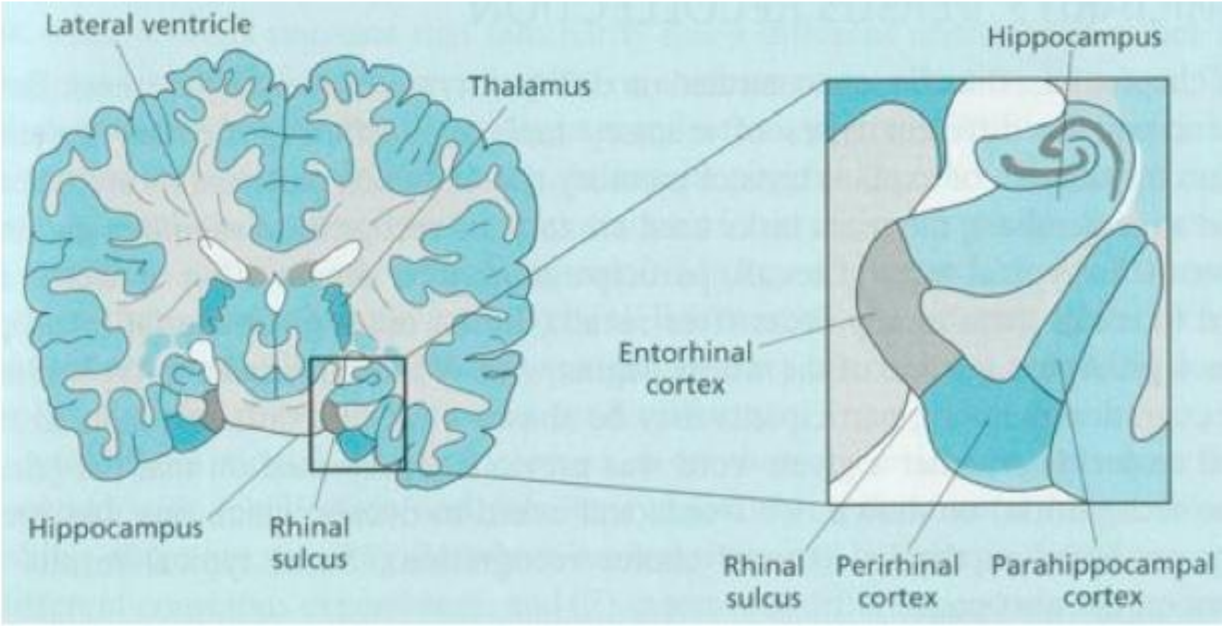
- *Hippocampus-dependent declarative memory*
- New memories are formed in the hippocampus/MTL

- Neurogenesis has been established in the **dentate gyrus** (part of the hippocampal formation) which is thought to contribute to the formation of **new episodic memories**

- Forming representations for new (significant)
  - **Episodes (time cells) (often)**
  - **Places (place cells) (often)**
  - **Entities (rare)**



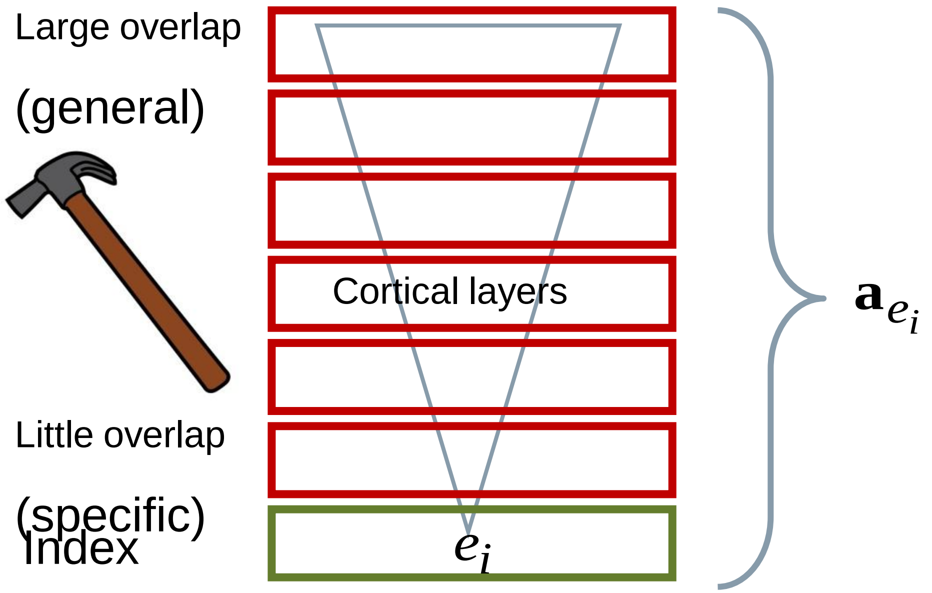


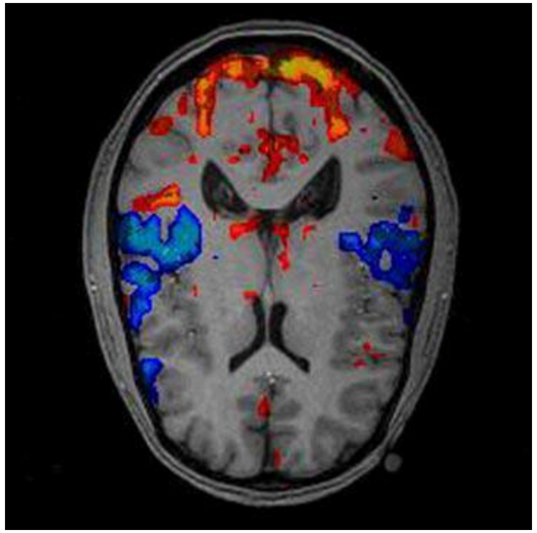


Representation of an Entity, a Concept, a Predicate

SIEMENS
Ingenuity for life

Hypothesis: in the same way that $\mathbf{h}_t$ / $\mathbf{a}_t$ represents the perception at time *t*, an entity e_i has a latent representation $\mathbf{a}_i$



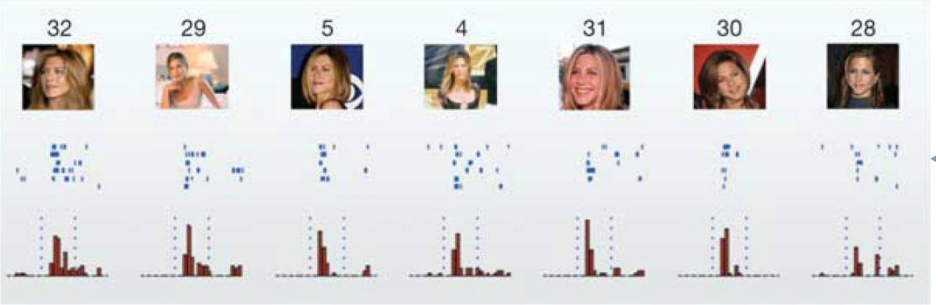


Max, hammer, Munich,
… …

Locality of Representations for Concepts

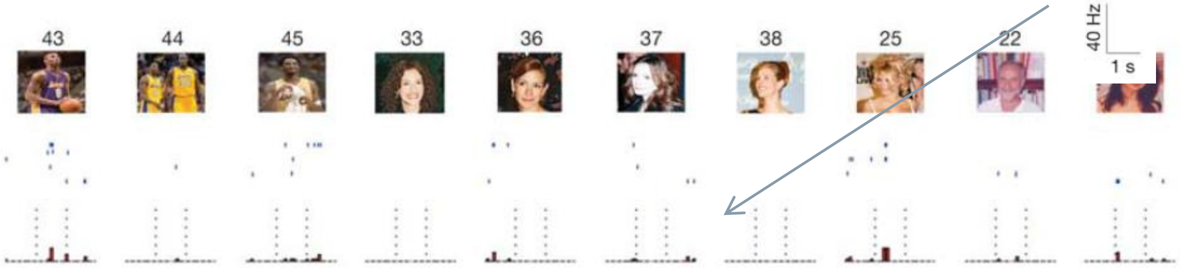
SIEMENS
Ingenuity for life

- Medial temporal lobe (MTL) neurons that are selectively activated by strikingly different **pictures of given individuals, landmarks or objects and in some cases even by letter strings with their names**
- “**Jennifer Aniston**“, „Halle Berry“ … concept cells



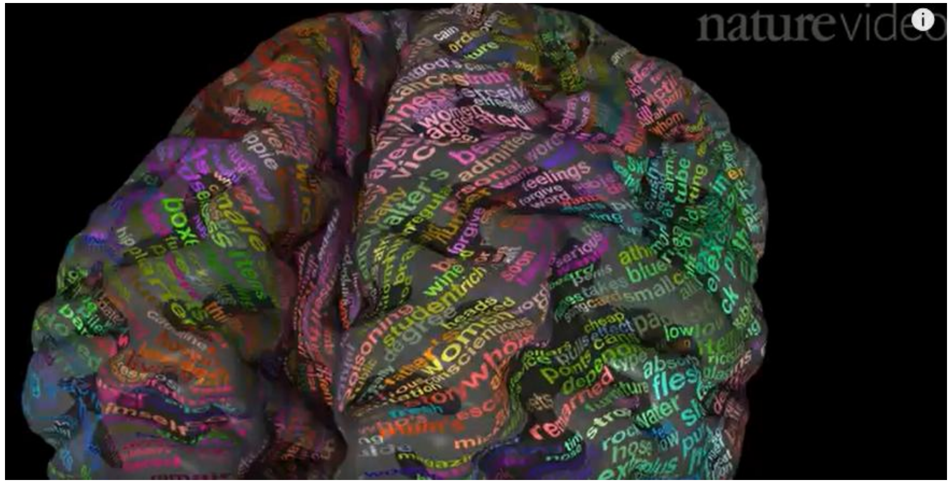
response

no response



Quiroga, Reddy, Kreiman, Koch, Fried. Invariant visual representation by single neurons in the human brain. Nature, 2005

Huth, de Heer, Griffiths, Theunissen, Gallant. Natural speech reveals the semantic maps that tile human cerebral cortex. Nature, 2016.



Semantic Memory from Episodic Memory?

SIEMENS
Ingenuity for life

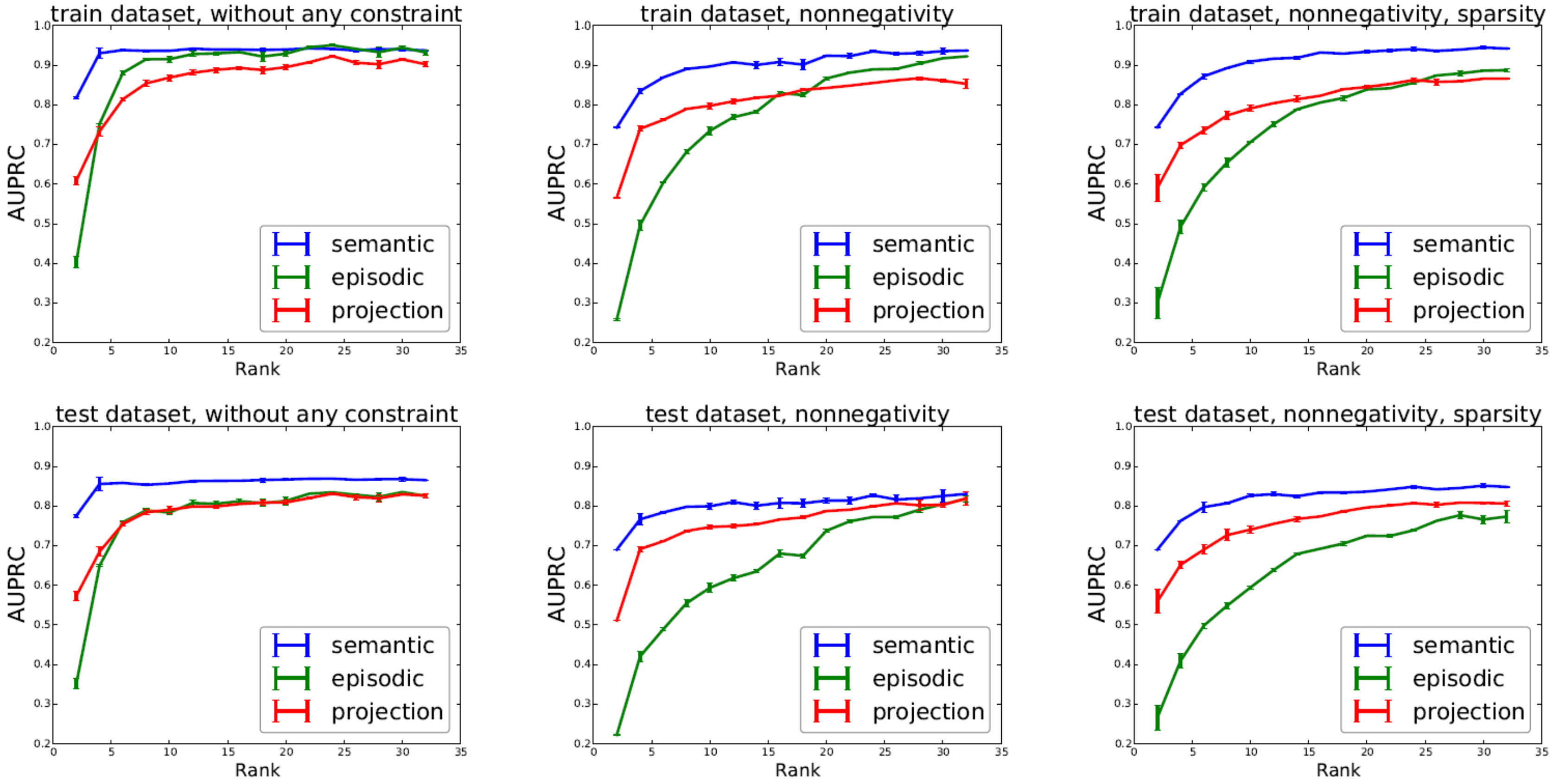


Fig. 3. AUPRC scores of the training and testing data sets for different model settings as a function of the rank.

Tresp, Ma, Baier, Yang. Embedding Learning for Declarative Memories. ESWC 2017

Memory Consolidation in the Real World

SIEMENS
Ingenuity for life

- No rapid eye movement!

Integrated Intelligence: Episodic Memory and Semantic Memory

Integrated Intelligence: "Man sieht nur, was man weiß“

Perception

(understand)

Integrated Intelligence: "Man sieht nur, was man weiß“

Episodic Memory

(remember)

Perception

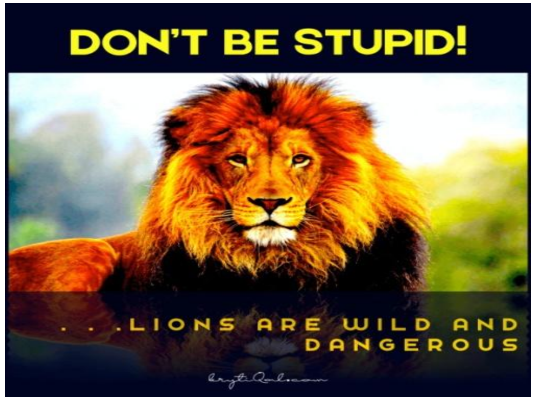
(understand)

Integrated Intelligence: "Man sieht nur, was man weiß“

Semantic Memory
(know)

Episodic Memory
(remember)

Perception
(understand)

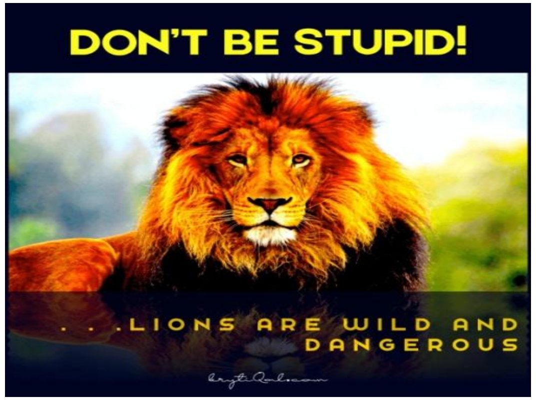


Integrated Intelligence: "Man sieht nur, was man weiß“

Semantic Memory
(know)

Episodic Memory
(remember)

Perception
(understand)



Integrated Intelligence: "Man sieht nur, was man weiß“

Declarative

Working Memory

- Prediction
- Decision Making, Reinforcement Learning
- Planning

Semantic Memory

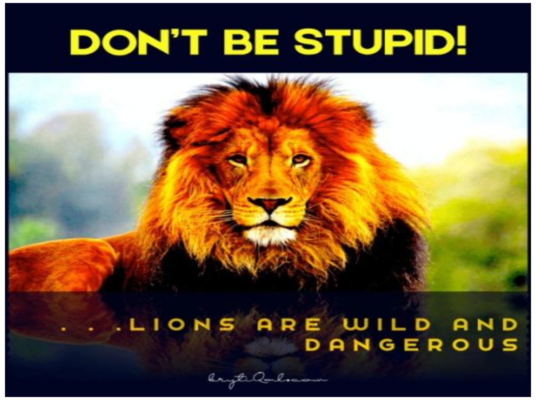
(know)

Episodic Memory

(remember)

Perception

(understand)



Integrated Intelligence: "Man sieht nur, was man weiß“

Working Memory

- Prediction
- Decision Making, Reinforcement Learning
- Planning

Semantic Memory

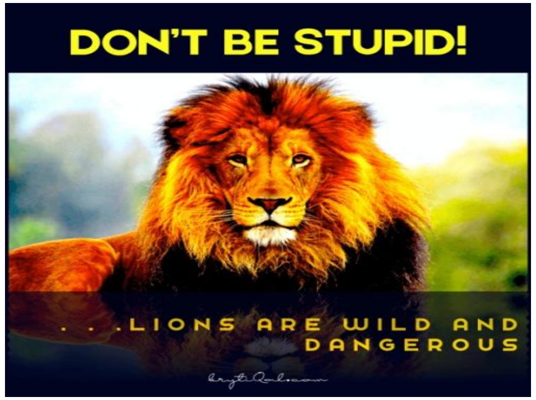
(know)

Episodic Memory

(remember)

Perception

(understand)

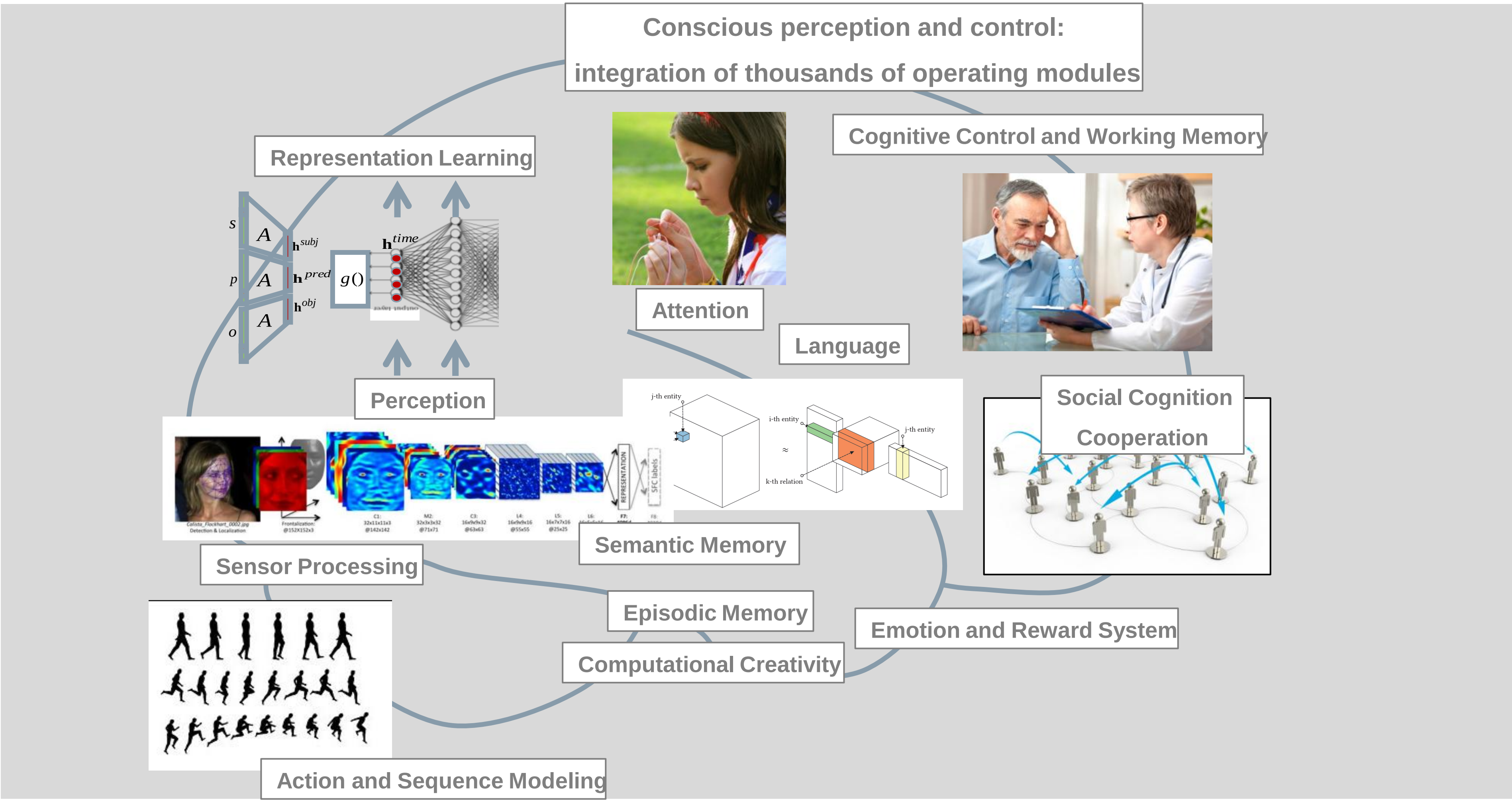


AI is Many Things

Synergies

Deep Learning with a Cognitive Perspective

Coordination of Thousands of Modules

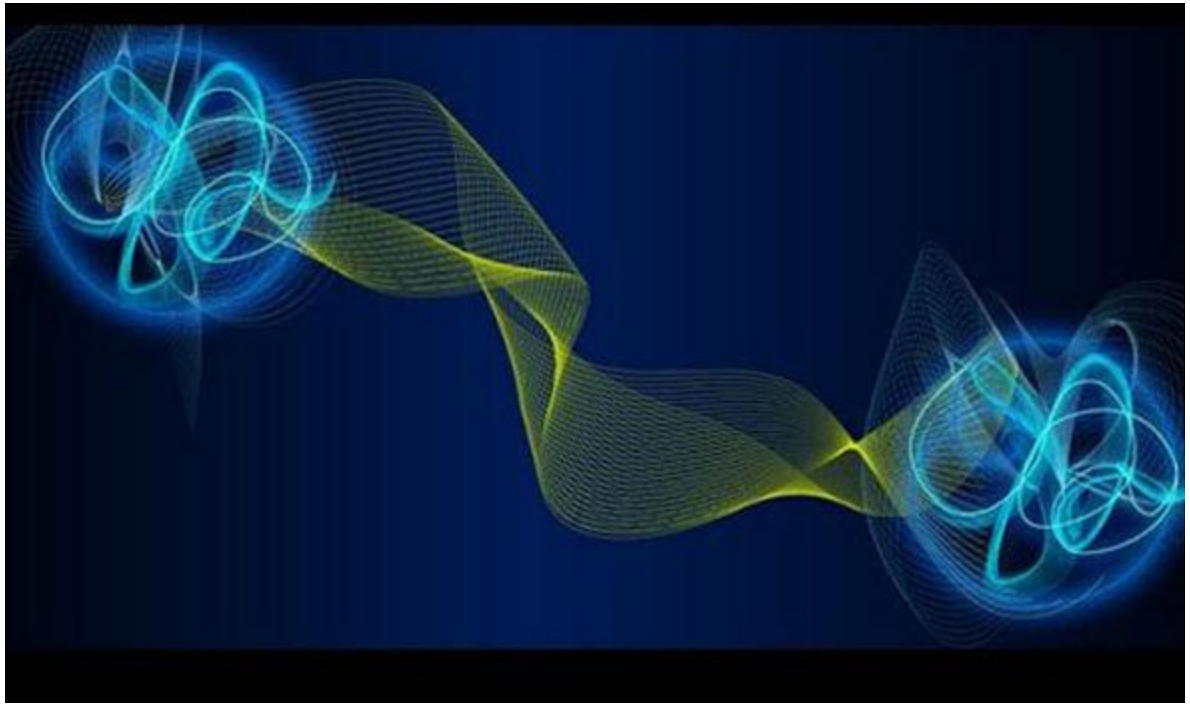


Gazzaniga. Organization of the human brain. Science, 245(4921), 1989.

The Future: Quantum Machine Learning?

- The quantum world computes
- …It always bothers me that, according to the laws as we understand them today, it takes **a computing machine an infinite number of logical operations** to figure out what goes on in **no matter how tiny a region of space, and no matter how tiny a region of time**. How can all that be going on in that tiny space?

  – *Richard Feynman*

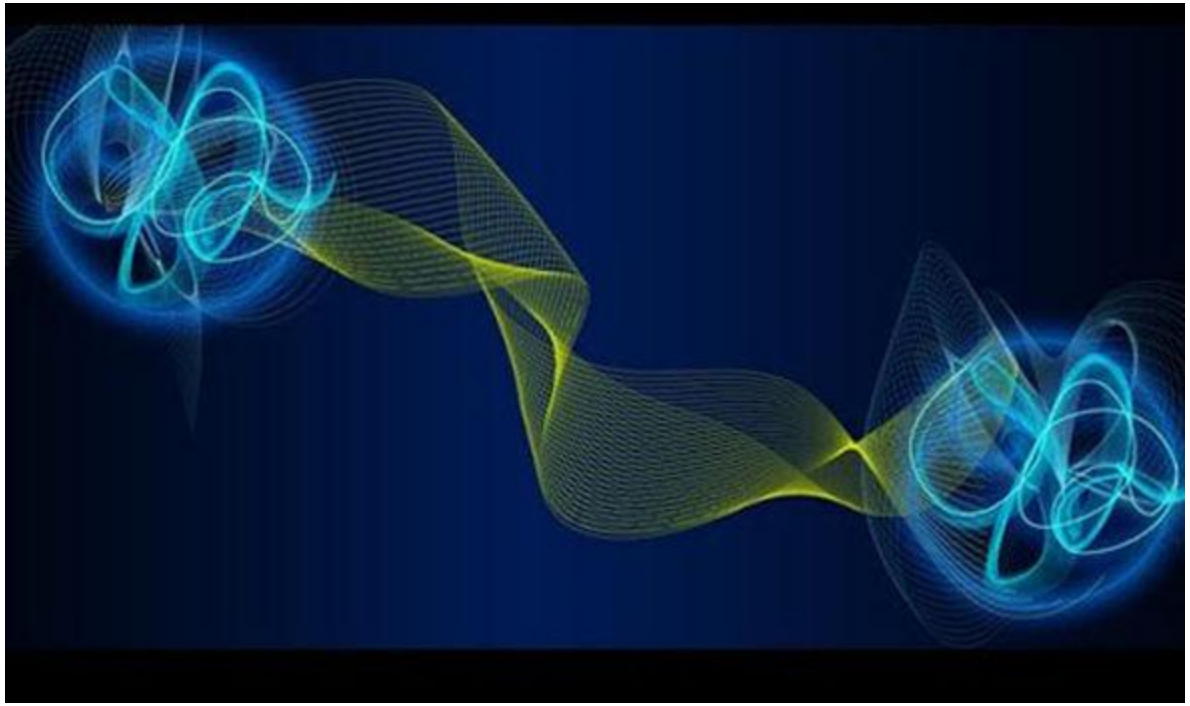


The Future: Quantum Machine Learning?

- The quantum world computes

- …It always bothers me that, according to the laws as we understand them today, it takes **a computing machine an infinite number of logical operations** to figure out what goes on in **no matter how tiny a region of space, and no matter how tiny a region of time**. How can all that be going on in that tiny space?

  – *Richard Feynman*



Scalable Quantum-Secured Blockchain

Quantum AI

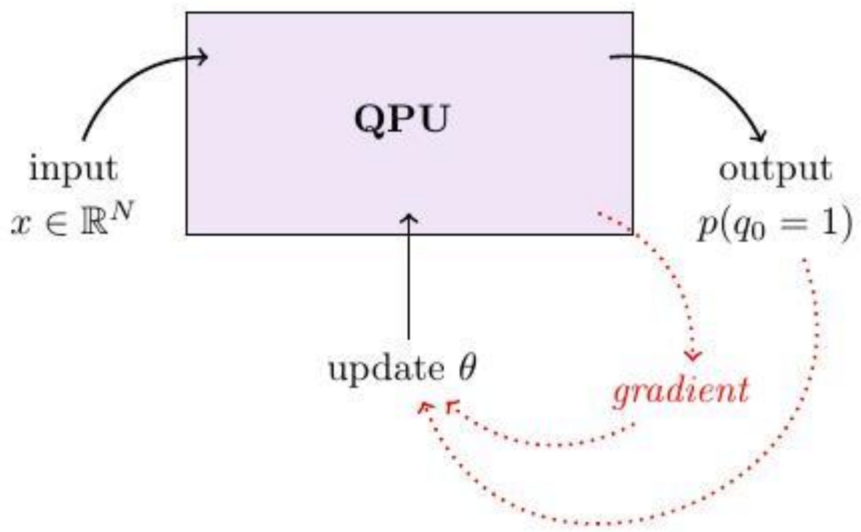
Quantum Computing for Knowledge Graphs

Modeling Knowledge Graphs with classical devices:

- Fast growing Knowledge Graphs
- Increasing number of entities
- Slow inductive inference

Modeling Knowledge Graphs with quantum devices:

- Quantum machine learning algorithms
- Noisy intermediate-scale quantum devices, e.g., 72-qubit-chip
- Speed-up inductive inference



Hybrid learning Method

- Combination of classical and quantum processing
    - Quantum units compute outputs
    - Classical units update parameters
- Prepare to be inferred entities in quantum state as superposition
- Simulation on GPUs
- Accelerated inference step $\mathcal{O}(N) \rightarrow \mathcal{O}(poly \log N)$

Thank you!

Volker Tresp

Distinguished Research Scientist at Siemens Corporate Technology and Professor at the Ludwig Maximilian University of Munich
volker.tresp@siemens.com