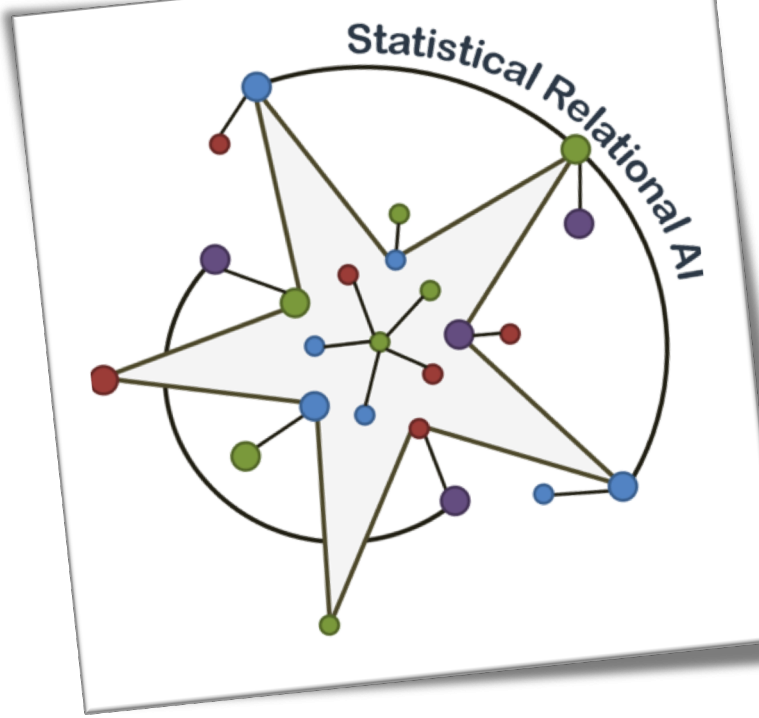
Wrap-up

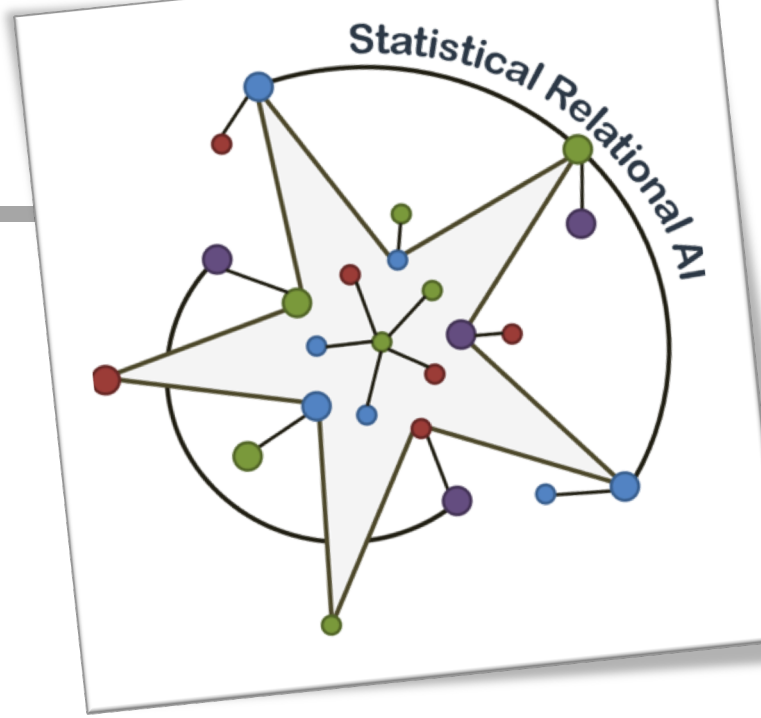
Statistical Relational AI

Tutorial at ICCS 2019



Tanya Braun and Marcel Gehrke, University of Lübeck

UNIVERSITÄT ZU LÜBECK

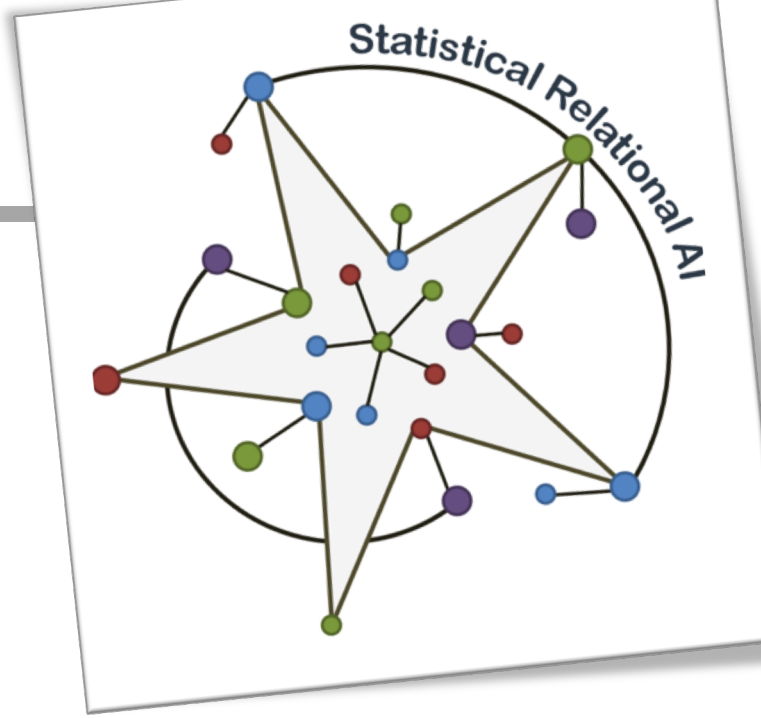


Mission for today

Providing an overview and an introduction into probabilistic inference with a focus on StaRAI

What have we seen?

- Probabilistic relational models
  - Treat objects identically if they are indistinguishable
  - Different flavours available
  - Compactly represent objects, relations, uncertainties
- Semantics
  - Grounding semantics
  - Goal: avoid grounding!
- Lifted Algorithms
  - Exact: LVE, LJT, LDJT
  - Lifted model from ground model: colour passing

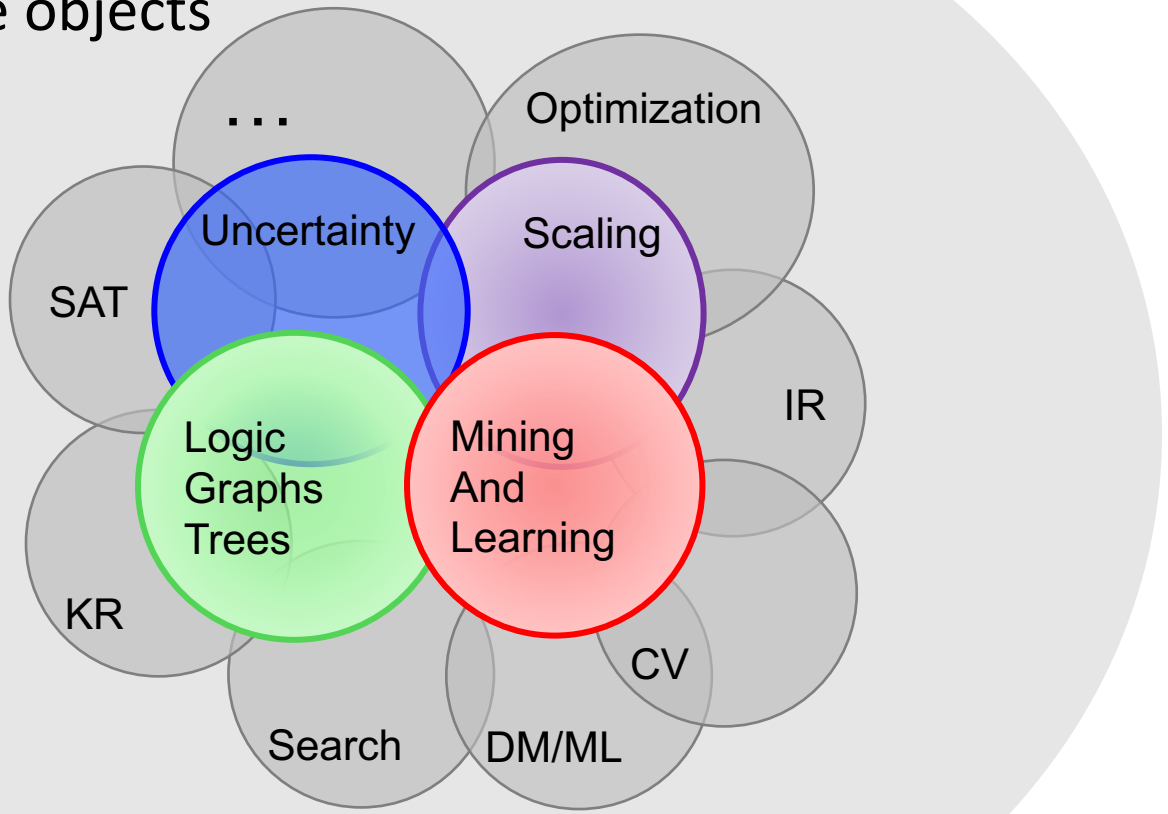


Goal: scalable, easy-to-use systems

Needs probabilities, graphs, logic, data bases, …

Statistical Relational Learning/AI

- Study and design intelligent agents that reason about and act in noisy worlds composed of objects and relations among the objects



[Getoor, Taskar MIT Press '07; De Raedt, Frasconi, Kersting, Muggleton, LNCS'08; Domingos, Lowd Morgan Claypool '09; Natarajan, Kersting, Khot, Shavlik Springer Brief'15; Russell CACM 58(7): 88-97 '15, Gogate, Domingos CACM 59(7):107-115 ´16]

It works

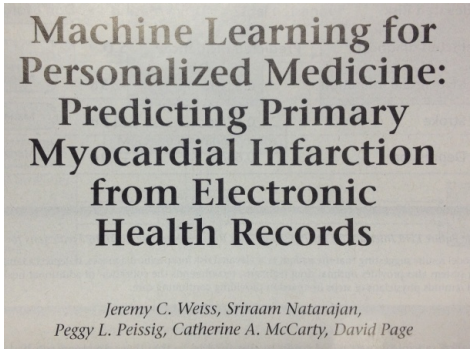
CARDIA EXAM COMPONENTS—ALL YEARS

Schedule of components in the core study, substudies, and ancillary studies by CARDIA exam

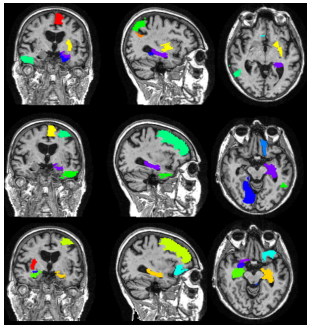
	Year/Exam[1]						
	1985	1987	1990	1992	1995	2000	2005	2010
	0	2	5	7	10	15	20	25
CORE STUDY								
BLOOD PRESSURE								
Resting	X	X	X	X	X	X	X	X
Standing	-	X	-	-	-	-	-	-
Reactivity	-	X	-	-	-	-	-	-
CHEMISTRIES								
Genetic								
DNA Storage	-	-	X	-	X	X	X	X
Stored Cells for Cell Immortalization	-	-	-	-	-	X	-	-
Plasma								
Lipids	X	X	X	X	X	X	X	X
Lipoproteins	X	X	X	X	X	X	X	-
Apoproteins	X	X	-	-	-	-	-	-
CBC	X	-	-	-	-	-	-	-
Lp(a)	-	-	X	-	-	-	-	-
Fibrinogen	-	-	X	-	-	-	X	-
ApoE Phenotype	-	-	-	X	-	-	-	-
Stored Plasma	-	X	X	X	X	X	X	X
C-Reactive Protein	-	-	-	X	-	X	X	X
Interleukin-6	-	-	-	-	-	-	X	-
Serum								
Cotinine	X	-	-	-	-	-	-	-
SMAC 12	X	-	-	-	-	-	-	-
Fasting Insulin	X	-	-	X	X	X	X	X
Fasting Glucose	X	-	-	X	X	X	X	X
Oral Glucose Tolerance Test	-	-	-	-	X	-	X	X
Stored Serum	X	X	X	X	X	X	X	X
GGT	X	-	-	-	X	-	-	X
Serum Creatinine	X	-	-	-	X	X	X	X
Uric Acid	X	-	-	-	X	X	-	-
Urine								
Urinary Creatinine	-	-	-	-	X	X	X	X
Albuminuria	-	-	-	-	X	X	X	X
ANTHROPOMETRY								
Height	X	X	X	X	X	X	X	X
Weight	X	X	X	X	X	X	X	X
Skinfolds	X	X	X	X	X	-	-	-
Chest Circumference	-	X	X	-	-	-	-	-
Waist Circumference	X	X	X	X	X	X	X	X
Hip Circumference	X	X	X	X	X	-	-	X
Thigh Circumference	-	-	-	X	-	-	-	-
Elbow Breadth	X	X	-	-	-	-	-	-
Shoulder Breadth	-	X	-	-	-	-	-	-
Sitting Height	-	X	-	-	-	-	-	-
Toenails	-	X	-	-	-	-	-	-
Eye Color	-	-	X	-	-	-	-	-
Skin Reflectance	-	-	-	X	-	-	-	-
MEDICAL HISTORY								
Medical History	X	X	X	X	X	X	X	X
Illicit Drug Use	X	X	X	X	X	X	X	X
Death Certificate	-	X	X	X	X	X	X	X
Mortal Events	-	-	-	X	X	X	X	X
Safety Questionnaire	-	X	-	-	-	-	-	-
Interim Hospitalization	-	X	X	X	X	X	X	X
Chest Pain/Palpitations	-	-	X	-	X	-	-	-
History of Lung Problems	X	X	-	-	X	X	X	-
Oral Contraceptive History	-	-	-	-	X	-	-	-
Women's Reproductive Health	-	-	-	-	-	X	X	X
Sleep Habits	-	-	-	-	-	X	X	-
Tobacco	X	X	X	X	X	X	X	X
Alcohol	X	X	X	X	X	X	X	X
Weight History	X	X	-	-	-	-	-	X
Sociodemographics	X	X	X	X	X	X	X	X
FAMILY HISTORY QUESTIONNAIRE								
Family History	X	-	X	-	X	-	-	X
PHYSICAL ACTIVITY/FITNESS								
7-Day Physical Activity	X	-	-	-	-	-	-	-
Physical Activity Questionnaire	X	X	X	X	X	X	X	X
Graded Exercise Test	X	-	-	X	-	-	-	-
Baecke Questionnaire	-	-	X	-	X	-	-	-
Household Chores	-	-	-	X	X	-	-	-
Sedentary Behavior Questionnaire	-	-	-	-	-	-	-	X

[1] year of study indicates when original data collection occurred; assay or coding may occur later

Cardiovascular study

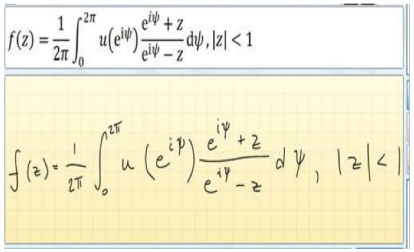


EHR



Alzheimer's

RTS Games



Handwriting Recognition

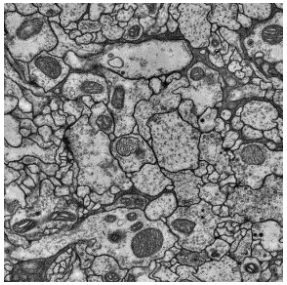
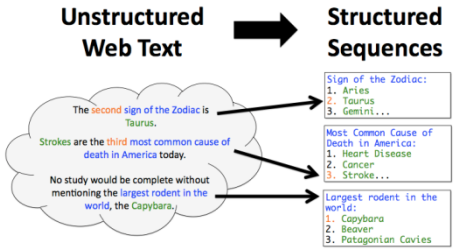
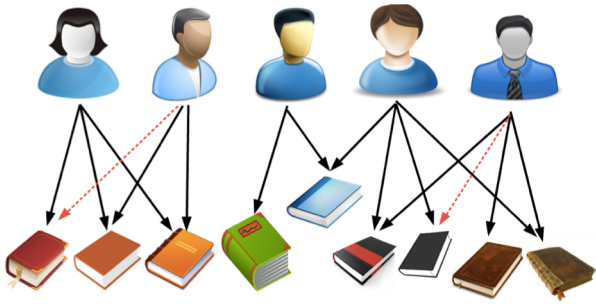


Image Segmentation/ Classification

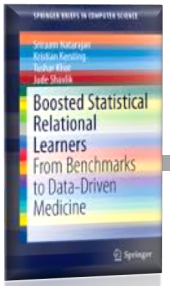


Information Extraction



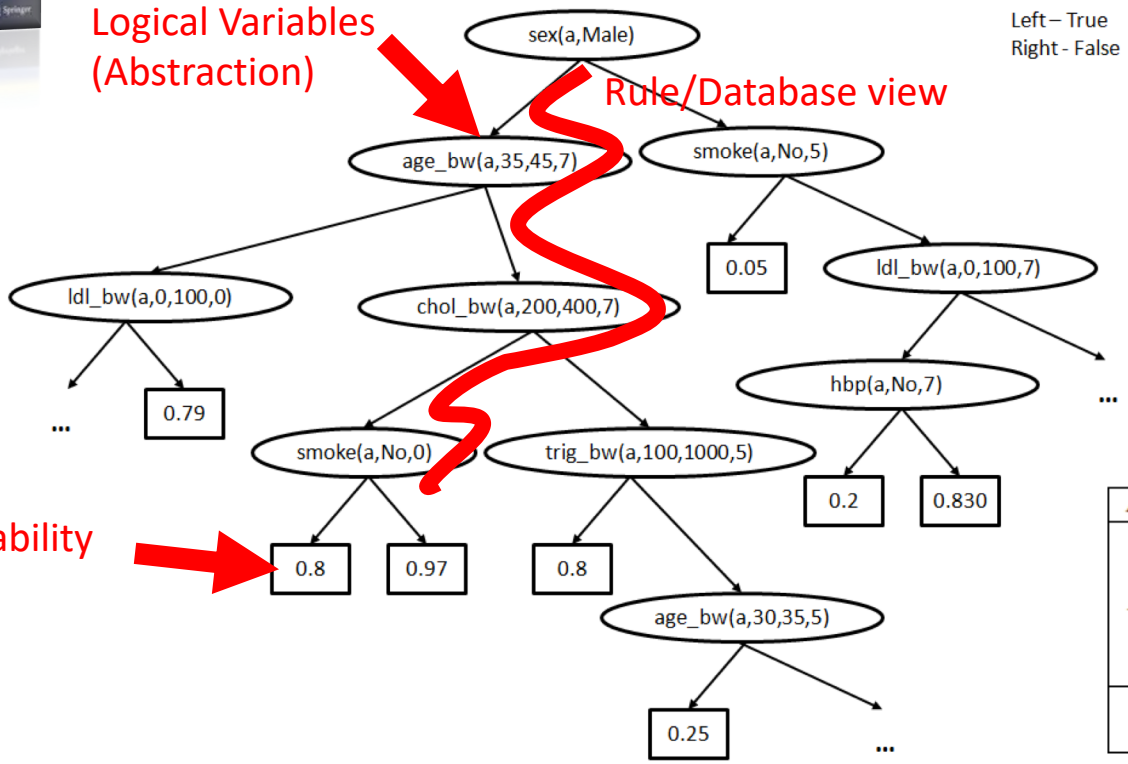
Recommendation System

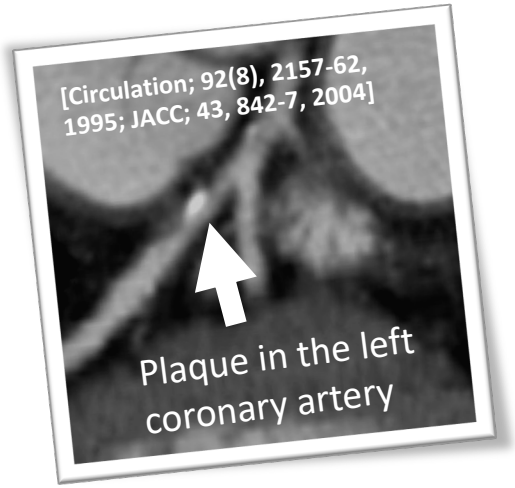
Weiss et al (2012,2013). Natarajan et al (2013,2012, 2014, 2015), Shivram et al (2014), Picado et al (2014)
Soni et al (2016), Viswanathan et al (2016), Odom et al (2014,2015a, 2015b), Yang et al (2017a, 2017b)



This "Deep AI" can understand EHRs

Atherosclerosis is the cause of the majority of Acute Myocardial Infarctions (heart attacks)





Algorithm	Accuracy	AUC-ROC
J48	0.667	0.607
SVM	0.667	0.5
AdaBoost	0.667	0.608
Bagging	0.677	0.613
NB	0.75	0.653
RPT	0.669*	0.778
RFGB	0.667*	0.819

The higher, the better

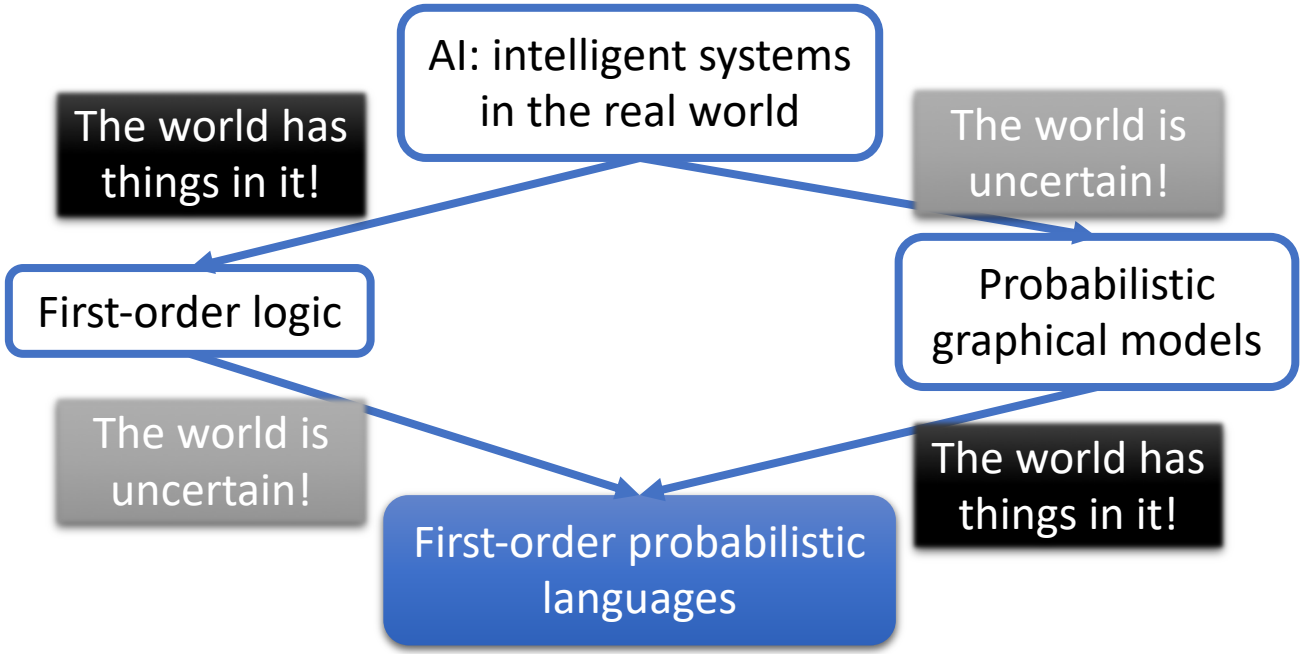
25%

Algorithm for Mining Markov Logic Networks	**Likelihood** The higher, the better	**AUC-ROC** The higher, the better	**AUC-PR** The higher, the better	**Time** The lower, the better
Boosting	0.81	0.96	0.93	**9s**
LSM	0.73	0.54	0.62	**93 hrs**

11% 78% 50% 37200x faster

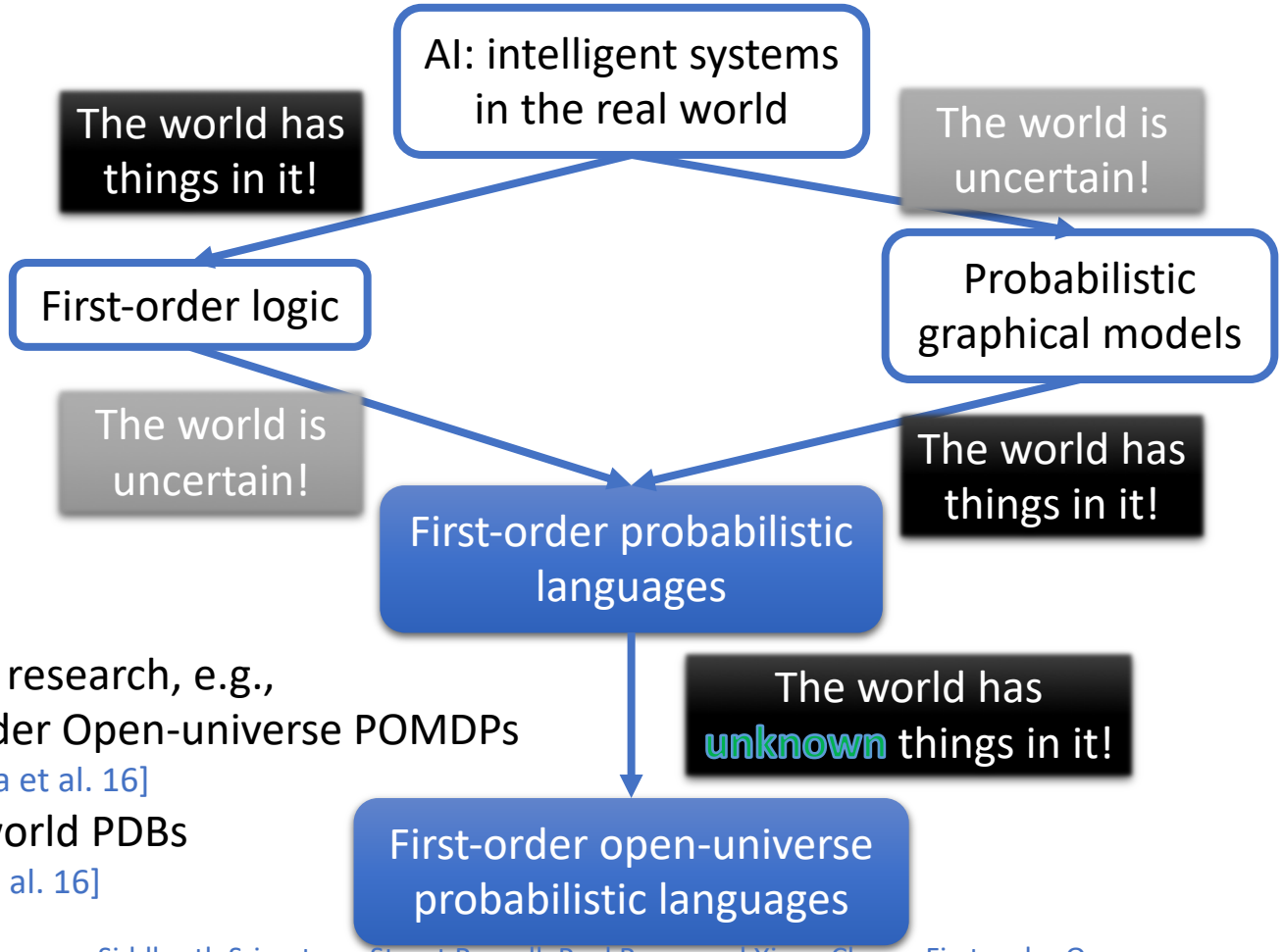
[Kersting, Driessens ICML´08; Karwath, Kersting, Landwehr ICDM´08; Natarajan, Joshi, Tadepelli, Kersting, Shavlik. IJCAI´11; Natarajan, Kersting, Ip, Jacobs, Carr IAAI `13; Yang, Kersting, Terry, Carr, Natarajan AIME ´15; Khot, Natarajan, Kersting, Shavlik ICDM´13, MLJ´12, MLJ´15]

Today!


AI: intelligent systems in the real world
The world has things in it!
The world is uncertain!
First-order logic
Probabilistic graphical models
The world is uncertain!
The world has things in it!
First-order probabilistic languages

Tomorrow?

Stuart Russell



Even more research, e.g.,

- First-order Open-universe POMDPs [Srivastava et al. 16]
- Open-world PDBs [Ceylan et al. 16]

Siddharth Srivastava, Stuart Russell, Paul Ruan, and Xiang Cheng. First-order Open-universe POMDPs. In: Proc. UAI-14, **2014**.

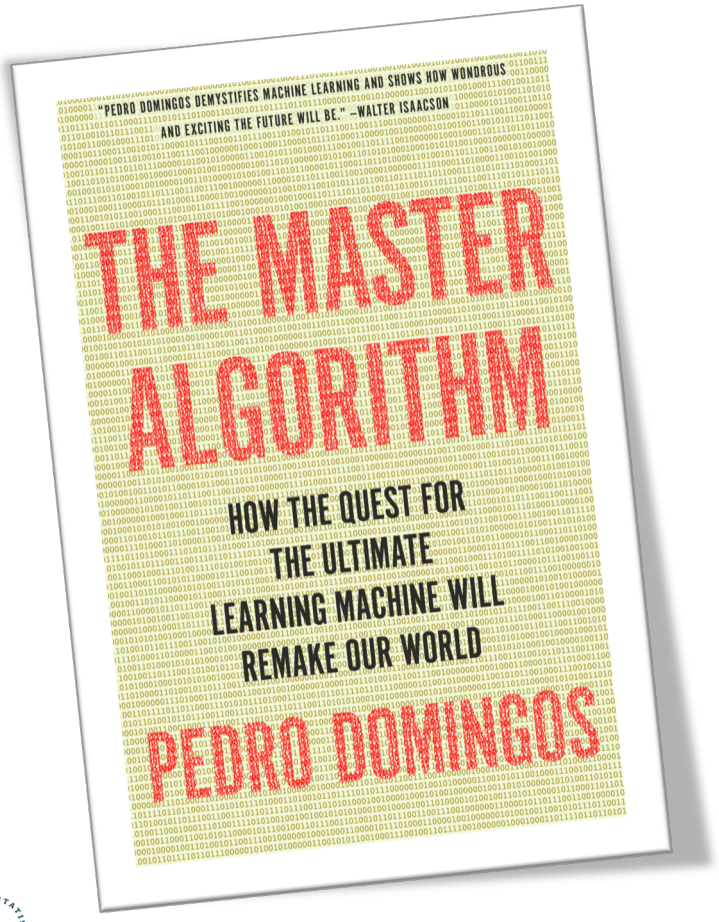
Ismail Ilkan Ceylan, Adnan Darwiche, and Guy Van den Broeck. Open-world Probabilistic Databases. In: IJCAI-16, **2016**.

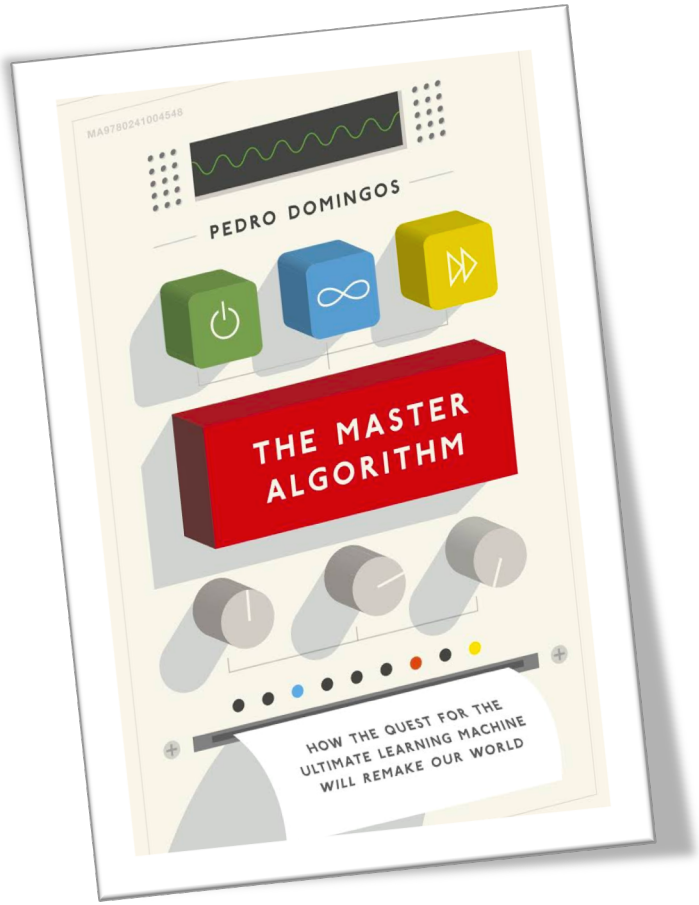
Open Problems

- In any field, say Electronic Health Records or Robotics, there are many open problems
  - Open world learning – new diseases, drugs, indicators
  - Multi-modal learning
  - Large-scale lifted inference
  - Large-scale learning
  - Evolving dynamics
  - Heterogeneous data and hybrid models
  - Expert knowledge elicitation
  - Planning & actions
  - Interactive learning
  - …

And there are popular science books about it.

In 2016, Bill Gates recommended the book, alongside Nick Bostrom's *Superintelligence*, as one of the two books everyone should read to understand AI.

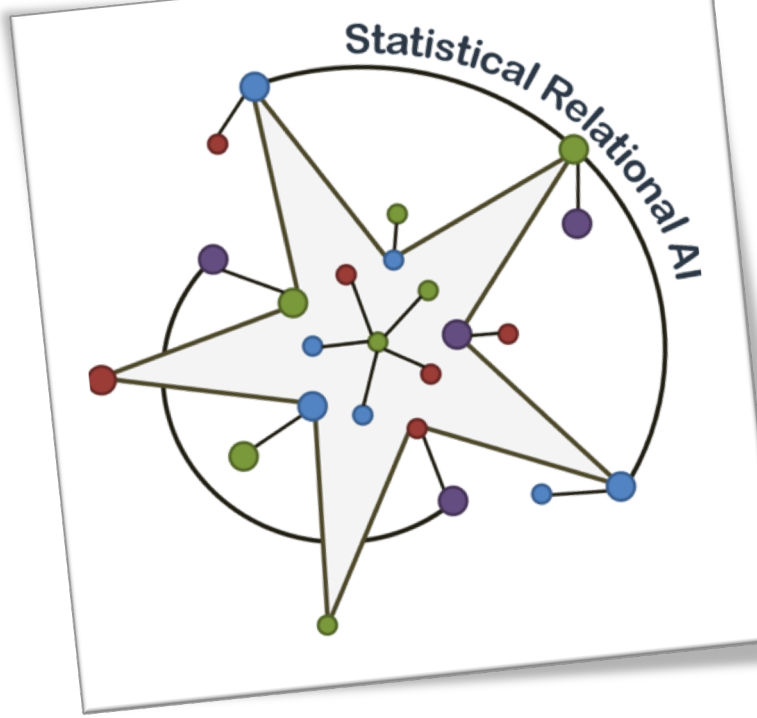




Slides

Statistical Relational AI

Tutorial at ICCS 2019



https://www.ifis.uni-luebeck.de/index.php?id=599

Tanya Braun and Marcel Gehrke, University of Lübeck

UNIVERSITÄT ZU LÜBECK

Thanks to Ralf Möller, Kristian Kersting, and many others
for making their slides publicly available

Logo: http://www.starai.org