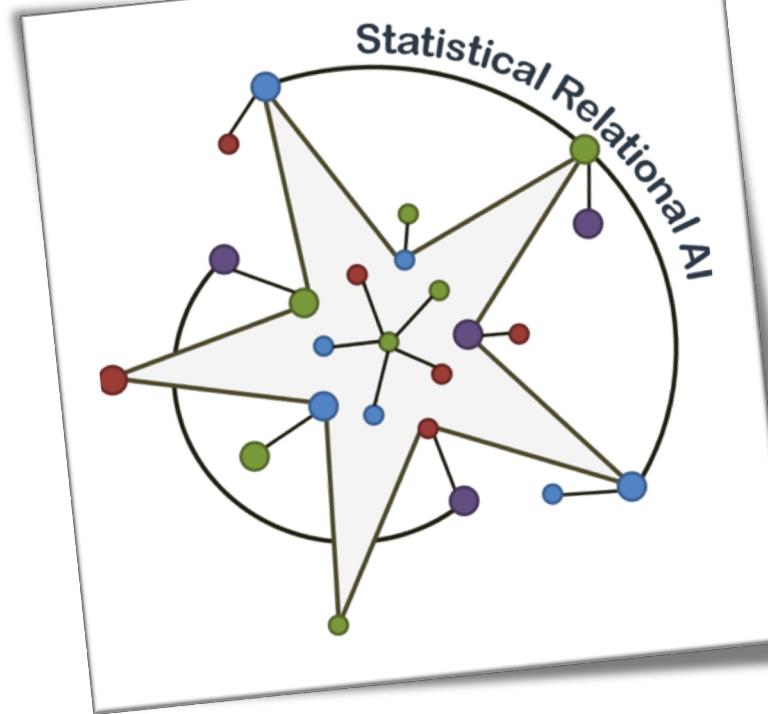


Wrap-up

Statistical Relational AI

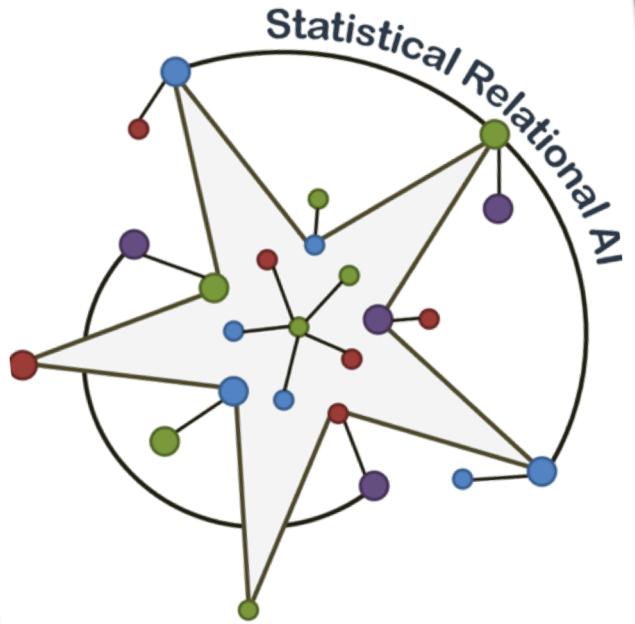
Tutorial at BTW 2019



Tanya Braun, University of Lübeck



UNIVERSITÄT ZU LÜBECK



Mission for today

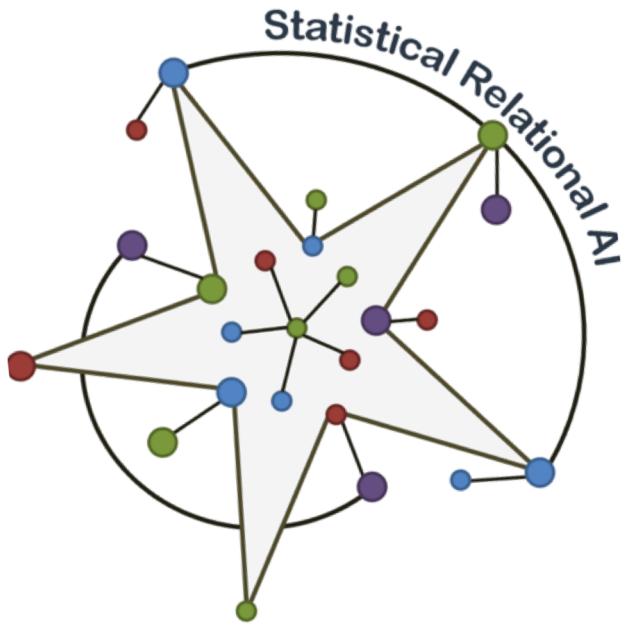
Providing an overview and a synthesis of StaRAI
to identify links between StaRAI and DBs and prospects for
collaboration for implementing scalable systems

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
 - Approximate: Belief propagation, sampling
 - Lifted model from ground model: colour passing

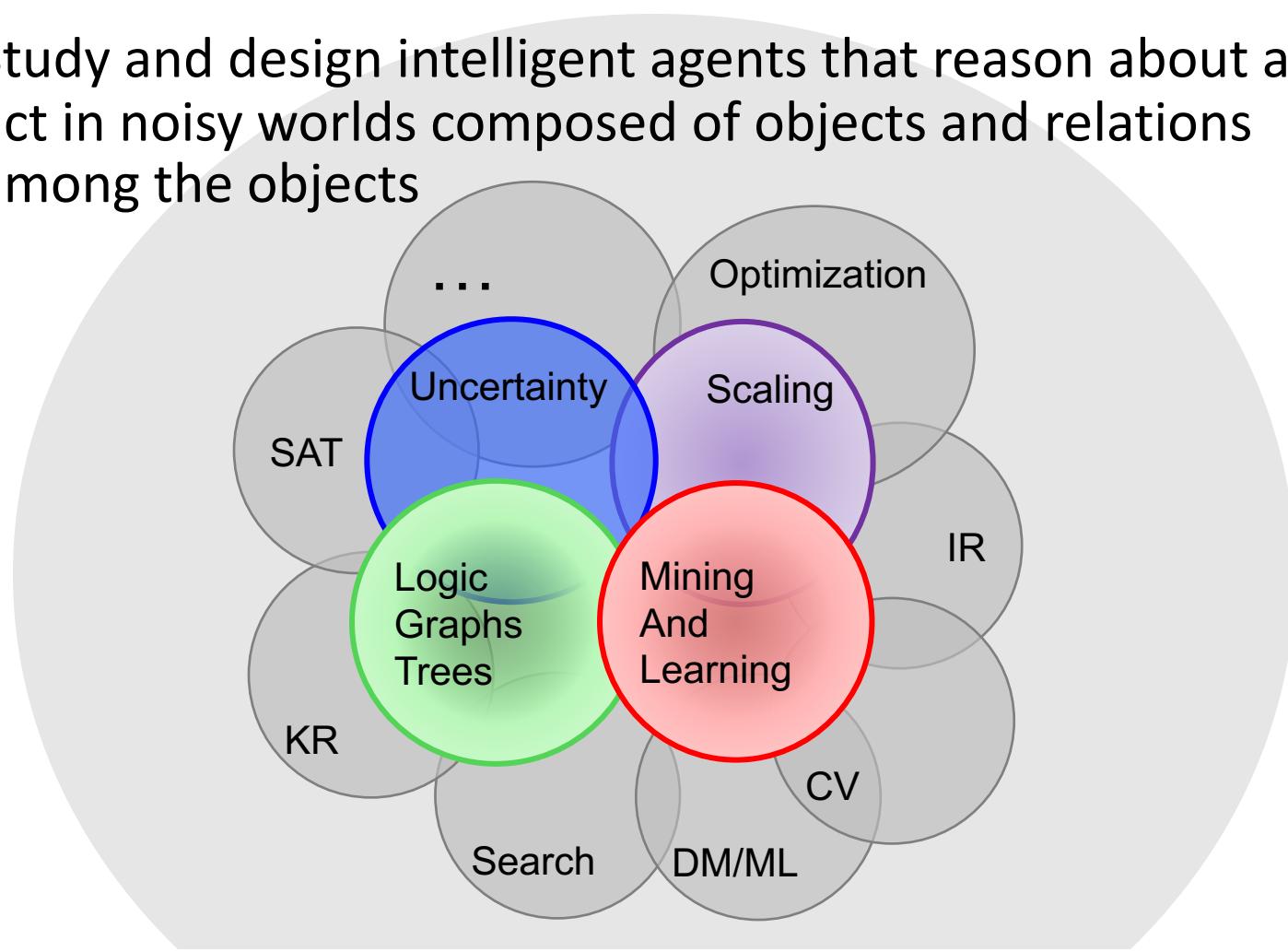
Goal: scalable, easy-to-use systems

Needs probabilities, graphs, logic, and ... 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

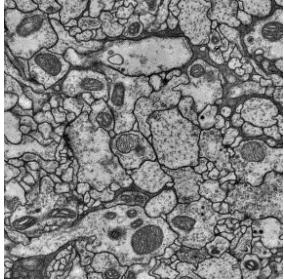
Schedule of components in the core study, substudies, and ancillary studies by CARDIA exam

	Year Exam*										Year Exam*												
	1985	1987	1990	1992	1995	2000	2005	2010	0	2	5	7	10	15	20	25	0	2	5	7	10	15	20
CORE STUDY																							
BLOOD PRESSURE	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Height	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Standing	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Body Mass Index	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CHOLESTEROL	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Genetic	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
DNA Storage	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Stained Cells for Cell Immunotherapy	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Plasma	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Lipids	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Lipoproteins	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Aspirin	X	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CBC	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Urea	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Lp(a)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Fibrinogen	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Anti-DP Serotype	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Sex Plasma	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
C-Reactive Protein	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Interleukin-6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Bilirubin	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Urine	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Urinary Creatinine	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Albuminuria	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ANTHROPOMETRY	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Height	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
*year of study indicates when original data collection occurred; assay or coding may occur later																							

Cardiovascular study

$$f(z) = \frac{1}{2\pi} \int_0^{2\pi} u(e^{i\psi}) \frac{e^{iz\psi} + z}{e^{i\psi} - z} d\psi, |z| < 1$$

$$\int (z) = \frac{1}{2\pi} \int_0^{2\pi} u(e^{i\psi}) \frac{e^{iz\psi} + 2}{e^{i\psi} - z} d\psi, |z| < 1$$



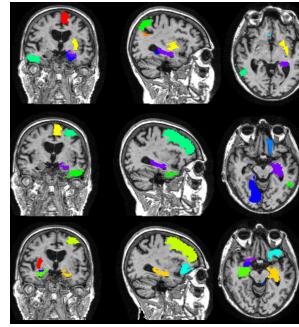
Handwriting Recognition

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)

Kristian Kersting. Statistical Relational AI. Tutorial at KI 2018.

Machine Learning for Personalized Medicine: Predicting Primary Myocardial Infarction from Electronic Health Records

Jeremy C. Weiss, Sriraam Natarajan,
Peggy L. Peissig, Catherine A. McCarty, David Page



EHR

Unstructured Web Text → Structured Sequences

The second sign of the Zodiac is Taurus.
Strokes are the third most common cause of death in America today.
No study would be complete without mentioning the largest rodent in the world, the Capybara.

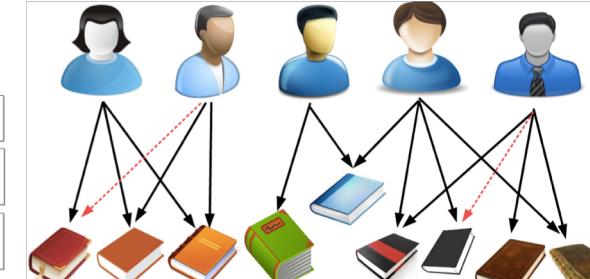
Sign of the Zodiac:
1. Aries
2. Taurus
3. Gemini...

Most Common Cause of Death in America:
1. Heart Disease
2. Cancer
3. Stroke...

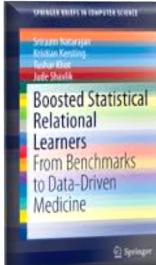
Largest rodent in the world:
1. Capybara
2. Beaver
3. Patagonian Cavies

Information Extraction

Alzheimer's

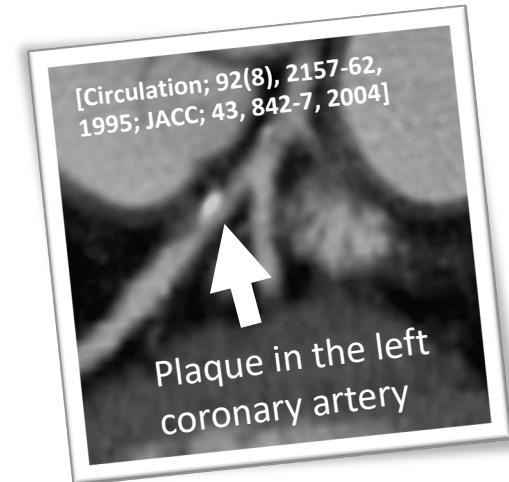
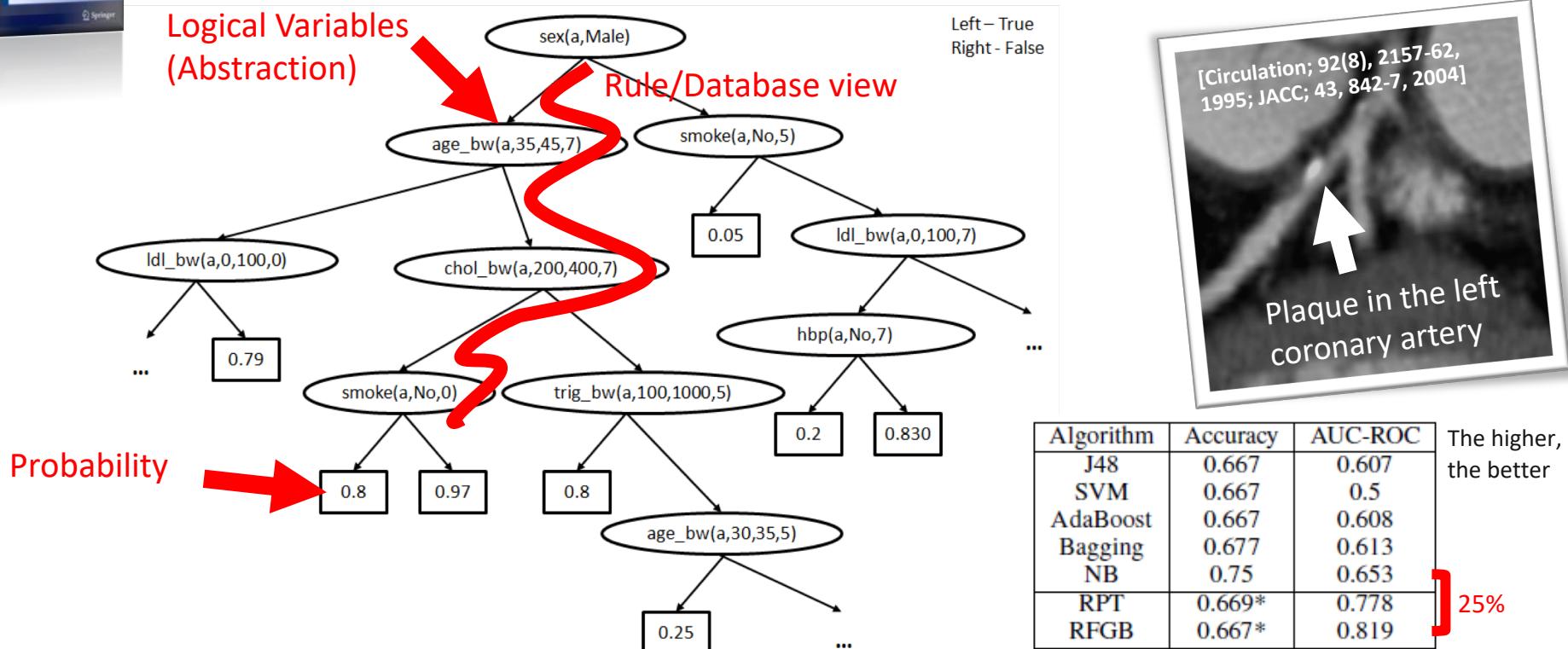


Recommendation System



This “Deep AI” can understand EHRs

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



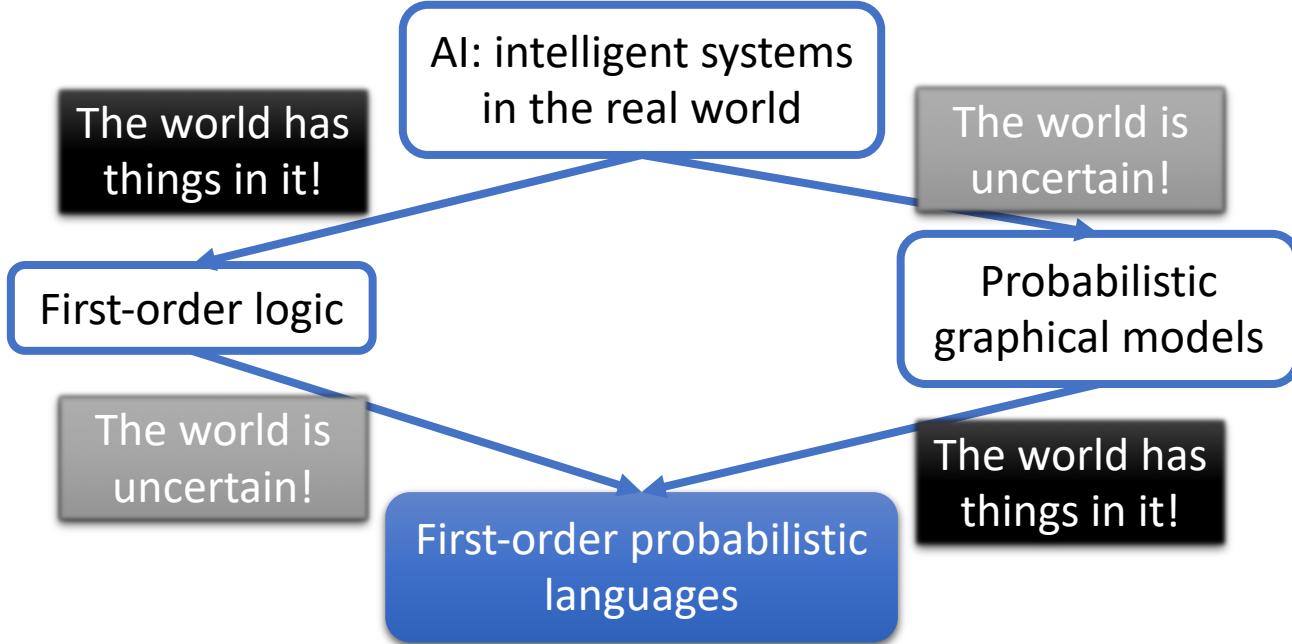
Algorithm	Accuracy	AUC-ROC	The higher, the better
J48	0.667	0.607	
SVM	0.667	0.5	
AdaBoost	0.667	0.608	
Bagging	0.677	0.613	
NB	0.75	0.653	
RPT	0.669*	0.778	
RFGB	0.667*	0.819	

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

The higher, the better

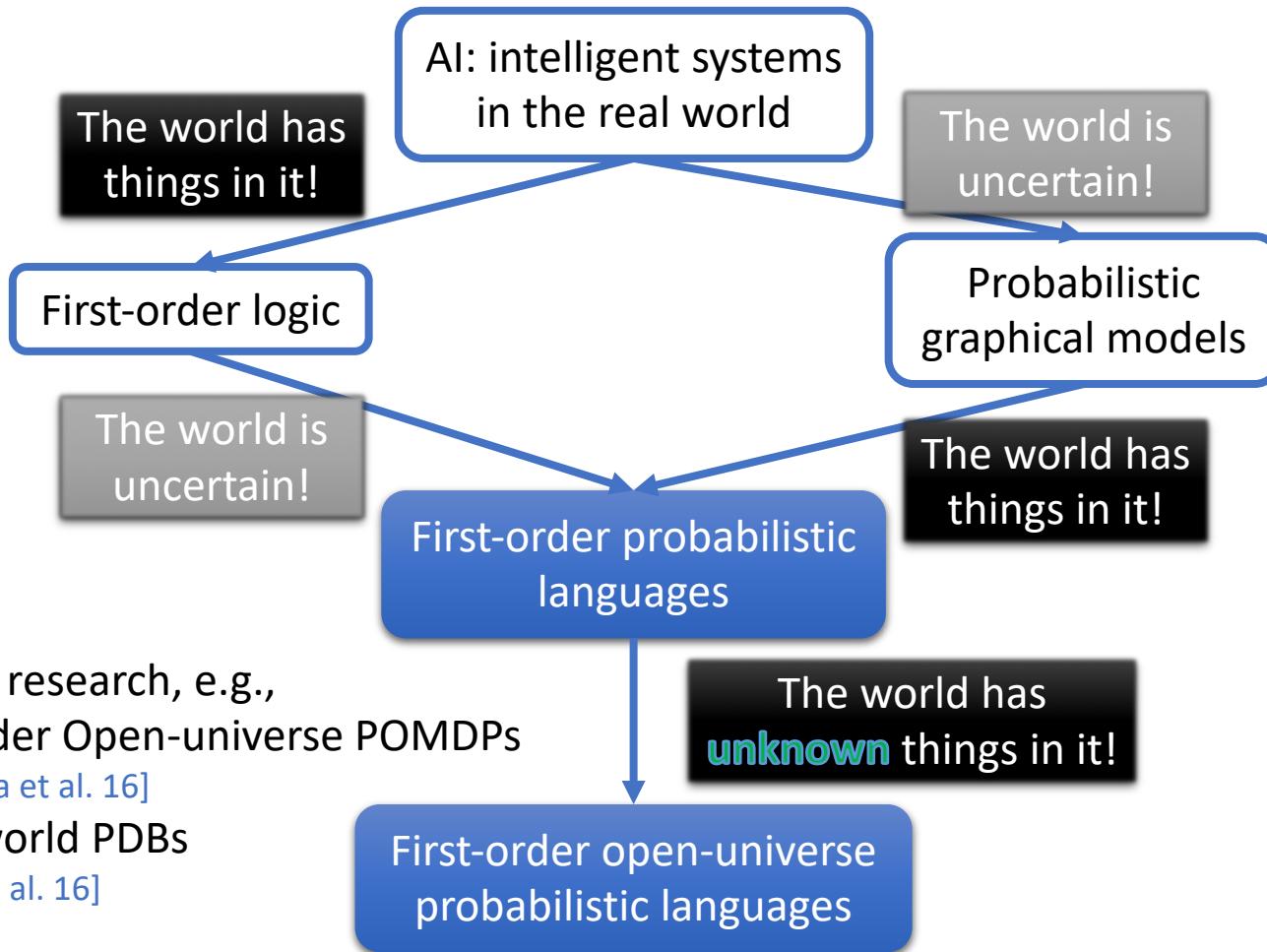
25%

Today!



Tomorrow?

Stuart Russell



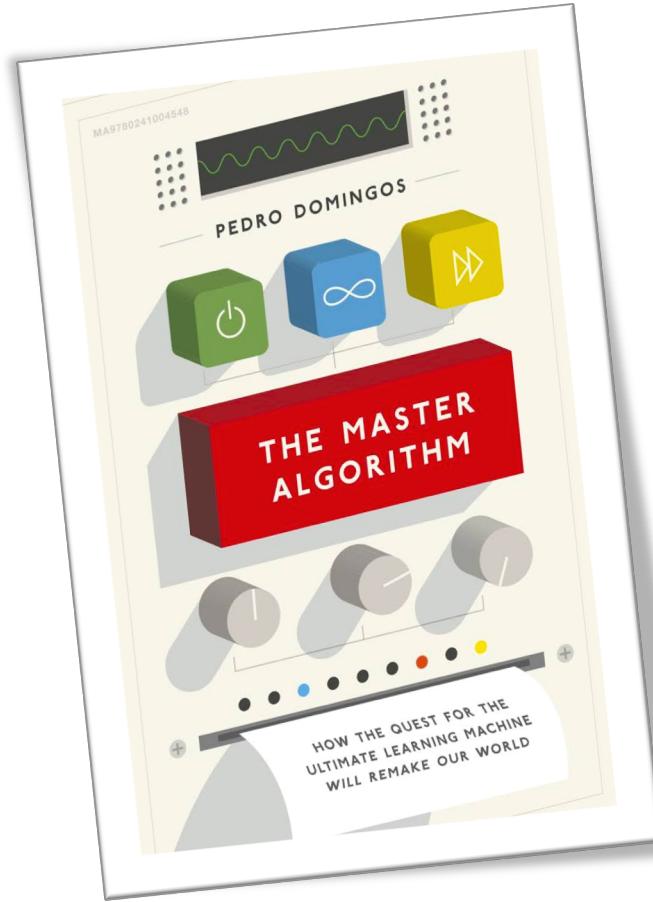
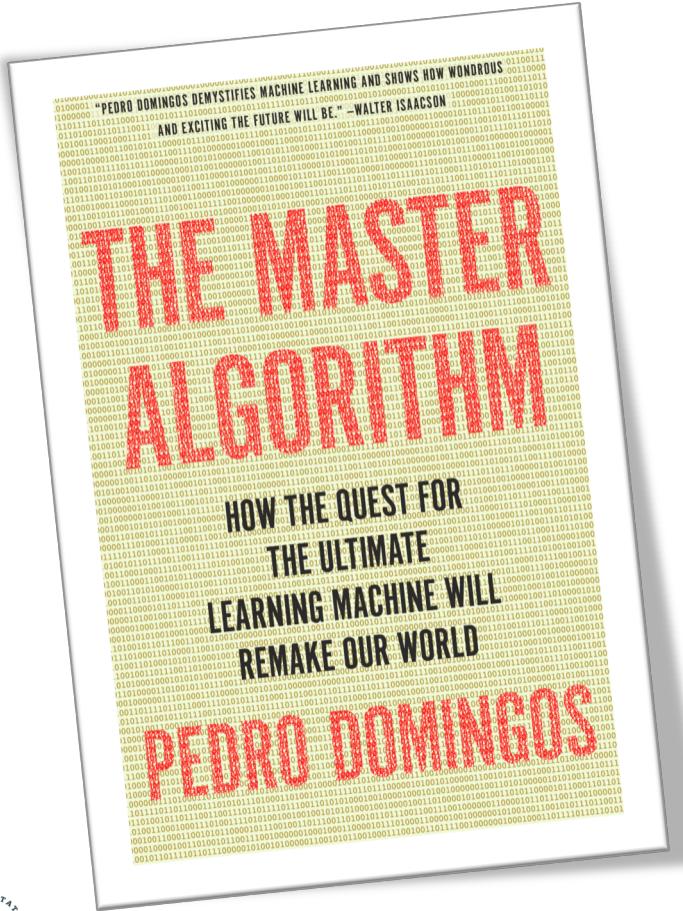
Siddharth Srivastava, Stuart Russell, Paul Ruan, and Xiang Cheng. First-order Open-universe POMDPs. In: Proc. UAI-14, 2014.
Ismail Ilkan Ceylan, Adnan Darwiche, and Guy Van den Broeck. Open-world Probabilistic Databases. In: IJCAI-16, 2016.

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.

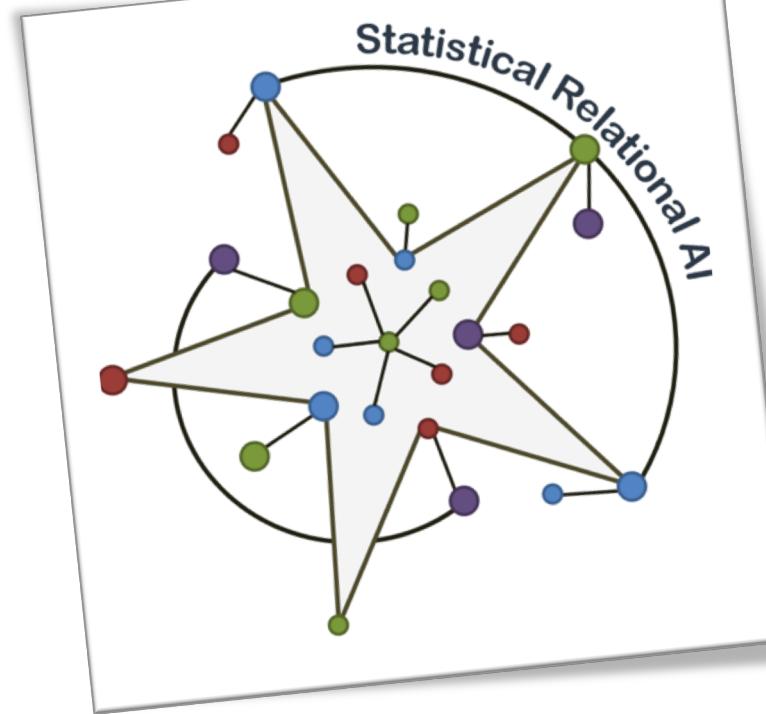
In 2016, Bill Gates recommended the book, alongside Nick Bostrom's *Superintelligence*, as one of the two books everyone should read to understand AI.



Slides

Statistical Relational AI

Tutorial at BTW 2019



<https://www.ifis.uni-luebeck.de/index.php?id=597>

Tanya Braun, 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>