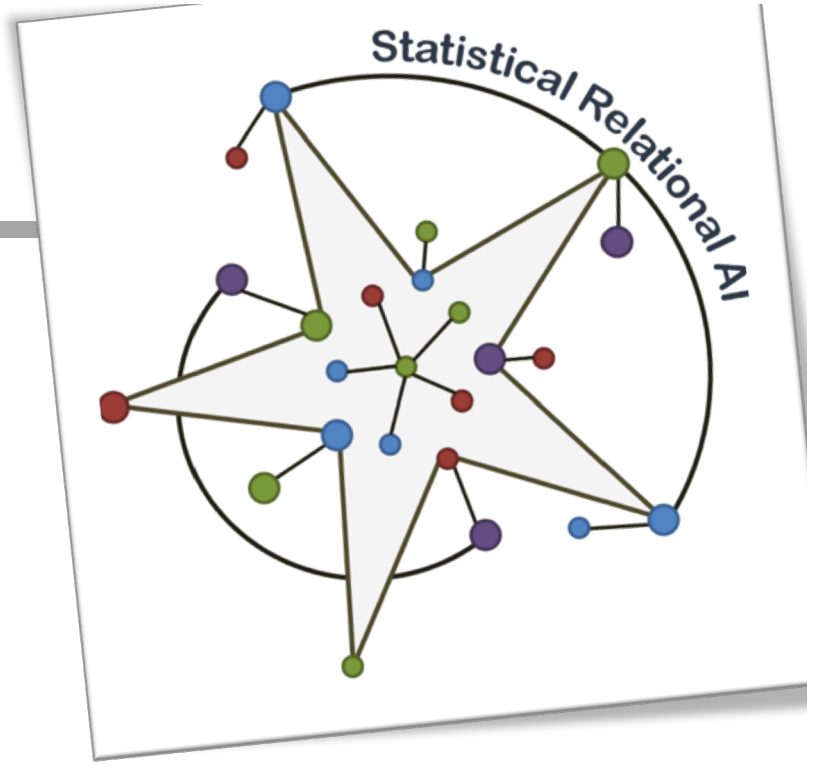
Wrap-Up

Statistical Relational AI

Tutorial at KI-2018



Tanya Braun, Universität zu Lübeck

Kristian Kersting, Technische Universität Darmstadt

Ralf Möller, Universität zu Lübeck

UNIVERSITÄT ZU LÜBECK

TECHNISCHE UNIVERSITÄT DARMSTADT

We need probabilities and optimization.

Artificial Intelligence Models

	Graphical models	**Logical models**	**Mathematical Programs**
Required by probability theory	No	Yes	No
Representable distributions	All (BNs) Positive (MNs)	All	-
Context-free decomposition	Some	All	Some
Context-specific decomposition	None	All	Some
Normalization constraints	Some	All	-
Representable optimization problems	Prob. inference	Logical inference	All propositional ones

Artificial Intelligence Models

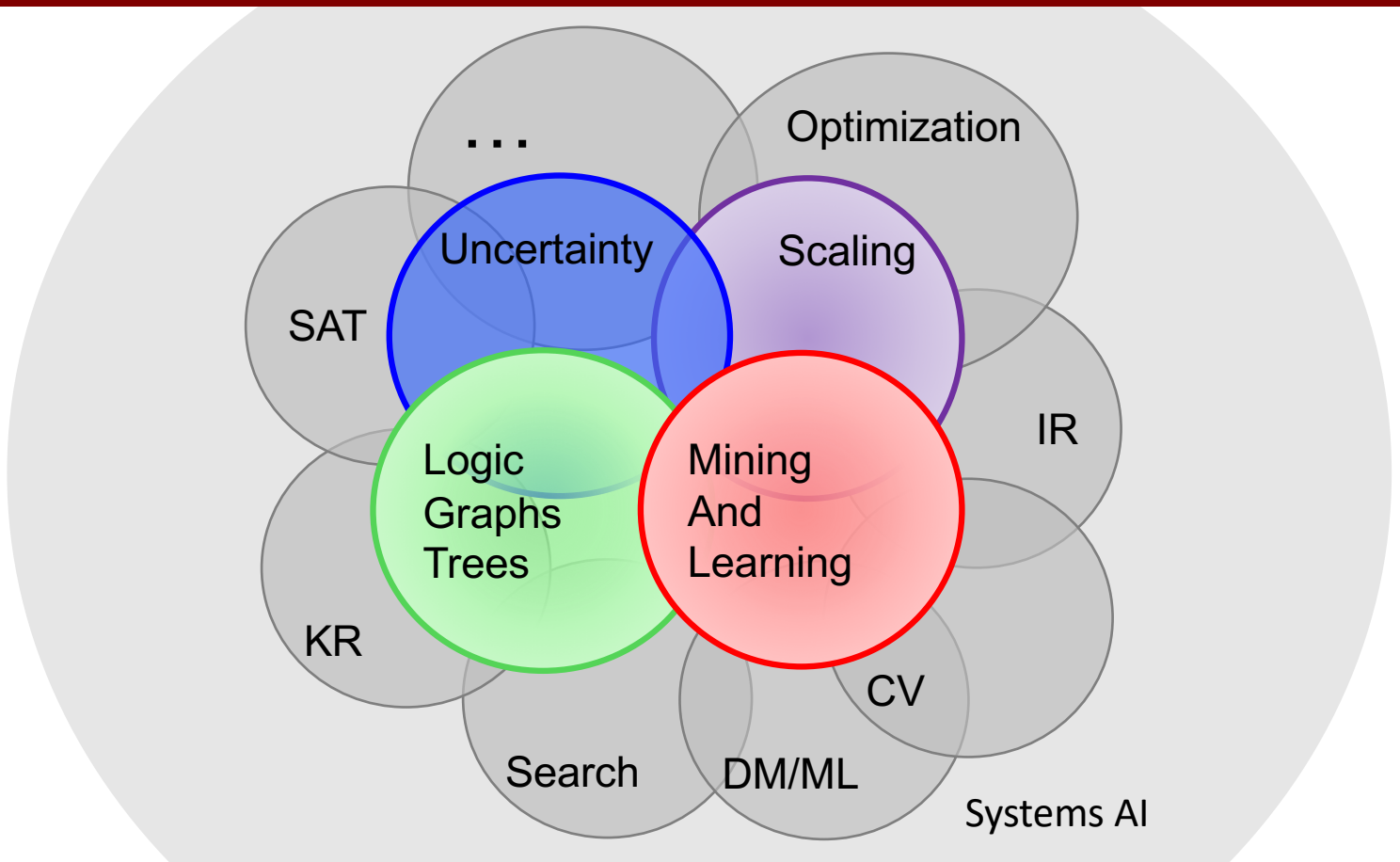
	Graphical models	**Logical models**	**Mathematical Programs**
Inference	Exp(treewidth)	Circuit/lifted complexity	?
Visual aid	Yes	No	No
Densely connected models	Unreadable	Readable	Unreadable
First-order	Plates	All	No
Lifted inference	No	Yes	Partly
Covers Statistical/ Deep ML	Partly	No	Partly
Covers AI	Partly	Partly	Partly
Available technology	Lots, used	Lots, unused	Lots, used

We need probabilities and optimization.

However, graphs are not enough, we need logic

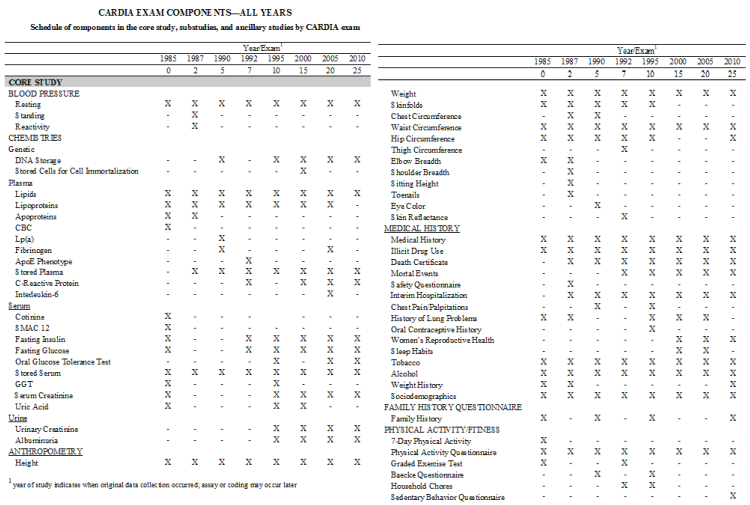
Statistical Relational Learning/AI

… 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

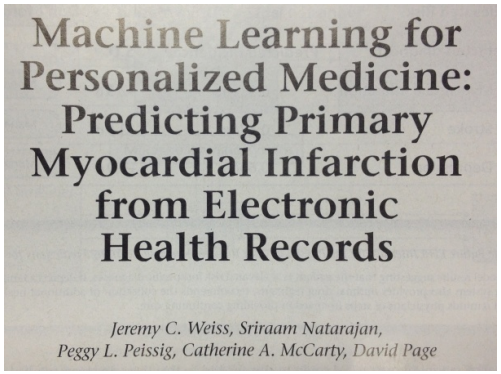
Schedule of components in the core study, substudies, and ancillary studies by CARDIA exam

	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	-	-	-
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

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

	Year/Exam[1]							
	1985	1987	1990	1992	1995	2000	2005	2010
	0	2	5	7	10	15	20	25
Weight	X	X	X	X	X	X	X	X
8 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

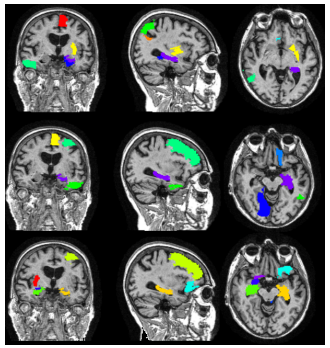
Cardiovascular study



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

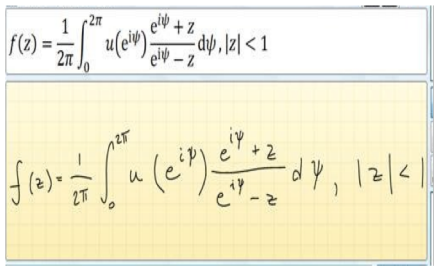
Jeremy C. Weiss, Sriraam Natarajan, Peggy L. Peissig, Catherine A. McCarty, David Page

EHR



Alzheimer's

RTS Games



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

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

Handwriting Recognition

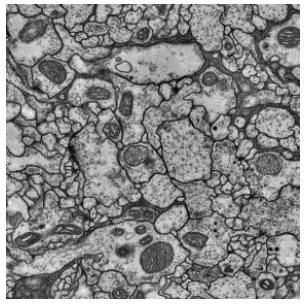
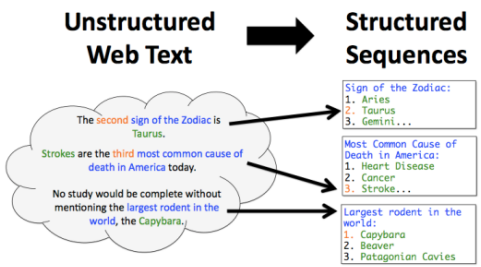
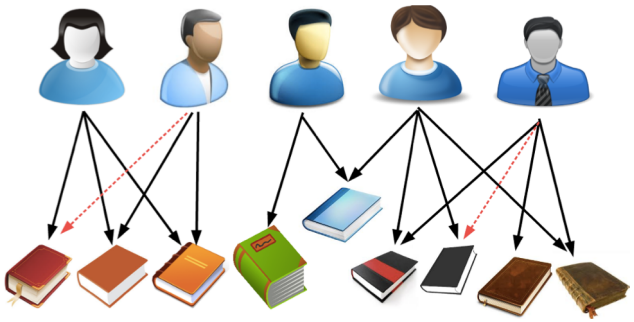


Image Segmentation/ Classification



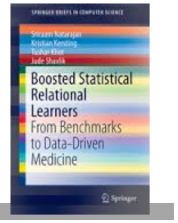
Information Extraction



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)

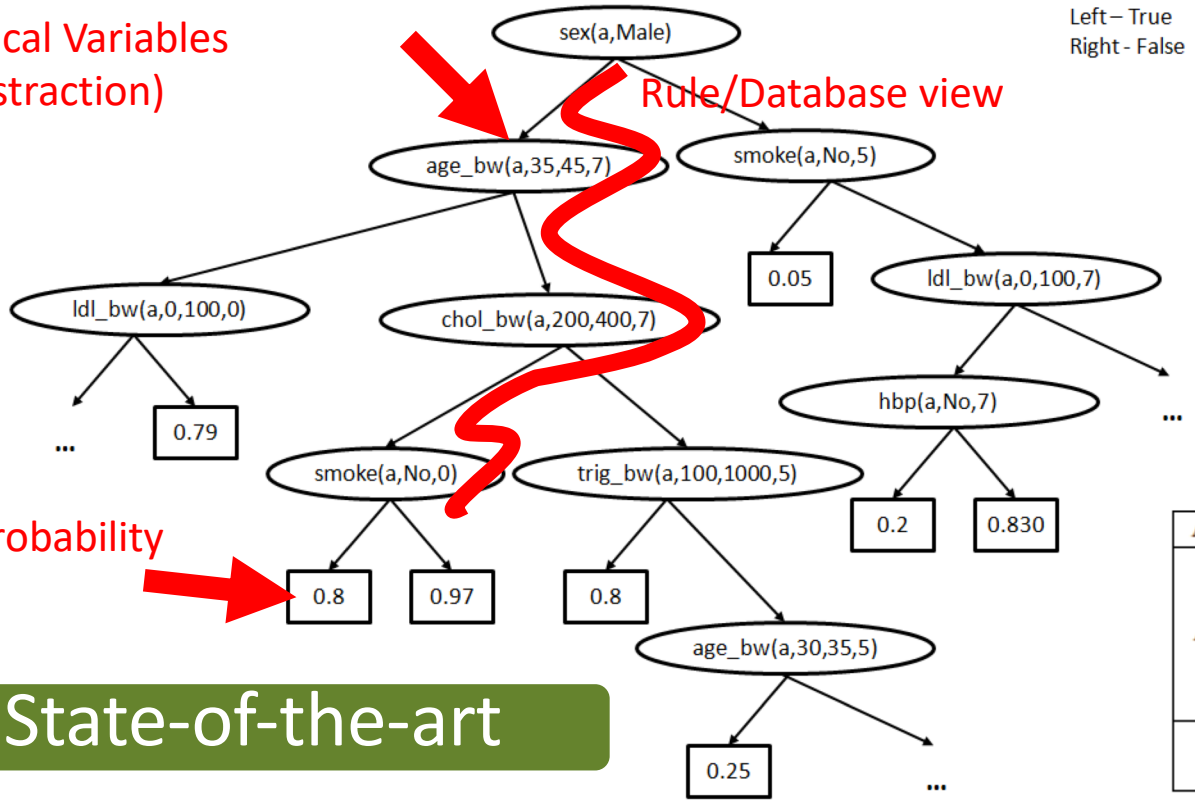
Natarajan, Khot, Kersting, Shavlik. Boosted Statistical Relational Learners. Springer Brief 2015



Mining Electronic Health Records

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

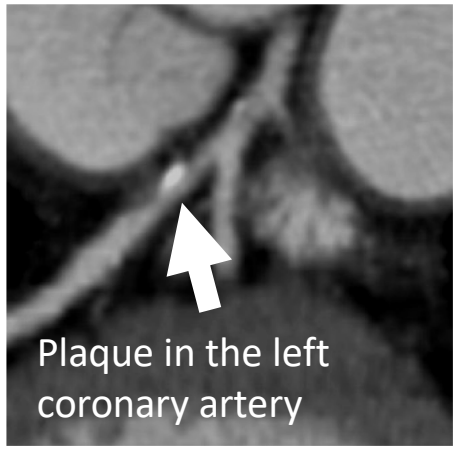
Logical Variables (Abstraction)



Left – True
Right - False

Rule/Database view

Probability



[Circulation; 92(8), 2157-62, 1995; JACC; 43, 842-7, 2004]

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%

State-of-the-art

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

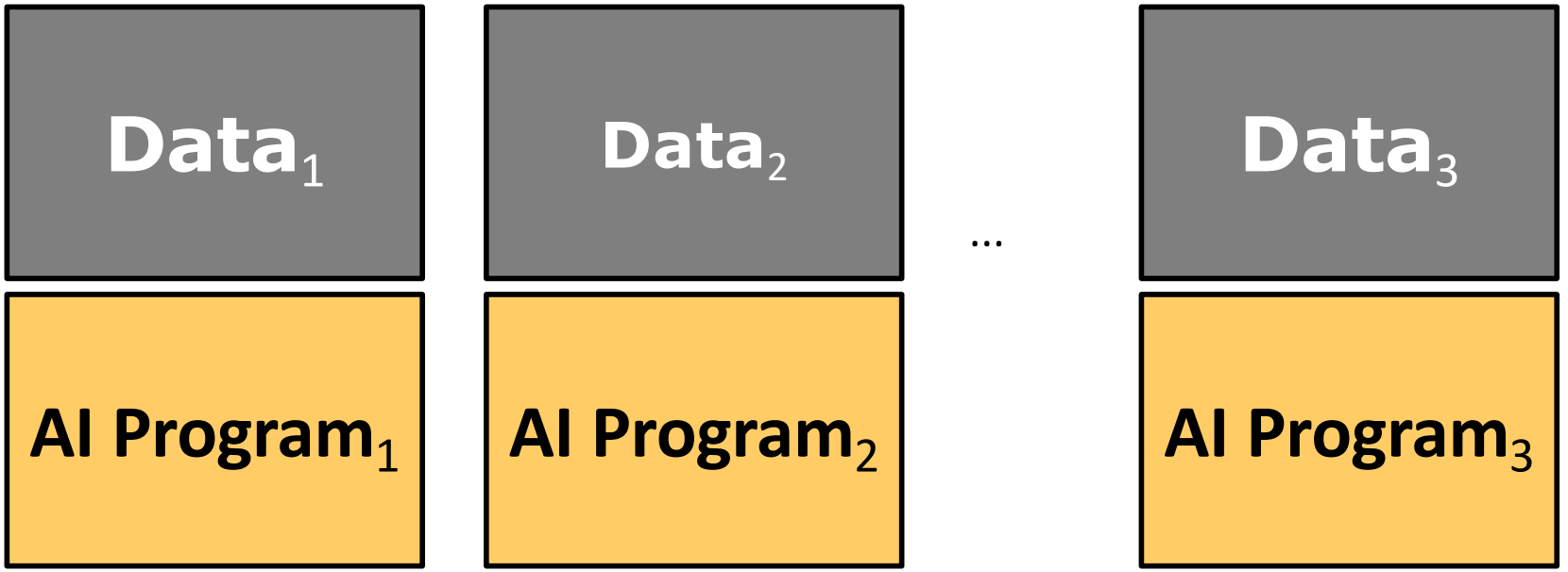
[Kersting, Driessens ICML´08; Karwath, Kersting, Landwehr ICDM´08; Natarajan, Joshi, Tadepelli, Kersting, Shavlik. IJCAI´11; Natarajan, Kersting, Ip, Jacobs, Carr IAAI `13; Yang, Kersting, Terry, Carr, Natarajan AIME ´15; Khot, Natarajan, Kersting, Shavlik ICDM´13, MLJ´12, MLJ´15]

Open Problems

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

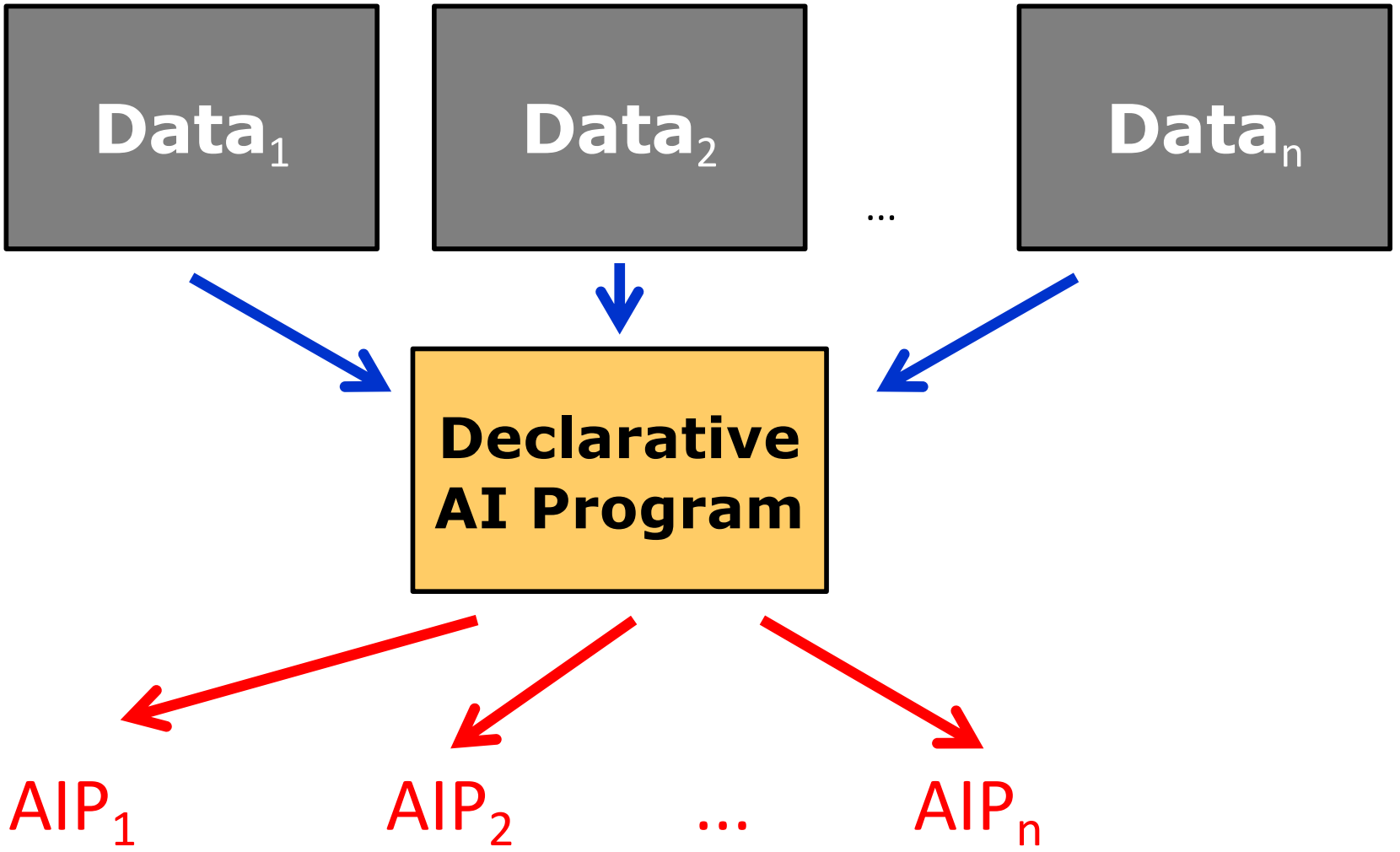
[Kersting, Mladenov, Tokmakov AIJ 2017, Kordjamshidi, Roth, Kersting IJCAI-ECAI 2018]

This puts forward a “Deep” Universal AI Modeling



[Kersting, Mladenov, Tokmakov AIJ 2017, Kordjamshidi, Roth, Kersting IJCAI-ECAI 2018]

This puts forward a "Deep" Universal AI Modeling



In general,

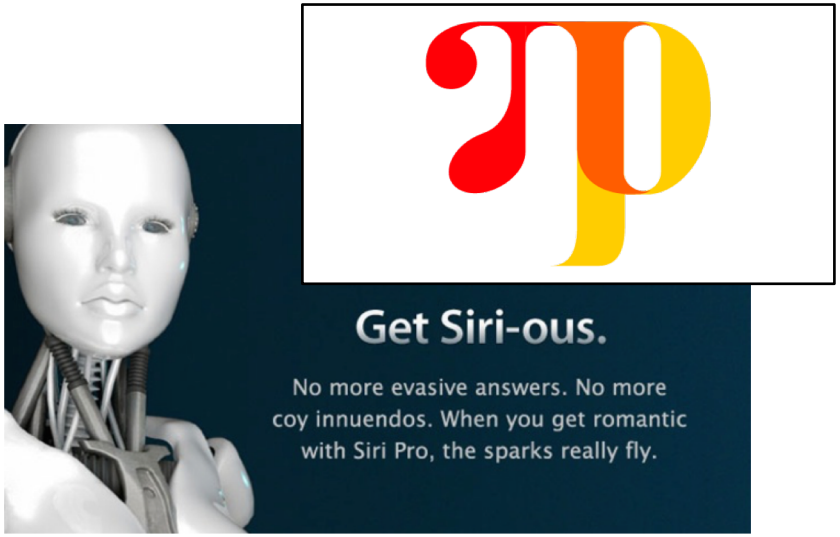
High-level languages for AI, ML and optimization are a step towards the …

Democratization of AI

- Reduces the level of expertise necessary to build AI applications, makes models faster to write and easier to communicate
- Facilitate the construction of sophisticated models with rich domain knowledge
- Speed up solvers by exploiting language properties, compression, and compilation

This „Deep AI“ excites industry

RelationalAI, LogicBlox, Apple, and Uber are investing hundreds of millions of US dollars



IBM CPLEX

GUROBI OPTIMIZATION

And it appears in industrial strength solvers such as CPLEX and GUROBI

And there are popular science books about it.

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