Aalto University

# Koneoppimisen eri muodot

**Samuel Kaski**

**4.5.2017**

# Machine Learning

**Deep Learning**

**With massive amounts of computational power, machines can now recognize objects and translate speech in real time. Artificial intelligence is finally getting smart.**

**MIT Technology Review**

AlphaGo

The New York Times

**Artificial Intelligence**

News about Artificial
Times.

Latest

**A Lawsuit Aga
Rush to Conqu**

The suit, brought by t
accuses Uber and a st
planning to steal trad

By MIKE ISAAC and DAISUK
Feb. 25, 2017

**Google Cousin
Flag Toxic Onl**

Researchers at the inc
help publishers filter
discourse civil.

By DAISUKE WAKABAYASHI
Feb. 23, 2017

**Ford to Invest
Intelligence Start-Up**

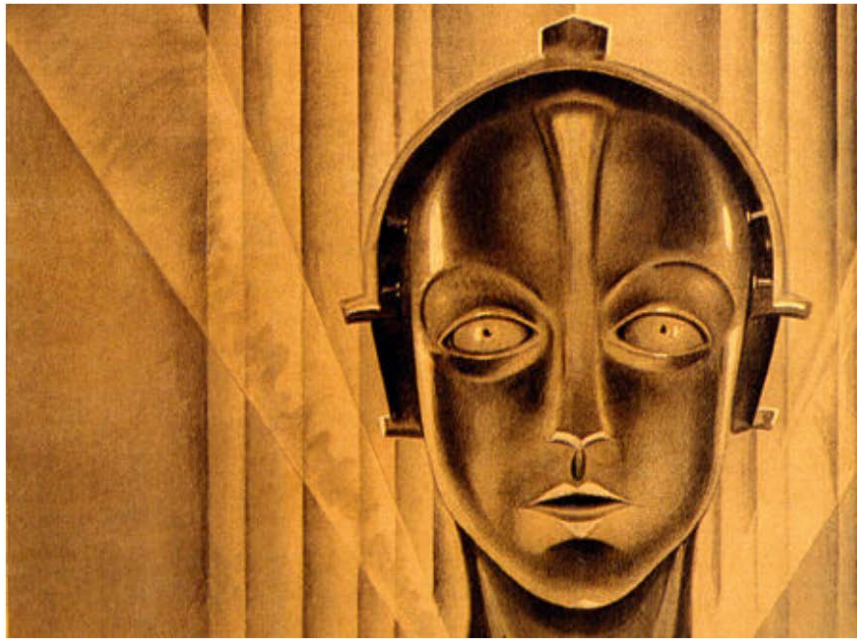
The automaker will be the majority shareholder Argo AI, which will focus on developing self-driving vehicle technology.

By MIKE ISAAC and NEAL E. BOUDETTE
Feb. 10, 2017

Artificial intelligence

# March of the machines

*What history tells us about the future of artificial intelligence—and how society should respond*



The Economist

TEKOÄLY | Mikko Metsämäki 7.2. klo 20:44

# Pääministeri haluaa Suomesta tekoälyn ykkösmaan

Kauppalehti


Aalto University

TEKOÄLY | Arto Klami, Samuel Kaski, Petri Myllymäki

# Tekoäly istuu jo suomalaisyhtiön johtoryhmässä - Suomi pärjää vain tekemällä ainutlaatuisia läpimurtoja

Suomi voi saavuttaa kilpailuetua vain tekemällä uniikkeja läpimurtoja, ei seurailemalla muita, tekoälytutkijat kirjoittavat.

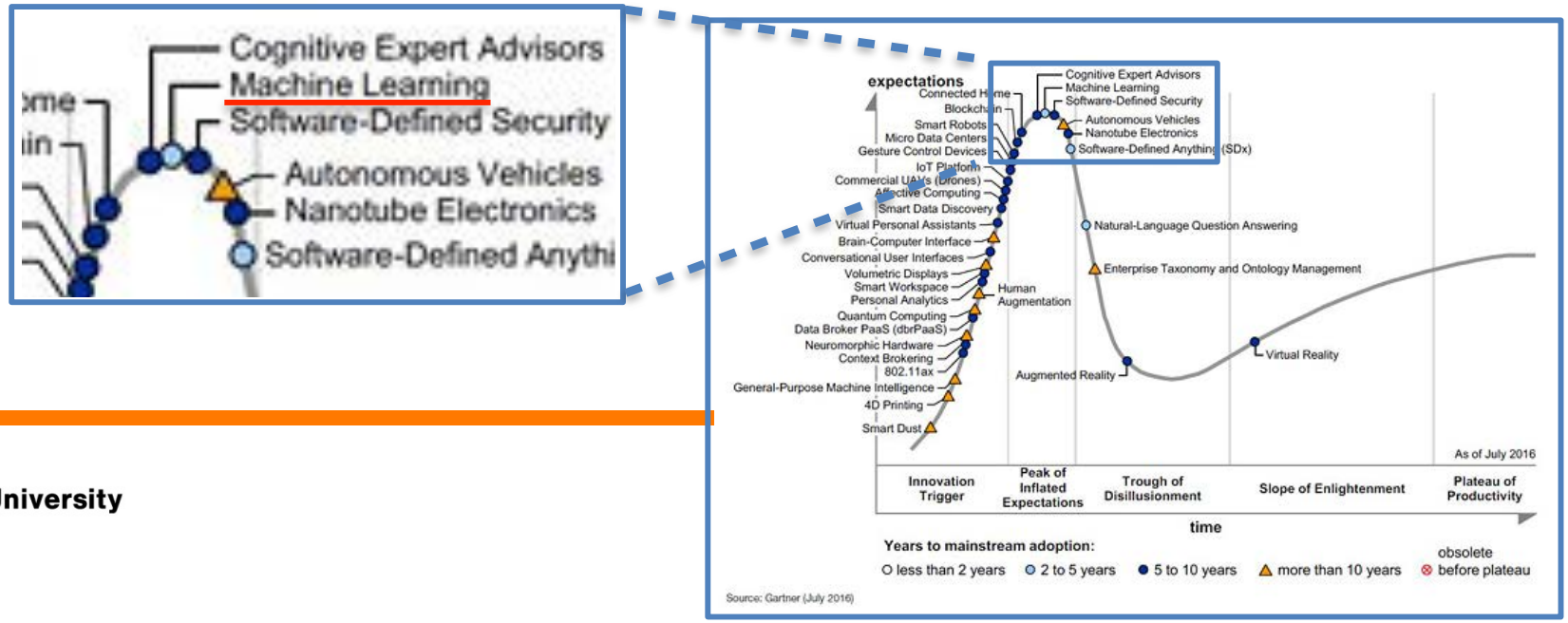
Eilen

tekniikka & talous

# But wait - that was about AI and we were supposed to talk about machine learning

Machine learning is the technology underlying the current AI revolution

It is also surprisingly a buzzword itself.



# Machine learning: Definition

Machine learning is the subfield of computer science that gives computers the ability to learn without being explicitly programmed (Arthur Samuel, 1959).

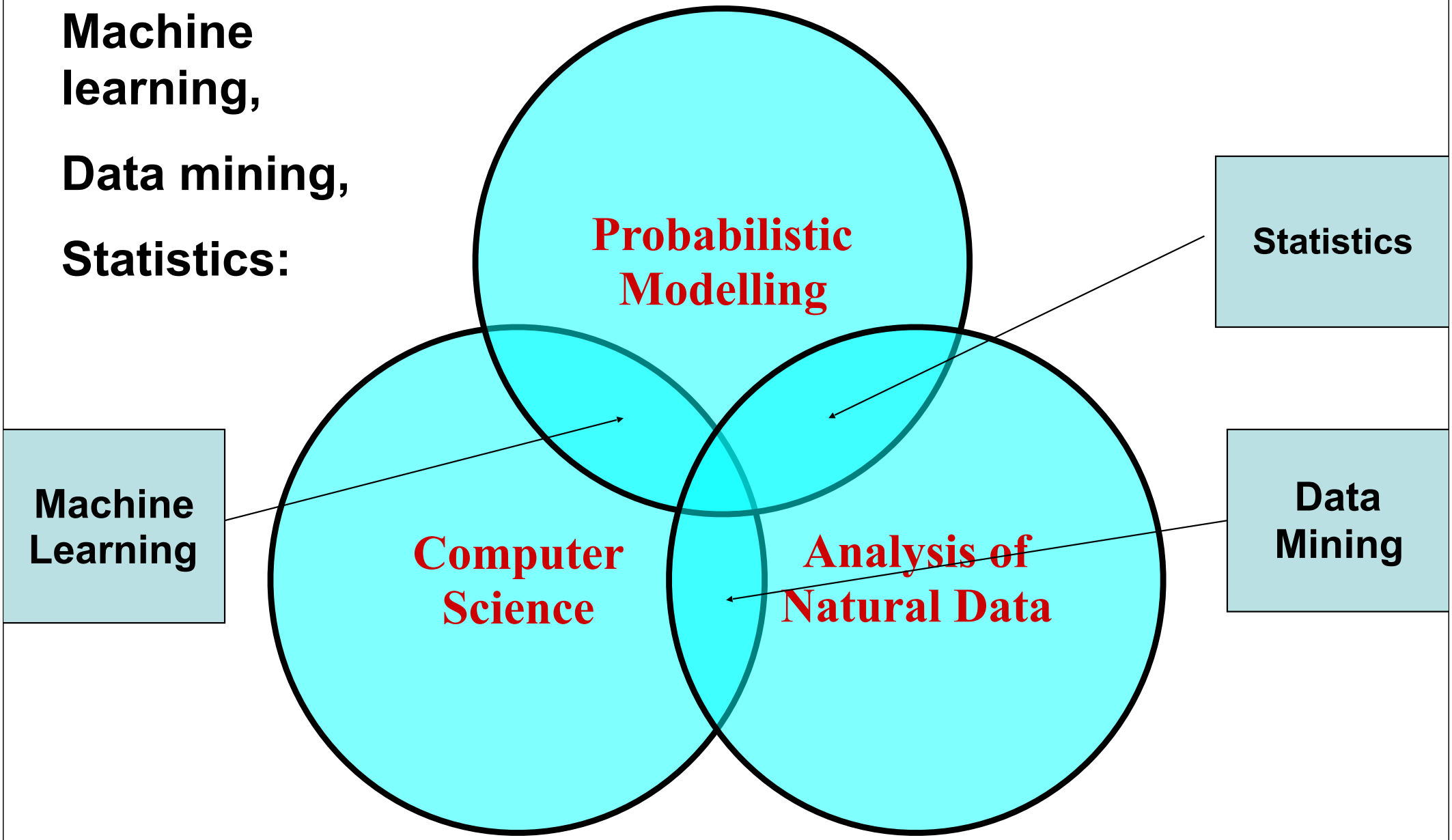
Related to:

- Artificial intelligence
- Computational statistics
- Pattern recognition
- Data mining
- Predictive analytics
- Data science

**Machine learning,**

**Data mining,**

**Statistics:**



# Main types, based on justification

Neural computation, deep neural networks

- one original motivation: take inspiration from brain
- current main motivation: effective with big data and big computational resources

Computational learning theory

- gives mathematical bounds on learning performance

Probabilistic machine learning

- justification as probabilistic modelling and statistical inference

# Probabilistic machine learning and artificial intelligence

Zoubin Ghahramani[1]

doi:10.1038/nature14541

There are two simple rules that underlie probability theory: the sum rule:

$$P(x) = \sum_{y \epsilon Y} P(x,y)$$

and the product rule:

$$P(x,y) = P(x)P(y \mid x).$$

For learning, we thus get:

$$P(\theta \mid D,m) = \frac{P(D \mid \theta,m)P(\theta \mid m)}{P(D \mid m)}$$

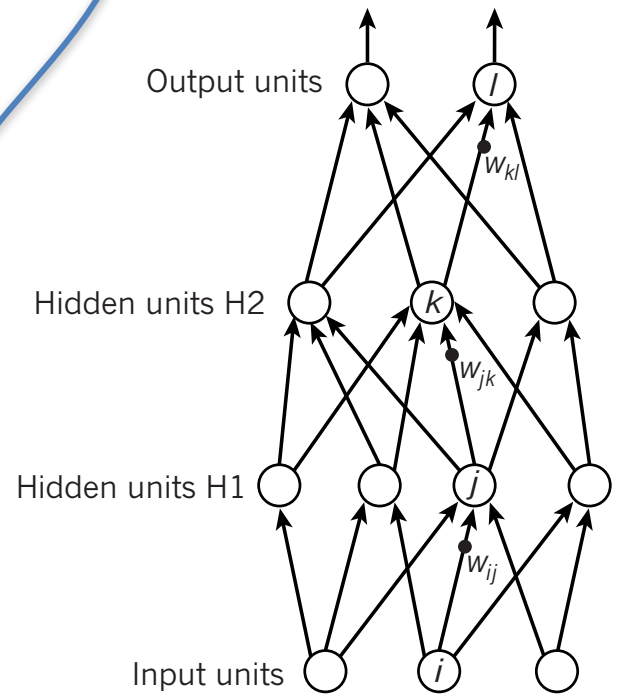
prediction:

$$P(D_{test} \mid D,m) = \int P(D_{test} \mid \theta,D,m)P(\theta \mid D,m)d\theta$$

# Deep learning

Yann LeCun[1,2], Yoshua Bengio[3] & Geoffrey Hinton[4,5]

doi:10.1038/nature14539




$$y_l = f(z_l)$$
$$z_l = \sum_{k \,\varepsilon\, H2} w_{kl}\, y_k$$

$$y_k = f(z_k)$$
$$z_k = \sum_{j \,\varepsilon\, H1} w_{jk}\, y_j$$

$$y_j = f(z_j)$$
$$z_j = \sum_{i \,\varepsilon\, Input} w_{ij}\, x_i$$

# Tensions

- black-box vs white-box
- heuristic vs rigorous (neats vs scruffies)

A bit funny dichotomies, because obviously we ultimately want a combination of both!

Example: Bayesian deep learning

# Main types, based on technical goals

Supervised learning

Unsupervised learning

Reinforcement learning

These are conceptual tools, meant to help think about machine learning tasks. Many complementary concepts exist:

- active learning
- interactive learning
- transfer learning …

Real operational systems usually combine many types and other elements: decision theory, control theory, …

# Main topics in ICML2017, 1/2

- Active Learning
- Approximate Inference
- Bayesian Nonparametric Methods
- Causal Inference
- Clustering
- Computational biology and bioinformatics
- Computational Learning Theory
- Computational Social Sciences
- Computer Vision
- Dimensionality Reduction
- Economics, Game Theory and Mechanism Design
- Ensemble Methods
- Feature Selection
- Gaussian Processes
- Generative Models
- Graphical Models
- Neuroscience and Cognitive Science
- Online Learning
- Optimization (Combinatorial)
- Optimization (Continuous)
- Other Applications
- Other Models and Methods
- Parallel and Distributed Learning
- Planning and Control
- Privacy, Anonymity, and Security
- Probabilistic Programming
- Ranking and Preference Learning
- Recommender Systems
- Reinforcement Learning
- Representation Learning
- Robotics
- Semi-Supervised Learning

# Main topics in ICML2017, 2/2

- Health Care
- Information Retrieval
- Information Theory
- Kernel Methods
- Large Scale Learning and Big Data
- Matrix Factorization
- Metric Learning
- Monte Carlo Methods
- Multi-Agent Learning
- Natural Language and Speech Processing
- Networks and Relational Learning
- Neural Networks and Deep Learning
- Society Impacts of Machine Learning
- Sparsity and Compressed Sensing
- Spectral Methods
- Statistical Learning Theory
- Structured Prediction
- Supervised Learning
- Sustainability, Climate, and Environment
- Systems and Software
- Time-Series Analysis
- Transfer and Multi-Task Learning
- Unsupervised Learning

# Why a machine learning boom right now?

- a lot of data in many problems, “big data”
- computational resources
- software and computational infrastructure
- fashionable to call various data analysis systems machine learning or AI

# Summary

Machine learning = techniques that learn based on data to solve problems

# Machine learning in Finland

Microsoft Academic Search (Beta)

Advanced Search

- Authors »
- Publications »
- Conferences »
- Journals »
- Keywords »
- Organizations »

Academic > Computer Science > Machine Learning & Pattern Recognition

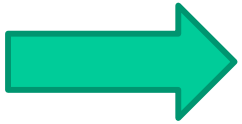
**Top organizations in machine learning & pattern recognition**

1–100 of 5,678 results

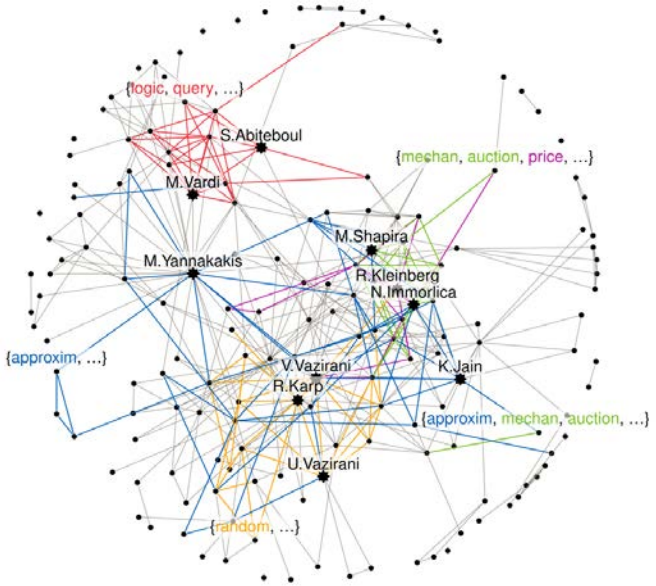
All Years | All Continents

1 2 3 4 5 6

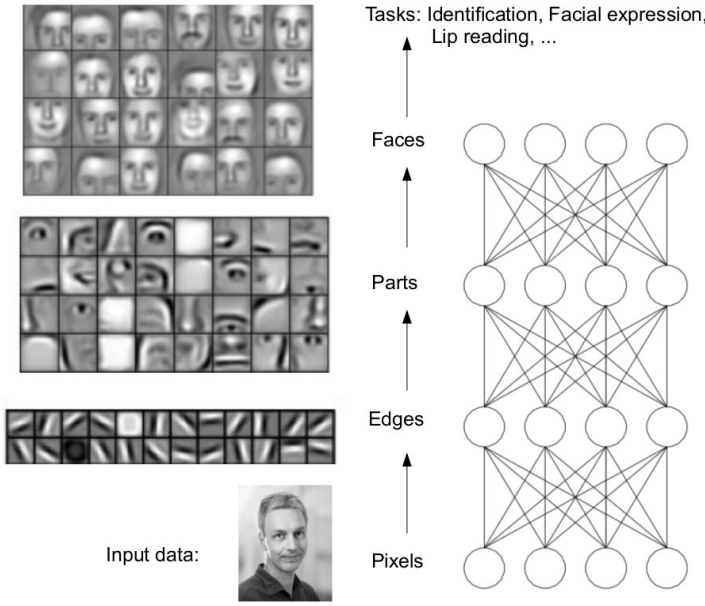
| Organizations | Publications | Citation |
|---|---|---|
| Microsoft | 2474 | 114275 |
| Massachusetts Institute of Technology | 2048 | 108683 |
| University of California Berkeley | 1644 | 99559 |
| Stanford University | 1606 | 79912 |
| University of California San Diego | 1274 | 74937 |
| Carnegie Mellon University | 2245 | 69519 |
| Princeton University | 562 | 53104 |
| IBM | 1483 | 41850 |
| Google Inc. | 865 | 40070 |
| Aalto University | 932 | 39748 |
| University College London | 1134 | 37166 |



# Samples of machine learning and data mining




Detection of dense communities in social networks and attributes that describe the communities



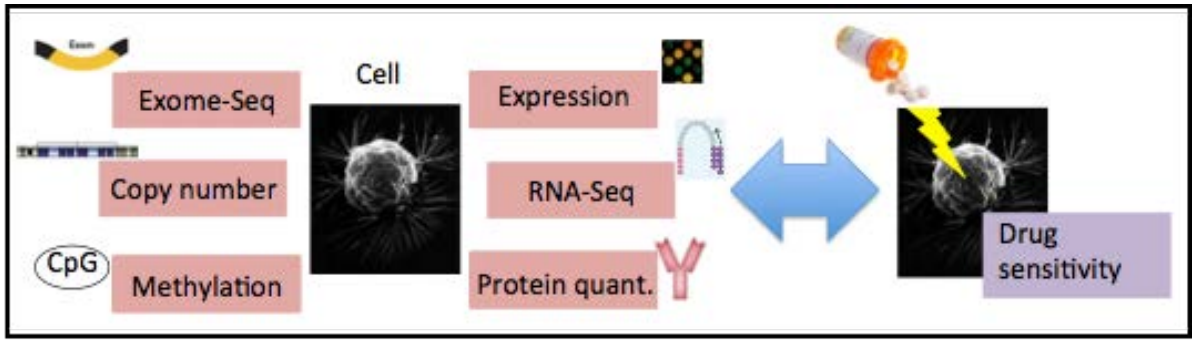

Deep Learning to derive abstract high-level features in a hierarchical manner that imitates human perception.

Google

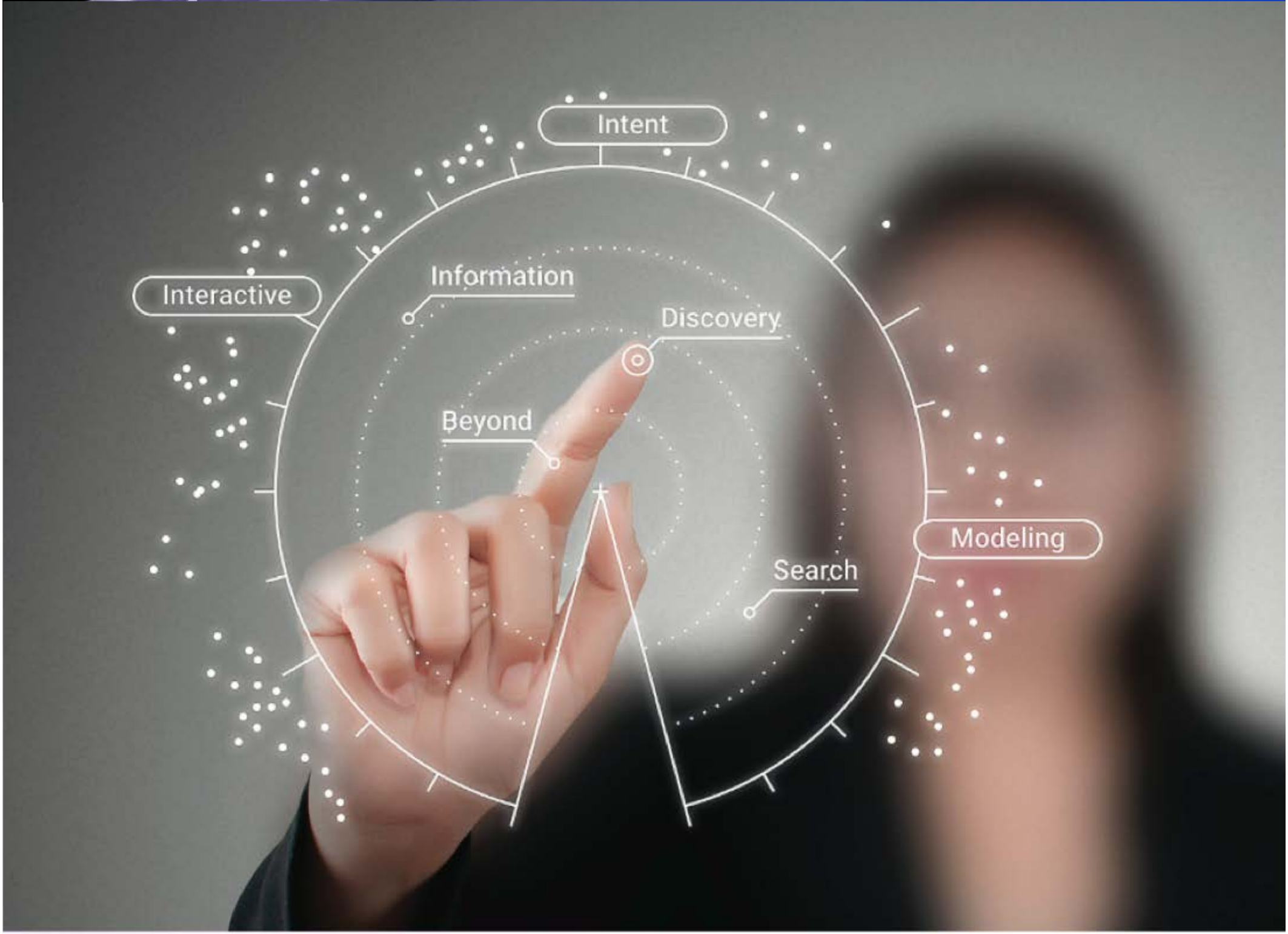
amazon

ETSIMO
discovery beyond search

Plenty of jobs



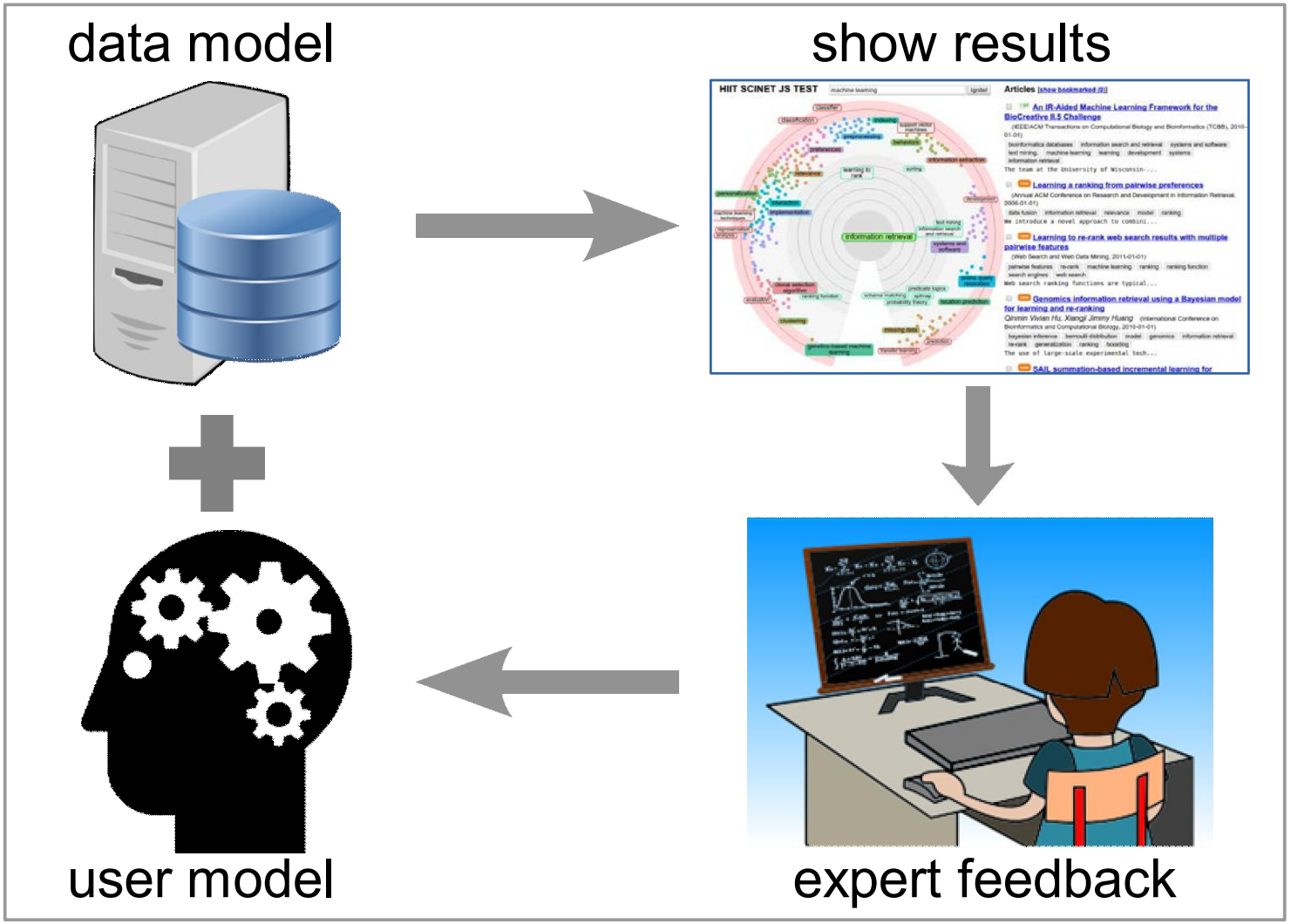

Personalized medicine to make individually tailored treatment decisions based on genomic and other data


Intent
Information
Interactive
Discovery
Beyond
Modeling
Search


HELSINKI INSTITUTE FOR INFORMATION TECHNOLOGY

Glowacka et al. IUI 2013,
Ruotsalo et al. Commun ACM 2015, …

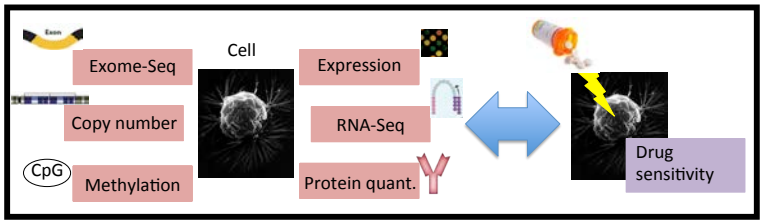
# Wanted: Interactive modelling




HELSINKI INSTITUTE FOR INFORMATION TECHNOLOGY

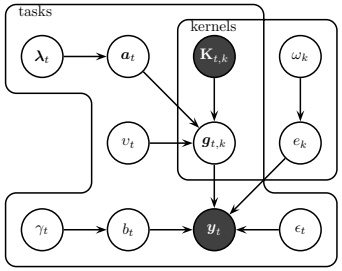
# Tietotekniikan menetelmillä tehostettu tutkimus

## Tietoaineistot




## Adaptiivisilla käyttöliittymillä mukaan saatu asiantuntemus

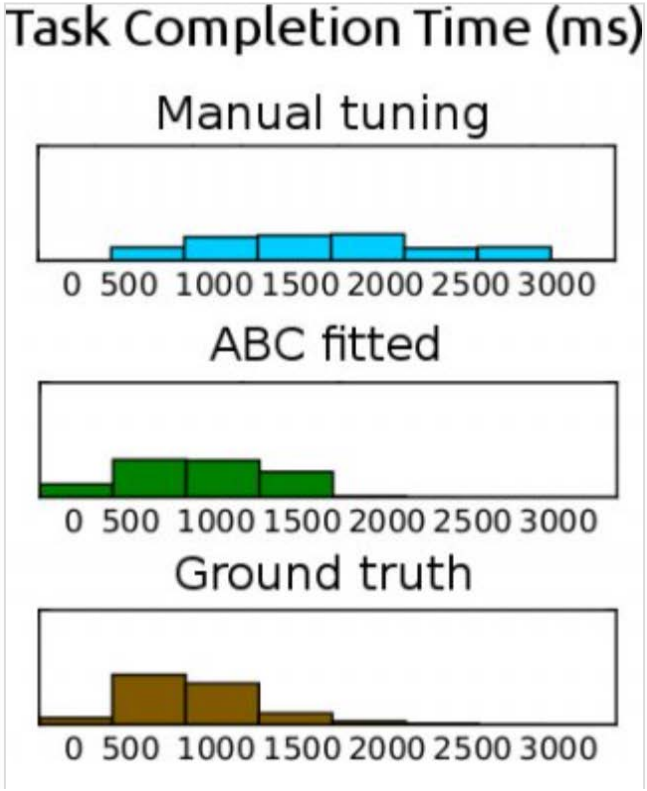

## Yhdistetään koneoppimismenetelmillä




$$\lambda_{t,i} \sim \mathcal{G}(\lambda_{t,i}; \alpha_\lambda, \beta_\lambda)$$
$$a_{t,i}|\lambda_{t,i} \sim \mathcal{N}(a_{t,i}; 0, \lambda_{t,i}^{-1})$$
$$v_t \sim \mathcal{G}(v_t; \alpha_v, \beta_v)$$
$$\boldsymbol{g}_{t,k}|\boldsymbol{a}_t, \mathbf{K}_{t,k}, v_t \sim \mathcal{N}(\boldsymbol{g}_{t,k}; \mathbf{K}_{t,k}\boldsymbol{a}_t, v_t^{-1}\mathbf{I})$$
$$\omega_k \sim \mathcal{G}(\omega_k; \alpha_\omega, \beta_\omega)$$
$$e_k|\omega_k \sim \mathcal{N}(e_k; 0, \omega_k^{-1})$$
$$\gamma_t \sim \mathcal{G}(\gamma_t; \alpha_\gamma, \beta_\gamma)$$
$$b_t|\gamma_t \sim \mathcal{N}(b_t; 0, \gamma_t^{-1})$$
$$\epsilon_t \sim \mathcal{G}(\epsilon_t; \alpha_\epsilon, \beta_\epsilon)$$
$$\boldsymbol{y}_t|b_t, e_\cdot, \boldsymbol{g}_{t,\cdot}, \epsilon_t \sim \mathcal{N}\left(\boldsymbol{y}_t; \sum_{k=1}^{K} e_k \boldsymbol{g}_{t,k} + b_t \mathbf{1}, \epsilon_t^{-1}\mathbf{I}\right)$$

Saadaan tarkempia ennusteita, joiden perusteella voidaan tehdä tarkempia päätöksiä (esim. valita hoidot yksilöllisemmin)

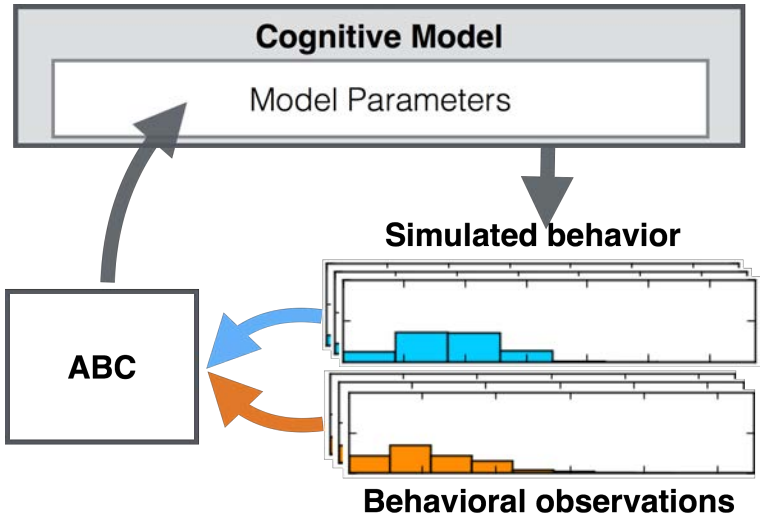
# Highlights




Wed, 03.05.2017

## Computers learn to understand humans better by modelling them

HIIT researchers collaborating with University of Birmingham and University of Oslo present results paving the way for computers to learn psychologically plausible models of individuals simply by observing them. In newly published CHI'17 article,...



Aalto University

# Koneoppimisen eri muodot

**Samuel Kaski**

**4.5.2017**