CHALLENGES IN PREDICTING THE FUTURE

Jan Larsen, Professor PhD
DTU Compute
Department of Applied Mathematics and Computer Science

DTU

LAW ENFORCEMENT

Discovering,
deterring
frustrating,
rehabilitating,
punishing

people who violate the law

Ref: wikipedia

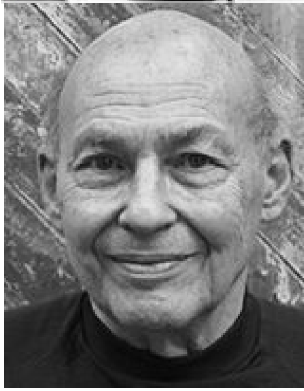
Predictive Policing is the usage of predictive and analytical techniques in law enforcement

- **predicting crimes**
- **predicting offenders**
- **predicting perpetrators‘ identities**
- **predicting victims of crime**

https://issuu.com/rutgerrienks/docs/predictive_policing_rienks_uk

Rutger Rienks, *Predictive Policing: Taking a Chance for a Safer Future*, 2015.

Brief history



Late 40's Allan Touring: theory of computation

1948 Claude Shannon: A Mathematical Theory of Communication

1948 Norbert Wiener: Cybernetics - *Control and Communication in the Animal and the Machine*

1950 The Touring test

1951 Marvin Minsky's analog neural networks (1st revolution)

1956 Dartmouth conference: Artificial intelligence with aim of human like intelligence

1956-1974 Many small scale "toy" projects in robotics, control and game solving

1974 Failure of success and Minsky's criticism of perceptron, lack of computational power, combinatorial explosion, Moravec's paradox: simple tasks are not easy to solve

1980's Expert systems useful in restricted domains

1980's Knowledge based systems – integration of diverse information sources

1980's The 2nd neural network revolution starts

Late 1980's Robotics and the role of embodiment to achieve intelligence

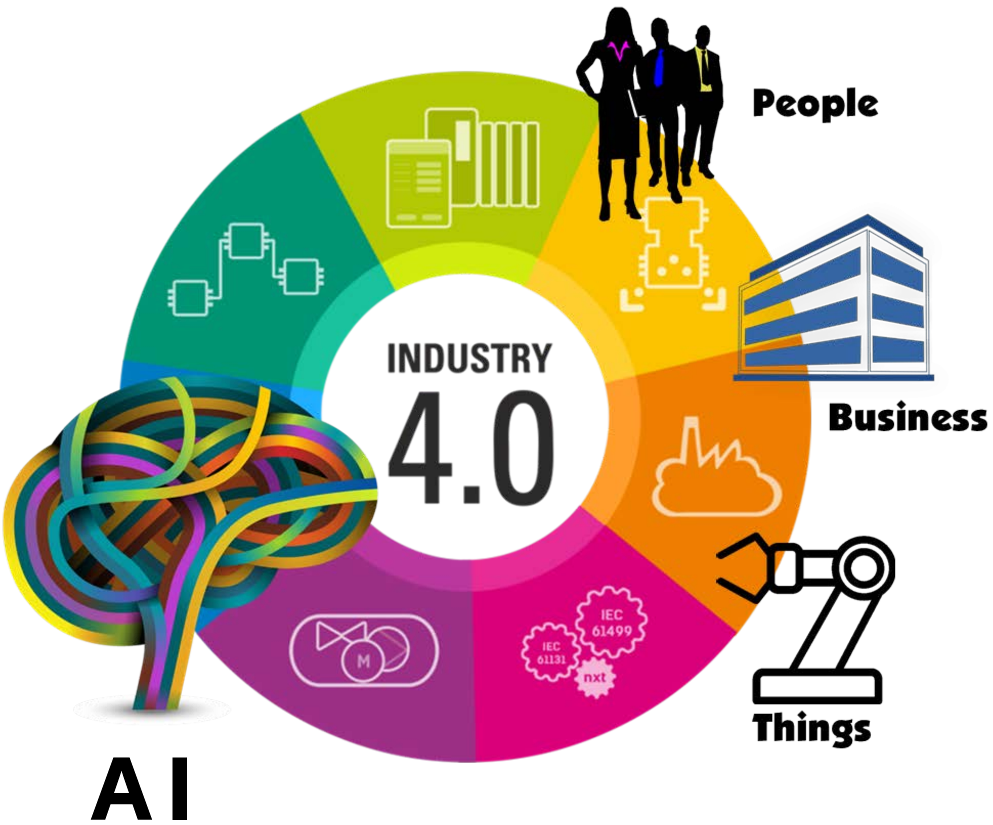
1990's AI and cybernetics research under new names such as machine learning, computational intelligence, evolutionary computing, neural networks, Bayesian networks, complex systems, game theory, deep neural networks (3rd generation) **cognitive systems**

2010's deep neural networks (4rd generation) and cognitive systems, large scale data and computational frameworks, ML is commoditized

http://en.wikipedia.org/wiki/Timeline_of_artificial_intelligence

http://en.wikipedia.org/wiki/History_of_artificial_intelligence

A copy of the physical world through digitization makes it possible for cyber-physical systems to communicate and cooperate with each other and with humans in real time and perform decentralized decision-making



https://en.wikipedia.org/wiki/Industry_4.0

B. Marr: Forbes, June 20, 2016, http://www.forbes.com/sites/bernardmarr/2016/06/20/what-everyone-must-know-about-industry-4-0/#4c979f804e3b

http://www.enterrasolutions.com/2015/10/industry-4-0-facing-the-coming-revolution.html

Industry 4.0 = Civilization 4.0

It is a cognitive revolution that could be even more disruptive than earlier as it concerns not only the industry but the whole way we live our lives.

AI

Artificial Intelligence

IA

Intelligence Augmentation

signal processing – processing of data

machine learning – ubiquitous learning from data

cognitive systems – making data relevant and understandable for people – and making people understand of the world

Modeling interaction and fusion of sensor signals (audio), related information, and information from humans

research focus

CoSound

Processing of audio data and related information, such as context, users' states, interaction, intention, and goals with the purpose of providing innovative services related to relevant societal challenges in

Transforming big audio data into semantically interoperable data assets and knowledge: enrichment and navigation in large sound archives such as broadcast

Experience economy and edutainment: new music services based on mood, optimization of sound systems

Healthcare: Music interventions to improve quality of life in relation to disorders such as e.g. stress, pain, and ADHD

User-driven optimization of hearing aids

research focus

MakeSense

Processing of sensor signals and related data streams with the purpose of fostering innovative systems addressing societal challenges in

Food: Grain analysis

Security: Explosives and drug detection

Health: blood and water analysis, intelligent drug delivery and sensing, e-health

Energy: wind mill maintenance

Environment: exhaust gas analysis, large diesel engine monitoring

Resource efficiency: waste sorting

Digital economy: event recommendation

What is machine learning?

Learning structures and patterns form from historical data to reliably predict outcome for new data.

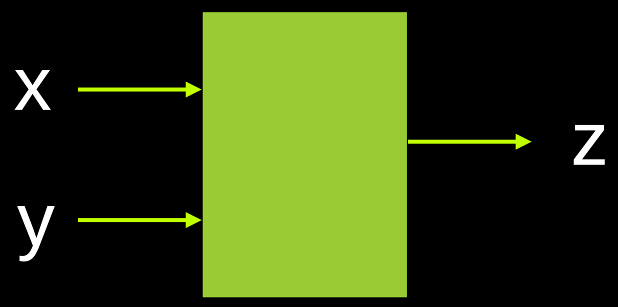
Computers only do what they are programmed to do. ML infers new relations and patterns, which were not programmed. They learn and adapt to changing environment.

1. Computer systems that automatically improve through experience, or learns from data.
2. Inferential process that operate from representations that encode probabilistic dependencies among data variables capturing the likelihoods of relevant states in the world.
3. Development of fundamental statistical computational-information-theoretic laws that govern learning systems - including computers, humans, and other entities.

M. I. Jordan and T. M. Mitchell. *Machine learning: Trends, perspectives, and prospects*. Science, July 2015.

Samuel J. Gershman, Eric J. Horvitz, Joshua B. Tenenbaum. *Computational rationality: A converging paradigm for intelligence in brains, minds, and machines.* Science, July 2015.

Learning from data: human and machine



x	y	z
7	3	41021
5	2	3710
17	8	925136

Mathematical model

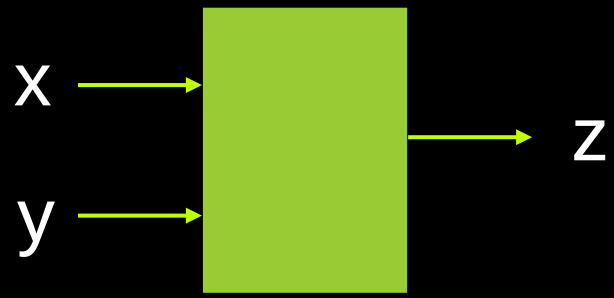
z=(x-y)*10^(floor(log10(x+y))+floor(log10(x)+log10(y))+2)

+(x+y)*10^(floor(log10(x)+log10(y))+1)

+(x*y), if x>y, and x>0, and y>0

Human assumptions and interpretation/description are maybe very different

Learning from data: human and machine



x	y	z
7	3	41021
5	2	3710
17	8	925136

How do we handle values outside observations: what happens if values are negative?

Does the machine have the right flexibility and capacity?

What is human prior knowledge?

How does context provide additional constraints?

Can we learn anything from very limited data?

Deep Learning: Automating Feature Discovery

Geoff Hinton, Yoshua Bengio, Yann LeCun, Deep Learning Tutorial, NIPS 2015, Montreal.

Deep learning is a disruptive technology

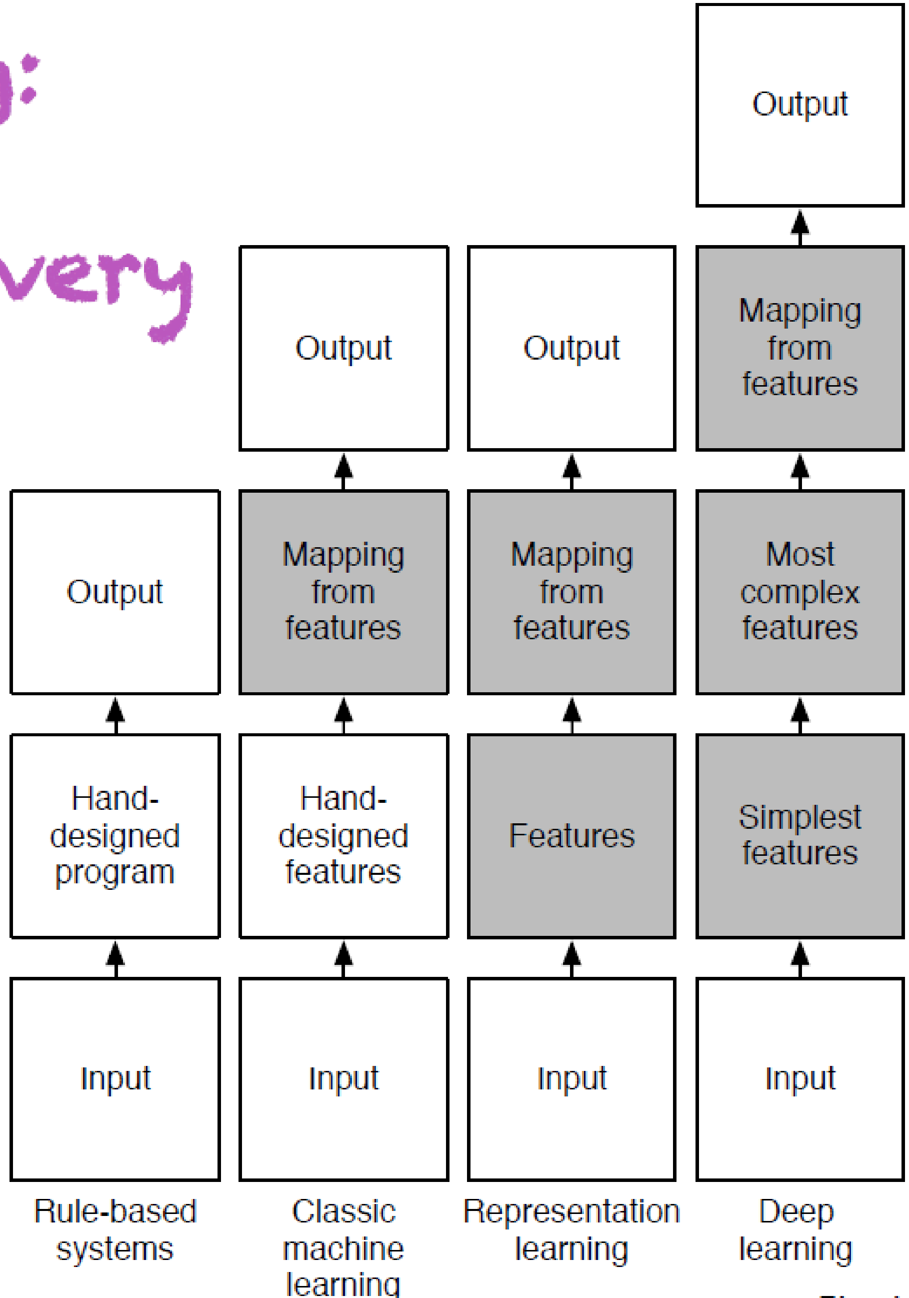
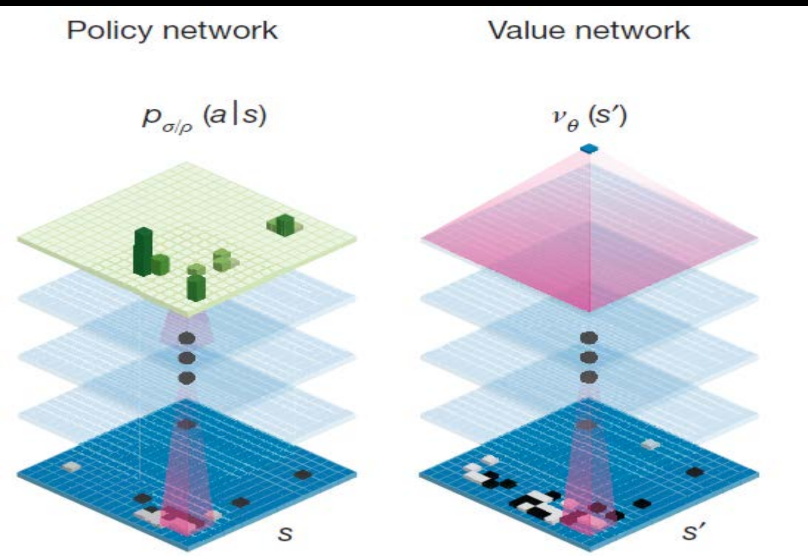


Fig: I. Goodfellow

Machine learning is very successful: playing GO

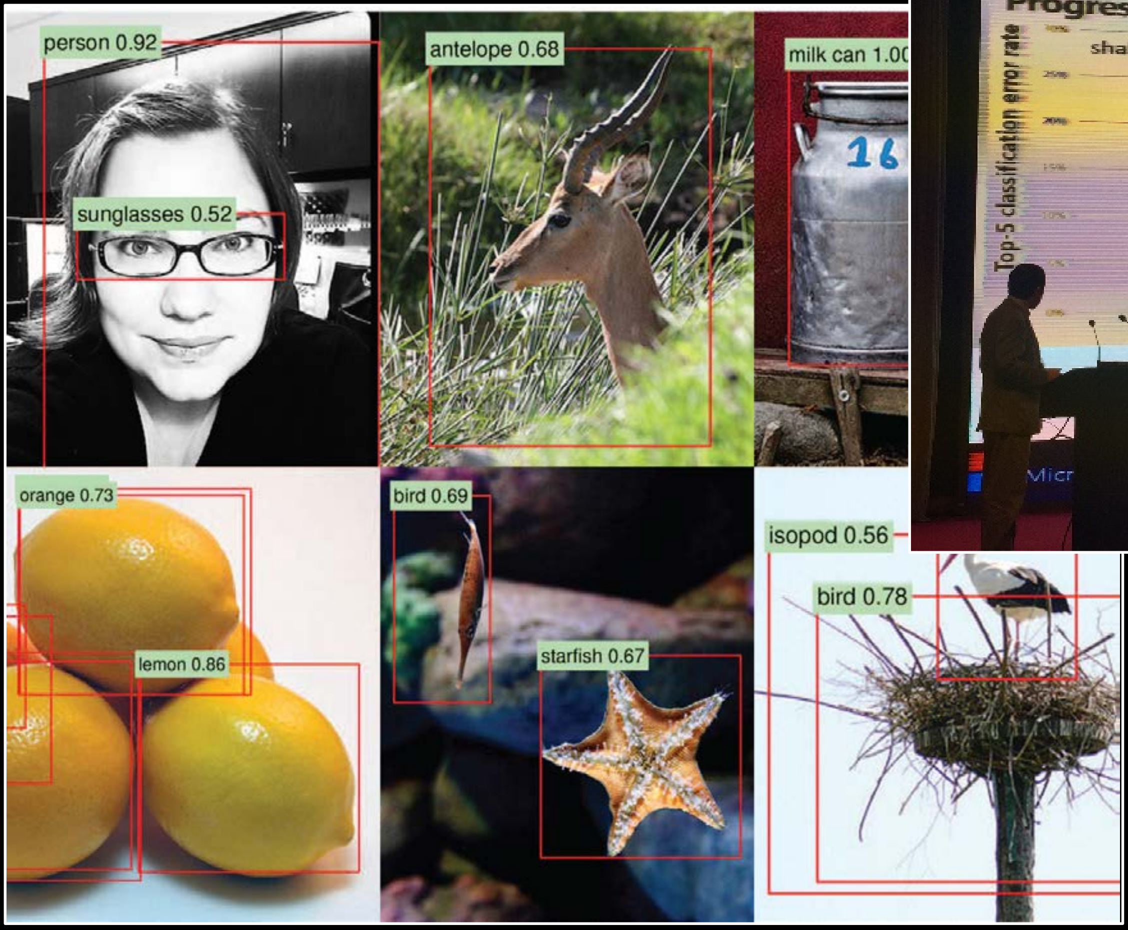


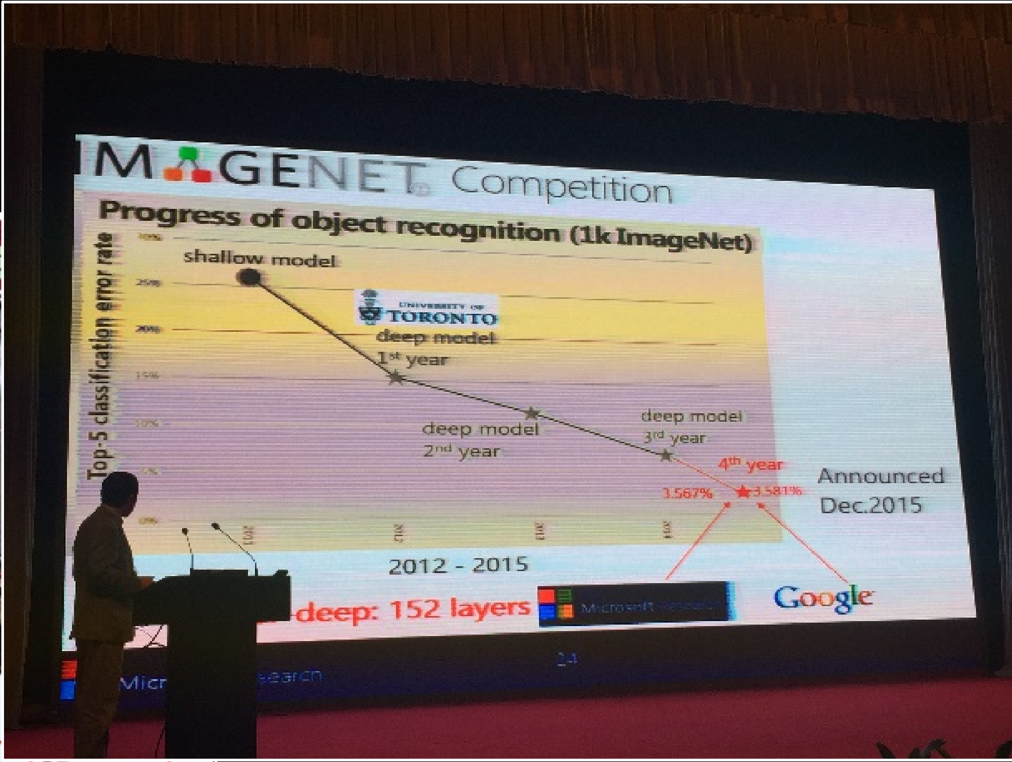
Deep neural 'value networks' evaluate board positions and other 'policy networks' select moves.

Networks are trained by a novel combination of supervised learning from human expert games, and reinforcement learning from games of self-play.

Silver, David; Huang, Aja; Maddison, Chris J.; Guez, Arthur; Sifre, Laurent; Driessche, George van den; Schrittwieser, Julian; Antonoglou, Ioannis; Panneershelvam, Veda. *Mastering the game of Go with deep neural networks and tree search. Nature 529(7587): 484–489, 2016*

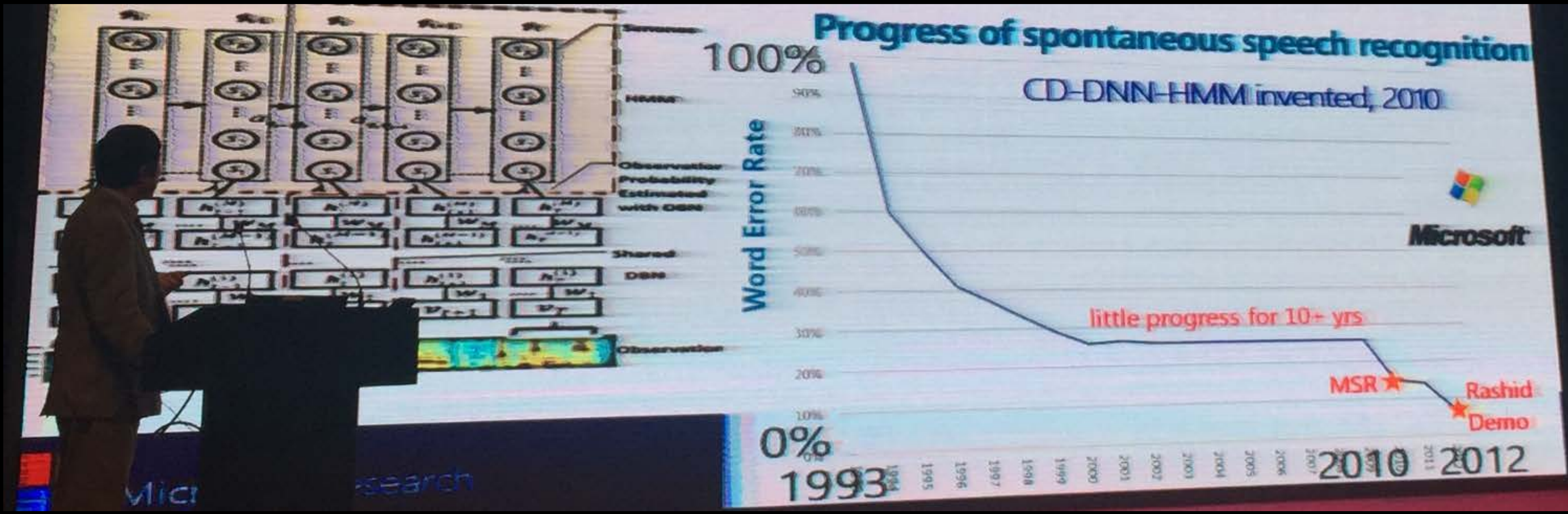
Machine learning is very successful: computer vision





M. I. Jordan and T. M. Mitchell. *Machine learning: Trends, perspectives, and prospects*. Science, July 2015.

Machine learning is very successful: speech recognition and chat bots



Human parity is achieved Feb/March 2017

Geoffrey Hinton, Li Deng, Dong Yu, George Dahl, Abdel-rahman Mohamed, Navdeep Jaitly, Andrew Senior, Vincent Vanhoucke, Patrick Nguyen, Tara Sainath, and Brian Kingsbury. *Deep Neural Networks for Acoustic Modeling in Speech Recognition.* IEEE Signal Processing Magazine, 82, Nov. 2012.

George Saon, Gakuto Kurata, Tom Sercu, Kartik Audhkhasi, Samuel Thomas, Dimitrios Dimitriadis, Xiaodong Cui, Bhuvana Ramabhadran, Michael Picheny, Lynn-Li Lim, Bergul Roomi, Phil Hall. *English Conversational Telephone Speech Recognition by Humans and Machines, https://arxiv.org/abs/1703.02136, March 2017*

W. Xiong, J. Droppo, X. Huang, F. Seide, M. Seltzer, A. Stolcke, D. Yu, G. Zweig. *Achieving Human Parity in Conversational Speech Recognition, https://arxiv.org/abs/1610.05256, October 2016.*

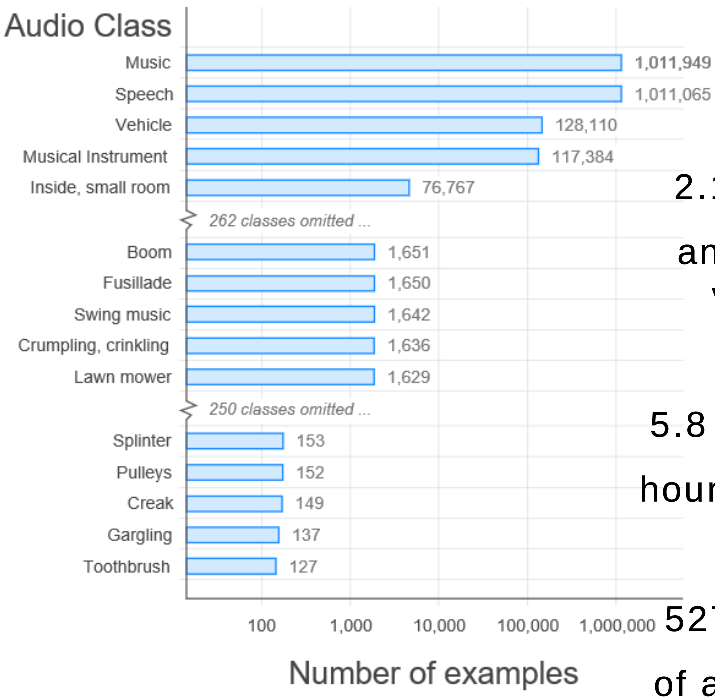
Private traits and attributes are predictable from digital records of human behavior

Facebook Likes, can be used to automatically and accurately predict a range of highly sensitive personal attributes including: sexual orientation, ethnicity, religious and political views, personality traits, intelligence, happiness, use of addictive substances, parental separation, age, and gender.

Michal Kosinski, David Stillwell and Thore Graepel

PNAS April 9, 2013. 110 (15) 5802-5805

Machine learning is very successful for audio classification



2.1 million annotated videos

5.8 thousand hours of audio

527 classes of annotated sounds

Table 2: Comparison of performance of several DNN architectures trained on 70M videos, each tagged with labels from a set of 3K. The last row contains results for a model that was trained much longer than the others, with a reduction in learning rate after 13 million steps.

Architectures	Steps	Time	AUC	d-prime	mAP
Fully Connected	5M	35h	0.851	1.471	0.058
AlexNet	5M	82h	0.894	1.764	0.115
VGG	5M	184h	0.911	1.909	0.161
Inception V3	5M	137h	**0.918**	**1.969**	0.181
ResNet-50	5M	119h	0.916	1.952	**0.182**
ResNet-50	17M	356h	**0.926**	**2.041**	**0.212**

Mean average precision mAP is low because of low class prior $<10^{-4}$.

AUC is the area under curve of true positive rate vs. false positive rate.

Jort F. Gemmeke, Daniel P. W. Ellis, Dylan Freedman, Aren Jansen, Wade Lawrence, R. Channing Moore, Manoj Plakal, Marvin Ritter. *Audio Set: An ontology and human-labeled dataset for audio events*, IEEE ICASSP 2017, New Orleans, LA, March 2017.

Shawn Hershey, Sourish Chaudhuri, Daniel P. W. Ellis, Jort F. Gemmeke, Aren Jansen, Channing Moore, Manoj Plakal, Devin Platt, Rif A. Saurous, Bryan Seybold, Malcolm Slaney, Ron Weiss, Kevin Wilson. *CNN Architectures for Large-Scale Audio Classification*, ICASSP 2017, New Orleans, LA, March 2017.

What are the issues?

- Massively missing data in specific applications.
- Almost always need for specific small data for personalization or adaptation to specific situation.
- Democratization of data: data should belong to and made available by the creator/user.
- Distributed storage and processing OpenPDS and SafeAnswers (Yves-Alexandre de Montjoye, Imperial College London)
- Privacy may be achieved though privacy aware learning e.g. using differential privacy constraints.

Current challenges in machine learning

- Better semi-supervised learning integrating unsupervised and unsupervised learning to lower requirements on number of data samples.
- Better regularization and incorporation prior information (compositionality, augmented data sets/dream networks).
- More efficient structures for learning to encoding relevant information (independent components, sparsity, autoencoders).
- New (network) more efficient architectures and handling of memory structure.
- More focus on robustness and sensitivity.
- Passive prediction is not enough to achieve real intelligent behavior that is more autonomous.
- Better ability to discover causation.
- Learning from few examples like humans (shared representations).

BLACK BOX OF AI

explanation of "system functionality" and explanation of the "rationale" of an individual decision

Objectives

Trust

Causality

Transferability

Decomposability

Informativeness

Counterfactual explanations

Legal issues

European Union regulations (GDPR) on algorithmic decision-making and a “right to explanation”

Davide Castelvecchi: http://www.nature.com/polopoly_fs/1.20731!/menu/main/topColumns/topLeftColumn/pdf/538020a.pdf, Nature, Vol. 538, 6 Oct. 2016

K.R. Müller and Wojciech Samek: *Explaining and Interpreting Deep Neural Networks, 02901 Advances Topics in Machine Learning, DTU 2017*

Z.C. Lipton: *The mythos of model interpretability*, arXiv:1606.03490, 2016.

Bryce Goodman, Seth Flaxman: *European Union regulations on algorithmic decision-making and a “right to explanation”*, https://arxiv.org/pdf/1606.08813v3.pdf

Adversarial learning

	blues	classical	country	disco	hiphop	jazz	metal	pop	reggae	rock	Pr
blues	92.0	0.0	0.0	0.0	0.0	4.0	0.0	0.0	4.0	8.0	85.2
classical	0.0	84.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0
country	0.0	4.0	92.0	0.0	0.0	0.0	0.0	0.0	0.0	4.0	92.0
disco	8.0	4.0	4.0	80.0	0.0	0.0	0.0	4.0	12.0	16.0	62.5
hiphop	0.0	0.0	0.0	0.0	76.0	0.0	0.0	0.0	8.0	0.0	90.5
jazz	0.0	8.0	0.0	0.0	0.0	92.0	0.0	0.0	4.0	0.0	88.5
metal	0.0	0.0	0.0	0.0	20.0	0.0	92.0	0.0	4.0	0.0	79.3
pop	0.0	0.0	0.0	12.0	0.0	4.0	0.0	92.0	0.0	16.0	74.2
reggae	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.0	64.0	8.0	84.2
rock	0.0	0.0	4.0	8.0	4.0	0.0	8.0	0.0	4.0	48.0	63.2
F	88.5	91.3	92.0	70.2	82.6	90.2	85.2	82.1	72.7	54.5	81.2

(a) Original Input

	blues	classical	country	disco	hiphop	jazz	metal	pop	reggae	rock	Pr
blues	16.0	4.0	0.0	4.0	12.0	4.0	12.0	0.0	12.0	20.0	19.0
classical	16.0	8.0	12.0	16.0	4.0	12.0	8.0	4.0	16.0	4.0	8.0
country	8.0	8.0	4.0	12.0	4.0	0.0	12.0	4.0	4.0	8.0	6.2
disco	8.0	16.0	8.0	8.0	4.0	8.0	12.0	12.0	16.0	12.0	7.7
hiphop	0.0	4.0	16.0	8.0	20.0	40.0	16.0	20.0	0.0	4.0	15.6
jazz	20.0	16.0	12.0	12.0	0.0	8.0	0.0	20.0	8.0	8.0	7.7
metal	4.0	12.0	20.0	12.0	12.0	12.0	4.0	8.0	4.0	12.0	4.0
pop	4.0	4.0	8.0	8.0	20.0	8.0	28.0	20.0	8.0	8.0	17.2
reggae	0.0	16.0	8.0	4.0	12.0	4.0	4.0	8.0	8.0	20.0	9.5
rock	24.0	12.0	12.0	16.0	12.0	4.0	4.0	4.0	24.0	4.0	3.4
F	17.4	8.0	4.9	7.8	17.5	7.8	4.0	18.5	8.7	3.7	10.0

(b) Input intercepted by adversary A1

Corey Kereliuk, Bob L. Sturm, Jan Larsen: Deep Learning and Music Adversaries**,** IEEE Transactions on Multimedia, Nov. 2015

Corey Kereliuk, Bob L. Sturm, Jan Larsen: Deep Learning, Audio Adversaries, and Music Content Analysis, 2015 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, Oct. 2015

Corey Kereliuk, Bob L. Sturm, Jan Larsen: ?El Caballo Viejo? Latin Genre Recognition with Deep Learning and Spectral Periodicity, Fifth Biennial International Conference on Mathematics and Computation in Music (MCM2015), 2015.

Adversarial learning

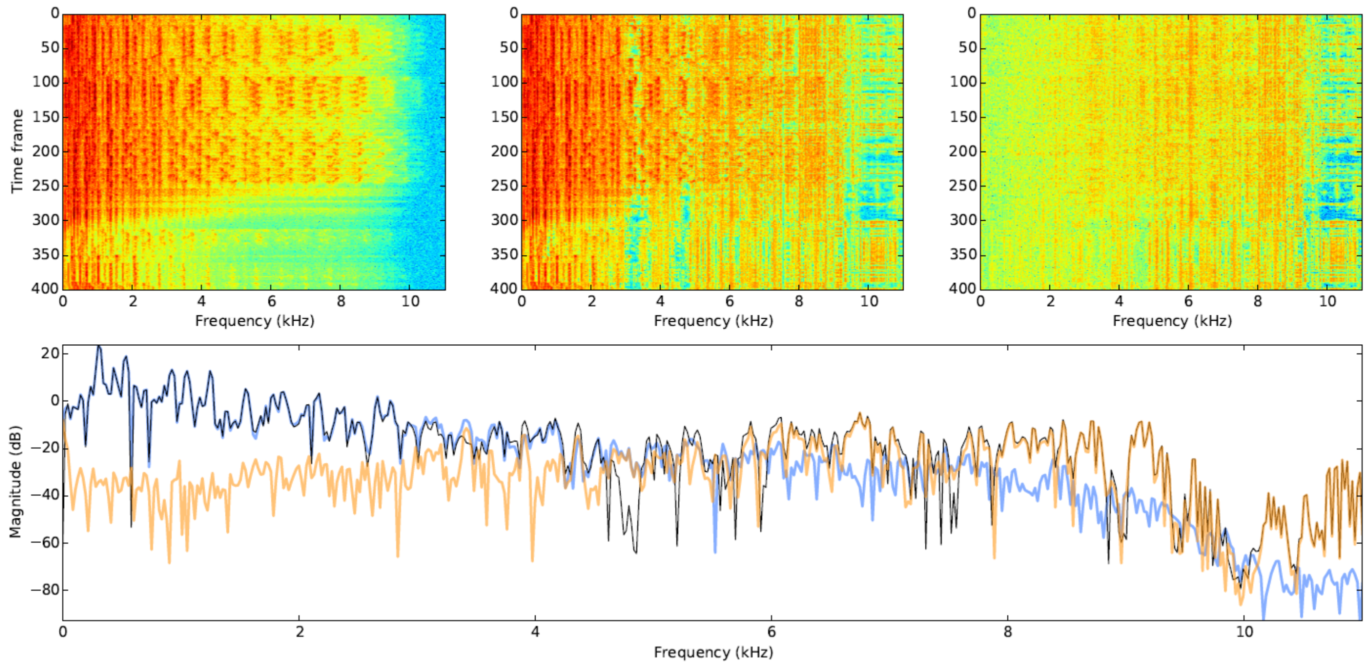


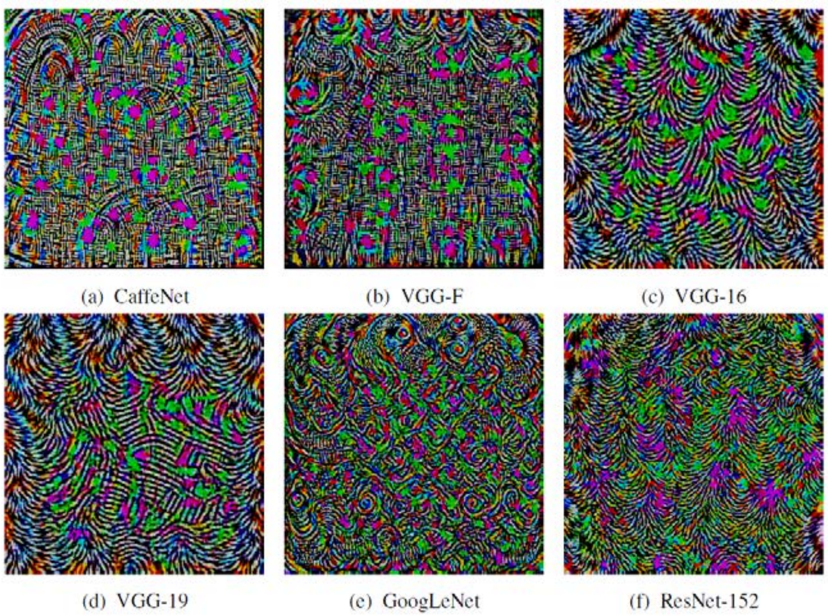
Fig. 5. Top left: spectrogram excerpt from *GTZAN* Classical "21" (Mozart, Symphony No. 39 Finale) that the DNN-based system in Fig. 2(b) classifies as *Classical*. Top middle: spectrogram of adversarial example classified as *Reggae*. Top right: spectrogram of the difference of the two. Bottom: magnitude spectrum of one frame (1024 samples) of the original (light blue), adversarial example (black), and difference (orange). Note that all excerpts in *GTZAN* have a sampling rate of 22050 Hz. The SNR = 21.1dB.

Corey Kereliuk, Bob L. Sturm, Jan Larsen: Deep Learning and Music Adversaries, IEEE Transactions on Multimedia, Nov. 2015

Corey Kereliuk, Bob L. Sturm, Jan Larsen: Deep Learning, Audio Adversaries, and Music Content Analysis, 2015 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, Oct. 2015

Corey Kereliuk, Bob L. Sturm, Jan Larsen: ?El Caballo Viejo? Latin Genre Recognition with Deep Learning and Spectral Periodicity, Fifth Biennial International Conference on Mathematics and Computation in Music (MCM2015), 2015.

Universal Adversarial Learning



Seyed-Mohsen Moosavi-Dezfooli, Alhussein Fawzi, Omar Fawzi, Pascal Frossard: Universal adversarial perturbations, arXiv:1610.08401. 2017

What defines simple and complex problems - and how do we solve them them?

passive

active and autonoumous

exploration and summarization

prediction

continuous learning

reflection

pro-activeness

engagement

experimentation

creativity

Unreasonable effectiveness of

Mathematics E. Wigner, 1960

Data Halevy, Norvig, Pereira, 2009

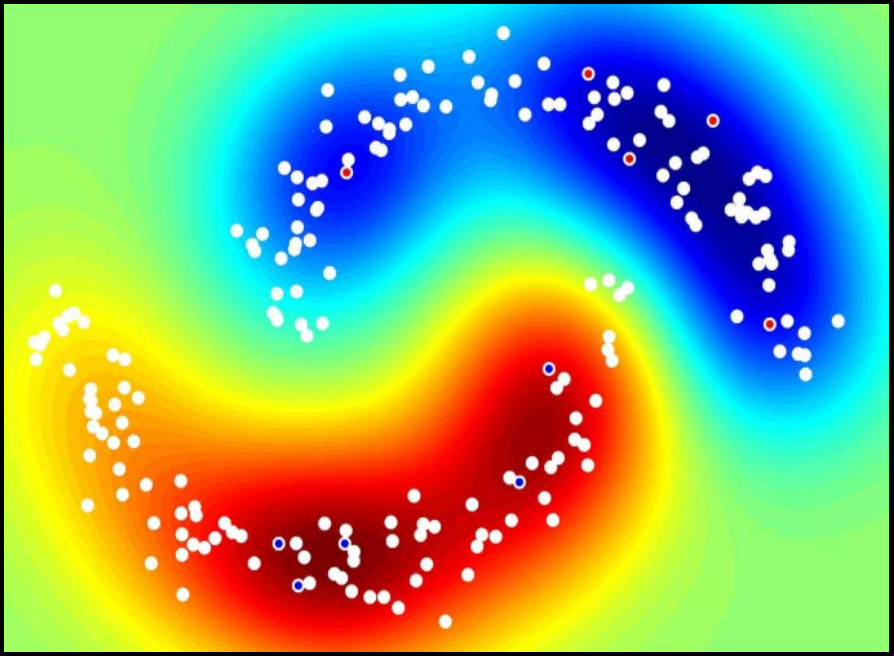
RNNs Karpathy, 2015

Experimentation and interaction through

users-in-the-loop

Unsupervised learning

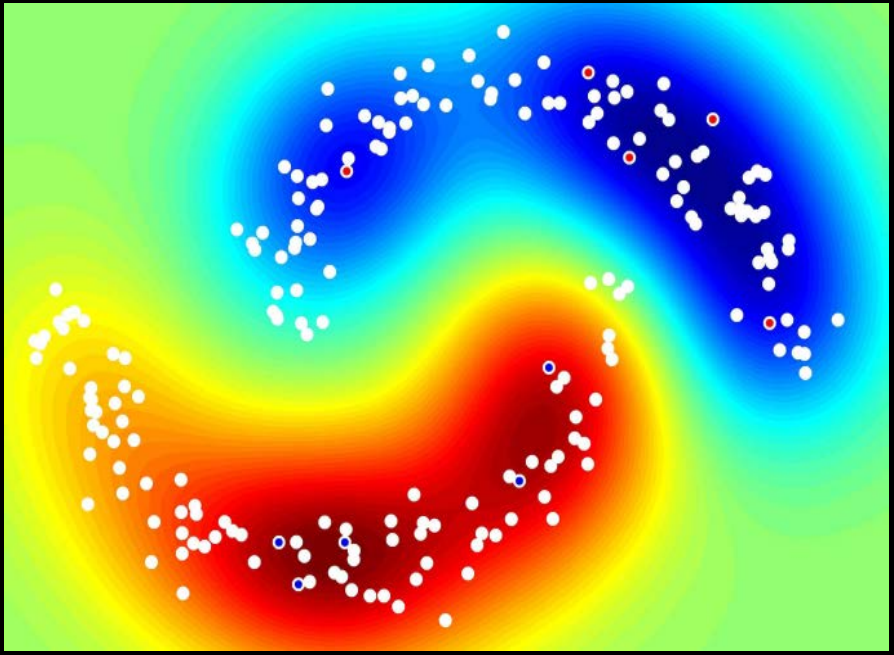
- Probabilistic modeling of structure in multivariate data
- Preprocessing, data reduction, outlier detection, noise reduction, de-convolution, anomaly detection
- Explorative - hypothesis generating



- Clustering
- Linear factor models (ICA, NMF)
- Kernel methods (nonlinear, non-parametric)
- Autoencoder deep neural networks

Supervised learning

- Predictive inference - from sensory features to decisions
- Bayesian hypothesis testing
- Learning from data set of simultaneous sensory input observations (features) and outcome (labels)



- Deep Neural networks
- Non-prametric Kernel machines
- Bayesian learning

Semi-supervised learning

- Learning from combined labeled and unlabeled data
- Optimal use of inexpensive unlabeled data
- Quantification of robustness

Active learning

- Active learning – relates to semi-supervised learning in which samples are initially unknown
- Methods help deciding which (expensive) samples improve learning the most

The power of human data

Humans as a measurement device - why

- With the purpose of individualization and dynamical response
- With the purpose of group studies and population models
- For eliciting perceptual, affective, and cognitive aspects
- For other aspects e.g. behavioral and physical
- For quality measurement and control
- Provide information which can not be verbalized

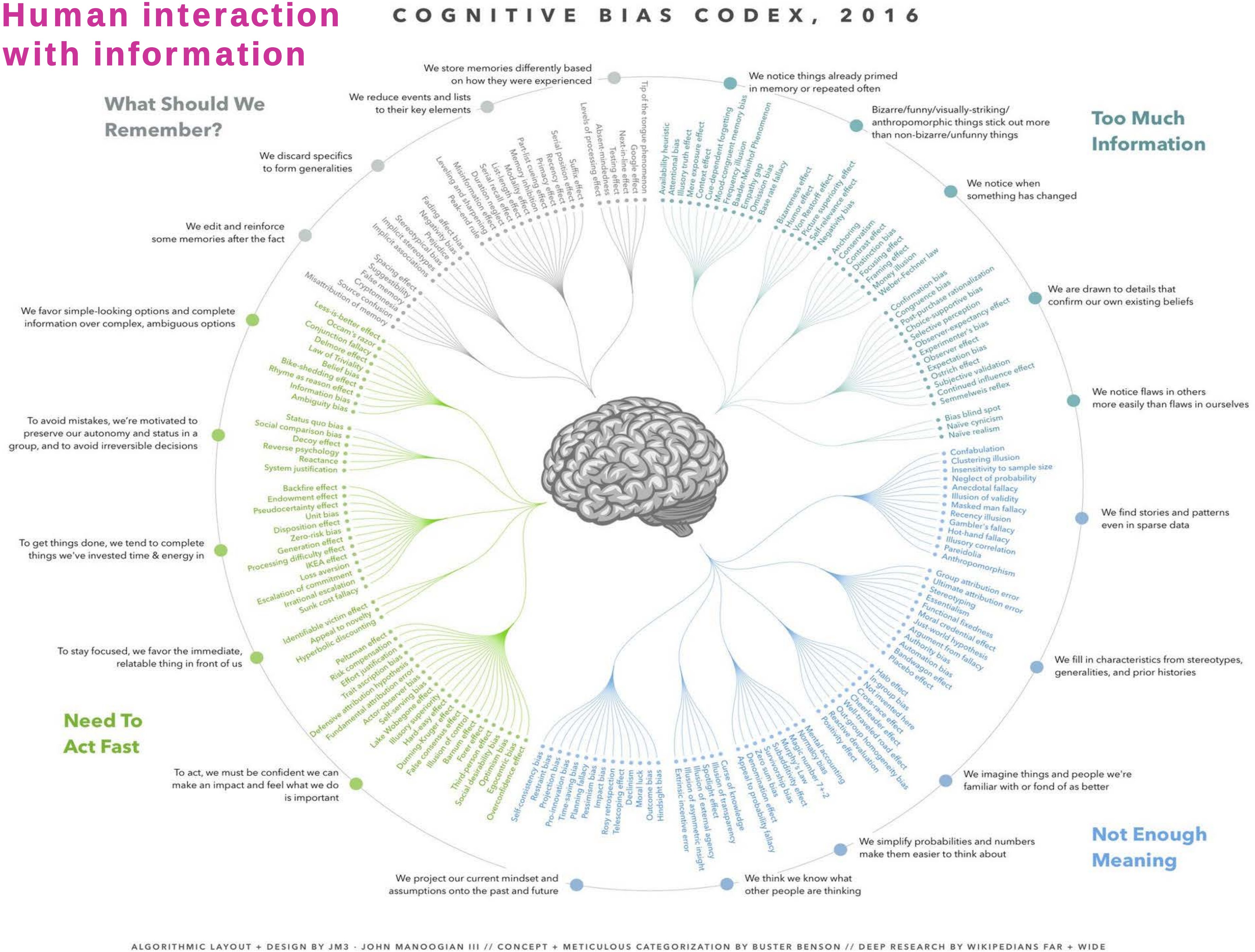
Humans in the loop – how

- Direct measurement of physiological, cognitive and behavior states from physical devices
- Indirect measurements from self-reports, experiments using direct, indirect and related scaling methods
- Indirect measurement of unconscious/uncontrolled behavior

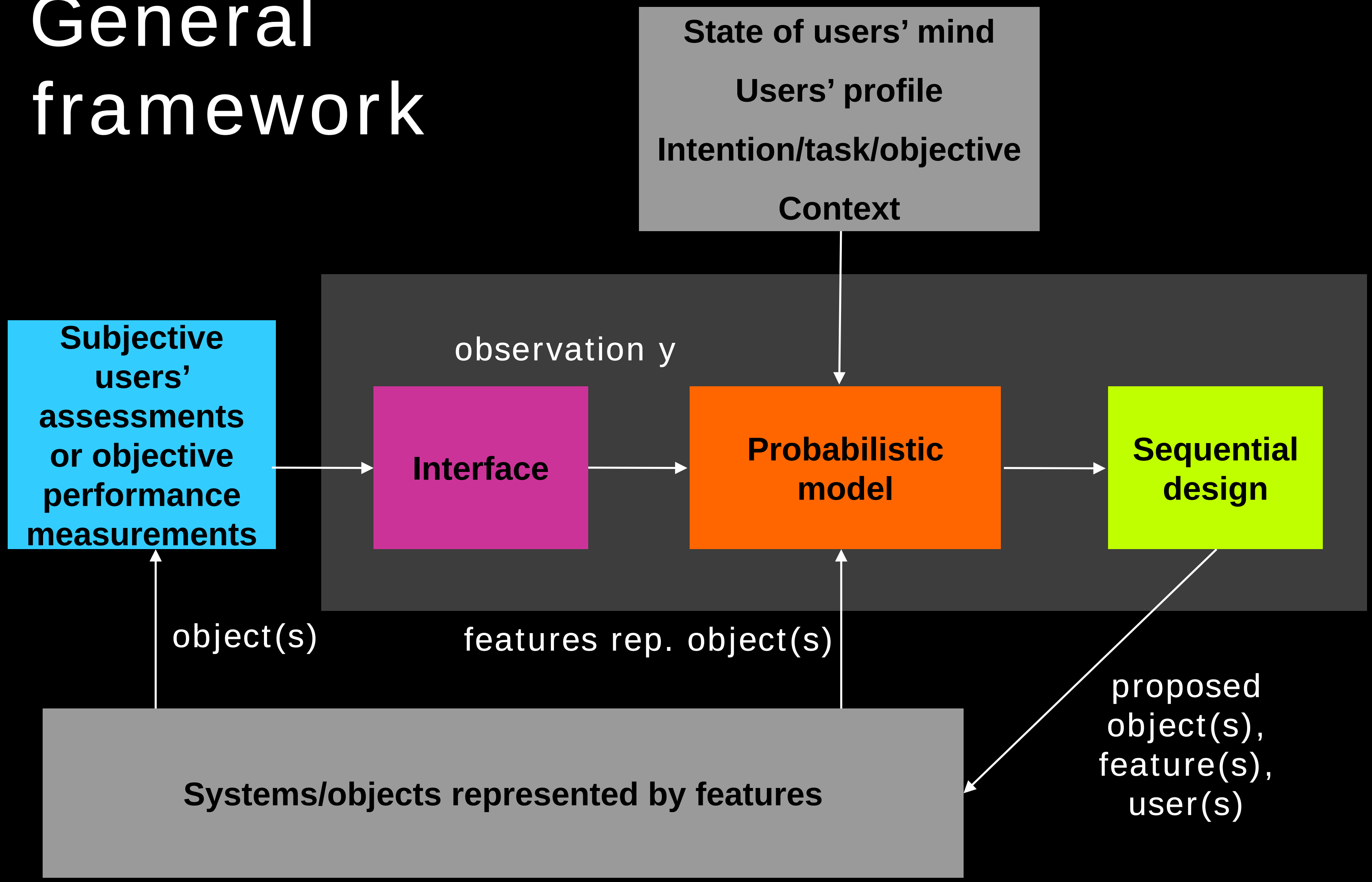
Humans in the loop - who

- End-user
- Experimenter
- Developer
- Expert user
- Collaborative, transfer learning for crowds of humans

Human interaction with information



General framework



Interactive Learning / Sequential Experimental Design

Generalization

Eliciting and learning the entire model / objective function.

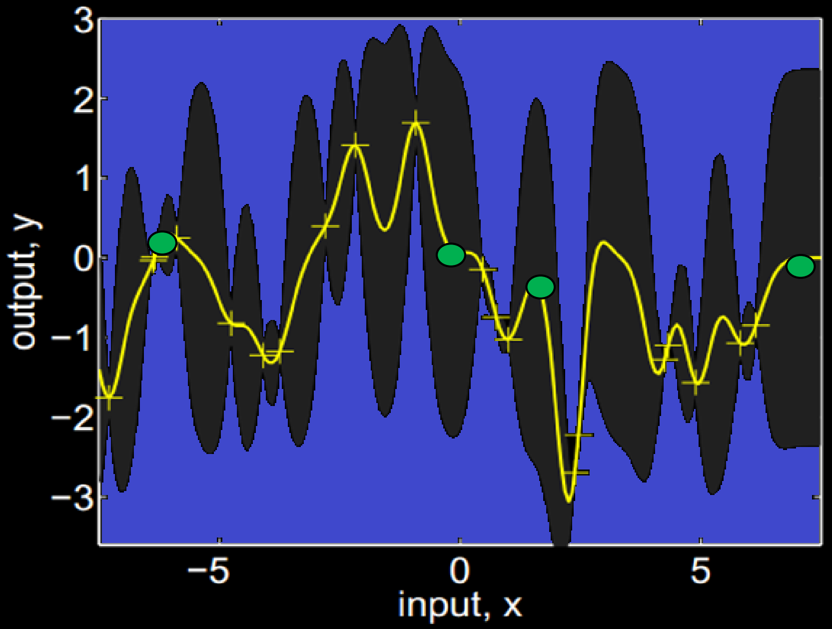
Expected change in relative entropy is derived from the posterior and predictive distribution.

Optimization

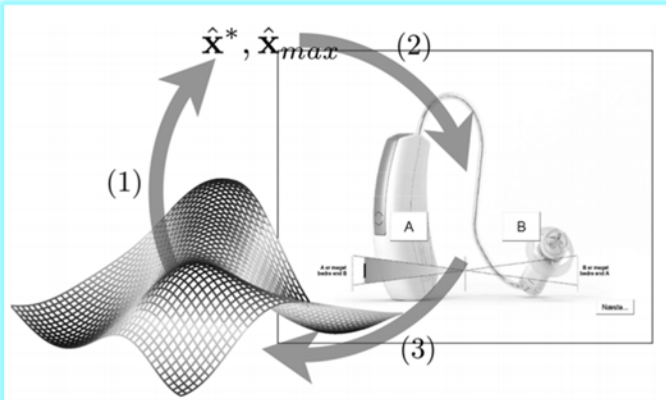
Learning and identifying optimum

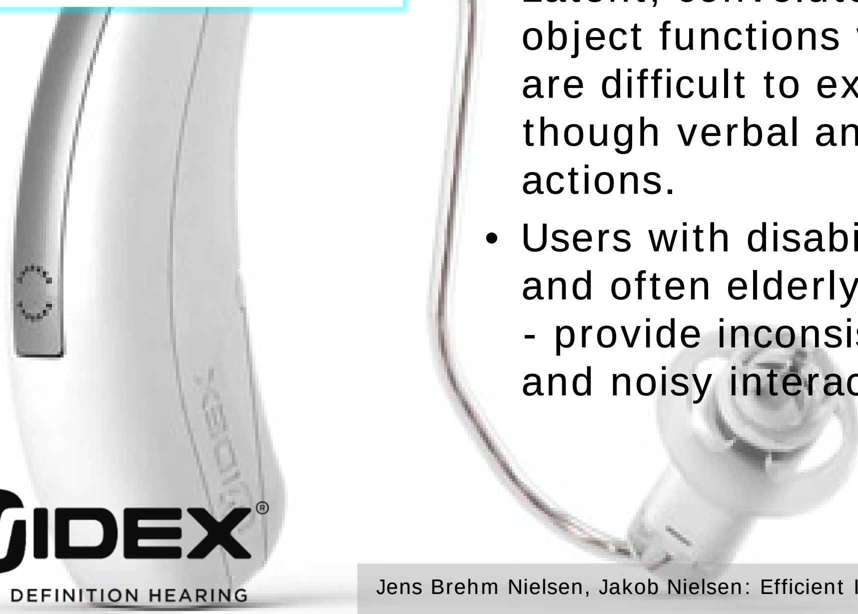
The Expected Improvement of a new candidate sample (green points) is derived from the predictive distribution.

Probabilistic Model is a Gaussian Process



Which of the four green parameters settings/products/interface, x, should the user assess (rate/ annotate/see/ hear) or where do we need objective performance measurements





Optimization of hearing aids using Bayesian optimization

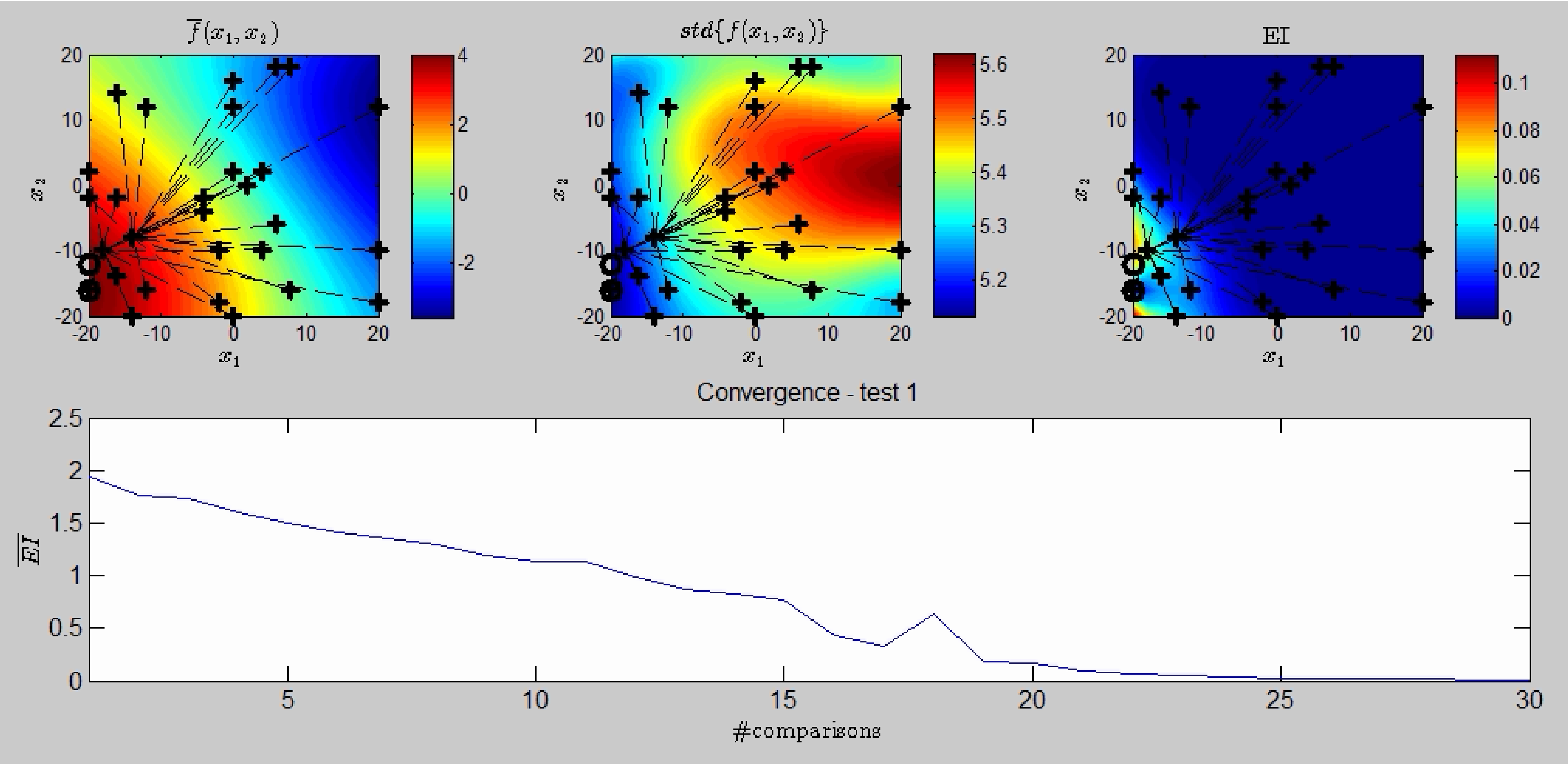
- Highly personalization needs.
- Dynamic environment and use with different needs.
- Latent, convoluted object functions which are difficult to express though verbal and motor actions.
- Users with disabilities – and often elderly people - provide inconsistent and noisy interactions.

Jens Brehm Nielsen, Jakob Nielsen: Efficient Individualization of Hearing and Processers Sound, ICASSP2013.

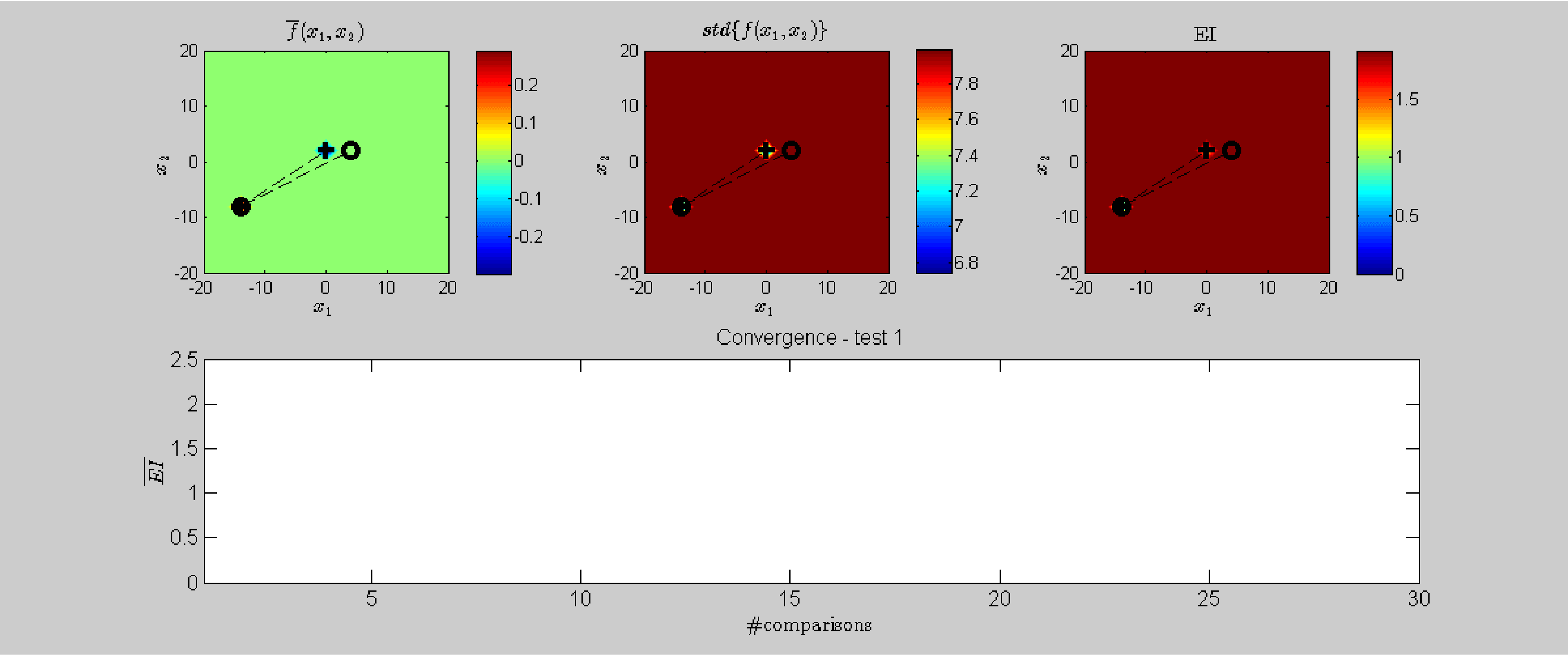
Jens Brehm Nielsen, Jakob Nielsen, Jan Larsen: Perception based Personalization of Hearing Aids using Gaussian Process and Active Learning, IEEE Trans. ASLP, vol. 23, no. 1, pp. 162 – 173, Jan 2015.

Maciej Korzepa, Michael Kai Petersen, Benjamin Johansen, Jan Larsen, Jakob Eg Larsen: Learning soundscapes from OPN sound navigator, poster 2017.

Pairwise (2AFC) personalization of hearing aids

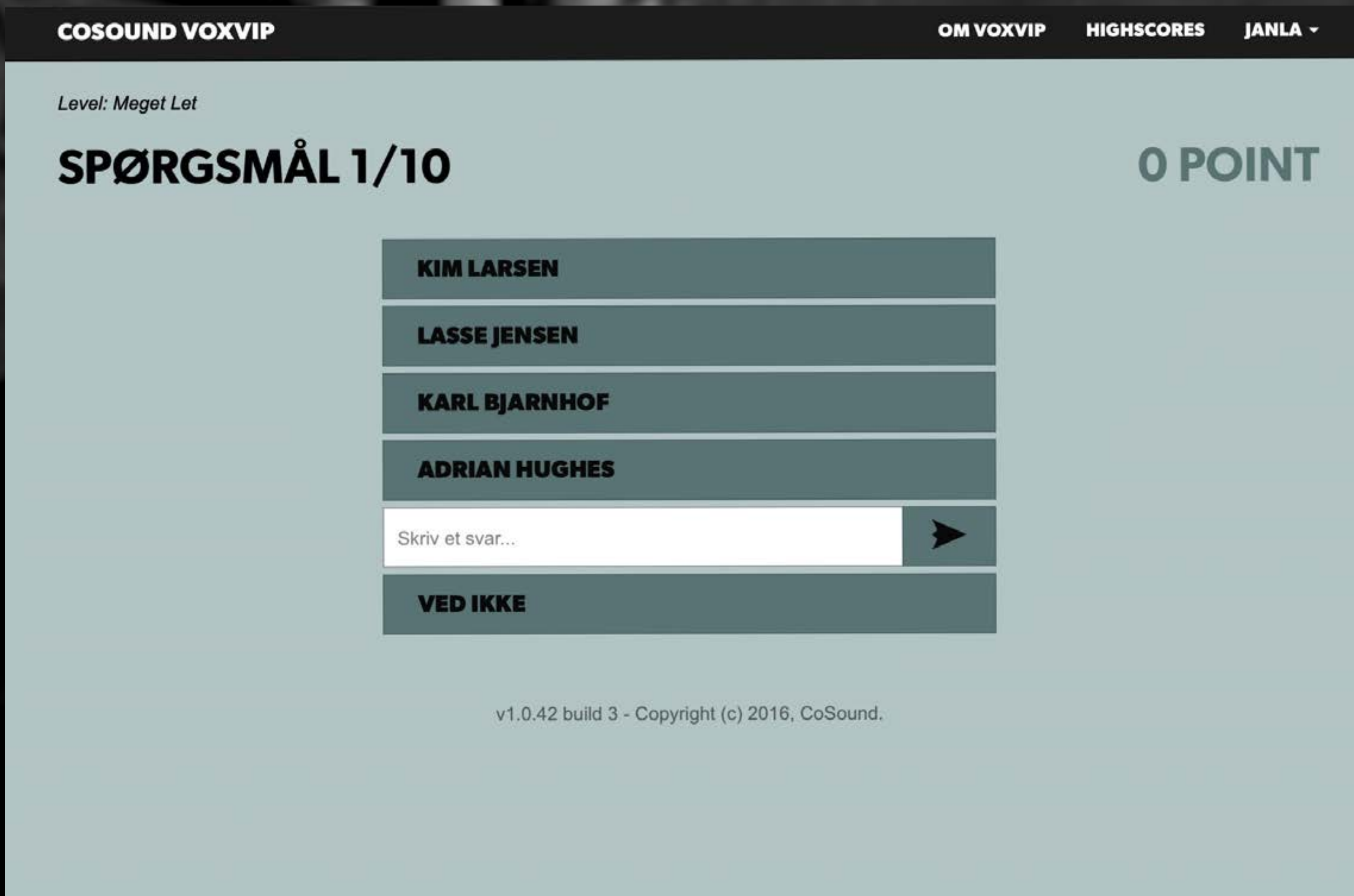


Pairwise (2AFC) personalization of hearing aids



J. B. B. Nielsen, J. Nielsen, J. Larsen, Perception-based Personalization of Hearing Aids using Gaussian Processes and Active Learning, *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 23(1), pp. 162-173, IEEE, 2015.

VOXVIP - intelligent crowdsourcing of the DR radio archive



voxvip.cosound.dk

Can smart crowdsourcing efficiently enrich radio archives with high quality metadata using machine learning and gamification?

Are model-based, active learning mechanisms suitable for smart crowdsourcing, and is optimal performance as regards time-use achieved?

Are age, sex, address relevant for recognition of specific voices?

Gamification: How does levels, difficulty and point assignment influence the quality and quantity of annotations?

What is meta information?

Infinite number of aspects provides information about the individual clip/object or similarity between such objects

Objective information

- Who is speaking
- What is the topic discussed?
- Which objects are present in the clip?

Subjective information

- Which emotions are expressed in the clip?
- What is the sound quality?
- Which clip is preferred?

How can meta information be created?

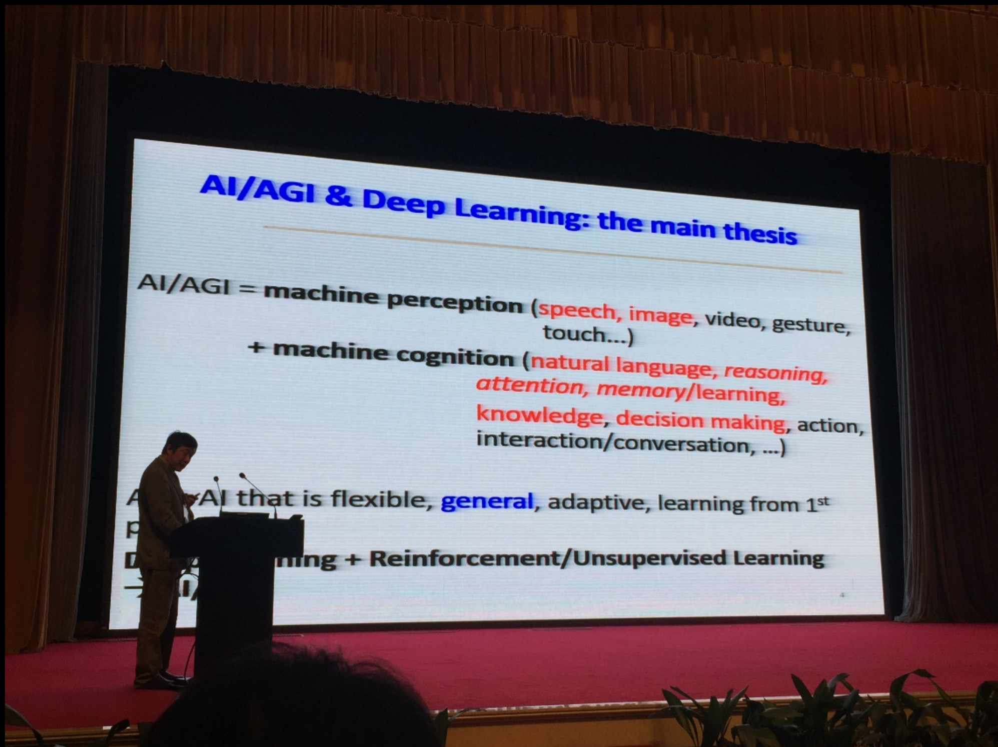
Lack of specific annotations requires prior knowledge

Manual annotation is limited or impossible due to the size of the archive, human resources, or annotators qualifications.

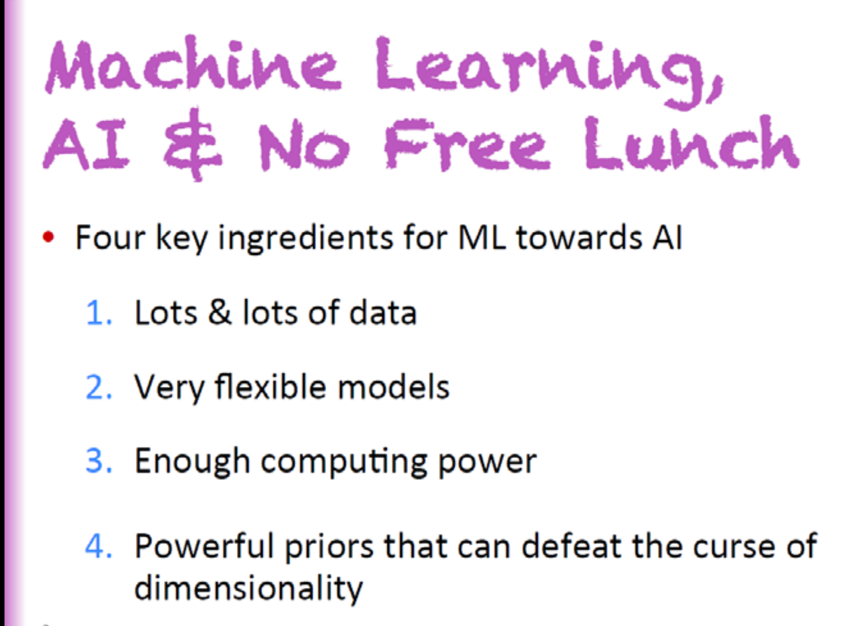
Semi-automatic machine learning can be used to predict information in the enture archive based on limited number of annotations.

Smart crowdsourcing exploits machine learning to predict information in the entire archive based on ‘crowd annotators’ annotations. The individual clip is selected based on uncertain information about the label, the annotators’ qualifications and engagement based on active learning mechanisms.

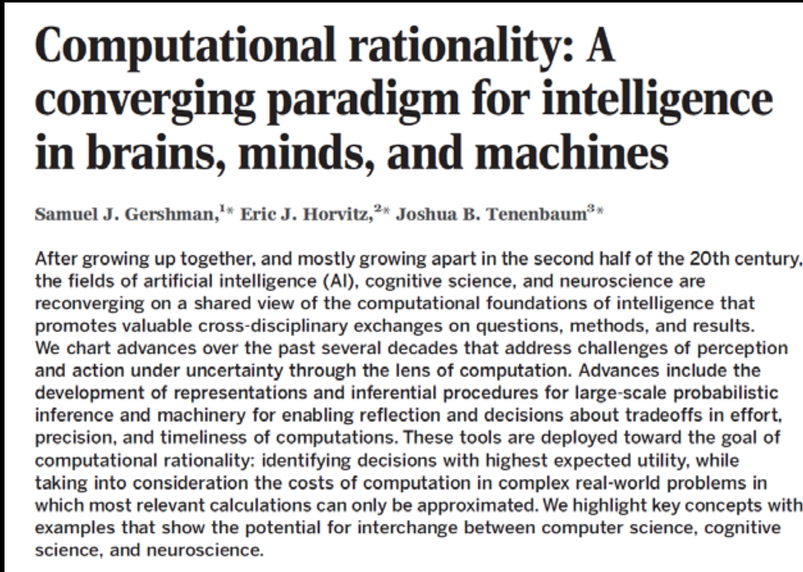
What is the solution?



Li Deng, Microsoft Research at ICASSP 2016, Shanghai.



Geoff Hinton, Yoshua Bengio & Yann LeCun, Deep Learning Tutorial, NIPS 2015, Montreal.



Computational rationality: A converging paradigm for intelligence in brains, minds, and machines

Samuel J. Gershman,[1*] Eric J. Horvitz,[2*] Joshua B. Tenenbaum[3*]

After growing up together, and mostly growing apart in the second half of the 20th century, the fields of artificial intelligence (AI), cognitive science, and neuroscience are reconverging on a shared view of the computational foundations of intelligence that promotes valuable cross-disciplinary exchanges on questions, methods, and results. We chart advances over the past several decades that address challenges of perception and action under uncertainty through the lens of computation. Advances include the development of representations and inferential procedures for large-scale probabilistic inference and machinery for enabling reflection and decisions about tradeoffs in effort, precision, and timeliness of computations. These tools are deployed toward the goal of computational rationality: identifying decisions with highest expected utility, while taking into consideration the costs of computation in complex real-world problems in which most relevant calculations can only be approximated. We highlight key concepts with examples that show the potential for interchange between computer science, cognitive science, and neuroscience.

Deep ANN, kernel methods, topic modeling/factor models

Ability to fuse noisy information and predict target parameters in changing environments under domain constraints and in simulated situations

Bayesian optimization

Ability to optimize system with incomplete or complex mechanisms

Goal-driven online learning communication systems

Ability to learning human interactions on all levels

Potentials

- Discovery of pattern in large unstructured data e.g. emails, social, behavioral, economical transaction, sound, images
- Anomaly detection
- Explaining causes, facts and sequences of events
- Robust and labor in-expensive predictive analysis and search for specific objects, events in multimodal data (audio, video, images etc.)
- Better involvement and integration of LEA personnel, general public, organizations and tasks (forensics, investigation, indictment, policing, intelligence, pro-activiness)
- Standardized tool but specialized solutions

On Collaboration – matching expectations

University & knowledge institutions

Primary objective

- international, open, independent knowledge production driven by curiosity
- focus on most difficult problems
- scientific publications: methods, principles, general/universal knowledge
- teaching incl. continuing education
- long term perspective

Secondary objectives

- innovation activities
- contribution to solving societal challenges aka scientific social responsibility
- communication and dissemination
- access to data, knowledge, and collaboration partners
- access to technology and facilities

LEAs

Primary objectives

- preventing crimes
- focus on relevant problems with high potential impact
- specific robust solutions with high quality
- shorter term perspective

Secondary objectives

- recruitment
- competence building
- international networks
- access, development and integration of newest methods, technology and tools