Mini Course
From Data Science
to interpretable AI

EXECUTIVE ACADEMY

Assoc. Prof. Dr. Andreas HOLZINGER (Medical University Graz)

Day 1 > Part 1 > Wednesday, 18.09.2019

Introduction to AI/Machine Learning

Remark

This is the version for printing and reading. The lecture version is didactically different.

Andreas Holzinger: Background

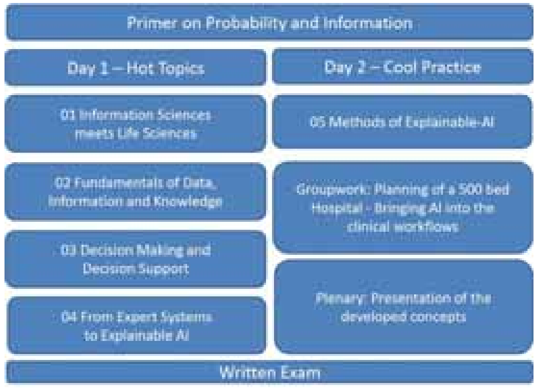
- Austrian Representative in IFIP TC 12 "Artificial Intelligence"
- Ordinary Member of Academia Europea: https://www.ae-info.org/ae/Member/Holzinger_Andreas
- PhD in Cognitive Science 1998
- Habilitation Computer Science 2003
- Lead Human-Centered AI (Holzinger Group)
- Personal Homepage: https://www.aholzinger.at
- Currently: Visiting Professor for explainable AI, Alberta Machine Intelligence Institute, Edmonton, CA
- Visiting Professor for Machine Learning in Health Informatics: TU Vienna, Univ. Verona, UCL London, RWTH Aachen
- Andreas Holzinger, Chris Biemann, Constantinos S. Pattichis & Douglas B. Kell, 2017. What do we need to build explainable AI systems for the medical domain? arXiv:1712.09923
- Andreas Holzinger, 2018. Explainable AI (ex-AI). Informatik-Spektrum, doi:10.1007/s00287-018-1102-5
- Andreas Holzinger et al., 2019. Causability and Explainability of AI in Medicine. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, doi:10.1002/widm.1312

Mini-Course Syllabus

- At the end of this mini course you will ...
- ... be fascinated to see our world in **data sets;**
- ... understand the differences between **data, information and knowledge**
- ... be aware of some problems and challenges in **biomedical informatics**
- ... understand the importance of the concept of **probabilistic information p(x)**
- ... know what **AI/Machine Learning** can (not) do
- ... have some fundamental insight into medical information science for **decision making**

Background Reading

Holzinger, A. 2014. Biomedical Informatics: Discovering Knowledge in Big Data, New York, Springer, doi:10.1007/978-3-319-04528-3.

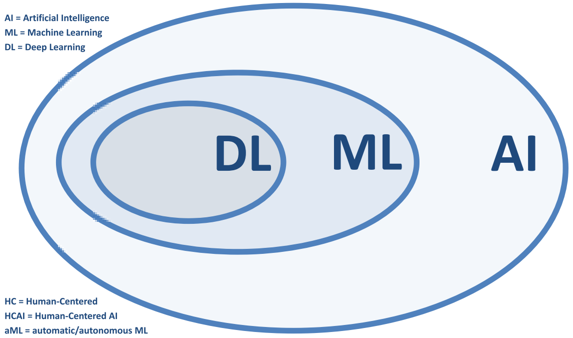


Agenda

- 01 What is HCAI ?
- 02 Application Area: Health Informatics
- 03 Probabilistic Information
- 04 Automatic Machine Learning
- 05 Interactive Machine Learning
- 06 Causality, Explainability, Interpretability

Abbreviations

AI = Artificial Intelligence
ML = Machine Learning
DL = Deep Learning

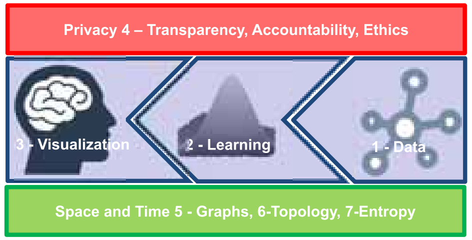


HC = Human-Centered
HCAI = Human-Centered AI
aML = automatic/autonomous ML
iML = interactive ML, interpretable ML
KDD = Knowledge Discovery from Data
ExAI = explainable AI

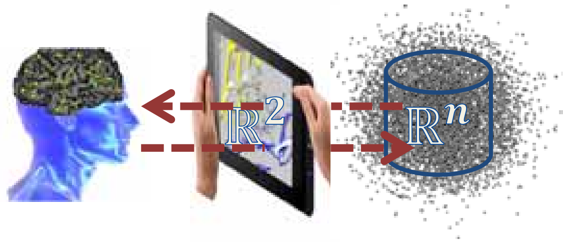
Andreas Holzinger, Peter Kieseberg, Edgar Weippl & A Min Tjoa 2018. Current Advances, Trends and Challenges of Machine Learning and Knowledge Extraction: From Machine Learning to Explainable AI. Springer Lecture Notes in Computer Science LNCS 11015. Cham: Springer, pp. 1-8, doi:10.1007/978-3-319-99740-7_1

01 What is HCAI

- **ML is a very practical field – algorithm development is at the core – however, successful ML needs a concerted effort of various topics …**

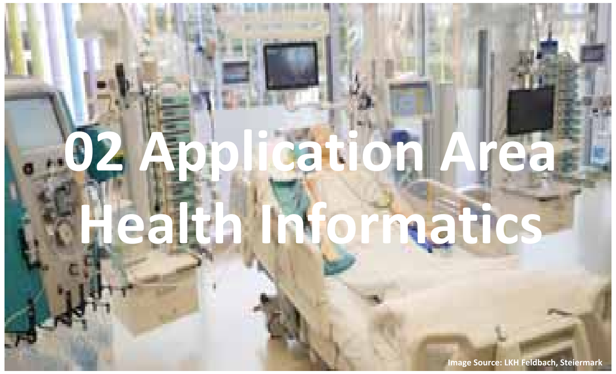


Andreas Holzinger 2013. Human–Computer Interaction and Knowledge Discovery (HCI-KDD): What is the benefit of bringing those two fields to work together? In: Multidisciplinary Research and Practice for Information Systems, Springer Lecture Notes in Computer Science LNCS 8127. Heidelberg, Berlin, New York: Springer, pp. 319-328, doi:10.1007/978-3-642-40511-2_22



Our goal is that human values are aligned to ensure responsible machine learning

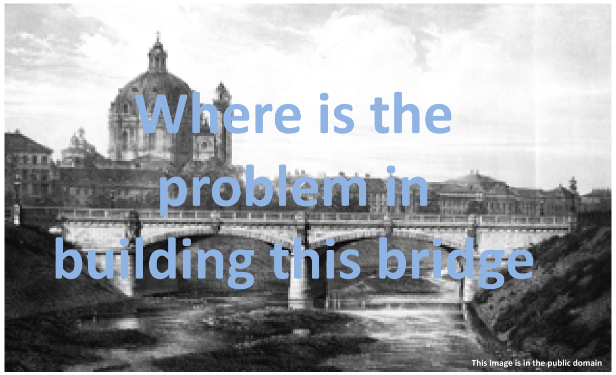
- 1) learn from prior data
- 2) extract knowledge
- 3) generalize, i.e. guessing where a probability mass function concentrates
- 4) fight the curse of dimensionality
- 5) disentangle **underlying explanatory factors of data**, i.e.
- 6) **understand** the data in the **context** of an application domain



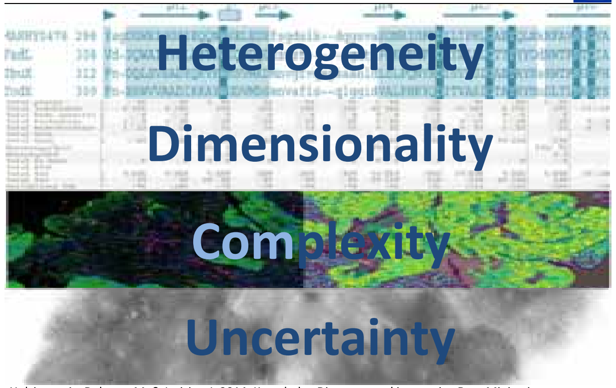
Why is this application area complex ?

Our central hypothesis: Information may bridge this gap

Holzinger, A. & Simonic, K.-M. (eds.) 2011. *Information Quality in e-Health. Lecture Notes in Computer Science LNCS 7058, Heidelberg, Berlin, New York: Springer.*

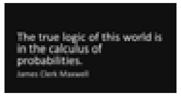


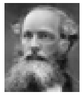
Main problems …



Holzinger, A., Dehmer, M. & Jurisica, I. 2014. Knowledge Discovery and interactive Data Mining in Bioinformatics - State-of-the-Art, future challenges and research directions. BMC Bioinformatics, 15, (S6), I1.

03 Probabilistic Learning



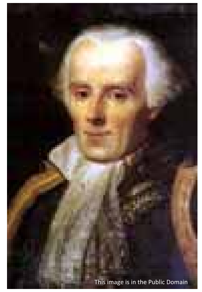


Maxwell, J. C. (1850). Letter to Lewis Campbell; reproduced in L. Campbell and W. Garrett, The Life of James Clerk Maxwell, Macmillan, 1881.

The foundation for modern machine learning …

- 1763: Richard Price publishes post hum the work of Thomas Bayes (see next slide)
- 1781: Pierre-Simon Laplace: Probability theory is nothing, but common sense reduced to calculation …
- 1812: Théorie Analytique des Probabilités, now known as Bayes' Theorem

- Hypothesis $h \in \mathcal{H}$ (uncertain quantities (Annahmen)
- Data $d \in \mathcal{D}$ … measured quantities (Entitäten)
- Prior probability $p(h)$ … probability that h is true
  Likelihood $p(d|h)$ … "how probable is the prior"
- Posterior Probability $p(h|d)$ … probability of h given d



Pierre Simon de Laplace (1749-1827)

$$p(h|d) \propto p(d|h) * p(h) \qquad p(h|d) = \frac{p(d|h)p(h)}{p(d)}$$

Repetition of Bayes - on the work of Laplace

What is the simplest mathematical operation for us?

$$p(x) = \sum_y (p(x,y)) \quad (1)$$

How do we call repeated adding?

$$p(x,y) = p(y|x) * p(y) \quad (2)$$

Laplace (1773) showed that we can write:

$$p(x,y) * p(y) = p(y|x) * p(x) \quad (3)$$

Now we introduce a third, more complicated operation:

$$\frac{p(x,y) * p(y)}{p(y)} = \frac{p(y|x) * p(x)}{p(y)} \quad (4)$$

We can reduce this fraction by $p(y)$ and we receive what is called Bayes rule:

$$p(x,y) = \frac{p(y|x) * p(x)}{p(y)} \qquad p(h|d) = \frac{p(d|h)p(h)}{p(d)} \quad (5)$$

Learning representations (θ, h) from observed data

Observed data: 

≈ Training data: $\mathcal{D} = x_{1:n} = \{x_1, x_2, ..., x_n\}$

Feature Parameter: θ or hypothesis h $h \in \mathcal{H}$

Prior belief ≈ prior probability of hypothesis h: $p(\theta)$ $p(h)$

Likelihood ≈ $p(x)$ of the data that h is true $p(\mathcal{D}|\theta)$ $p(d|h)$

Data evidence ≈ marginal $p(x)$ that h = true $p(\mathcal{D})$ $\sum_{h \in \mathcal{H}} p(d|h) * p(h)$

Posterior ≈ $p(x)$ of h after seen ("learn") data d $p(\theta|\mathcal{D})$ $p(h|d)$

$$posterior = \frac{likelihood * prior}{evidence} \qquad p(\theta|\mathcal{D}) = \frac{p(\mathcal{D}|\theta) * p(\theta)}{p(\mathcal{D})}$$

$$p(h|d) = \frac{p(d|h) * p(h)}{\sum_{h \in H} p(d|h)\, p(h)}$$

Learning and Probabilistic Inference (Prediction)

04

d … data $\quad \mathcal{H}$ … $\{H_1, H_2, ..., H_n\}$ $\quad \forall\, h, d$ …

h … hypotheses

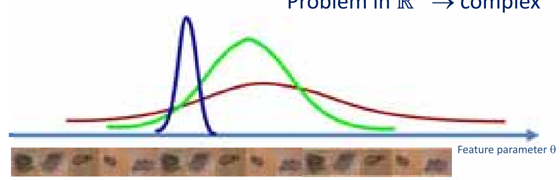
Likelihood

Prior Probability

$$\boxed{p(h|d)} = \frac{\boxed{p(d|h)} * \boxed{p(h)}}{\sum_{h \in H} p(d|h)\, p(h)}$$

Posterior Probability

Problem in $\mathbb{R}^n \rightarrow$ complex



Why is this relevant for medicine?

Reasoning under uncertainty: Decision Making

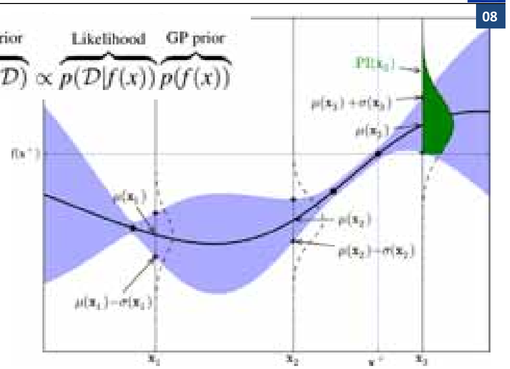
- Take patient information, e.g., observations, symptoms, test results, -omics data, etc. etc.
- Reach conclusions, and **predict** into the future, e.g. how likely will the patient be …
- Prior = belief before making a particular observation
- Posterior = belief after making the observation and is the prior for the next observation – intrinsically incremental

$$p(x_i|y_j) = \frac{p(y_j|x_i)p(x_i)}{\sum p(x_i, y_j)p(x_i)}$$

GP = distribution, observations occur in a cont. domain, e.g. t or space

08

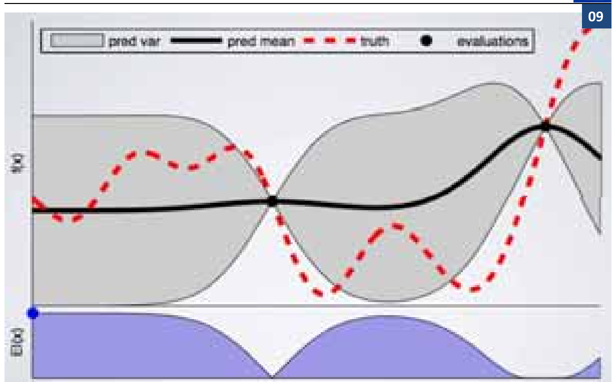
$$\overbrace{p(f(x)|\mathcal{D})}^{\text{GP posterior}} \propto \overbrace{p(\mathcal{D}|f(x))}^{\text{Likelihood}}\overbrace{p(f(x))}^{\text{GP prior}}$$



Brochu, E., Cora, V. M. & De Freitas, N. 2010. A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. arXiv:1012.2599.

Learning from data

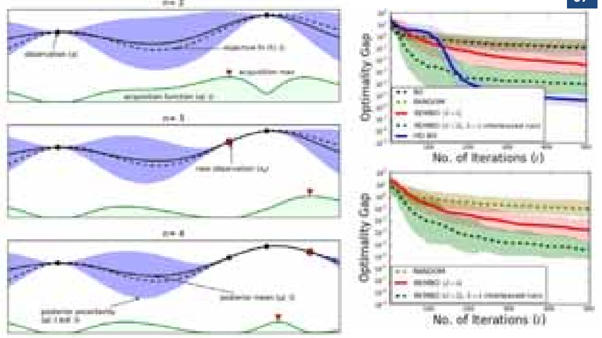
09



Snoek, J., Larochelle, H. & Adams, R. P. Practical Bayesian optimization of machine learning algorithms. Advances in neural information processing systems, 2012. 2951-2959.

Scaling to high-dimensions is the holy grail in ML

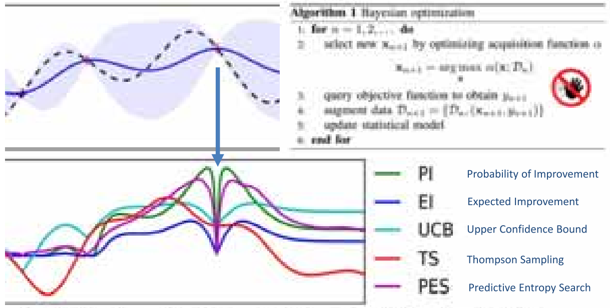
07



Wang, Z., Hutter, F., Zoghi, M., Matheson, D. & De Feitas, N. 2016. Bayesian optimization in a billion dimensions via random embeddings. Journal of Artificial Intelligence Research, 55, 361-387, doi:10.1613/jair.4806.

Fully automatic → Goal: Taking the human out of the loop

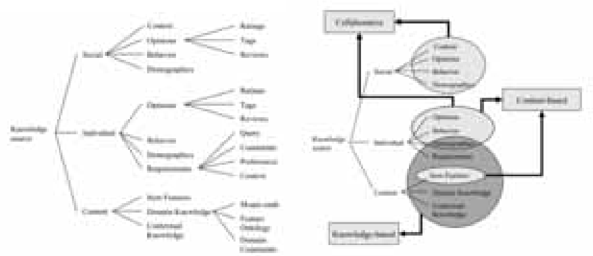
07



Shahriari, B., Swersky, K., Wang, Z., Adams, R. P. & De Freitas, N. 2016. Taking the human out of the loop: A review of Bayesian optimization. *Proceedings of the IEEE*, 104, (1), 148-175, doi:10.1109/JPROC.2015.2494218.

04 aML

Example for aML: Recommender Systems



Francesco Ricci, Lior Rokach & Bracha Shapira 2015. Recommender Systems: Introduction and Challenges. Recommender Systems Handbook. New York: Springer, pp. 1-34, doi:10.1007/978-1-4899-7637-6_1.

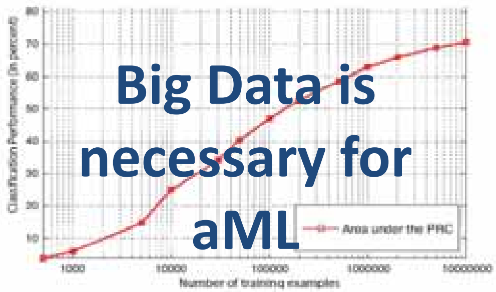
Fully automatic autonomous vehicles ("Google car")

Guizzo, E. 2011. How Google's self-driving car works. IEEE Spectrum Online, 10, 18.

… and thousands of industrial aML applications …

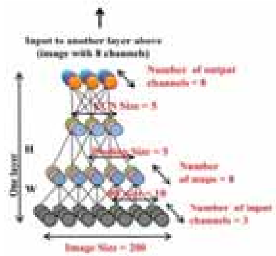
Seshia, S. A., Juniwal, G., Sadigh, D., Donze, A., Li, W., Jensen, J. C., Jin, X., Deshmukh, J., Lee, E. & Sastry, S. 2015. Verification by, for, and of Humans: Formal Methods for Cyber-Physical Systems and Beyond. Illinois ECE Colloquium.

Big Data is necessary for aML !



Sonnenburg, S., Rätsch, G., Schäfer, C. & Schölkopf, B. 2006. Large scale multiple kernel learning. Journal of Machine Learning Research, 7, (7), 1531-1565.

10 million 200 x 200 px images downloaded from Web



$$x^* = \arg\min_{x} f(x; W, H), \text{ subject to } ||x||_2 = 1.$$

Le, Q. V., Ranzato, M. A., Monga, R., Devin, M., Chen, K., Corrado, G. S., Dean, J. & Ng, A. Y. 2011. Building high-level features using large scale unsupervised learning. arXiv preprint arXiv:1112.6209.

Le, Q. V. 2013. Building high-level features using large scale unsupervised learning. *IEEE Intl. Conference on Acoustics, Speech and Signal Processing ICASSP.* IEEE. 8595-8598, doi:10.1109/ICASSP.2013.6639343.

When does aML fail …

- Sometimes we **do not have “big data”**, where aML-algorithms benefit.
- Sometimes we have
  - **Small amount of data sets**
  - Rare Events – **no training samples**
  - **NP-hard problems, e.g.**
    - Subspace Clustering,
    - k-Anonymization,
    - Protein-Folding, …

Holzinger, A. 2016. Interactive Machine Learning for Health Informatics: When do we need the human-in-the-loop? Springer Brain Informatics (BRIN), 3, (2), 119-131, doi:10.1007/s40708-016-0042-6.

Consequently …

Sometimes we (still) need a human-in-the-loop

05 iML

Definition of iML (Holzinger – 2016)

A6

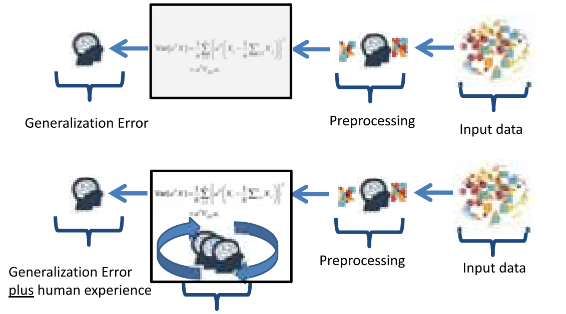
- iML := algorithms which interact with agents***)** and can optimize their learning behaviour through this interaction

***) where the agents can be human**

Holzinger, A. 2016. Interactive Machine Learning (iML). Informatik Spektrum, 39, (1), 64-68, doi:10.1007/s00287-015-0941-6.

A crowd of people-in-the-loop

iML: bringing in human intuition



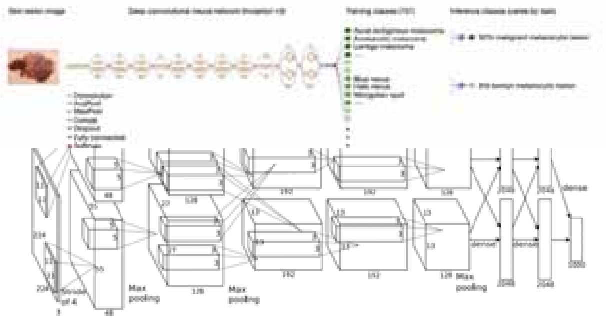
iML = human inspection – bring in human intuition

Andreas Holzinger et al. (2017) A glass-box interactive machine learning approach for solving NP-hard problems with the human-in-the-loop. arXiv:1708.01104.

06 Why Explainability?

Deep Convolutional Neural Network Pipeline

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M. & Thrun, S. 2017. Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542, (7639), 115-118, doi:10.1038/nature21056.

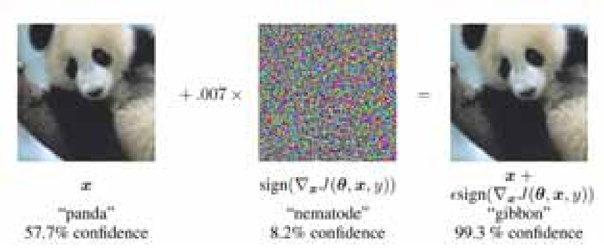


Krizhevsky, A., Sutskever, I. & Hinton, G. E. Imagenet classification with deep convolutional neural networks. In: Pereira, F., Burges, C. J. C., Bottou, L. & Weinberger, K. Q., eds. Advances in neural information processing systems (NIPS 2012), 2012 Lake Tahoe. 1097-1105.

Deep Learning “It's not complicated, it's just a lot of it”

- **Non-convex:** difficult to set up, to train, to optimize, needs a lot of expertise, error prone
- **Resource intensive** (GPU’s, cloud CPUs, federated learning, …)
- **Data intensive,** needs often millions of training samples …
- **Transparency lacking,** do not foster trust and acceptance among end-user, legal aspects make “black box” difficult

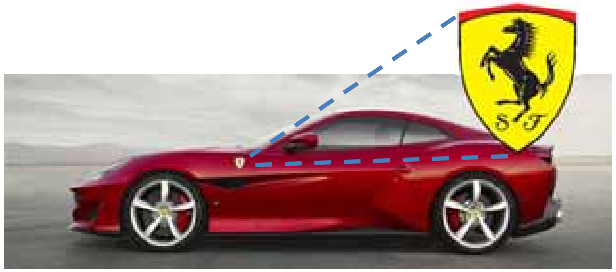
Example: Adversarial examples



See also: Ian J. Goodfellow, Jonathon Shlens & Christian Szegedy 2014. Explaining and harnessing adversarial examples. arXiv:1412.6572, and see more examples: https://imgur.com/a/K4RWn

Example: Classifier Errors

- Result of the classifier: **This is a horse**
- **Why is this a horse?**



Source: Image is in the public domain

Example: (Automatic) Context understanding …

Image Captions by deep learning:
State-of-the-Art of the Stanford Machine Learning Group

Andrej Karpathy, Justin Johnson & Li Fei-Fei 2015. Visualizing and understanding recurrent networks. arXiv:1506.02078.

Benefits …

Verify that algorithms/classifiers work as expected
Wrong decisions can be costly and dangerous …

Understanding the weaknesses and errors
Detection of bias – bring in human intuition to know the error …

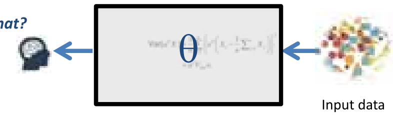
Scientific replicability and causality
The "why" is often more important than the prediction …

Andreas Holzinger 2018. Explainable AI (ex-AI). Informatik-Spektrum, 41, (2), 138-143, doi:10.1007/s00287-018-1102-5.

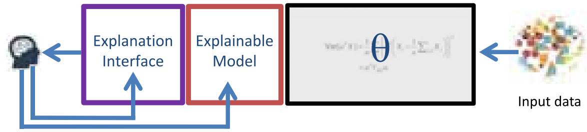
Conclusion

We need effective Human-AI mapping

Why did the algorithm do that?
Can I trust these results?
How can I correct an error?



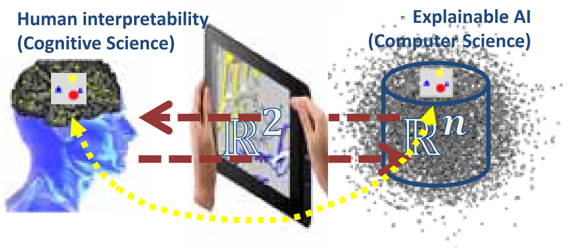
We contribute to …



The domain expert can understand why …
The domain expert can learn and correct errors …
The domain expert can re-enact on demand …

Explainable AI needs effective concept mapping !!!

- **Causability := a property of a person (Human)**
- **Explainability := a property of a system (Computer)**



Andreas Holzinger, Georg Langs, Helmut Denk, Kurt Zatloukal & Heimo Mueller 2019. Causability and Explainability of AI in Medicine. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, doi:10.1002/widm.1312.

Thank you!

Appendix

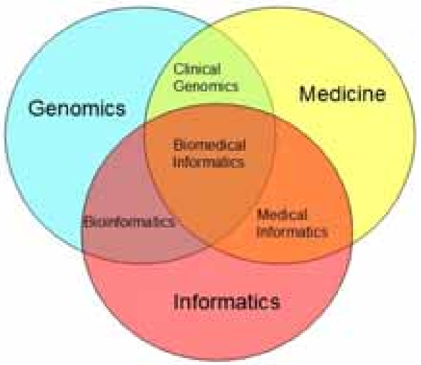
What is biomedical informatics?

INFORMATICS PROFESSIONALS. LEADING THE WAY.

- ***Biomedical informatics* (BMI) is the interdisciplinary field that studies and pursues the effective use of biomedical data, information, and knowledge for scientific problem solving, and decision making, motivated by efforts to improve human health**

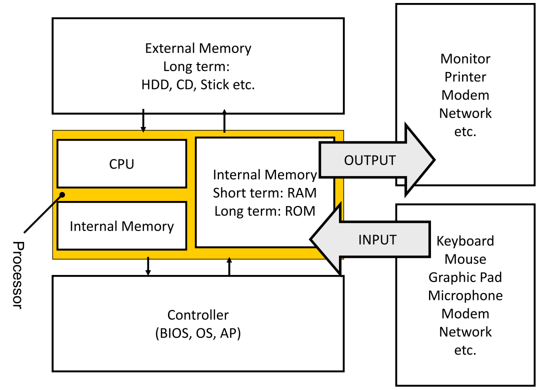
Shortliffe, E. H. (2011). Biomedical Informatics: Defining the Science and its Role in Health Professional Education. In A. Holzinger & K.-M. Simonic (Eds.), *Information Quality in e-Health. Lecture Notes in Computer Science LNCS 7058 (pp. 711-714). Heidelberg, New York: Springer.*

Computational Sciences meet Life Sciences



http://www.bioinformaticslaboratory.nl/twiki/bin/view/BioLab/EducationMIK1-2

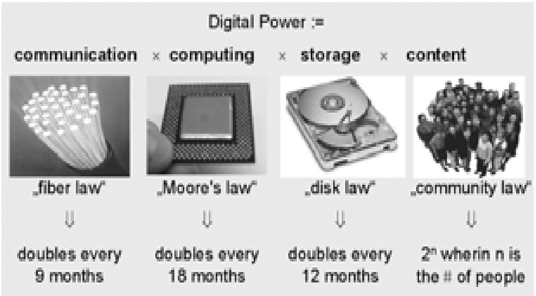
Computer: Von-Neumann Architecture



Holzinger (2002), 90-134

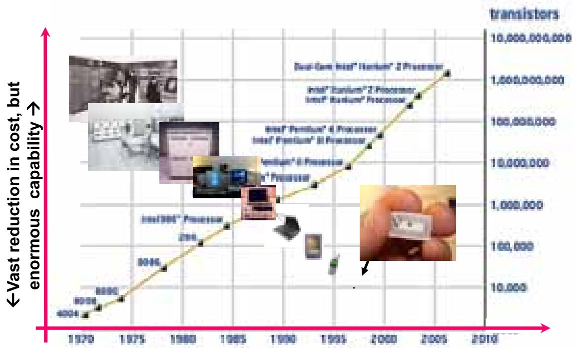
Technological Performance / Digital Power

Gordon E. Moore (1965, 1989, 1997)



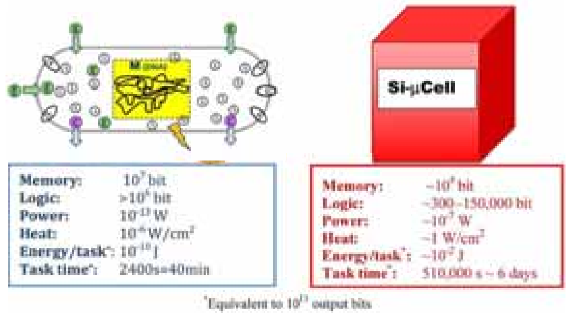
Holzinger, A. 2002. *Basiswissen IT/Informatik Band 1: Informationstechnik. Das Basiswissen für die Informationsgesellschaft des 21. Jahrhunderts, Wuerzburg, Vogel Buchverlag.*

Computer cost/size versus Performance



Cf. with Moore (1965), Holzinger (2002), Scholtz & Consolvo (2004), Intel (2007)

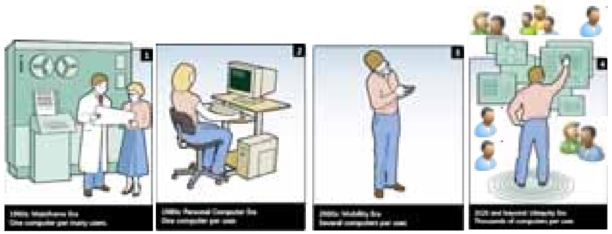
Beyond Moore's Law -> biological computing



Cavin, R., Lugli, P. & Zhirnov, V. 2012. Science and Engineering Beyond Moore's Law. *Proc. of the IEEE, 100, 1720-49* (L=Logic-Protein; S=Sensor-Protein; C=Signaling-Molecule, E=Glucose-Energy)

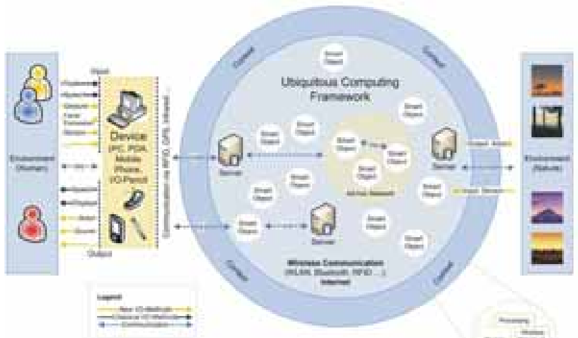
From mainframe to Ubiquitous Computing

- ... using technology to augment human capabilities for structuring, retrieving and managing information



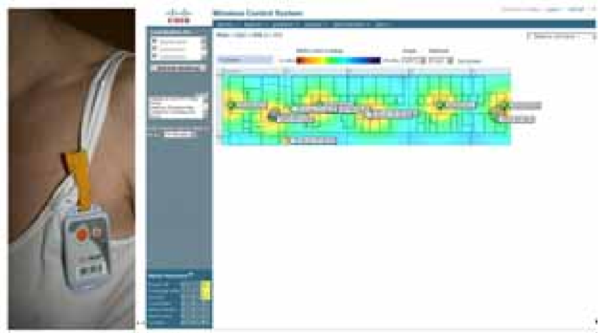
Harper, R., Rodden, T., Rogers, Y. & Sellen, A. (2008) *Being Human: Human-Computer Interaction in the Year 2020. Cambridge, Microsoft Research.*

Ubiquitous Computing – Smart Objects



Holzinger, A., Nischelwitzer, A., Friedl, S. & Hu, B. (2010) Towards life long learning: three models for ubiquitous applications. *Wireless Communications and Mobile Computing, 10, 10, 1350-1365.*

Slide 1-34 Example: Pervasive Health Computing

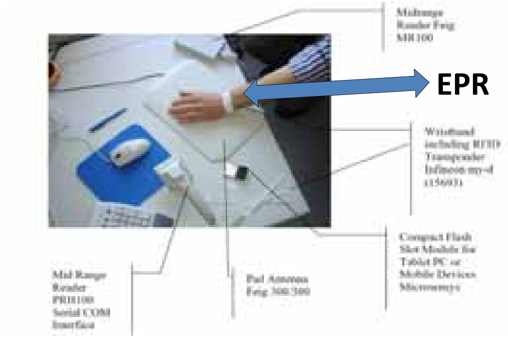


Holzinger, A., Schaupp, K. & Eder-Halbedl, W. (2008) An Investigation on Acceptance of Ubiquitous Devices for the Elderly in an Geriatric Hospital Environment: using the Example of Person Tracking In: *Lecture Notes in Computer Science (LNCS 5105). Heidelberg, Springer, 22-29.*

Ambient Assisted Living - pHealth

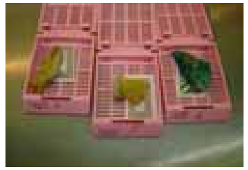
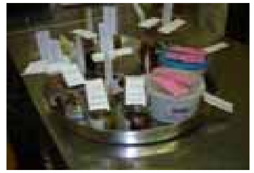
Alagoez, F., Valdez, A. C., Wilkowska, W., Ziefle, M., Dorner, S. & Holzinger, A. (2010) From cloud computing to mobile Internet, from user focus to culture and hedonism: The crucible of mobile health care and Wellness applications. *5th International Conference on Pervasive Computing and Applications (ICPCA). IEEE, 38-45.*

Example Pervasive Computing in the Hospital



Holzinger, A., Schwaberger, K. & Weitlaner, M. (2005) Ubiquitous Computing for Hospital Applications: RFID-Applications to enable research in Real-Life environments *29th Annual IEEE International Computer Software & Applications Conference (IEEE COMPSAC), 19-20.*

Smart Objects in the pathology

Images taken by Andreas Holzinger (2005)

The medical world is mobile (Mocomed)

Holzinger, A., Kosec, P., Schwantzer, G., Debevc, M., Hofmann-Wellenhof, R. & Frühauf, J. 2011. Design and Development of a Mobile Computer Application to Reengineer Workflows in the Hospital and the Methodology to evaluate its Effectiveness. *Journal of Biomedical Informatics, 44**, 968-977.***

1970 Turning Knowledge into Data

Photo by Institute of Medical Informatics, Graz (1970)