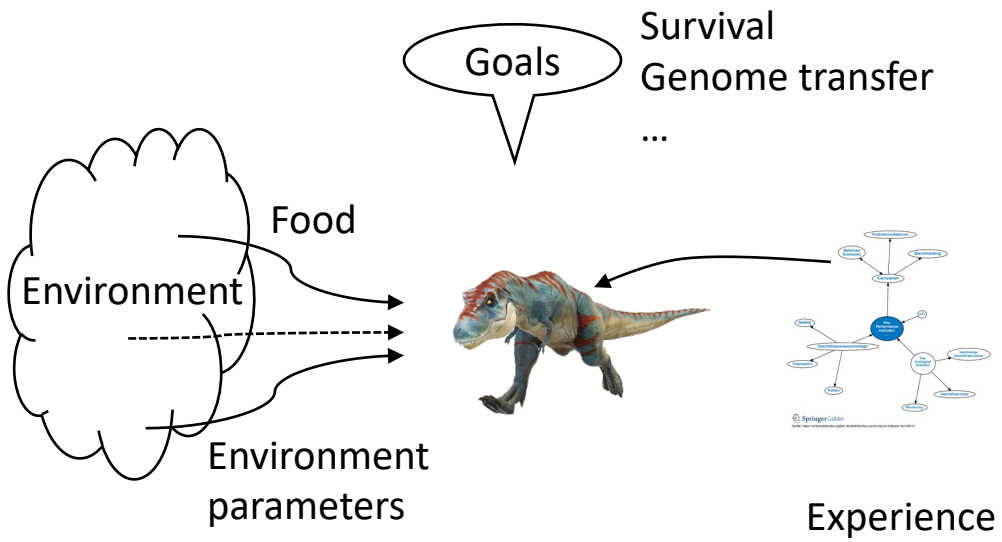
The Digital World – Challenges for CIOs
September 28, 29, 2023
Silk Road Samarkand

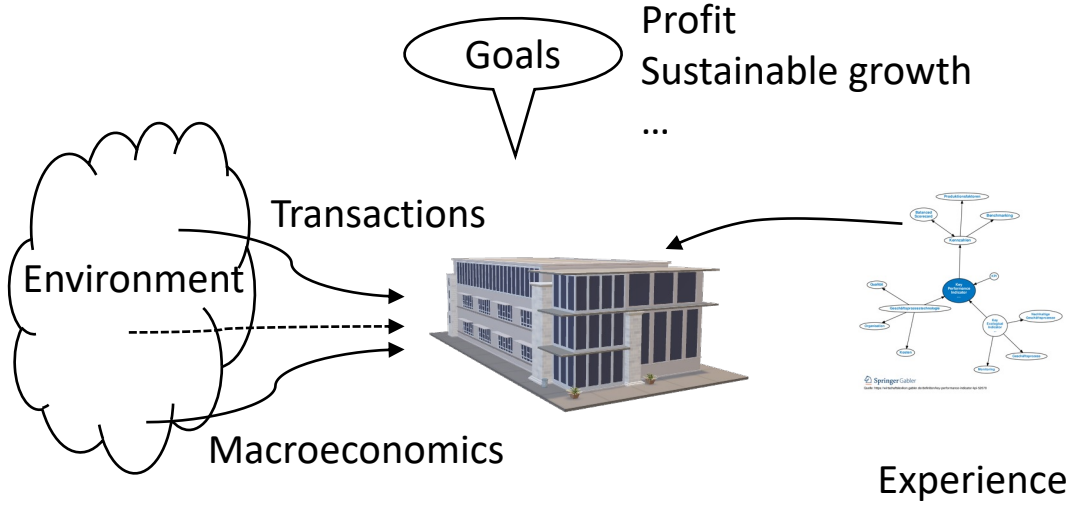
The opportunities and and limitations of hybrid intelligence in the analysis of data related to medicine and healthcare

Prof. Alexander. Ryjov
Dept. Intelligent Information Technologies, MSU
School of IT-management, RANEPA
alexander.ryjov@gmail.com

Any system uses data and knowledge to achieve its goals

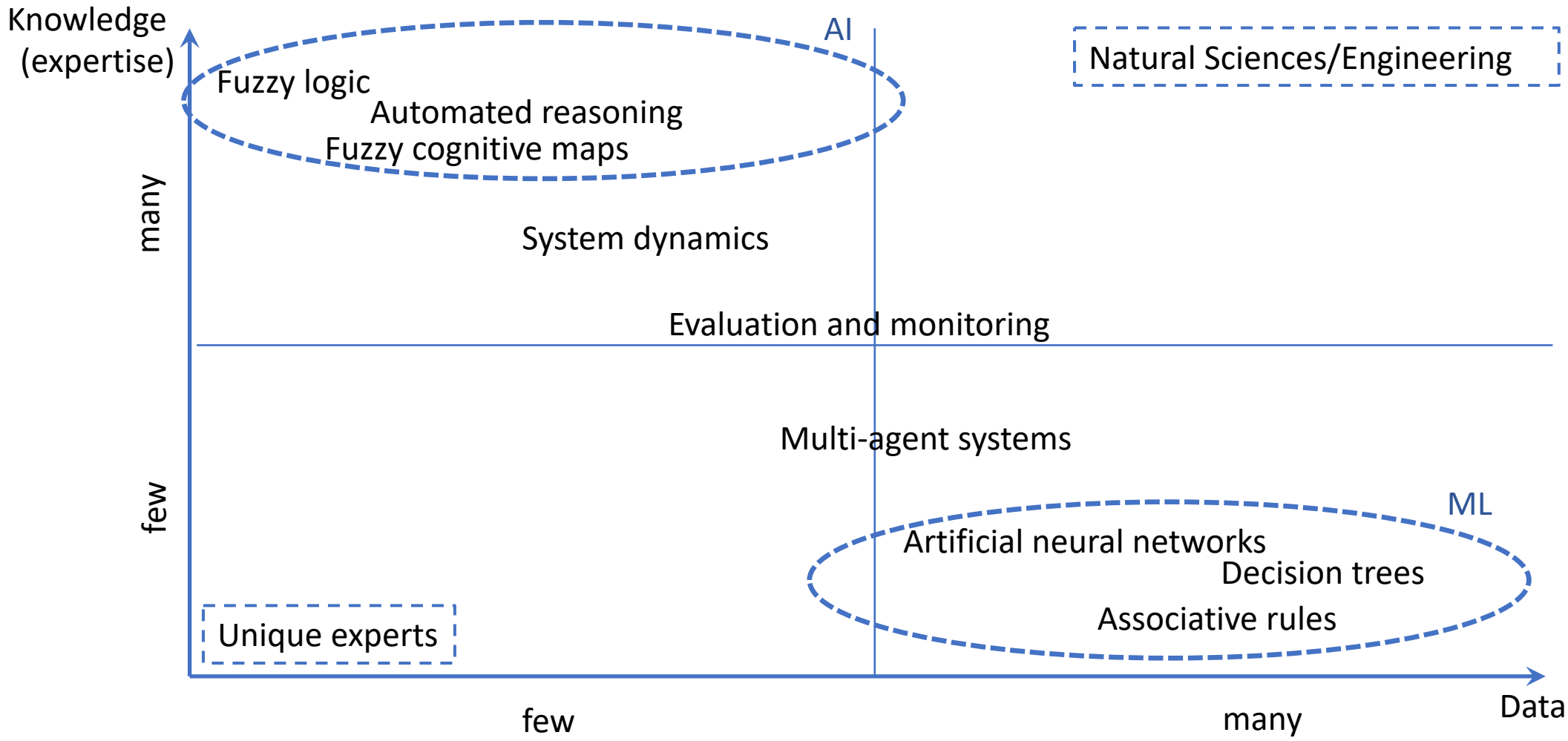


Task: how to achieve goals as efficiently as possible now based on input data and expertise?



Task: how to achieve goals as efficiently as possible now based on input data and expertise?

Data and knowledge are processed by different tools



Example 1: analysis of clinical data

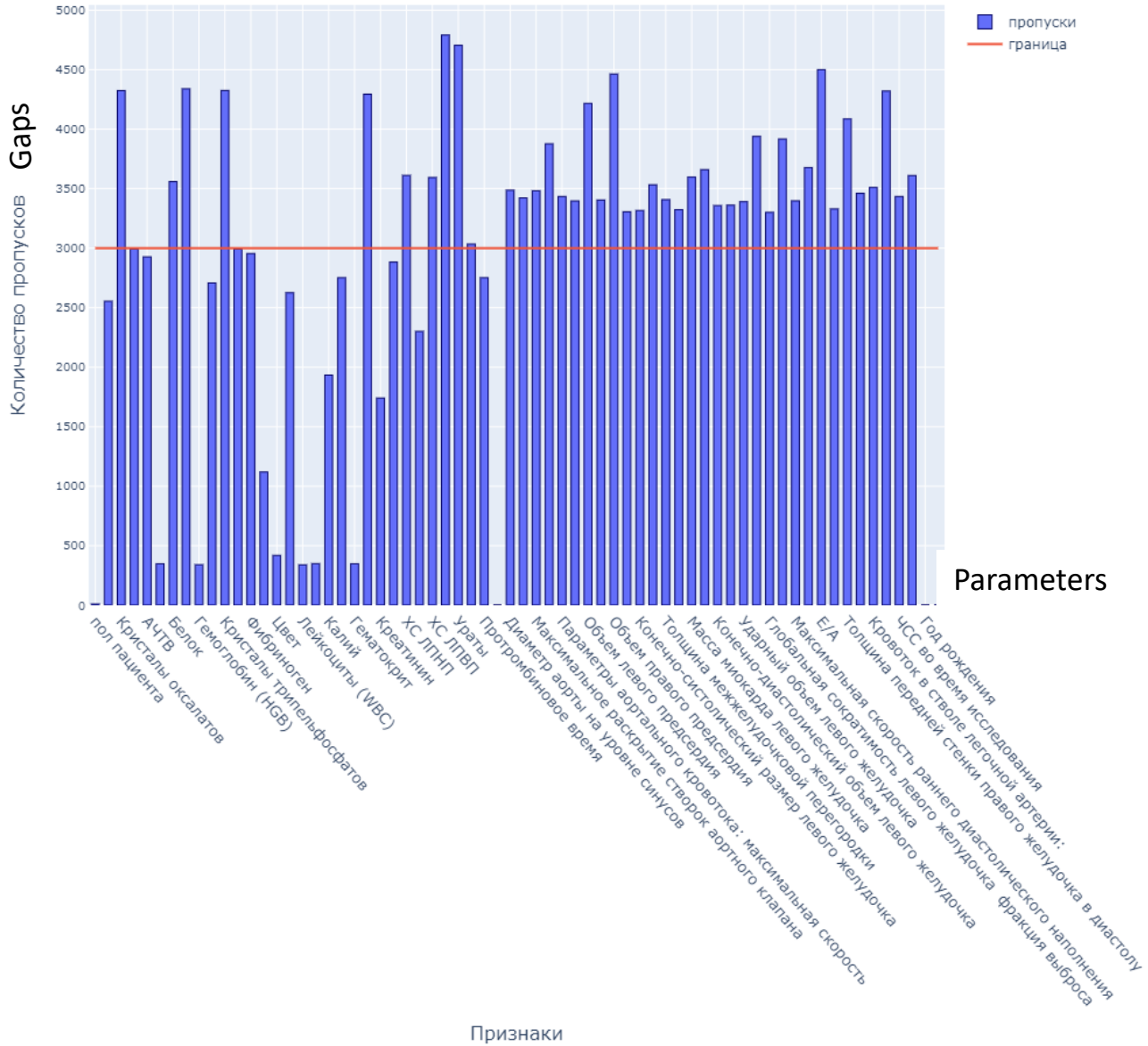
Task: prediction of the risk of an adverse clinical event (ACE)

Initial dataset:

- 79 clinical, demographic and laboratory parameters
- 5062 patients who underwent high-tech endovascular (60%) and interventional arrhythmological interventions (40%); 15% of operations were performed for emergency indications
- 58% of data gaps

Used/cleaned data set :

- 23 parameters (29%)
- 3146 patients в (62%)
- 16% gaps



Example 1: analysis of clinical data

Methods:

- Artificial neural networks
- Decision trees
- Associative rules induction

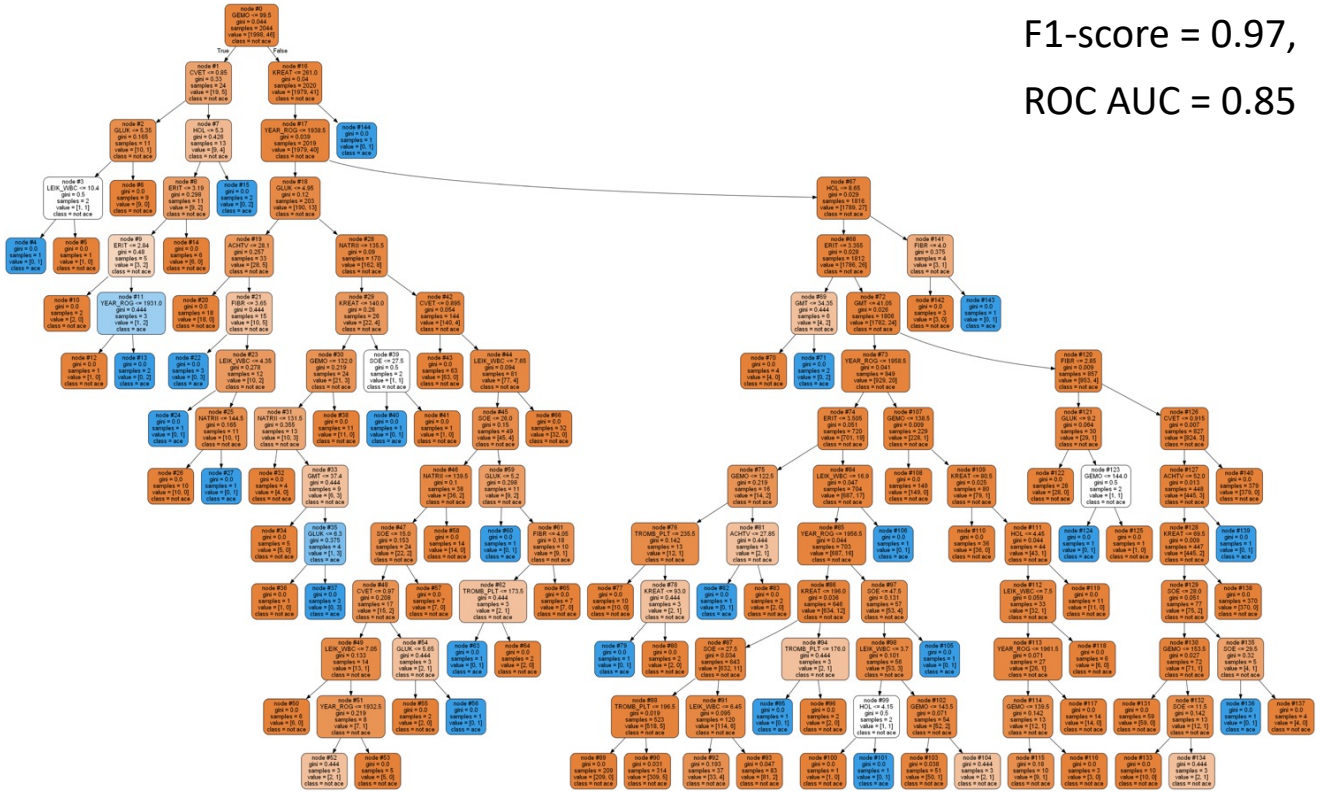
Results:

F1-score = 0.97,

ROC AUC = 0.85

	index	field_name	feature_importances_rand	field_description
0	8.0	NATRII	0.120946	Натрий
1	3.0	SOE	0.118923	СОЭ(Скорость оседания эритроцитов)
2	10.0	TRIG	0.101272	Триглицериды
3	0.0	ERIT	0.088649	Эритроциты (RBC)
4	4.0	KREAT	0.083006	Креатинин
5	12.0	GLUK	0.071385	Глюкоза
6	13.0	LEIK_WBC	0.062446	Лейкоциты (WBC)
7	7.0	HOL	0.059461	Холестерин
8	11.0	ACHTV	0.056378	АЧТВ
9	1.0	GMT	0.046284	Гематокрит
10	6.0	TROMB_PLT	0.045132	Тромбоциты (PLT)
11	9.0	CVET	0.043179	Цвет
12	2.0	YEAR_ROG	0.041890	NaN
13	5.0	GEMO	0.040107	Гемоглобин (HGB)
14	14.0	KALII	0.020941	Калий

roc_auc = 0.63; f1_score = 0.96



Feature importance

Горный Б. Э., Рыжов А. П., Строгалов А. С., Журавлев А. Д., Хусаенов А. А., Шергин И. А., Фещенко Д. А., Абдуллаев А. М., Концевая А. В. Оценка риска неблагоприятного клинического исхода методами углубленного анализа данных. *Интеллектуальные Системы Теория и приложения*. Т. 25, Вып. 2, 2021, с. 23-45.
https://www.mathnet.ru/links/dad6cce1b11ebeb2ce10efdfd90397e9/ista301.pdf

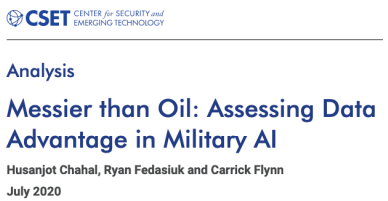
Example 1: Lessons

1. Data mining/ ML methods work in the field of clinical data analysis
2. Formal quality of work is good (comparable to scoring systems in banks)
3. Problem: Data quality
   - The data is not entered into the system
   - The data is entered in the wrong place (for example, in an attached pdf file)
   - The data is entered "for the bosses"
4. Problem: Transparency (explicability, interpretability) of the forecast
   - “I don't need a "black box", even a very good one”
   - ”I have to understand the forecast and trust it”
   - ” I am responsible for the patient's treatment (up to criminal liability)”

Harvard Business Review

Bad Data Costs the U.S. $3 Trillion Per Year
Thomas C. Redman, September 22, 2016
https://hbr.org/2016/09/bad-data-costs-the-u-s-3-trillion-per-year?fbclid=IwAR3M26u_lsKM9iTuMolnlD6Ek3Xv8zzPEWDTGqhW6Wvlyoq3JY7yxW5A9oc

$136 *billion per year* – estimation of the size of the big data market, worldwide (IDC)
$3.1 *trillion* – estimation of the yearly cost of poor quality data, in the US alone (IBM)


CSET CENTER for SECURITY and EMERGING TECHNOLOGY

Analysis

Messier than Oil: Assessing Data Advantage in Military AI

Husanjot Chahal, Ryan Fedasiuk and Carrick Flynn

July 2020

https://cset.georgetown.edu/research/messier-than-oil-assessing-data-advantage-in-military-ai/

Data is the new garbage (in Russian):
https://expert.ru/2020/08/6/dannyie---eto-novyij-musor/?fbclid=IwAR3JNQQBqQiBylgbSBXdpTJ7hWooIQR6qMe9FJ6xOwoomxCNJiPmApo197I

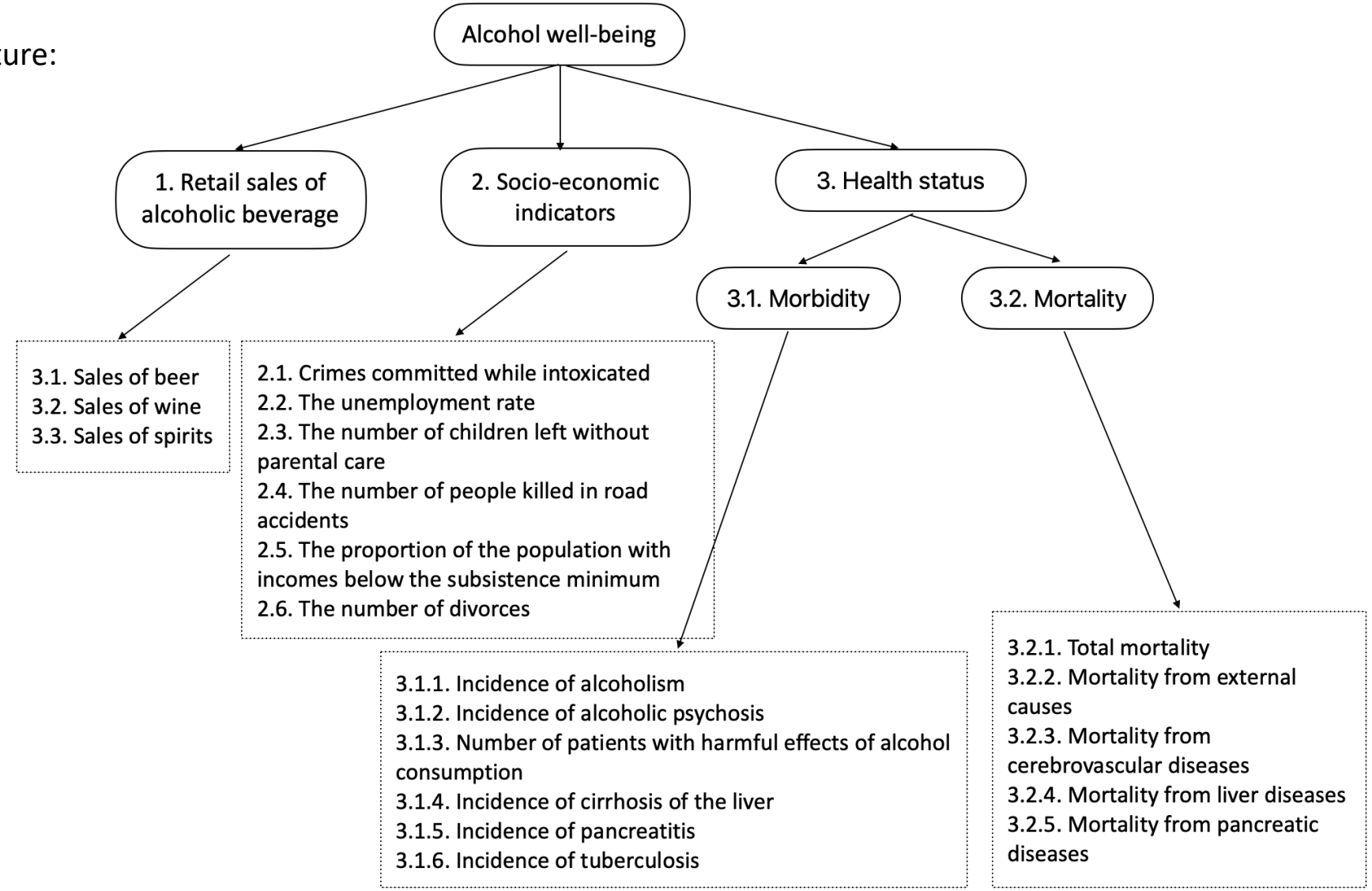
Example 2: Population data analysis

Task: To develop an index of alcohol well-being based on statistical data (data + knowledge)

- According to WHO estimates, excessive alcohol consumption caused about 3.3 million deaths in 2012, or 5.9% of their total. In Russia, 66% of serious offenses, 50% of murders, 40% of assaults, 24% of traffic offenses are associated with excessive alcohol consumption. The total costs of the state (direct medical, direct non-medical and indirect) related to alcohol abuse in 2011 amounted to 843.51 billion rubles. Per year, the costs per person abusing alcohol amount to 150 thousand 845 rubles, and the costs associated with treatment – 23 thousand 813 rubles.
- Statistical indicators that are directly and indirectly related to alcohol well-being:
  - indicators of retail sales of alcoholic beverages: sales of beer, wine, spirits;
  - socio-economic indicators: crimes committed while intoxicated; the unemployment rate; the number of children left without parental care; the number of people killed in road accidents; the proportion of the population with incomes below the subsistence minimum; the number of divorces;
  - morbidity rates: alcoholism, alcoholic psychosis, cirrhosis of the liver, pancreatitis, tuberculosis;
  - mortality rates: total mortality; mortality from external causes; mortality from cerebrovascular diseases; mortality from liver and pancreatic diseases.
- Experts

Example 2: Index structure and rules

Index structure:



Examples of rules :

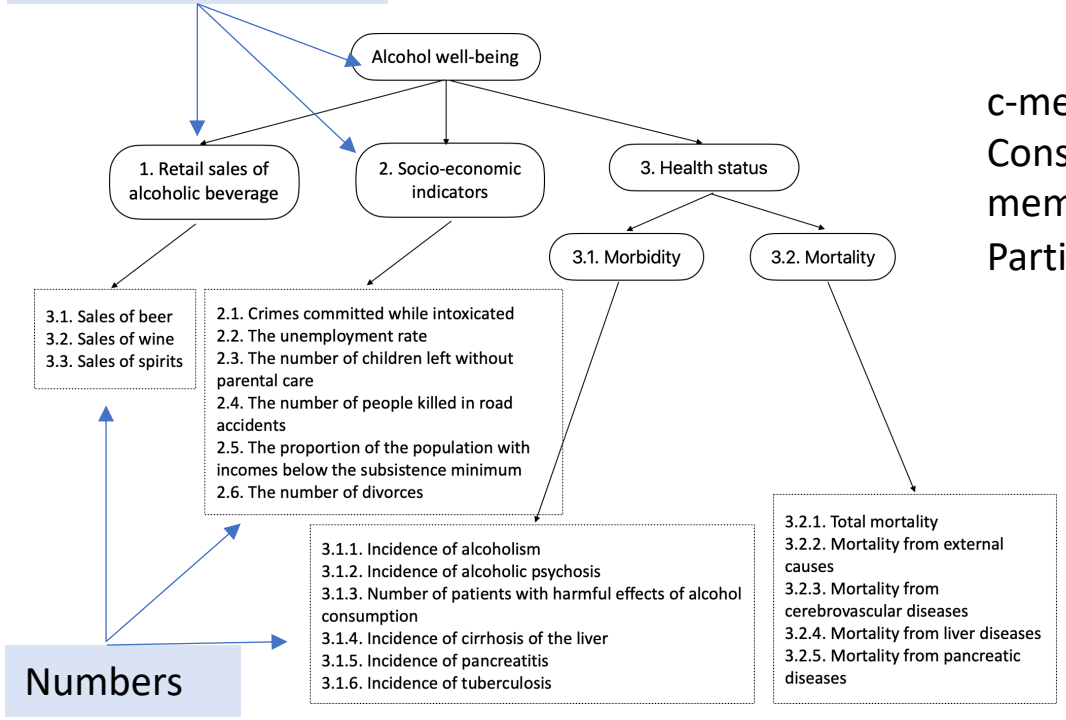
…

Rule # 3: If "Strong alcohol sales" = relatively low and
"Wine sales" = relatively low and
"Beer sales" = relatively high, then
"Retail sales of alcoholic beverages" = relatively low;

…

Example 2: Implementation

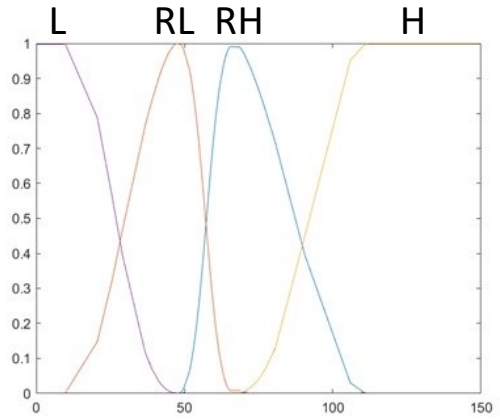
Concepts/ words {low, relatively low, relatively high, high}



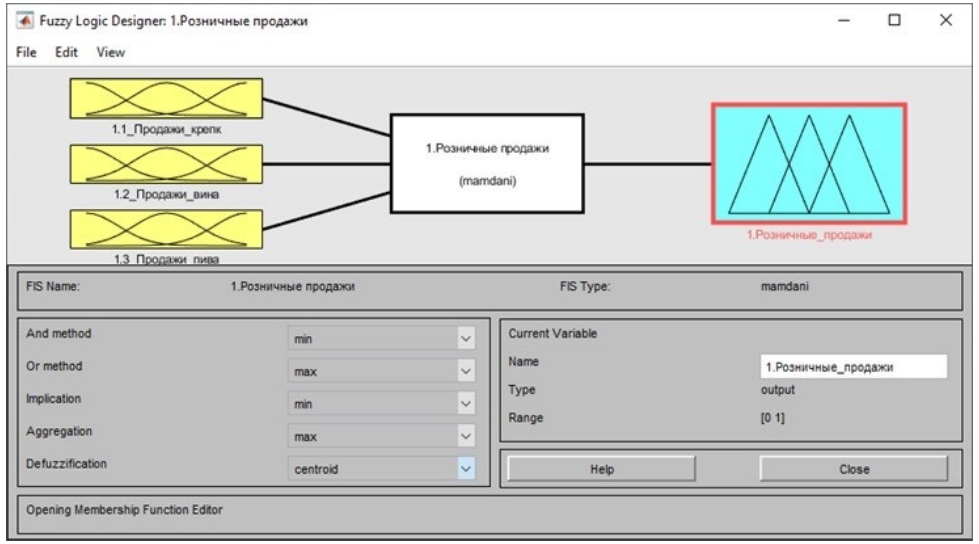
Numbers

c-means (Matlab, fcm): Construction of membership functions/ Partitioning into clusters

```
1
2 -   load 'data.dat';
3 -   q=sort(data);
4 -   [center, U, obj_fcm] = fcm( q,4);
5 -   y1=U(1,:);
6 -   y2=U(2,:);
7 -   y3=U(3,:);
8 -   y4=U(4,:);
9 -   plot(q,y1);
10 -  hold on;
11 -  plot(q,y2);
12 -  hold on;
13 -  plot(q,y3);
14 -  hold on;
15 -  plot(q,y4);
16 -  hold on;
```



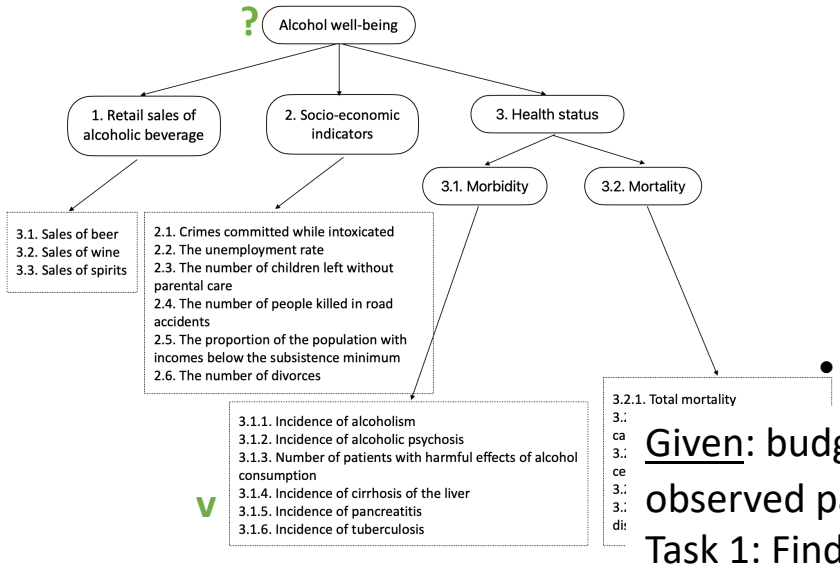
Matlab/ Fuzzy Logic Toolbox: Example of evaluation calculation (node 1)



Rule # 3: If "Strong alcohol sales" = relatively low
and "Wine sales" = relatively low
and "Beer sales" = relatively high,
then "Retail sales of alcoholic beverages" = relatively low

Example 2: Results

1. Rating of the regions (cities/ countries – depends on statistical data)
2. Additional ”bonuses”
   - Modeling (direct problem)



Регион	Индекс алкогольного благополучия
Республика Дагестан	0.88660341863593
Москва	0.88474556281952
Кабардино-Балкарская Республика	0.87696102915250
Республика Ингушетия	0.87167331310612
Карачаево-Черкесская Республика	0.86638541143022
Чеченская Республика	0.85932103574478
Санкт-Петербург	0.85913289650087
Краснодарский край	0.84766829610611
Белгородская область	0.84352087605548
Тюменская обл.без данных по Ханты-Мансийскому и	0.84031173165260
Республика Северная Осетия - Алания	0.83673986408290
Воронежская область	0.83508050933716
Ханты-Мансийский автономный округ - Югра (Тюменская	0.82478134424218
Псковская область	0.60239497298516
Республика Хакасия	0.55386442938895
Республика Адыгея (Адыгея)	0.34865472979340
Магаданская область	0.33974503389647
Амурская область	0.33502755439053
Республика Алтай	0.33295494930709
Сахалинская область	0.33114749703571
Республика Карелия	0.31505565253775
Ненецкий автономный округ (Архангельская область)	0.30698089123564
Республика Коми	0.27352962628243
Камчатская край	0.26851494620093
Еврейская автономная область	0.26809932791151
Чукотский автономный округ	0.11532854922210

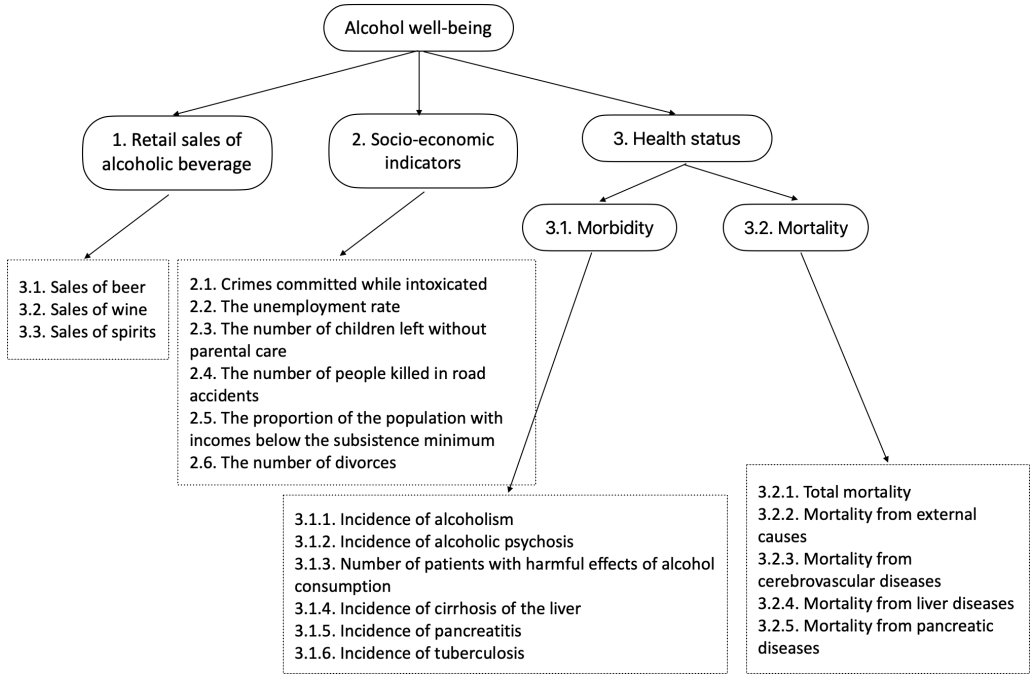
Anti-alcohol policy in Russia: results of the work of the State Duma
December 20, 2022, 11:00 Moscow
https://tass.ru/press/18497

- Management (inverse problems)

Given: budget X, the cost of changing the source data c_j $(j = 1, \cdots, N)$, where N is the number of observed parameters. Denote the change of the target vertex of the index structure by Δa_0.
Task 1: Find a set of parameters $\{i_1, \cdots, i_n\}$ $(n \leq N)$: $\Delta a_0 \to max$, $\sum_{j=1}^{n} c_{i_j} \leq X$ (maximal effect within the budget).
Let's say a target change in alcohol well-being is set q (for example, $q = 10\%$)
Task 2: Find a set of parameters $\{i_1, \cdots, i_n\}$ $(n \leq N)$: $\sum_{j=1}^{n} c_{i_j} \to min$, $\Delta a_0 \geq q$.
This is the task of finding the minimum budget to achieve the desired effect.

Example 2: Lessons

1. Fuzzy logic methods work in the field of population data analysis
2. Quality of work is good (comparable with expert committees)
3. Problem: Data sufficiency and adequacy
   - Data is needed to calculate all the parameters of the model (Retail sales of alcoholic beverages, Socio-economic indicators, etc.)
4. Problem: Availability of qualified and motivated experts
   - The expert should be able to link the available data and the target parameter of the model
5. It is possible to develop a wide range of similar indicators for healthcare (drug well-being, life expectancy, etc.) and other areas when solving problems 3 and 4.
6. Additional features (bonuses) allow you to calculate the effectiveness of decisions and optimize budgets to achieve goals



Example 3: Evaluation and monitoring of cardiovascular disease risks

Task: evaluate the cardiovascular disease risks and develop personal recommendations for risks minimization.

Problem: we can not measure all the factors (information is missing; information is expensive to collect and process, etc.); doctor can evaluate/ “measure” the status of the factor.

Idea: use hybrid intelligence approach.

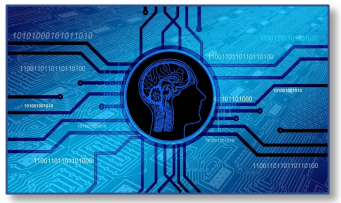
Two poles:

People doing all

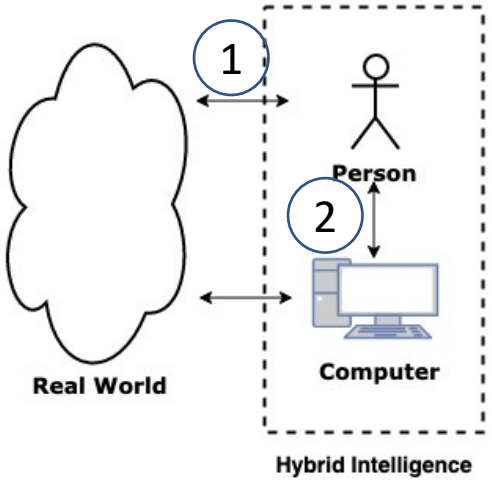
Not scalable
Not reliable
…

Scalable
Reliable
Realistic

AI doing all



Not realistic for now



- Problem 1 (Perception modelling):
  How we describe objects from the real world? Can we describe the objects by the most reliable and the most effective for further computing way?
- Problem 2 (Perception-base computing):
  How we can manipulate with perception-based information (for example, search or generalize)? Can we optimize these calculations?

Alexander Ryjov. A Note on Hybrid Intelligence. *CIO and Digital Innovation. IAC Online Journal*, N2, 2019, pp. 23 - 29.
https://iacio.org/wp-content/uploads/2019/04/IAC-Journal-of-CIO-and-Digital-Innovation-2018-9.pdf

Hybrid Intelligence today

McKinsey. Automation of knowledge work - *"These capabilities not only extend computing into new realms …, but also create **new relationships between knowledge workers and machines**. It is increasingly possible to interact with a machine the way one would with a **coworker**"*
(McKinsey Global Institute. *Disruptive technologies: Advances that will transform life, business, and the global economy*. (p.41) Retrieved from http://www.mckinsey.com/insights/business_technology/disruptive_technologies)

Why hybrid intelligence is the future of artificial intelligence at **McKinsey**
- https://www.mckinsey.com/about-us/new-at-mckinsey-blog/hybrid-intelligence-the-future-of-artificial-intelligence

Hybrid intelligence. Power of technology + power of people
https://www.mckinsey.com/business-functions/quantumblack/our-approach

NSF (NSF's 10 big ideas - https://www.nsf.gov/news/special_reports/big_ideas/index.jsp, Big Idea #1 https://www.nsf.gov/news/special_reports/big_ideas/human_tech.jsp (Building the **human-technology partnership**, **Augmenting** human performance))

DARPA: Третья волна ИИ
https://www.darpa.mil/work-with-us/ai-next-campaign
(DARPA research and development in **human-machine symbiosis** sets a **goal to partner with machines**)

Национальная стратегия США в области ИИ https://www.nitrd.gov/pubs/National-AI-RD-Strategy-2019.pdf
(Increase understanding of how to create AI systems that **effectively complement and augment human capabilities**)

IBM What is human-centered AI?
https://research.ibm.com/blog/what-is-human-centered-ai

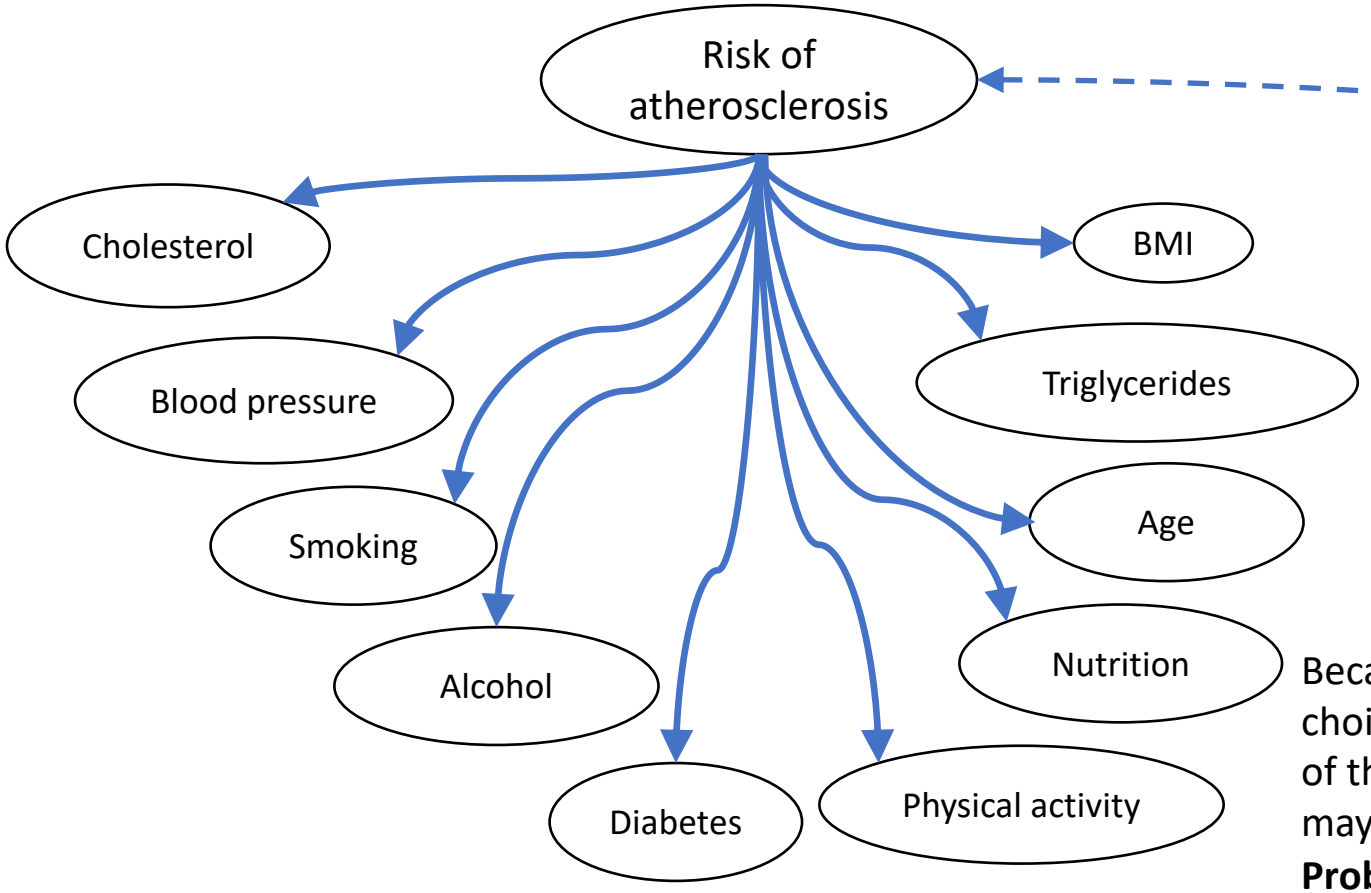
Google People + AI Research (PAIR)
https://research.google/teams/brain/pair/

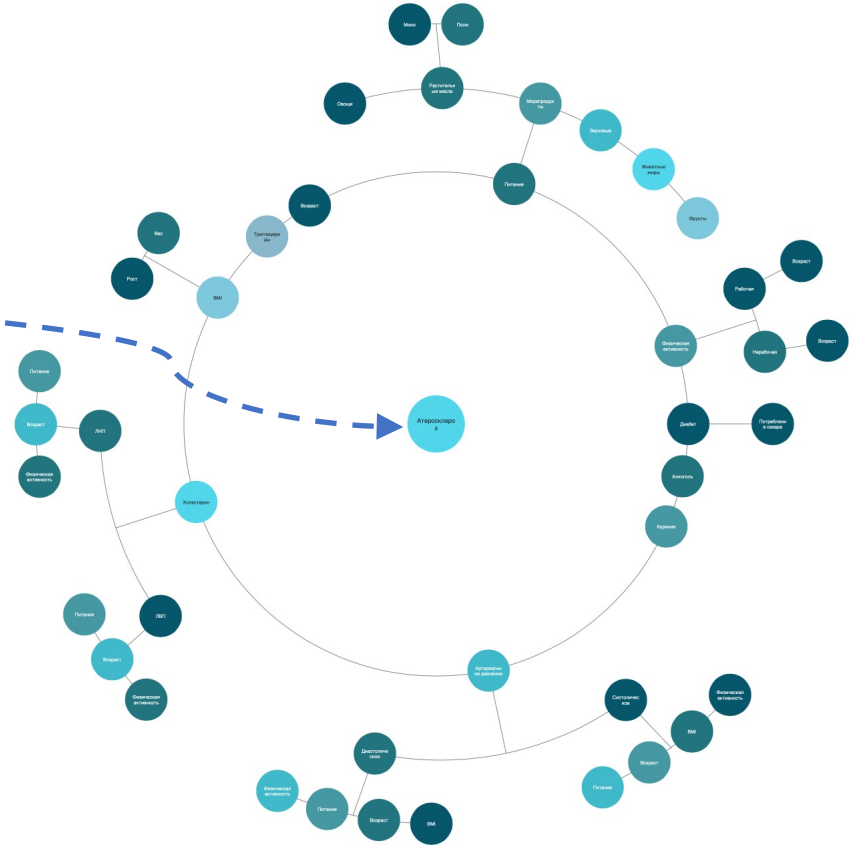
Apple. The Human-Centric AI Podcast
https://podcasts.apple.com/us/podcast/the-human-centric-ai-podcast/

Microsoft Guidelines for Human-AI Interaction
https://www.microsoft.com/en-us/research/project/guidelines-for-human-ai-interaction/

Google partners with NSF to fund human-AI research initiative
A new institute will study how humans and AI working together can make better decisions than either on their own.
https://www.cnet.com/science/google-partners-with-nsf-to-fund-human-ai-research-initiative/

The model



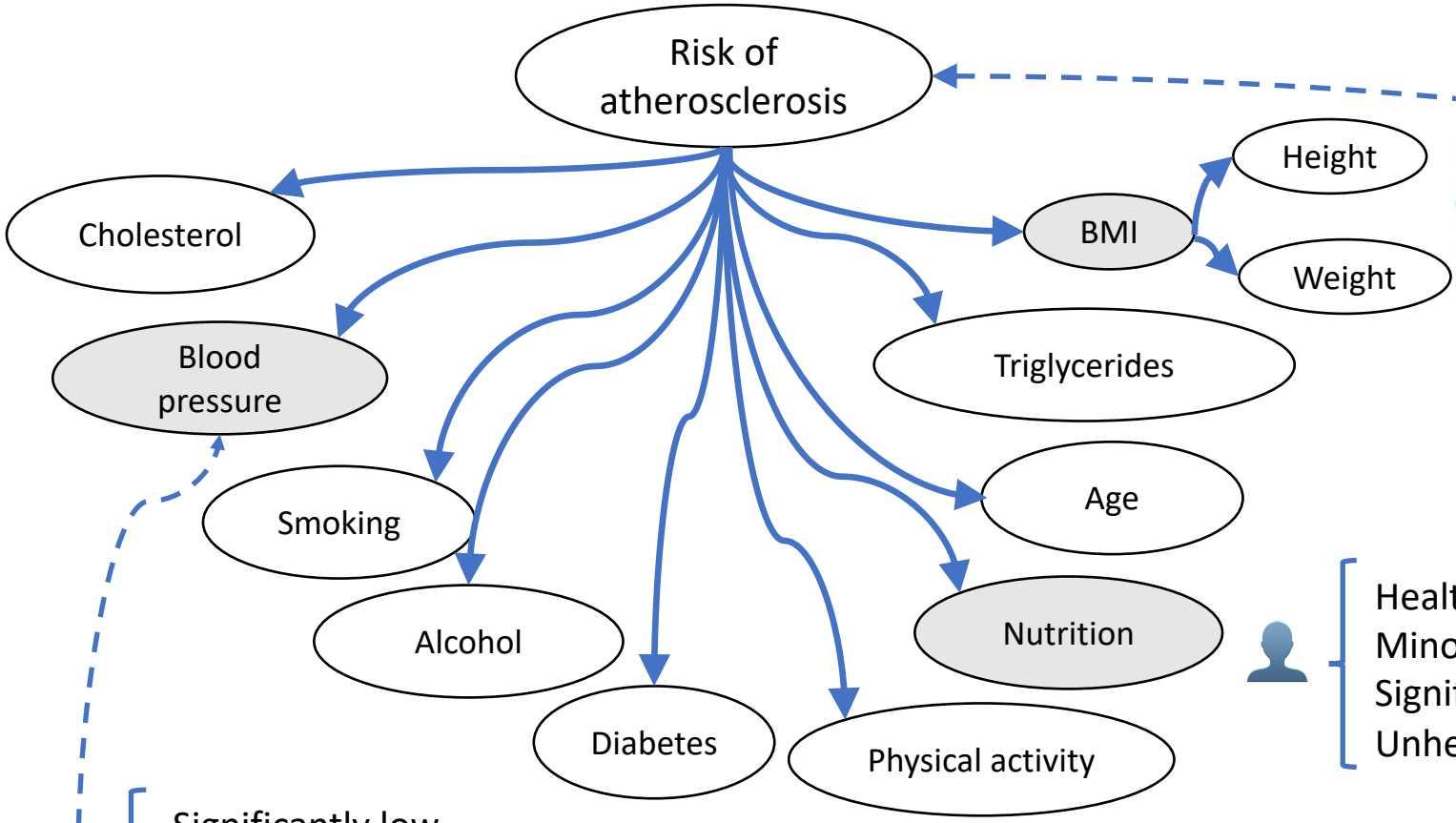


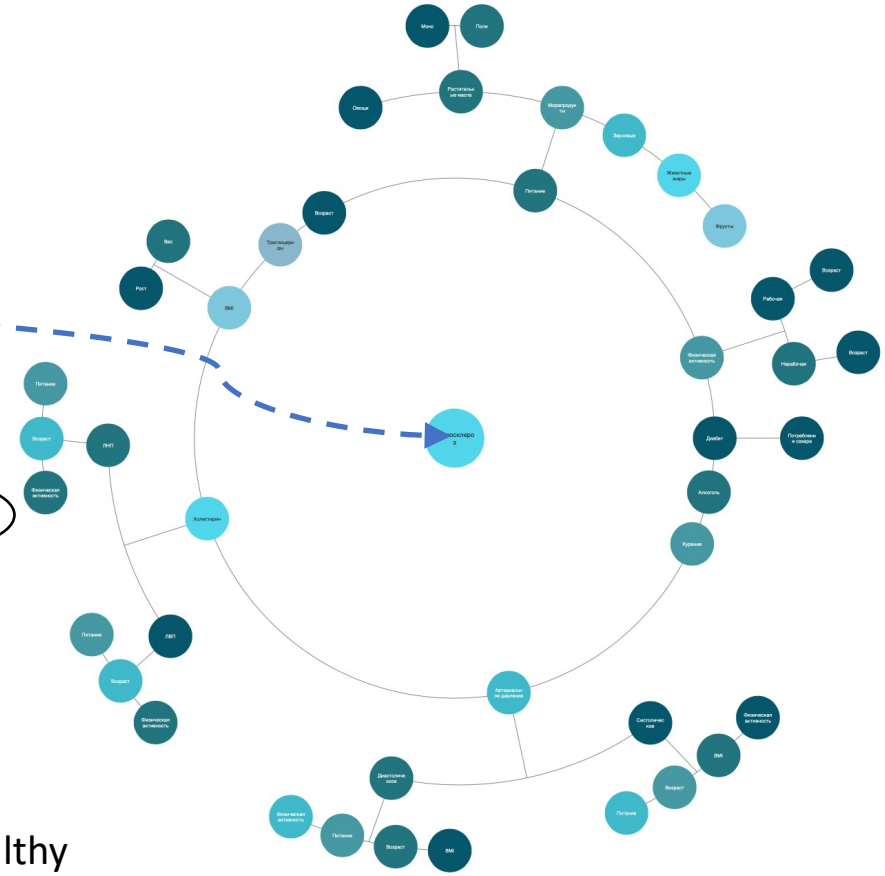
Because model of the problem/process have hierarchical stricture, choice and selection (tuning) of aggregation operators for the nodes of the model is one more important issue in development SEM. We may formulate this problem as follows:

Problem 3. Is it possible to propose the procedures of information aggregation in fuzzy hierarchical dynamic systems which allow us to minimize contradictoriness in the model of problem/process in IMS?

Ryjov A. Information aggregation in fuzzy hierarchical systems. Intelligent Systems. V.6, Issue 1-4, Moscow, 2001, p. 341-364
http://www.intsys.msu.ru/magazine/archive/v6(1-4)/ryzhov.pdf

Measurements





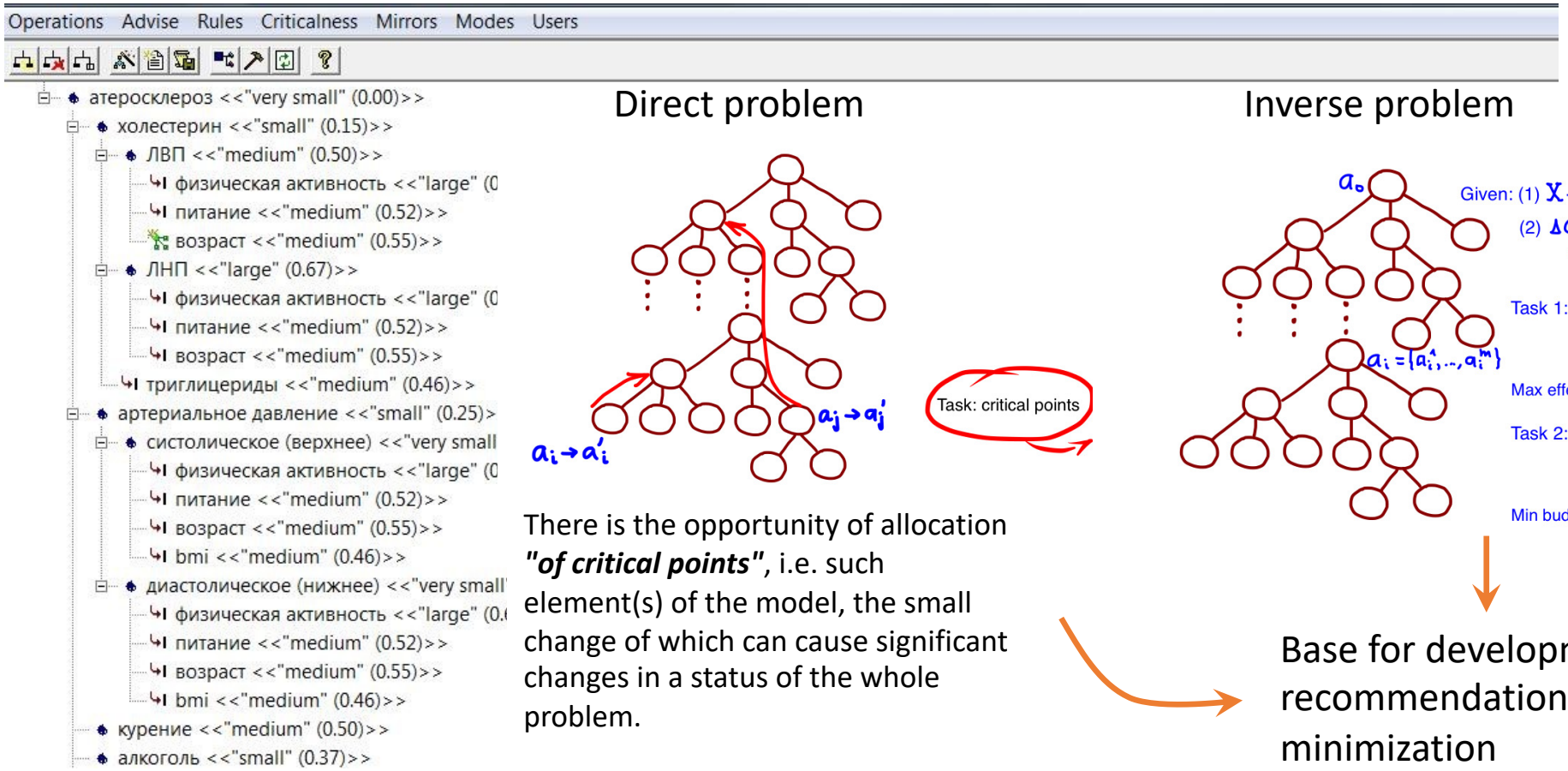
Significantly low
Low
Norm
High
Significant violations

Healthy
Minor violations
Significant violations
Unhealthy

We are not interested in the specific values of the parameter, we are interested in the integral estimation. Such an integral assessment depends on the time of observation, specific parameter changes (for example, day/night), age, eating style, physical activity and other parameters of a particular patient.

It is difficult (and maybe impossible) to develop a mathematical model for calculating the integral estimate; collecting all the data for the model is practically an unsolvable task.

Example 3: Results



Direct problem

Task: critical points

Inverse problem

Given: (1) X - budget

(2) $\Delta c_i = C(a_i^k \to a_i^{k+1})$ - cost of changing

Task 1: Find $\{a_{i_1}, .., a_{i_n}\}$:

$\Delta a_0 \to max$,

$\sum_{i=1}^{n} \Delta c_{ij} \le X$

Max effect for given budget

Task 2: Find $\{a_{i_1}, .., a_{i_n}\}$:

$\sum_{i=1}^{n} \Delta c_{ij} \to min$

$\Delta a_0 \ge \Delta a_0^*$

Min budget for given effect

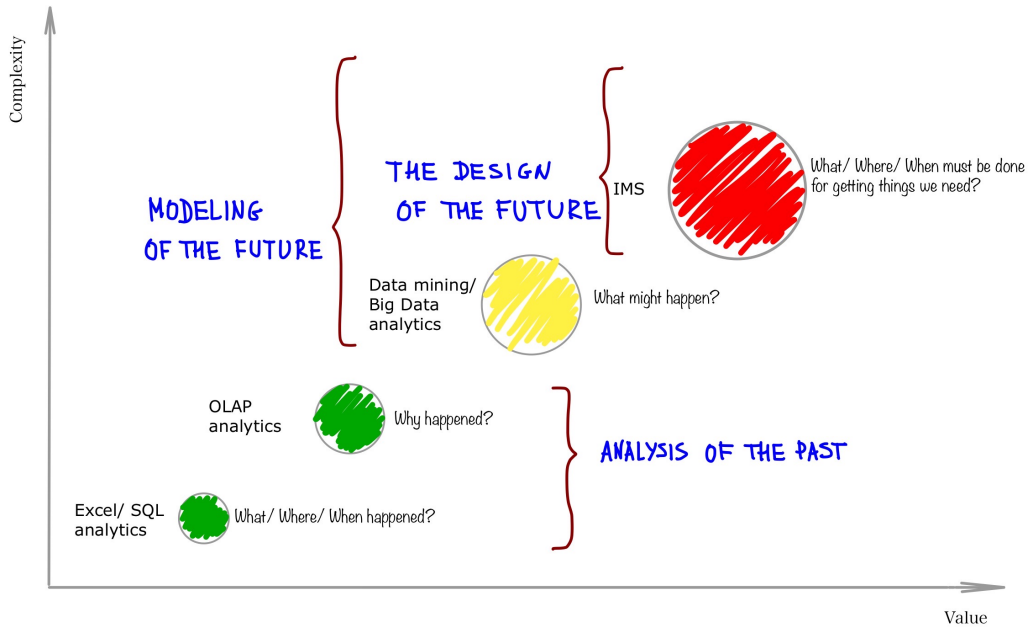
There is the opportunity of allocation ***"of critical points"***, i.e. such element(s) of the model, the small change of which can cause significant changes in a status of the whole problem.

Base for development the personal recommendations for risks minimization

Ahkmedzhanov N.M., Zhukotcky A.V., Kudrjavtcev V.B., Oganov R.G., Rastorguev V.V., Ryjov A.P., Stogalov A.S. System for evaluation and monitoring of risks of cardiovascular disease. *Intelligent Systems*. V.7, Issue 1-4, Moscow, 2003, p. 5-38.
http://intsys.msu.ru/magazine/archive/v7(1-4)/ryzhov-005-038.pdf

Example 3: Lessons

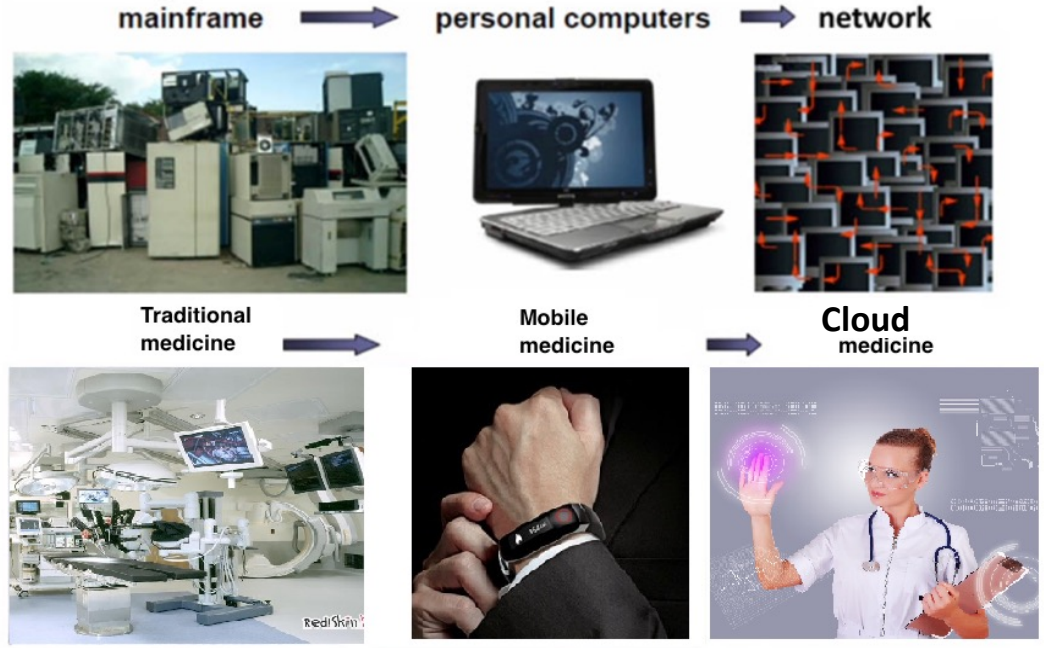
1. Hybrid Intelligence work in the field of healthcare/ medicine
2. Quality of work is good (comparable with experts)
3. Hybrid Intelligence (HI) is effective:
   - When there is no / it is impossible to build a mathematical model of the process in the form of equations, automata, etc.
   - When there are specialists (doctors) who solve the task of evaluation and monitoring on a systematic basis
4. The development of HI is possible:
   - When it is possible to build a semantic model of the process in the form of a set of concepts and their interrelations
   - Real information is received and analyzed - training and customization are possible
5. It is possible to develop optimal HI system from the point of view of:
   - input convenience;
   - consistency of estimates
   - of input information support and modeling

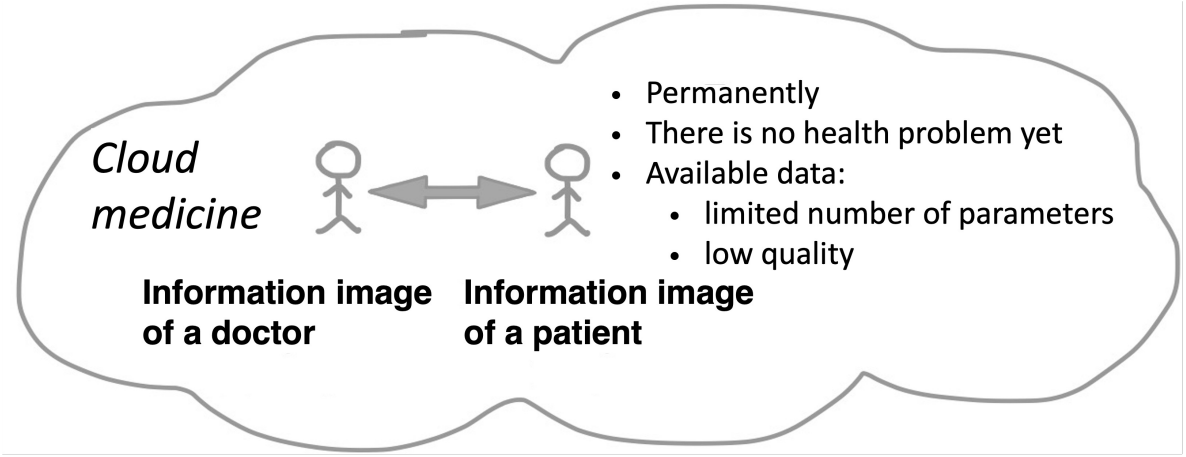


6. It is possible to develop a wide range of similar applications for healthcare, medicine, and other areas.
7. Additional features (bonuses) allow you to calculate the effectiveness of decisions and optimize budgets to achieve goals.

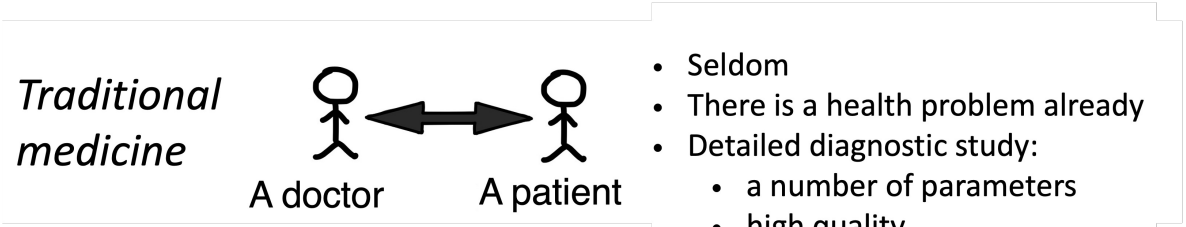
Next step: cloud medicine

- Widespread adoption and continuous improvement of health trackers
- The ability to accumulate and process large amounts of data





- Permanently
- There is no health problem yet
- Available data:
  - limited number of parameters
  - low quality



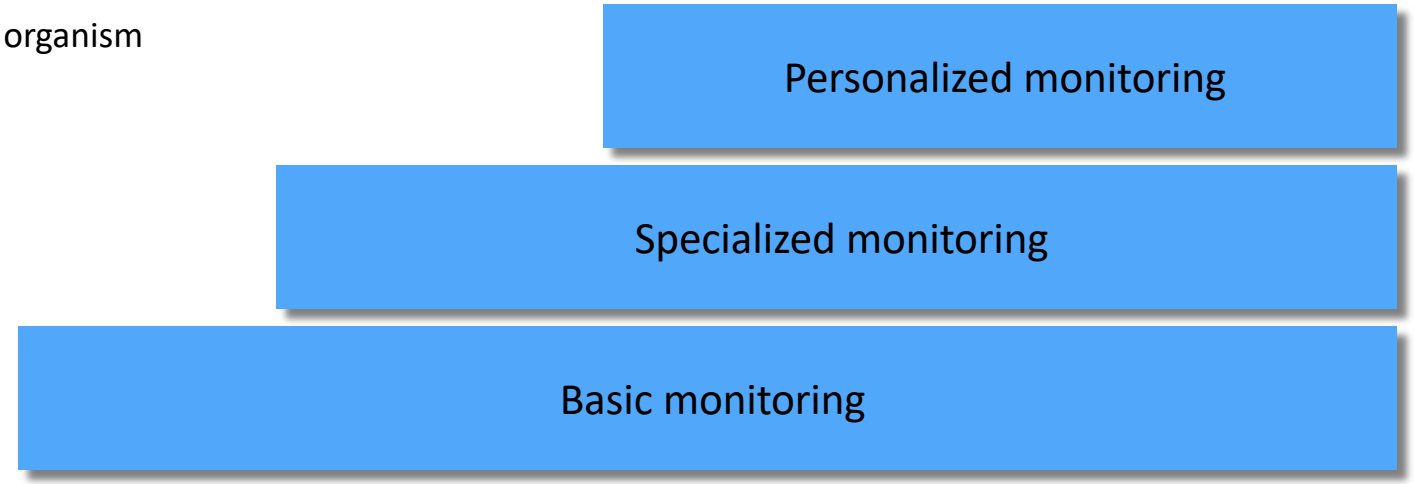
- Seldom
- There is a health problem already
- Detailed diagnostic study:
  - a number of parameters
  - high quality

Levels of the cloud medicine

Goal: genome-based evaluation and monitoring for organism
Segment: VIP

Goal: evaluation and monitoring for systems
(cardiovascular, prostate, etc.)
Segment: everybody who have a problem

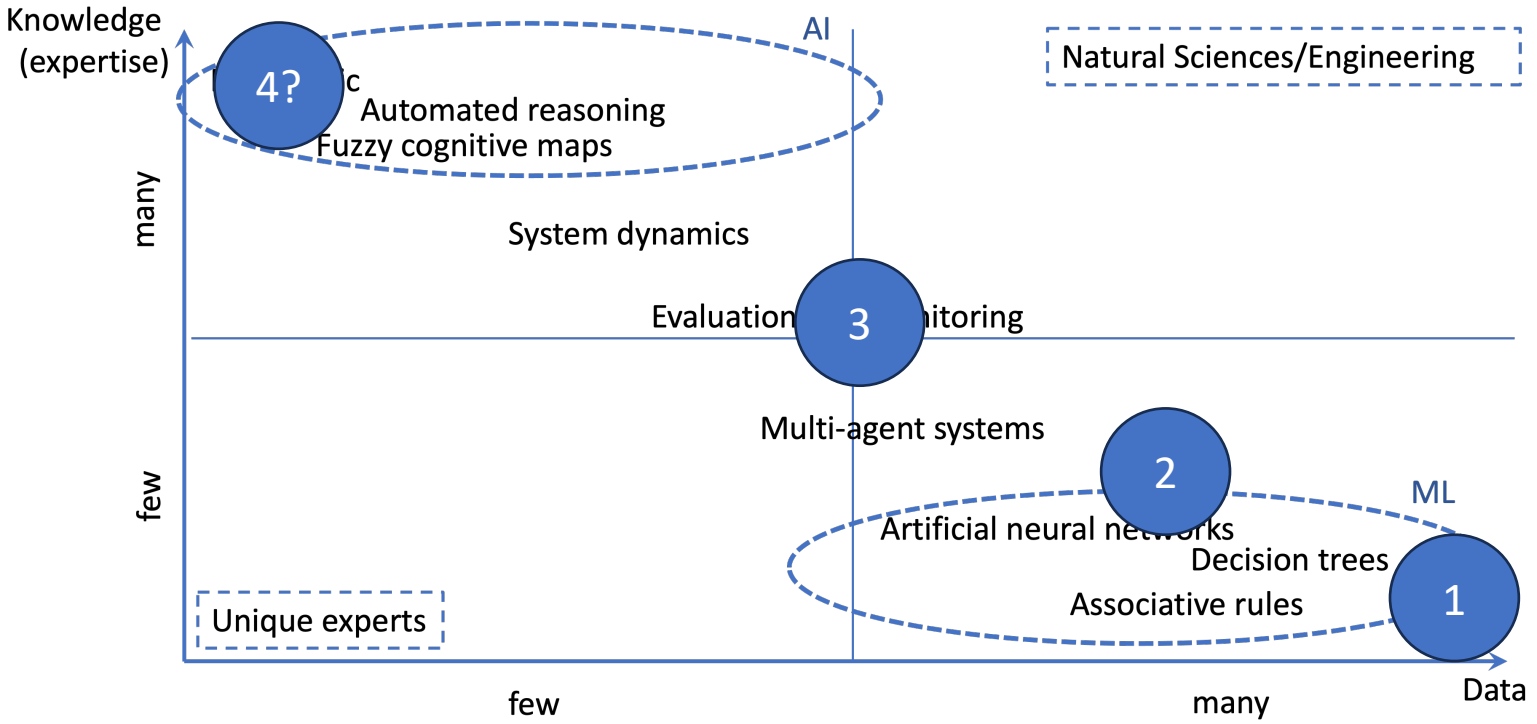
Goal: fitness
Segment: everybody who care about health



Basic	Checking of available parameters	Rule-based general recommendations	"Traffic lights"	Quick: Standard methods
Specialized	Checking of available parameters, new gadgets	System for evaluation and monitoring of complex processes	Evaluation the status and monitoring the progress of target parameters	Several months: There are theory and prototypes *)
Personalized	Analysis of the genome, cluster («similar genomes») analysis	Drugs, life-style, new opportunities	Monitoring of the organism, «computational medicine»	Several years: availability of the analysis of the genome, new theory and tools

Ahkmedzhanov N.M., Zhukotcky A.V., Kudrjavtcev V.B., Oganov R.G., Rastorguev V.V., Ryjov A.P., Stogalov A.S. System for evaluation and monitoring of risks of cardiovascular disease. *Intelligent Systems*. V.7, Issue 1-4, Moscow, 2003, p. 5-38. (http://www.intsys.msu.ru/magazine/archive/v7(1-4)/ryzhov-005-038.pdf)

Resume



1. Data mining/ ML methods work in the field of clinical data analysis. Formal quality of work is good.
   Limitations are: (1) data quality/ motivation; (2) results transparency (explicability, interpretability).
2. Fuzzy logic methods work in the field of population data analysis. Quality of work is good (comparable with experts)
   Limitations are: (1) data sufficiency and adequacy; (2) availability of qualified and motivated experts
   Opportunity: additional features (bonuses) allow to calculate the effectiveness of decisions and optimize budgets to achieve goals.
3. Hybrid Intelligence work in the field of healthcare/ medicine. Quality of work is good (comparable with experts).
   Limitation: availability of qualified and motivated experts.
   Opportunity: (1) treatment personalization and optimization; (2) technological base for ”cloud medicine”

Thank you!

http://itm.ranepa.ru/node/566

Prof. Alexander Ryjov
+7.916.323.4499
alexander.ryjov@gmail.com
http://intsys.msu.ru/en/staff/ryzhov/

HI and healthcare (Ru) -- https://www.youtube.com/watch?v=zvsl0nF4aNc

HI and digital med (Ru)-- https://www.youtube.com/watch?v=W9LnYJGme-0&list=FLoNS93hOJn0DZXzjZom_mlA&index=3

HI theory and application scenarios (En) -- https://www.youtube.com/watch?v=Td6BxQBIBj8