10. Big data a Artificial Intelligence

6MTSN1

Technické a strategické nástroje ve zdravotnictví

Ing. Daniela Kandilaki, Ph.D.

doc. Ing. Peter Pažitný, MSc. PhD.

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání

MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

Obsah

Big data:

- Definície
- Použitie v zdravotníctve

Artificial Intelligence

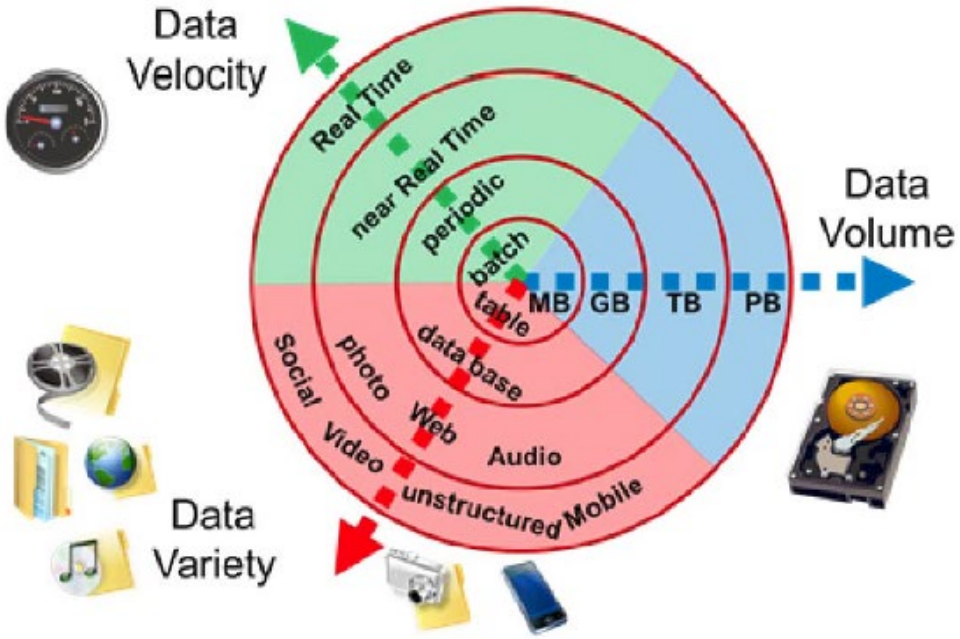
- Definície
- AI všeobecne
- AI v zdravotníctve

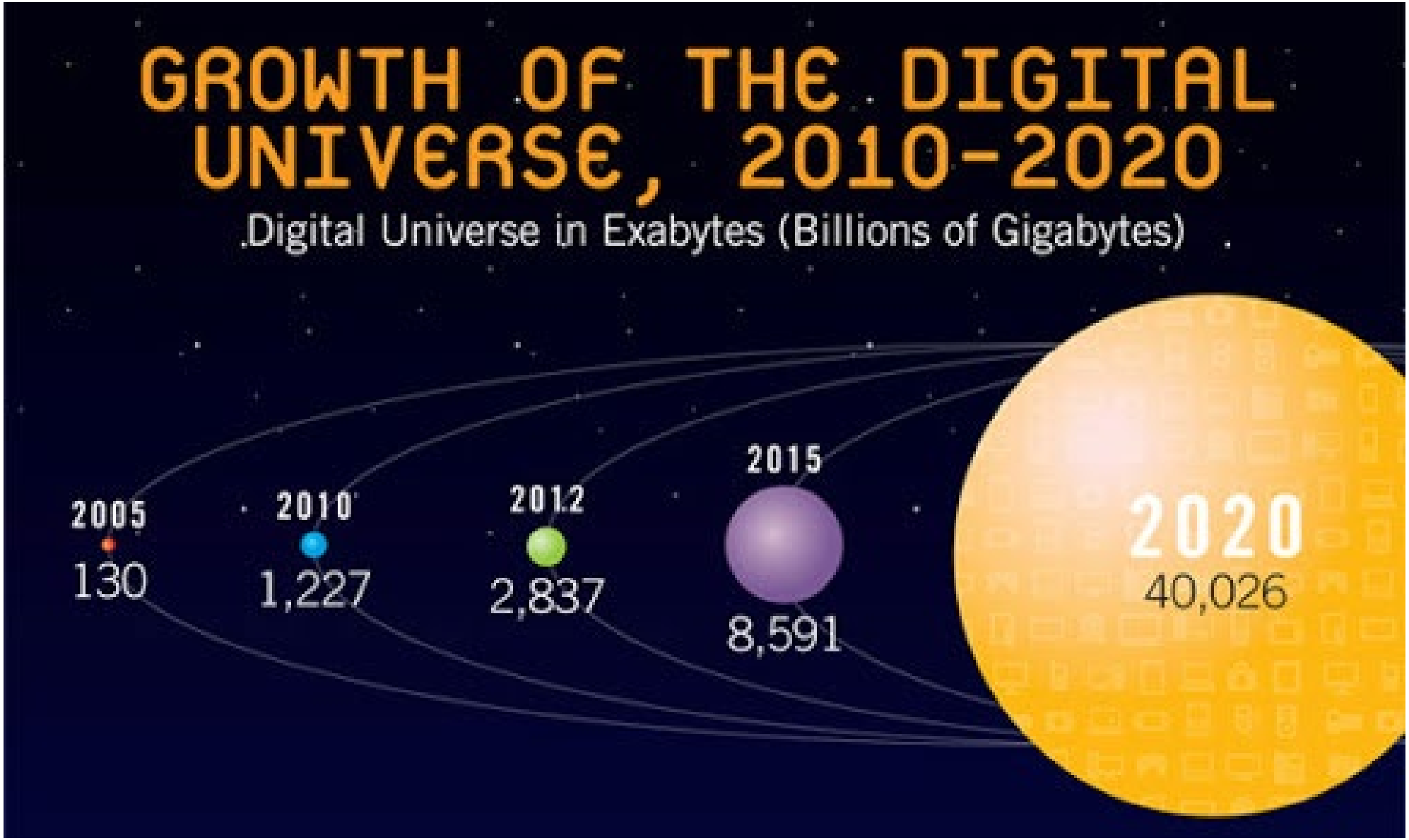
Big data

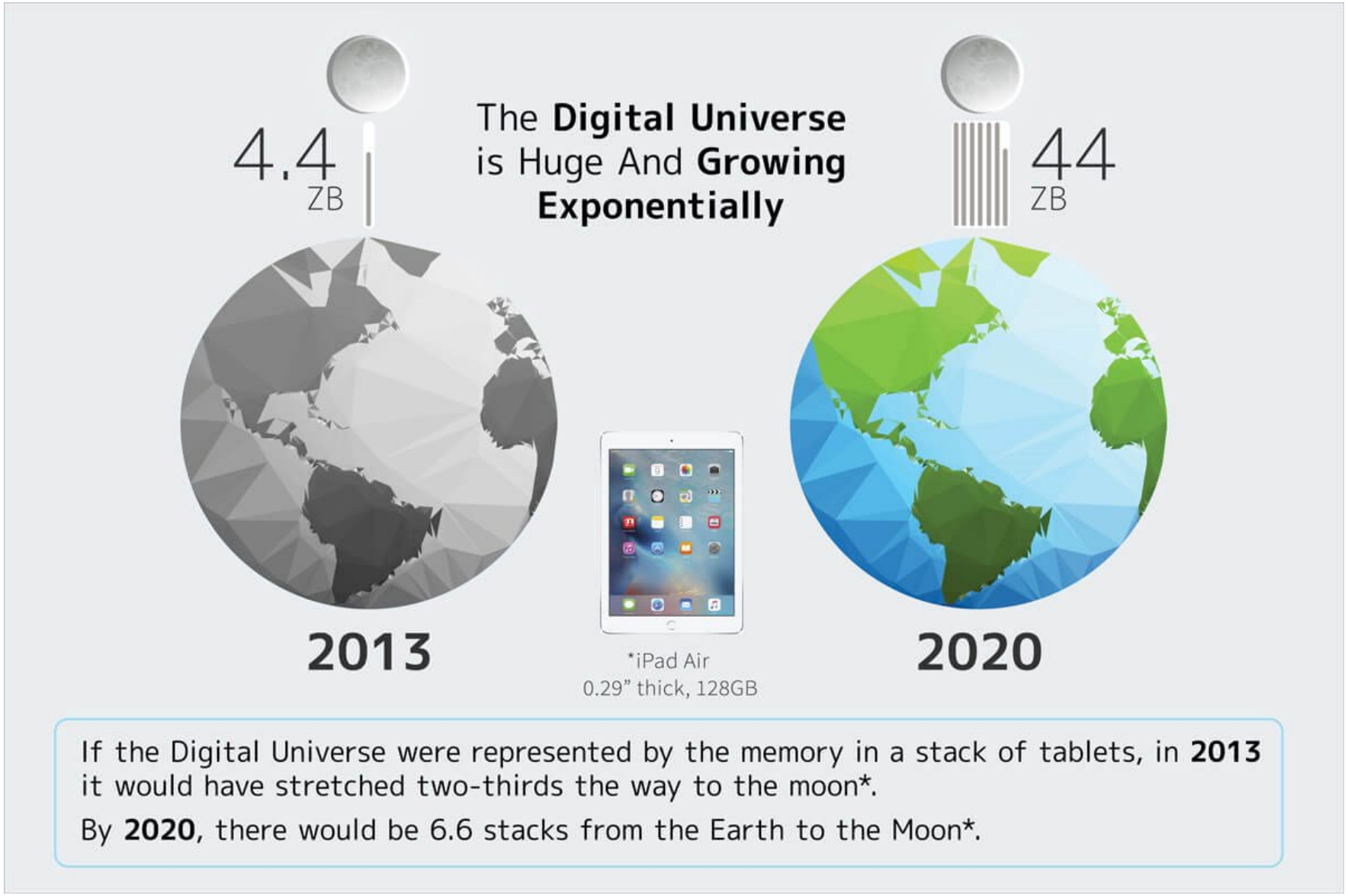
- Predstavuje oblasť skutočne veľkého množstva údajov
- V princípe obsahuje datasety, ktoré prekračujú schopnosti a možnosti bežných softvérov
- Z hľadiska definície je to „pohyblivý terč“, lebo objem dát aj výpočtová kapacita bežných softvérov rastie
- Business intelligence – používa aplikovanú matematiku a deskriptívnu štatistiku na dátach s vysokou informačnou hustotou
- Big data – používa matematickú analýzu, optimalizáciu, induktívnu štatistiku a koncept z nelineárnych systémov, aby boli vyvodené zákony z veľkého množstva dát s nízkou informačnou hustotou, aby boli odhalené vzťahy a závislosti alebo aby boli vytvorené predikcie výsledkov a správania

Big data

- **Volume**
  - Objem vygenerovaných a uložených dát
- **Variety**
  - Typ a pôvod dát.
- **Velocity**
  - Rýchlosť ktorou sú data generované a sprocesované
- **Veracity**
  - Kvalita dát a ich hodnota



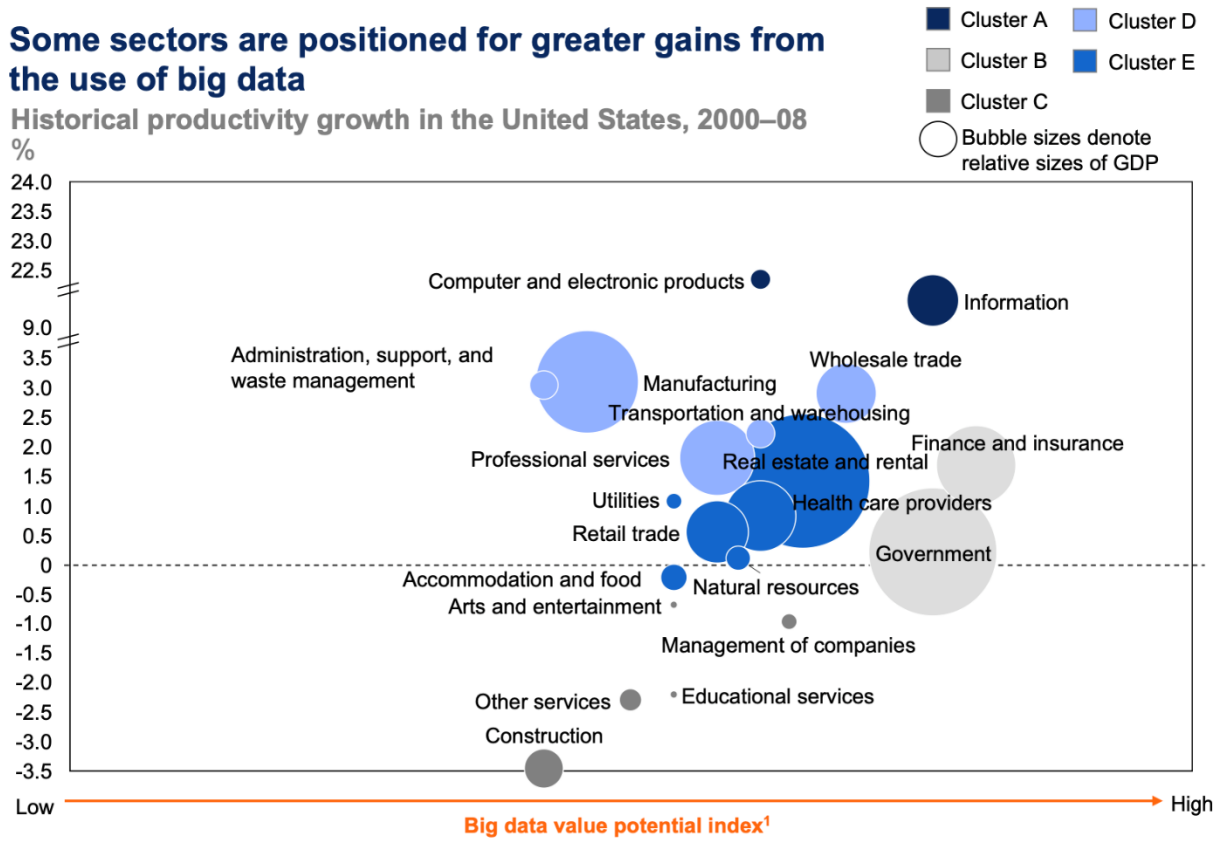

GROWTH OF THE DIGITAL UNIVERSE, 2010-2020
Digital Universe in Exabytes (Billions of Gigabytes)
2005
130
2010
1,227
2012
2,837
2015
8,591
2020
40,026


4.4
ZB
The Digital Universe is Huge And Growing Exponentially
44
ZB
2013
*iPad Air
0.29" thick, 128GB
2020
If the Digital Universe were represented by the memory in a stack of tablets, in 2013 it would have stretched two-thirds the way to the moon*.
By 2020, there would be 6.6 stacks from the Earth to the Moon*.

NAME	EQUAL TO	SIZE(IN BYTES)
Bit	1 Bit	1/8
Nibble	4 Bits	1/2 (rare)
Byte	8 Bits	1
Kilobyte	1024 Bytes	1024
Megabyte	1, 024 Kilobytes	1, 048, 576
Gigabyte	1, 024 Megabytes	1, 073, 741, 824
Terrabyte	1, 024 Gigabytes	1, 099, 511, 627, 776
Petabyte	1, 024 Terabytes	1, 125, 899, 906, 842, 624
Exabyte	1, 024 Petabytes	1, 152, 921, 504, 606, 846, 976
Zettabyte	1, 024 Exabytes	1, 180, 591, 620, 717, 411, 303, 424
Yottabyte	1, 024 Zettabytes	1, 208, 925, 819, 614, 629, 174, 706, 176

Zdravotnictví a big data

Exhibit 2



1 See appendix for detailed definitions and metrics used for value potential index.
SOURCE: US Bureau of Labor Statistics; McKinsey Global Institute analysis

Zdravotnictví a big data

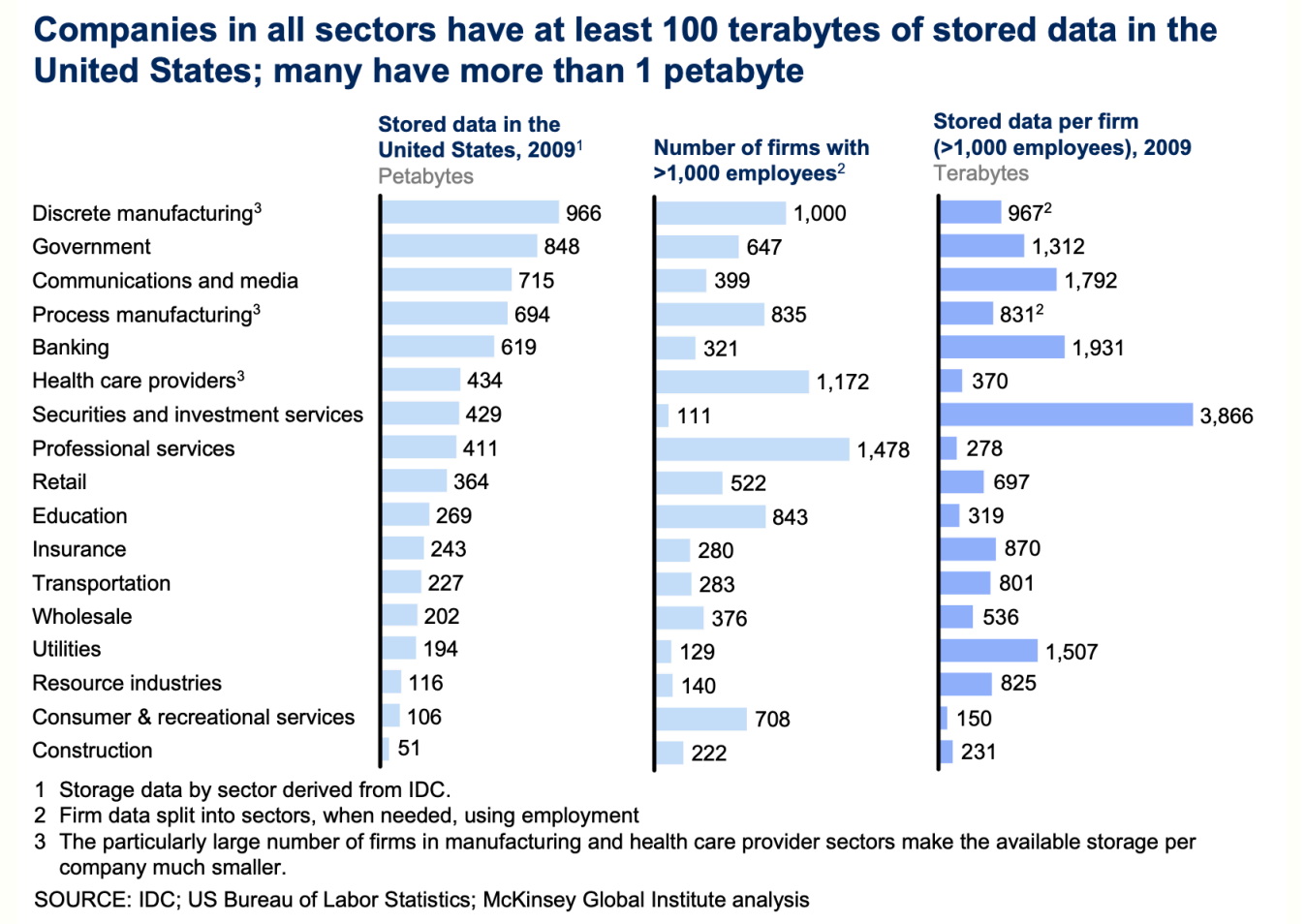
A heat map shows the relative ease of capturing the value potential across sectors

Top quintile (easiest to capture)
2nd quintile
3rd quintile
4th quintile
Bottom quintile (most difficult) to capture)
No data available

Cate-gories	Sectors	Overall ease of capture index[1]	Talent	IT intensity	Data-driven mind-set	Data availability
Goods	Manufacturing					
	Construction					
	Natural resources					
	Computer and electronic products					
	Real estate, rental, and leasing					
	Wholesale trade					
	Information					
Services	Transportation and warehousing					
	Retail trade					
	Administrative, support, waste management, and remediation services					
	Accommodation and food services					
	Other services (except public administration)					
	Arts, entertainment, and recreation					
	Finance and Insurance					
	Professional, scientific, and technical services					
	Management of companies and enterprises					
Regulated and public	Government					
	Educational services					
	Health care and social assistance					
	Utilities					

1 See appendix for detailed definitions and metrics used for each of the criteria.
SOURCE: McKinsey Global Institute analysis

Zdravotnictví a big data



Zdravotnictví a big data

The type of data generated and stored varies by sector[1]

	Video	Image	Audio	Text/ numbers
Banking				
Insurance				
Securities and investment services				
Discrete manufacturing				
Process manufacturing				
Retail				
Wholesale				
Professional services				
Consumer and recreational services				
Health care				
Transportation				
Communications and media[2]				
Utilities				
Construction				
Resource industries				
Government				
Education				

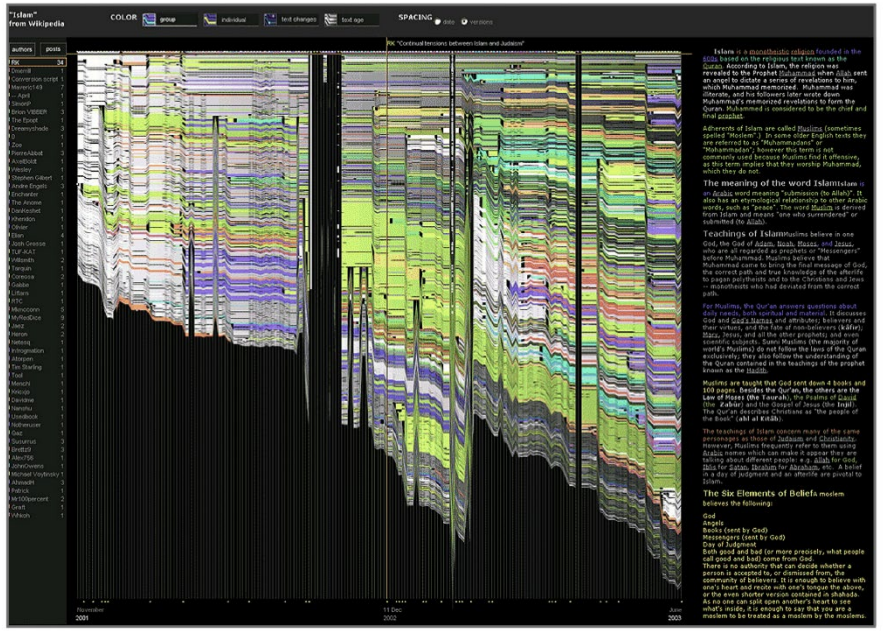
Penetration
- High
- Medium
- Low

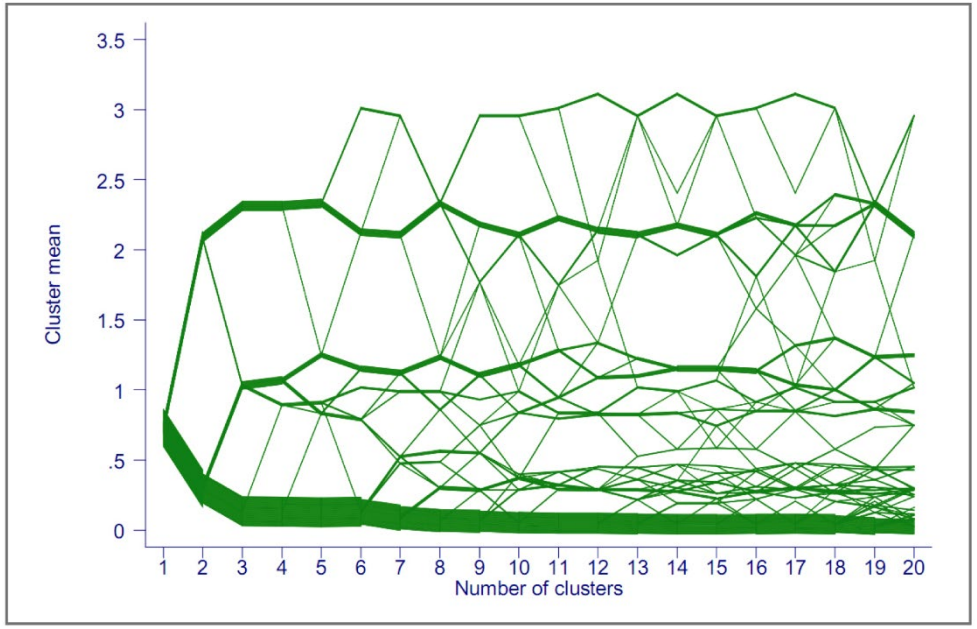
1 We compiled this heat map using units of data (in files or minutes of video) rather than bytes.
2 Video and audio are high in some subsectors.
SOURCE: McKinsey Global Institute analysis

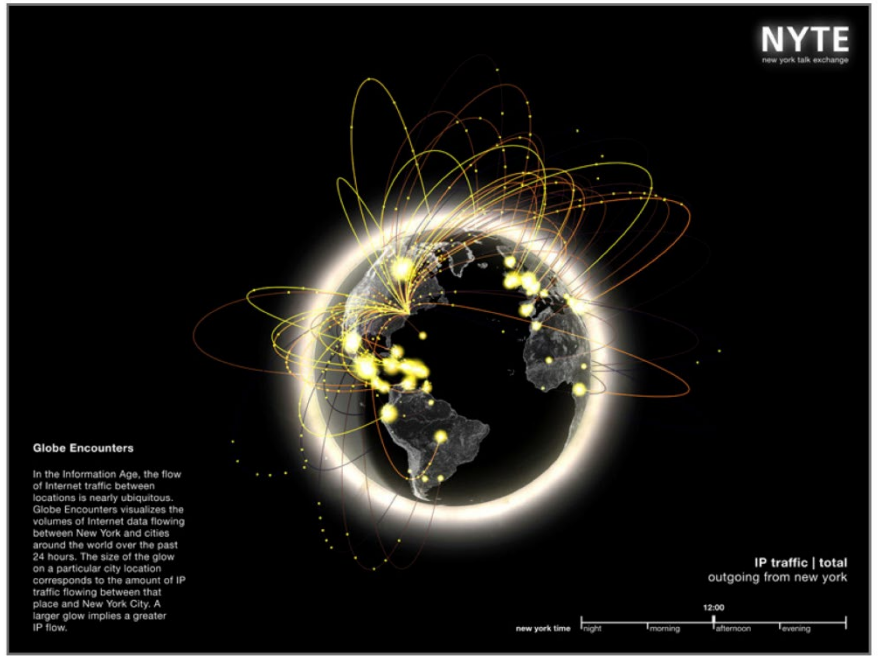
Techniky big data

- Techniky analýzy dát
  (data mining, genetic alghoritms, ..., machine learning, neural networks, ..., pattern recognition, statistics)
- Big data technológie
  (cloud computing, ..., hadoop, mashup, ..., stream processing)
- Vizualizácia – najväčšia výzva
  (tag cloud, cluster gram, history flow, spatial informatin flow)

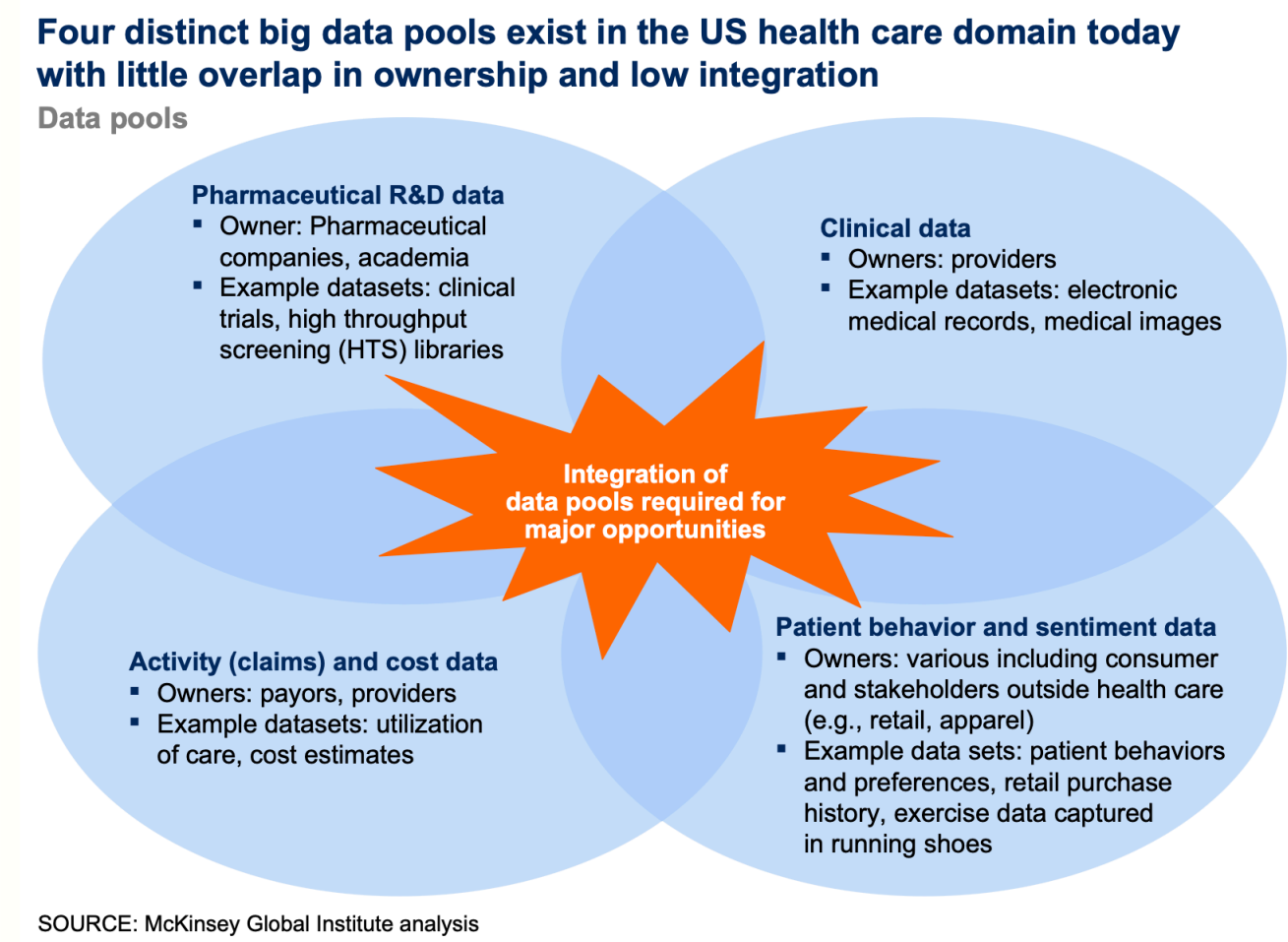
EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY



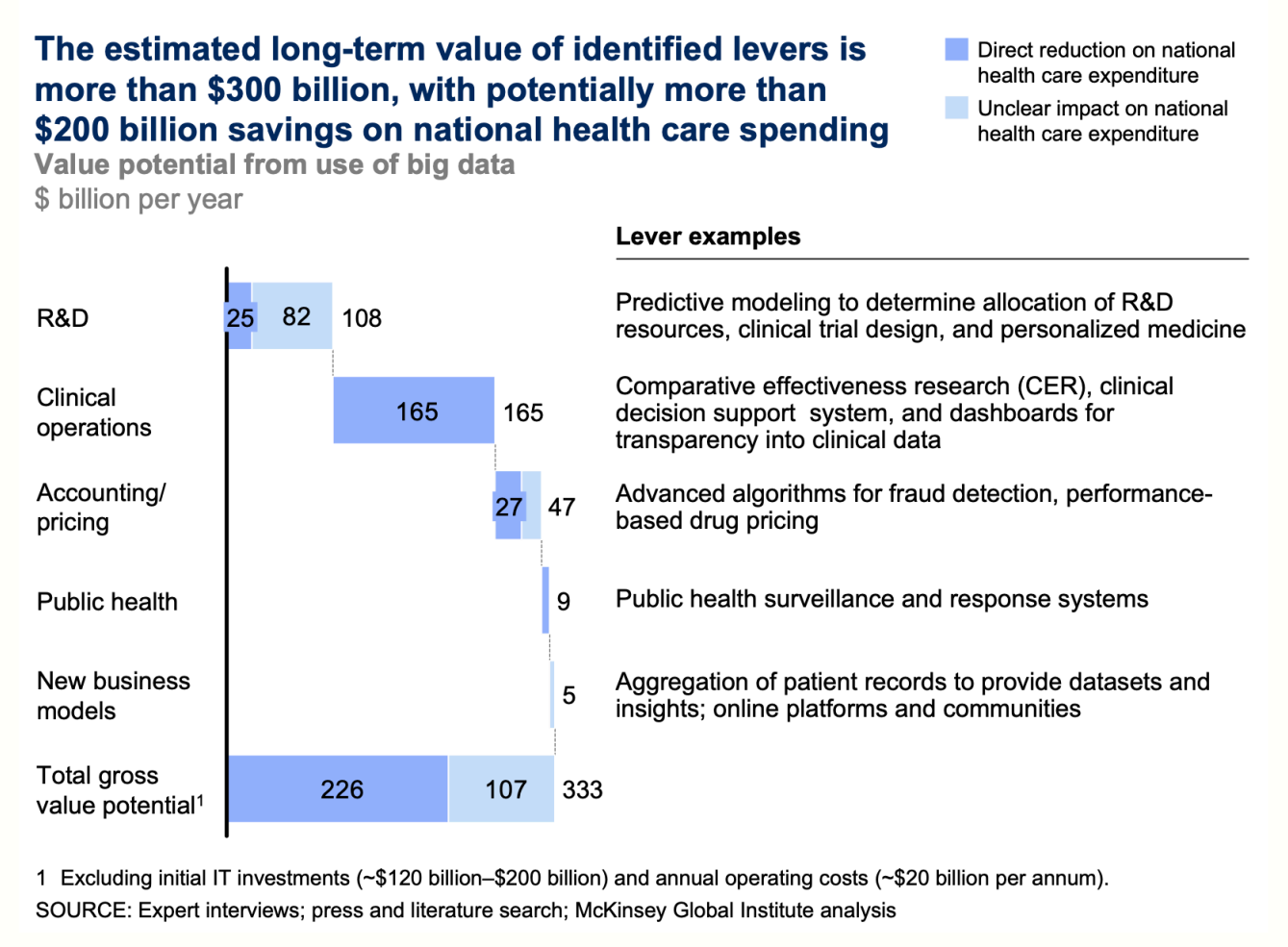




Big data ve zdravotnictví – zdroje dat



Big data ve zdravotnictví – přidaná hodnota



Kto je schopný tieto obrovské objemy dát generované v reálnom čase spracovať?

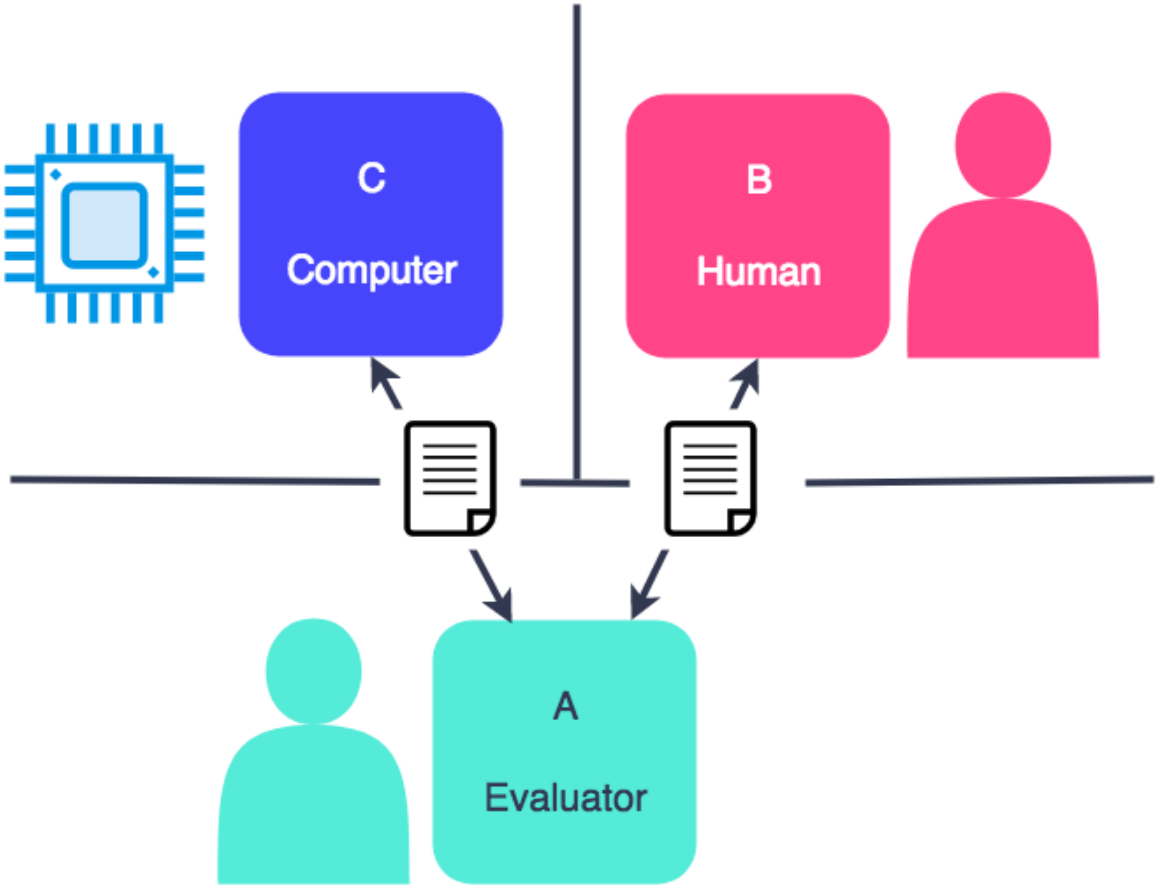
Umelá inteligencia

Umelá inteligencia

Schopnosť stroja imitovať inteligentné ľudské chovanie

- **Reaktívny stroj**
  Stroj môže reagovať len na aktuálny scenár bez schopnosti využiť skúsenosti z predchádzajúcich rozhodnutí
- **2. Limitovaná pamäť**
  Stroj môže vykonávať pozorovania z jeho bezprostredného okolia a spracovať ich pri tvorbe svojho následného rozhodnutia
- **3. Teória mysle**
  Schopnosť porozumieť myšlienkam a emóciám a byť schopný ich zahrnúť do svojho rozhodovacieho procesu a vytvoriť adekvátnu reakciu na svet okolo
- **4. Seba-uvedomená AI**
  Najrozvinutejšia forma AI. Definuje stroj, ktorý má svoje vlastné vedomie.

Turingov test (1950)



Deep blue (1997)

AlphaGo (2015)

AlphaZero vs. Stockfish (2017)

- 24 hodinový „tréning“
- Hranie samého so sebou
- Žiadny prístup k literatúre „otvorení“, či „ukončení“
- Analyzuje 80 000 pozícií za sekundu, ale iba tie najsľubnejšie

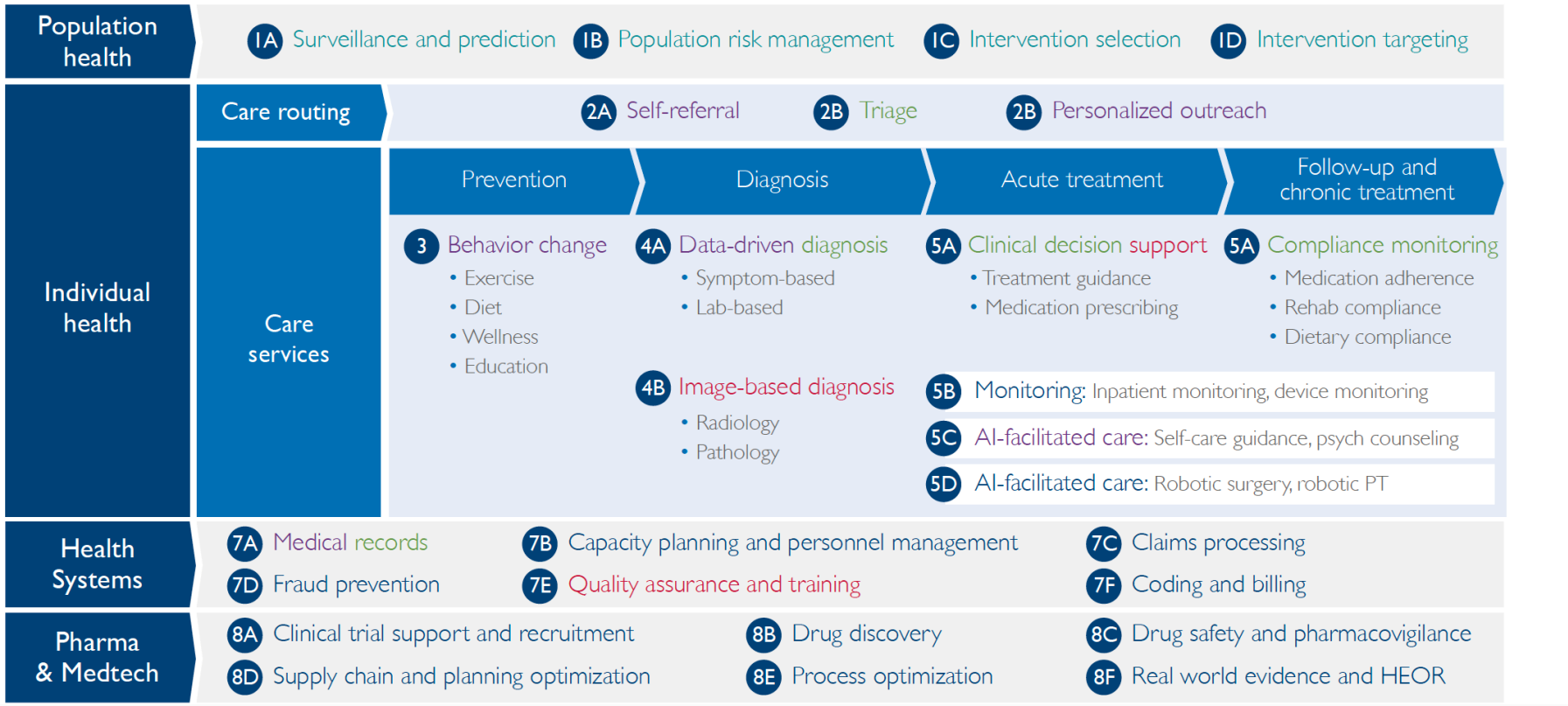
28 výhier 72 remíz 0 výhier

- V tom čase najvýkonnejší šachový program naprogramovaný človekom
- Analyzuje 70 000 000 pozícií za sekundu (všetky)

Film „Ex Machina“ (2014)

AI v zdravotníctve

Figure 3: Framework of all Artificial Intelligence Use Cases in Healthcare Categorized into Four Key Groupings



Population risk management

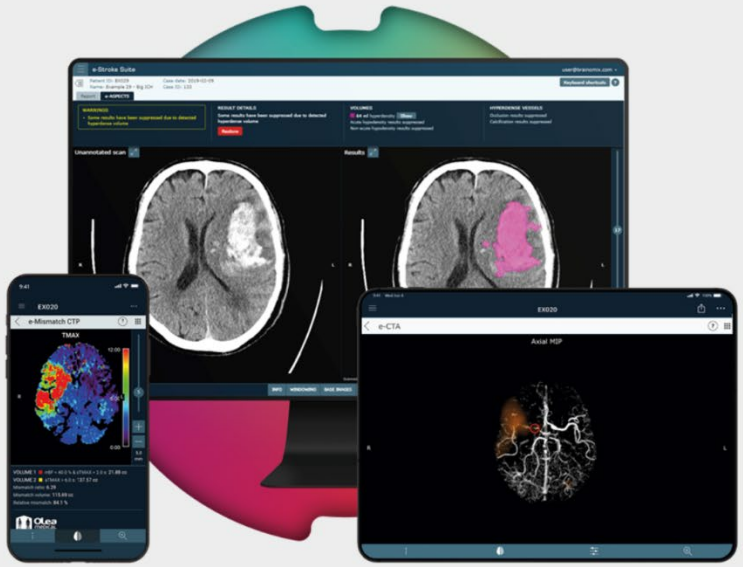
Table 3. Top 10 risk factor variables for CVD algorithms listed in descending order of coefficient effect size (ACC/AHA; logistic regression), weighting (neural networks), or selection frequency (random forest, gradient boosting machines). Algorithms were derived from training cohort of 295,267 patients.

ACC/AHA Algorithm		Machine-learning Algorithms			
Men	**Women**	**ML: Logistic Regression**	**ML: Random Forest**	**ML: Gradient Boosting Machines**	**ML: Neural Networks**
Age	Age	Ethnicity	Age	Age	Atrial Fibrillation
Total Cholesterol	HDL Cholesterol	Age	Gender	Gender	Ethnicity
HDL Cholesterol	Total Cholesterol	SES: Townsend Deprivation Index	Ethnicity	Ethnicity	Oral Corticosteroid Prescribed
Smoking	Smoking	Gender	Smoking	Smoking	Age
Age x Total Cholesterol	Age x *HDL Cholesterol*	Smoking	*HDL cholesterol*	*HDL cholesterol*	Severe Mental Illness
Treated Systolic Blood Pressure	Age x Total Cholesterol	Atrial Fibrillation	HbA1c	Triglycerides	SES: Townsend Deprivation Index
Age x Smoking	Treated Systolic Blood Pressure	Chronic Kidney Disease	Triglycerides	Total Cholesterol	Chronic Kidney Disease
Age x *HDL Cholesterol*	Untreated Systolic Blood Pressure	Rheumatoid Arthritis	SES: Townsend Deprivation Index	HbA1c	*BMI missing*
Untreated Systolic Blood Pressure	Age x Smoking	Family history of premature CHD	BMI	Systolic Blood Pressure	Smoking
Diabetes	Diabetes	COPD	Total Cholesterol	SES: Townsend Deprivation Index	Gender

Italics: Protective Factors

https://doi.org/10.1371/journal.pone.0174944.t003

Brainomix



The Most Comprehensive Stroke Imaging Solution

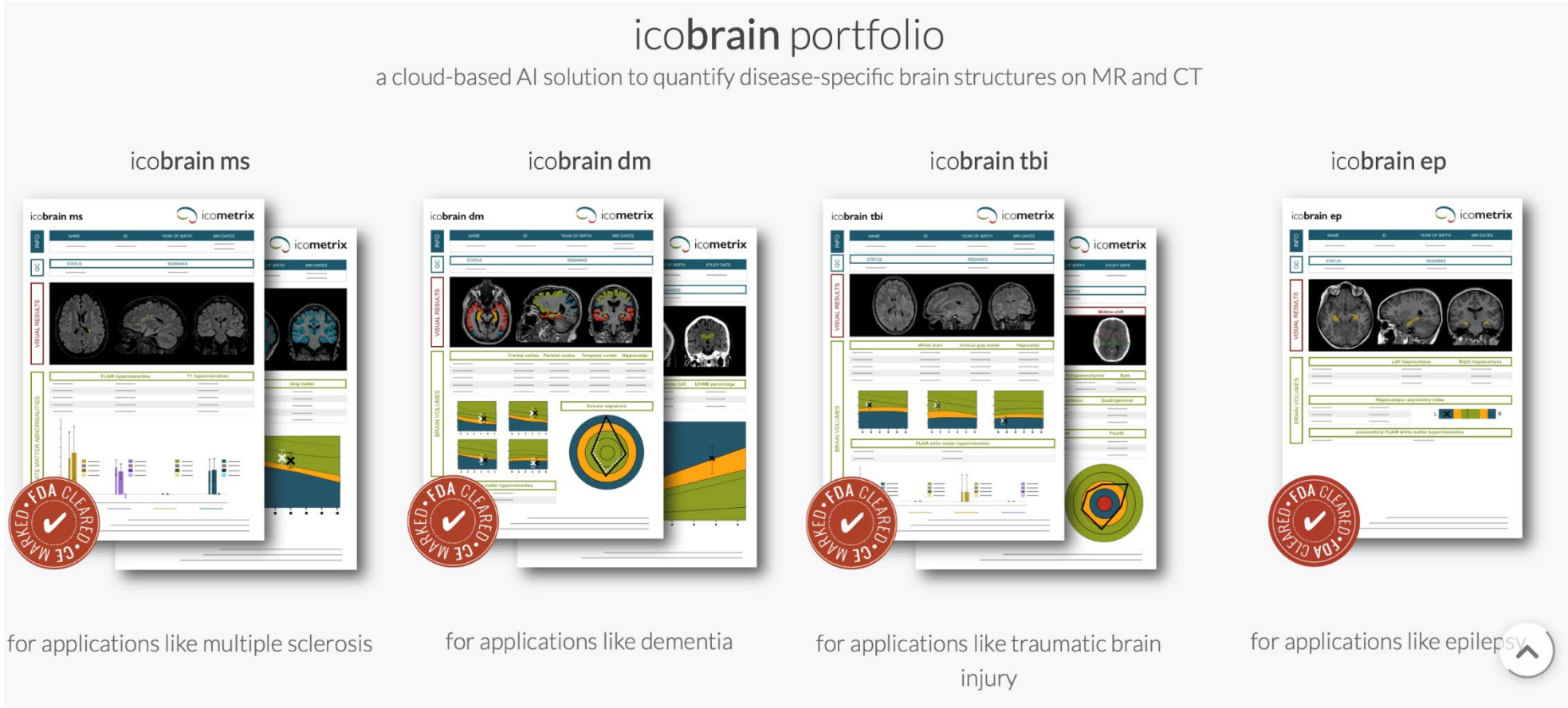
e-Stroke is a collection of tools that use our state-of-the-art AI algorithms to support doctors by providing real-time interpretation of brain scans to help guide treatment and transfer decisions for stroke patients, allowing **more patients to get the right treatment, in the right place, at the right time**.

Advancing the Value of Simple Imaging

Our proprietary software can uniquely generate critical information from simple brain scans that can help **expand patient access to life-saving stroke treatments**.

Our AI-powered advanced technology has been studied in a number of publications, providing technical validation and demonstrating **health economic benefit**, and has also been shown to **improve stroke treatment rates**, both thrombectomy and thrombolysis. Click here to learn more.

IcoBrain


icobrain portfolio
a cloud-based AI solution to quantify disease-specific brain structures on MR and CT
icobrain ms
icobrain dm
icobrain tbi
icobrain ep
for applications like multiple sclerosis
for applications like dementia
for applications like traumatic brain injury
for applications like epilepsy

Da Vinci chirurgický robot

Robotická chirurgia

Robodoc

Cyberknife

Human vs. Robot Surgeons					
0 No Automation Traditional surgery: Human surgeon performs all tasks.	1 Some Assistance e.g. Intraoperative image guidance; human still physically performs all surgery.	2 Partial Automation e.g. ROBODOC hip arthroplasty robot: reduced level of human input required but range of procedures.	3 Conditional Automation e.g. Cyberknife, automated pre-operative planning and radiosurgery (but not technically "surgery").	4 High Automation e.g. Robot capable of performing most, if not all parts of a complex procedure. Negligible human input required.	5 Full Automation e.g. Robot for deep space exploration? Fully autonomous, versatile; no human assistance needed.

Fraud detection

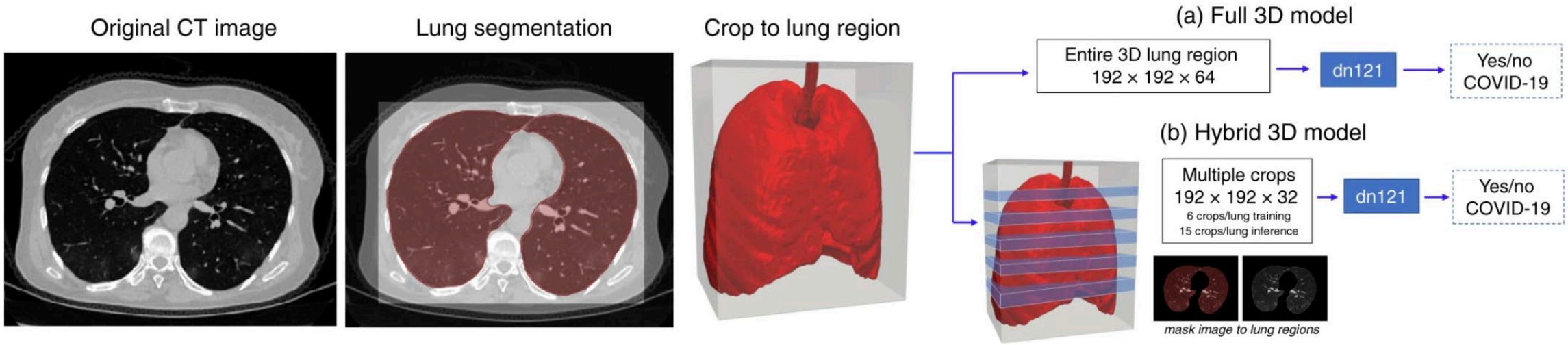
Vendor Comparison: AI for Fraud Detection in Health Insurance

emerj

	FRISS	H2O.ai	SAS	LexisNexis
Company Name	Friss	H20.ai	SAS	LexisNexis
Year founded	2006	2011	1976	1970
Number of employees	130	138	14,052	Year founded
Application description	Anomaly detection	Anomaly detection fraud detection	Predictive analytics for fraud detection	Predictive analytics
Business problem software solves	Reducing loss ratio and detect more fraudulent claims	Reduce money loss on fraudulent claims	Reduce fraud by detecting it as it comes, reduce false positives Also waste and abuse	Identify key components and people of a fraud event, combat fraud at different medical/ pharmaceutical levels
Venture capital raised	$17 MM	$73.6 MM	Existing enterprise	$30 MM
Prominent investors	Aquiline Technology Growth	Nvidia GPU Ventures, New York Life Insurance	n/a	Schoolhouse Partners, Hambrecht
Number of listed case studies	2	0	2	1
Number of listed prominent enterprise clients for this product	3	2	3	2-3
Name of the AI lead at company	Folksam	Kaiser Permanante, HCA	Highmark health, CZ	None mentioned by name
Prominent enterprise clients include	Jeroen Morrenhof	Arno Candel	Jim Goodnight	Jeff Rehil
Highest degree of AI lead, and field of study	Doctorate, Business information systems	PhD, Computational physics	PhD, Statistics	MS, Computer science
Year the AI lead earned their highest degree	1995 University of Amsterdam	2005 ETH Zurich	1972 NC State University	1987 The Johns Hopkins University

AI a Covid-19

From: Artificial intelligence for the detection of COVID-19 pneumonia on chest CT using multinational datasets



All CT images under lung segmentation for localization to chest cavity region. Following cropping to lung region, two methods were considered for differentiation of COVID-19 from other clinical entities. **a** Full 3D Model resampled the cropped lung region of CT to a fixed size (192 × 192 × 64 voxels) for input to algorithm. **b** Hybrid CT resampled the cropped lung region of CT to fixed resolution (1mm × 1mm × 5mm) and sampled multiple 3D regions (192 × 192 × 32) for input to algorithm. At training, 6 regions/patient were used. At inference 15 regions/patient were used and results were averaged to produce final probability of COVID-19.

Zhrnutie

- **Big data** sú prezentované datasety, ktoré prekračujú schopnosti a možnosti bežných softvérov. Sú špecifikované za pomoci 4 základných charakteristík:
  - Volume
  - Variety
  - Velocity
  - Veracity
- Obrovské množstvo dát, ktorá big data predstavujú, sú schopné zpracovávať **systémy umelej inteligencie**. Umelá inteligencia je schopnosť stroja imitovať inteligentné ľudské správanie.
- Big data a umelá inteligencia pomáhajú ve zdravotníctve k přesnejšej a rychlejšej diagnostike, i pri následnej liečbe a starostlivosti o pacientov. V starostlivosti a liečbe je možne využíť i napríklad aj robotickú chirurgiu.

Zdroje

- HARMON A. Stephanie et al, 2020. *Artificial Intelligence for the pneumonia on COVID-19 pneumonia on chest CT using multinational datasets*. Nature communications, Aug 14, (11)1.
- McKinsey Global Institute analysis McKinsey Global Institute | McKinsey & Company
- US Bureau of Labor Statistics U.S. Bureau of Labor Statistics (bls.gov)
- https://www.coworthtechnologies.com/blog/the-digital-universe-how-will-it-impact-your-business/
- https://trak.in/tags/business/2014/04/15/digital-data-universe-expansion-2020/
- *https://theconversation.com/understanding-the-four-types-of-ai-from-reactive-robots-to-self-aware-beings-67616*
- *https://botsociety.io/blog/2018/03/the-turing-test/*
- *https://www.chess.com/sv/lessons/ibm-deep-blue-vs-kasparov*
- *https://sites.google.com/site/uirqfuiirnfe/ai-applications/alphago*
- https://www.chess.com/blog/SamCopeland/a-new-kind-of-chess-top-10-of-the-2010s-alphazero-vs-stockfish-2017
- *https://fangirlish.com/2021/01/03/a-z-movie-reviews-ex-machina/*
- *Artificial Intelligence in Global Health, US AID, 2019*
- *PANESAR et al.: Artificial Intelligence and the Future of Surgical Robotics, Annals of Surgery, volume 270, Number 2, 2019*
- www.brainomix.com
- www.icometrix.com
- www.davincisurgery.com
- Harmon, S.A., Sanford, T.H., Xu, S. *et al.* Artificial intelligence for the detection of COVID-19 pneumonia on chest CT using multinational datasets. *Nat Commun* **11,** 4080 (2020). https://doi.org/10.1038/s41467-020-17971-2

Děkuji za pozornost