WWW.IARIA.ORG

PANEL
…on Software and Data

Theme: Designing Complex Systems
Topic: Machine Learning and Visual Data

MODERATOR:
Petre Dini, IARIA

petre@iaria.org

Petre

- (partially) retired | (mostly) volunteering | enjoying (affordable) traveling
- Past topics: mathematics, VLSI, formal languages, protocol V&V, real-time embedded systems, nomadic code, active routing, software arch, (AT&T) IMS, (Cisco) IOS, Fault Mgr, Performance Mgr, …Data, Learning,….
- Academics: U de Montreal, Concordia U, McGill U | (AT&T) Stanford, (Cisco) Berkeley, China Space Agency Center - Beijing | Cisco-IBM Coop Dir., AT&T-Cisco Coop. Dir.
- Industrial/research: BNR (Nortel R&D), CRIM (R&D), AT&T, Cisco Systems, Inc.
- Salient: 19 US Cisco patents, (co)supervisor: 38 Master&PhD
- Current hobbies: self-x, systems/apps adaptation, conflicts in decision policies, crowd-in-the-middle, reflective architectures, new trends in software development (apps)
- Open: how can a piece of software can realize by itself (or by any other means) that a copy of that piece was (? illegally) made somewhere [see CLOUD]
- Still learning …from you!
- Yet, playing: 4 grandkids, my neighbors

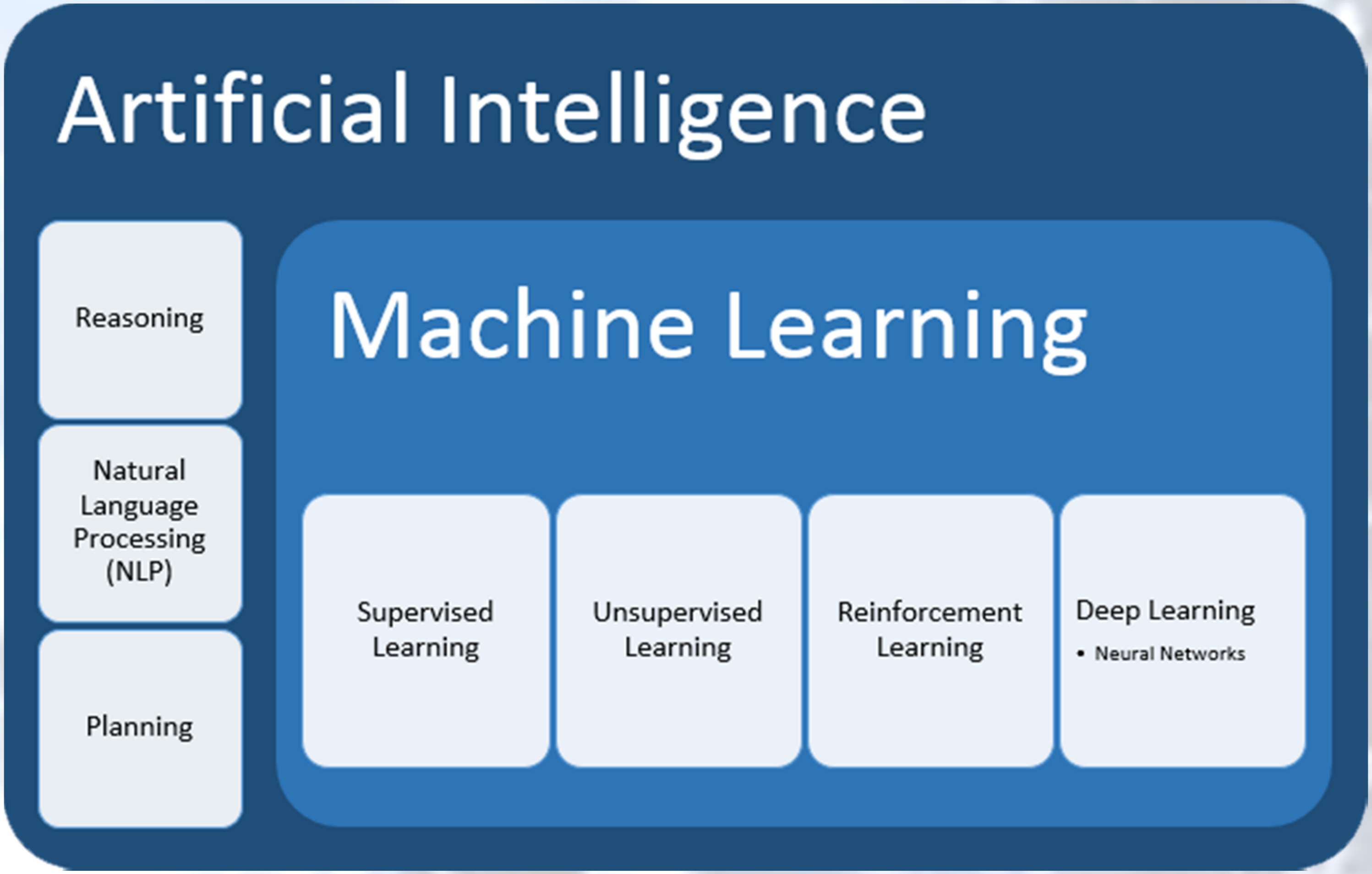
Keywords Invasion

- **Machine Learning**
- **Deep Learning (Neuronal Networks, Sandwitch, etc.)**
- **Data Visualization (Uni-, Multi-variables, Links)**
- **Semantic Gap/Semantic Matching**
- **Ontologies/Taxonomies**
- **Prediction/Intuition/Data Analytics/Data Science**
- **Advanced Artificial Intelligence**
- **Accelerated Deep Learning/Deep Thinking**
- **Data Orchestration (open data, data sets, visible/hidden links)**

➔

- **Brain-Like computing (*neuromorphic computing* mimics the brain's structure)**
- **Intelligent-Analysis-as-a-Service**
- **Prediction-as-a-Service**
- **Data Science**

Data and Learning



credit: https://www.ibm.com/analytics/machine-learning

BIG | the Vs | 3v, 5v, 7v, 10v, …. ?

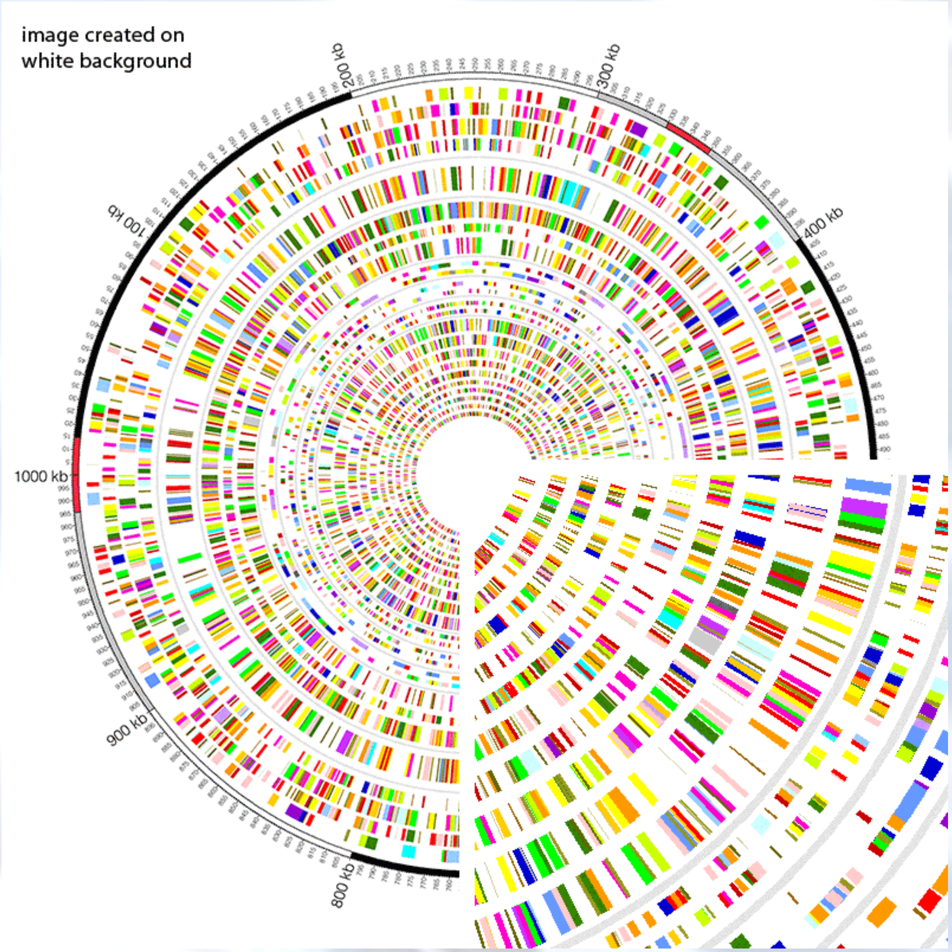
- **Volume** (length of a records, # of records) (entity-relationship databases)(datasets) || BIG vs. HUGE
- **Variety** (types: strings, pictures, voice, etc.) (structured, non-structured)
- **Veracity** (precision and accuracy of data)
- **Velocity** (of change)
- **Value** (as a business/service) IMPACT
- **Volatility** (temporary; quick action)
- **Vasting resources**
  (storage, computation, transfer)
- **Viability** (are data still useful?)
- **Visibility** (open, hidden, ..)
- **Validity**
  (are there still valid/updated data?)
  (in context validity)
  (e-government datasets)

- incomplete
- redundant
- inconsistent
- noisy

quality of data

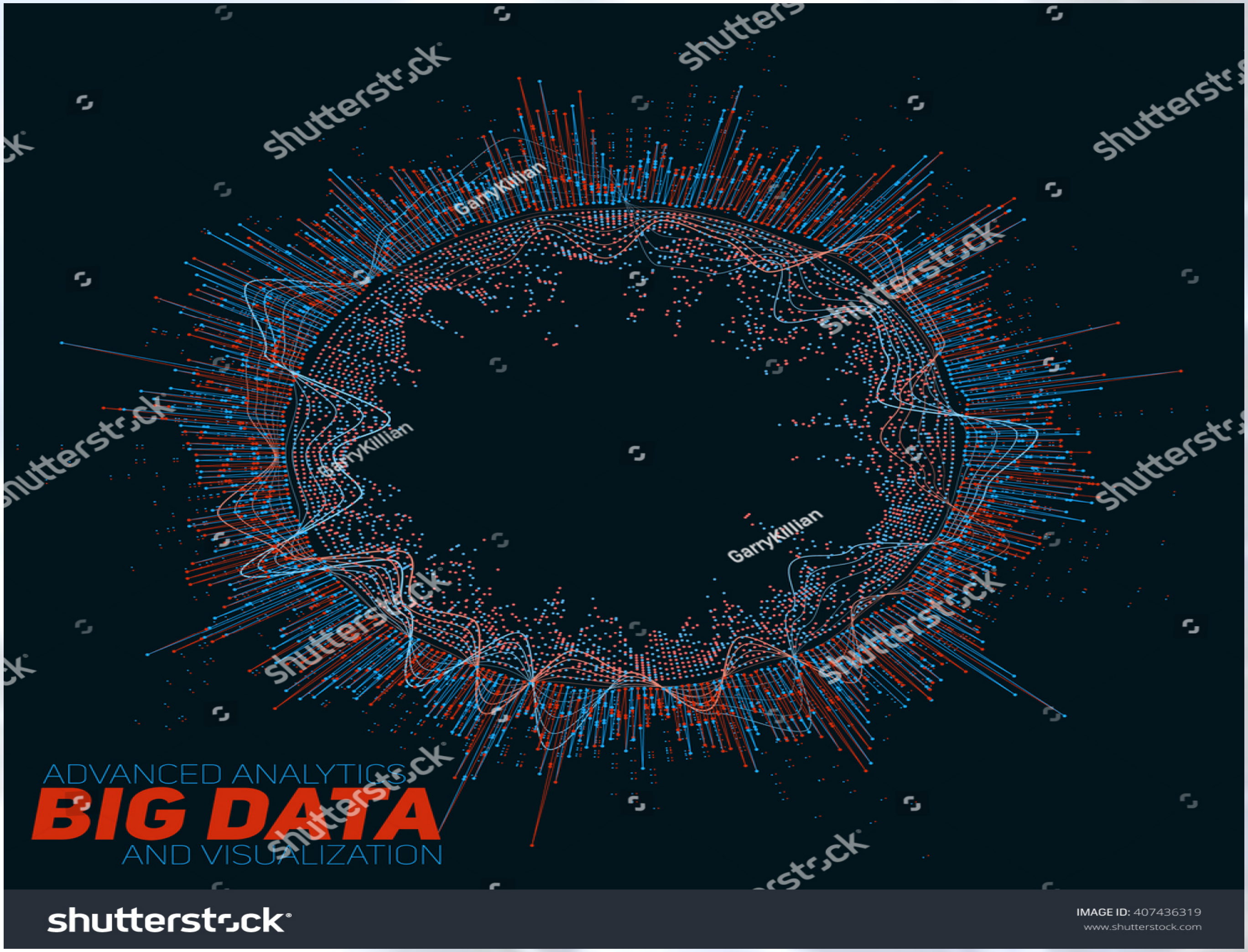
filling missing values with estimated values calculated for complete records of the same dataset

Big/Huge Data Visualization



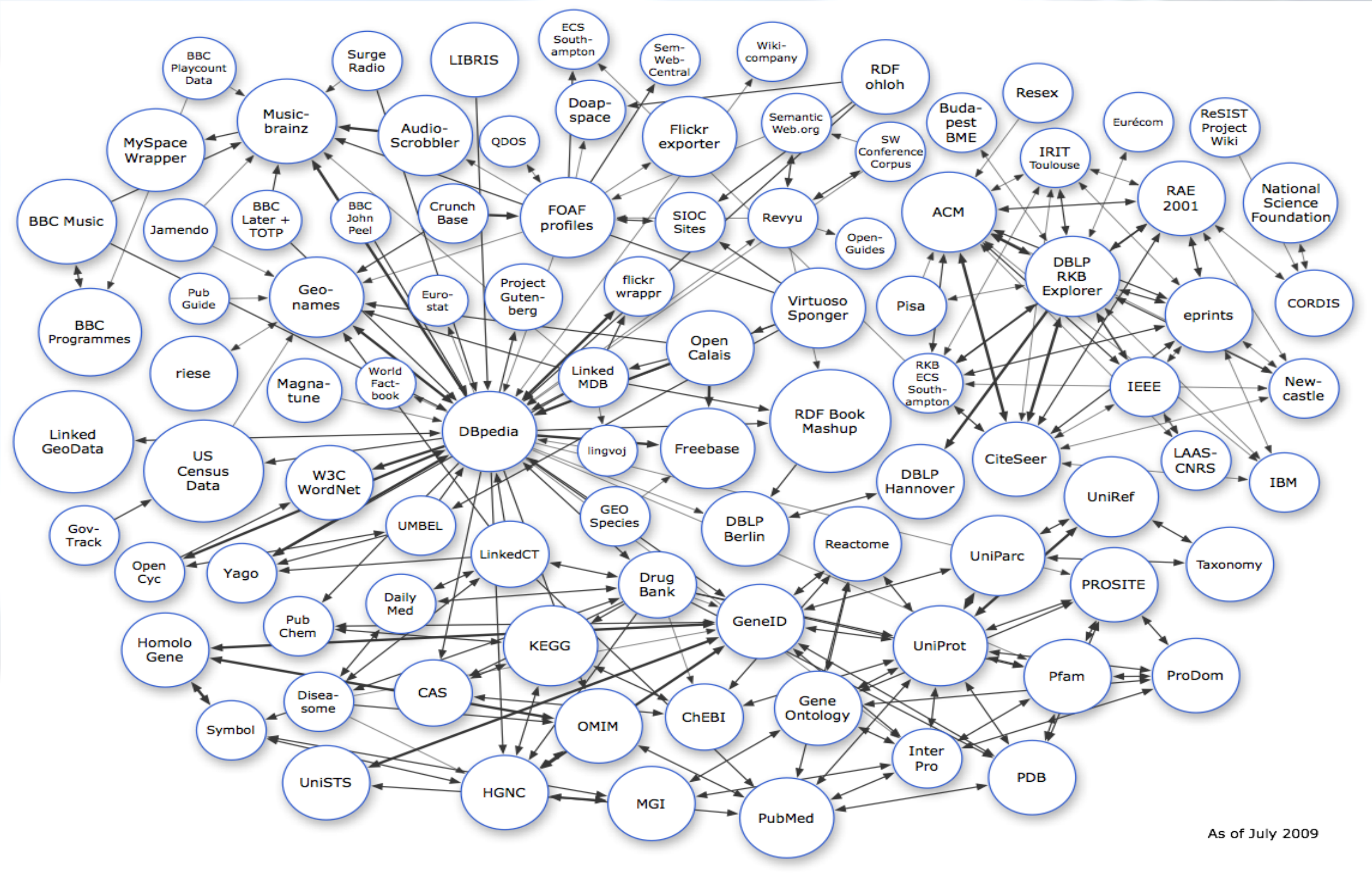
https://www.google.com/search?source=hp&ei=taOBXOv8B9Kt5wLkkLbwCA&q=data+visualization+images&oq=data+visualization+images&gs_l=psy-ab.1.0.0l2j0i22i30l5.1769.7553..10480...0.0..0.175.3377.5j20......0....1..gws-wiz.....0..0i131.BzVu5eI6-MI

Big Data Clusters



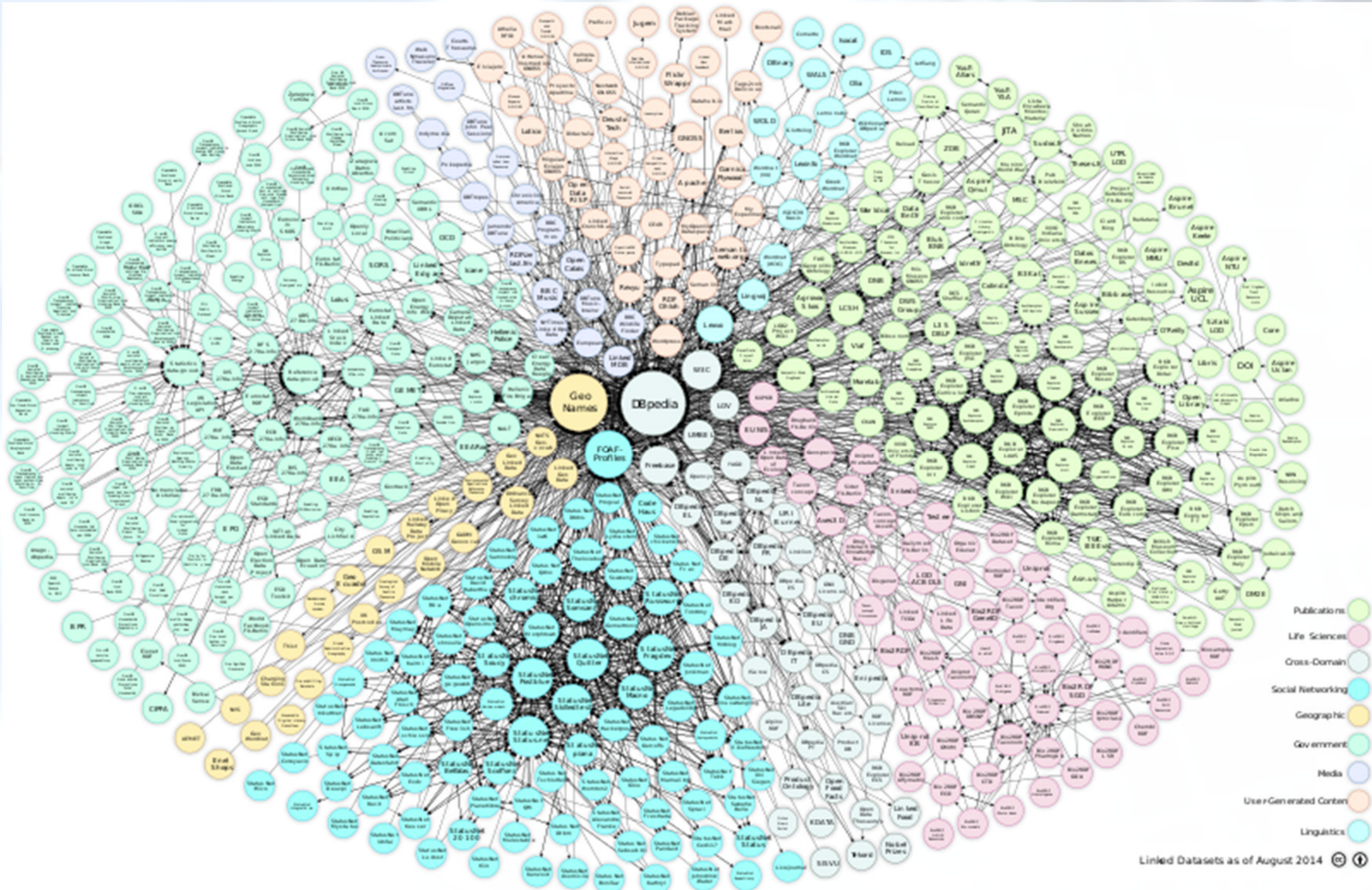
Credit: https://www.shutterstock.com/search/data+visualization?studio=1

Linked Big Data



credit: https://neuinfo.org/about/nifblog/521

Linking Open Data datasets



credit https://en.wikipedia.org/wiki/Linked_data#/media/File:LOD_Cloud_2014-08.svg

Panelists

Moderator

Petre Dini, IARIA, USA

Panelists

- Maaike de Boer, TNO, The Netherlands
  Explainable Artificial Intelligence
- Marco La Cascia, University of Palermo, Italia
  Deep Learning and the Semantic Gap
- Tsan-Sheng Hsu, Institute of Information Science, Academia Sinica, Taiwan

  Deep Learning and Knowledge underneath the Data **| P**laying Rules and "Intelligence" Learned in Complex Applications
- Hiroshi Ishikawa, Tokyo Metropolitan University, Japan
  From Data Science to Science-disciplined Data Analysis

Planning

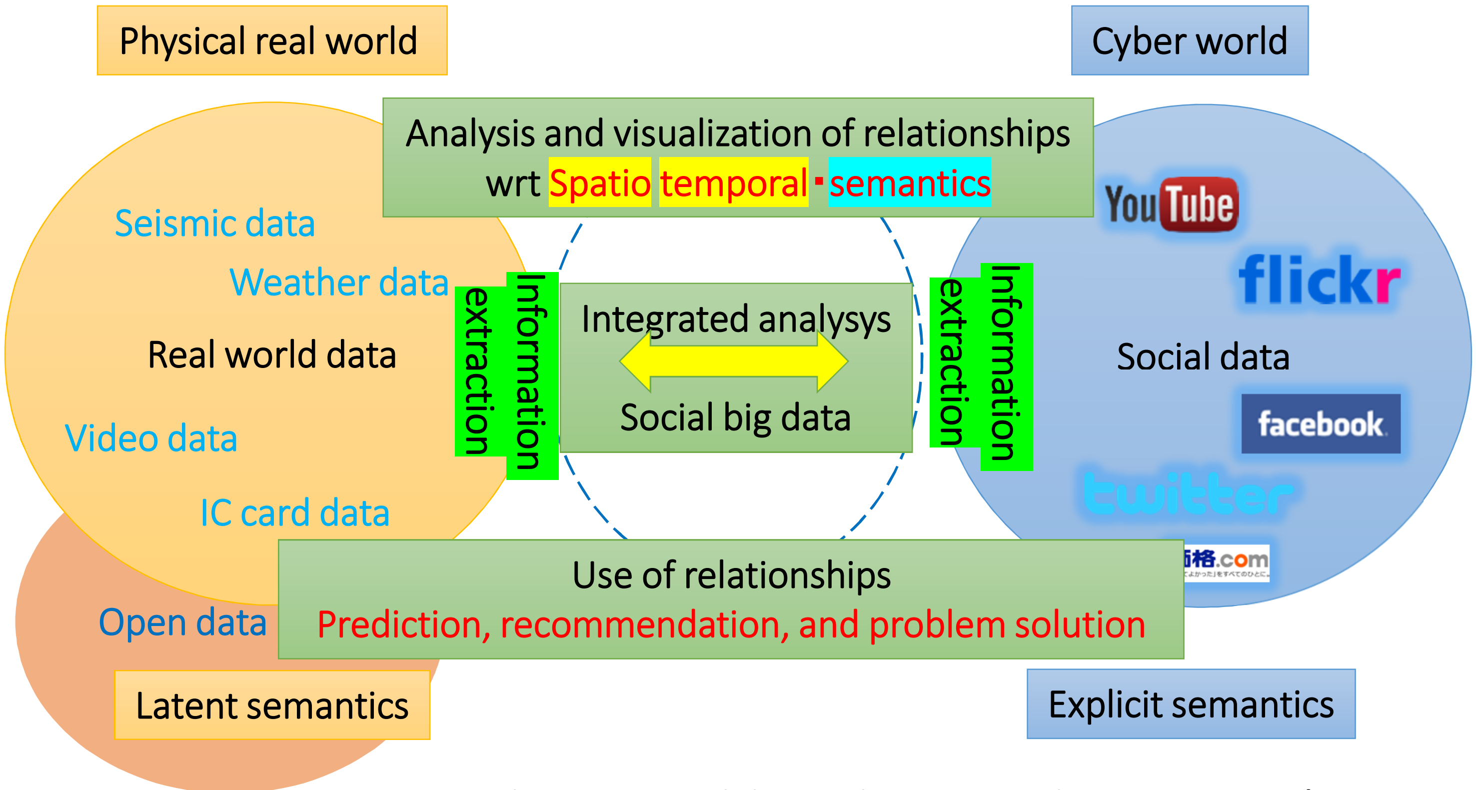
- Stage for the panelists
- Open discussion
- Concluding remarks
- Next panel topic | end of Feb, 2020, Lisbon

WWW.IARIA.ORG

From data science to science-disciplined data analysis

Hiroshi Ishikawa
Director, Research center for social big data
Tokyo Metropolitan University
Tokyo, Japan

Integrated analysis of real world data and open data, social data “Ishikawa’s concepts” (Olshannikova et al 2017)

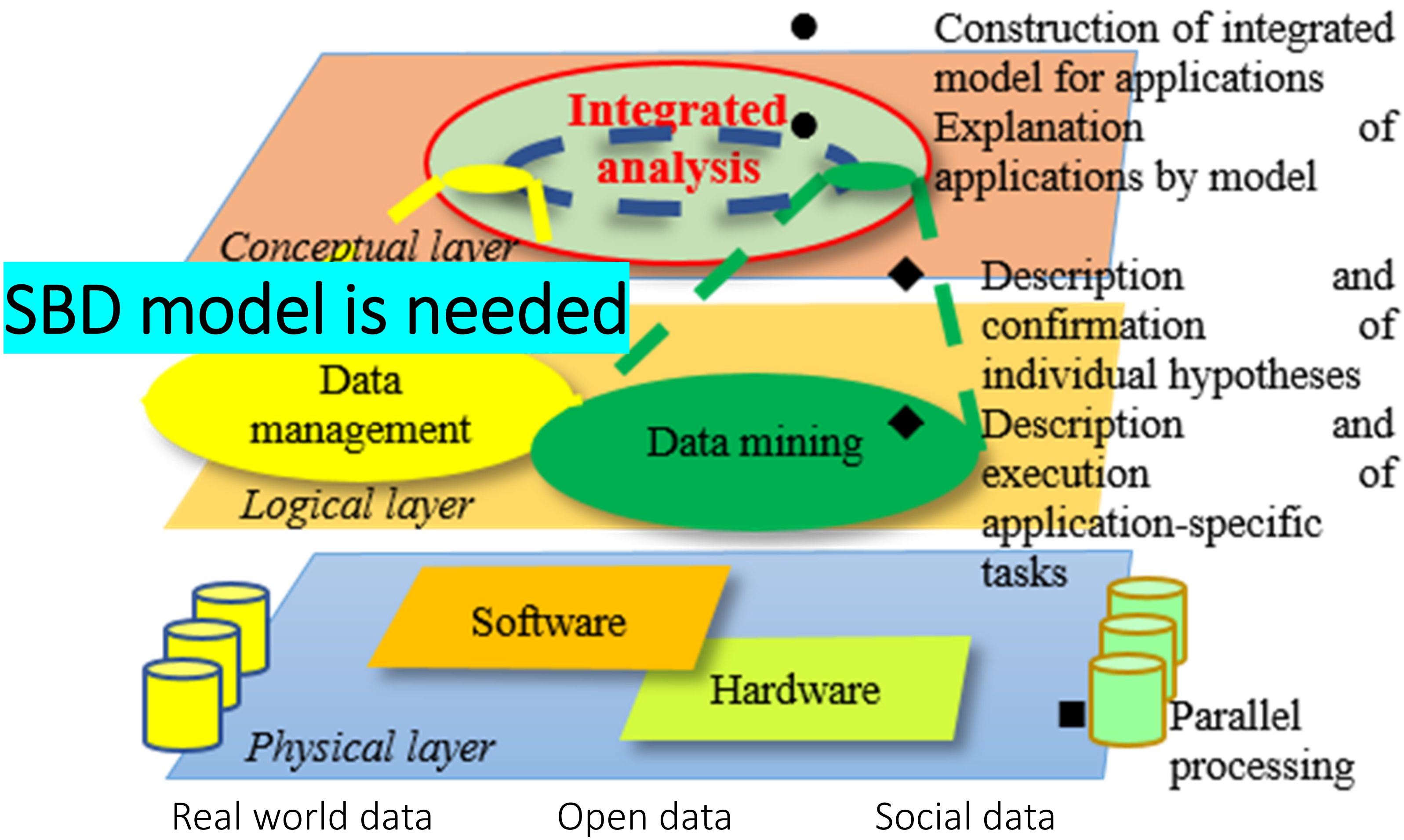


- Ekaterina Olshannikova et al, **Conceptualizing Big Social Data**, Journal of big data, Springer, Vol.4, No.3, pp.1-19, 2017.

Social big data (SBD) in summary

- In social big data applications, especially, cases where two or more data sources including at least one social data source are involved, are more interesting from a viewpoint of usefulness to businesses.
- If more than one data source can be analyzed by relating them to each other, and by paying attention to the interactions between them, it may be possible to understand what cannot be understood, by analysis of only either of them.
- By collecting those articles and images based on conditions specified with respect to locations and time intervals and counting them for each grid (i.e., unit location), probabilities that users post such data at the locations can be basically computed.
- By using such probabilities, for example, human activities can be analyzed such as probabilities of foreigners staying at specific spots or those moving from one spot to another.

Reference architecture for social big data



Requirements of our SBD model

- Description of SBD applications must be as independent from individual programming languages and frameworks (e.g. Spark and MLI) as possible.
  - it is not always possible for all researchers to access the same data and tools that the authors have used.
  - In other words, by enabling the mapping from description of applications by an abstract *SBD model* to individual tools available for the other researchers, reproducibility [Südhof 2016] can be realized even if the tools are not the same with the original one.
- Both data management and data mining must be described in an integrated manner.
  - In SBD applications, a lot of time is spent on development and execution of data management including preprocessing and postprocessing in addition to data mining.
  - Further, data management and data mining cannot be always separated in a crisp manner.
  - Rather, most SBD applications require hybrid processes mixed with data management and data mining.

Thomas C. Südhof, Truth in Science Publishing: A Personal Perspective *PLOS* August 26, 2016.

Whole processes of SBD applications in general require explanation

Prepare	Prepare data *How*
Analyze	Analyze data *How, Why*
Visualize	Visualize result *Why*

Necessity of an integrated framework for explanation

In order for social big data to widely be used, it is necessary to explain the user the application system.

Both microscopic description, that is, interpretation of the analytical model and explanation of individual decisions and macroscopic description, that is, description of the whole processes including the data manipulation and the model construction are required.

From the development experiences of multiple use cases, we have come to think that both the macro explanation based on the proposed data model in this paper and the micro explanation emerging in AI are urgently needed.

A macro explanation is necessary for the following reasons.

- In order for social big data applications to be accepted by users, it is necessary to ensure at least their reliability. Since information science is one area of science, we should guarantee reproducibility as science. In other words, it is necessary to ensure that third parties can prepare and analyze data according to given explanation and can get the same results.
- In addition, in order for the service to be operatable, it is necessary for the final user of the service to be convinced of how the service processes and uses the personal information.
- If the users can be convinced of the description of way of using the personal information, the progress of data portability can be advanced based on the EU's GDPR law on personal information protection and Japan-based information bank to promote the use of personal information.

A micro explanation is necessary for the following reasons.

- In order for analysts of social big data and field experts using the data to accept decisions made by the constructed model, it is assumed that they must understand the structure, actions and grounds of the model and are satisfied with them as well.
- Up to now, the authors have been involved in the development of a wide range of social big data use cases ranging from tourism, disaster prevention to lunar and planetary science.
- In the course of these processes, from the users of the use cases, we have often received questions as to what kind of data are processed, what kind of model are created as the core of analysis, and furthermore, what are the grounds for the decisions.

EXPLAINABLE DATA SCIENCE

Maaike de Boer, TNO

TNO innovation for life

WHY EXPLAINABLE AI?

- With the recent advances in deep learning, it becomes more and more important to create transparency in order to gain trust in the AI systems.

- This can be done by
  - Opening up black box models
  - Using model-agnostic algorithms
  - And many more methods

DISCUSSION POINTS

- We should have explainability at all cost, even if it implies lower performance
- The human in the loop is the key to explainable AI
- Using model agnostic algorithms is better than creating new deep learning algorithms that are more transparent


THANK YOU FOR YOUR ATTENTION
Take a look:
TNO.NL/TNO-INSIGHTS
TNO innovation for life

Deep learning and the semantic gap

Marco La Cascia

Università degli Studi di Palermo

Dipartimento di Ingegneria

Data, information and knowledge

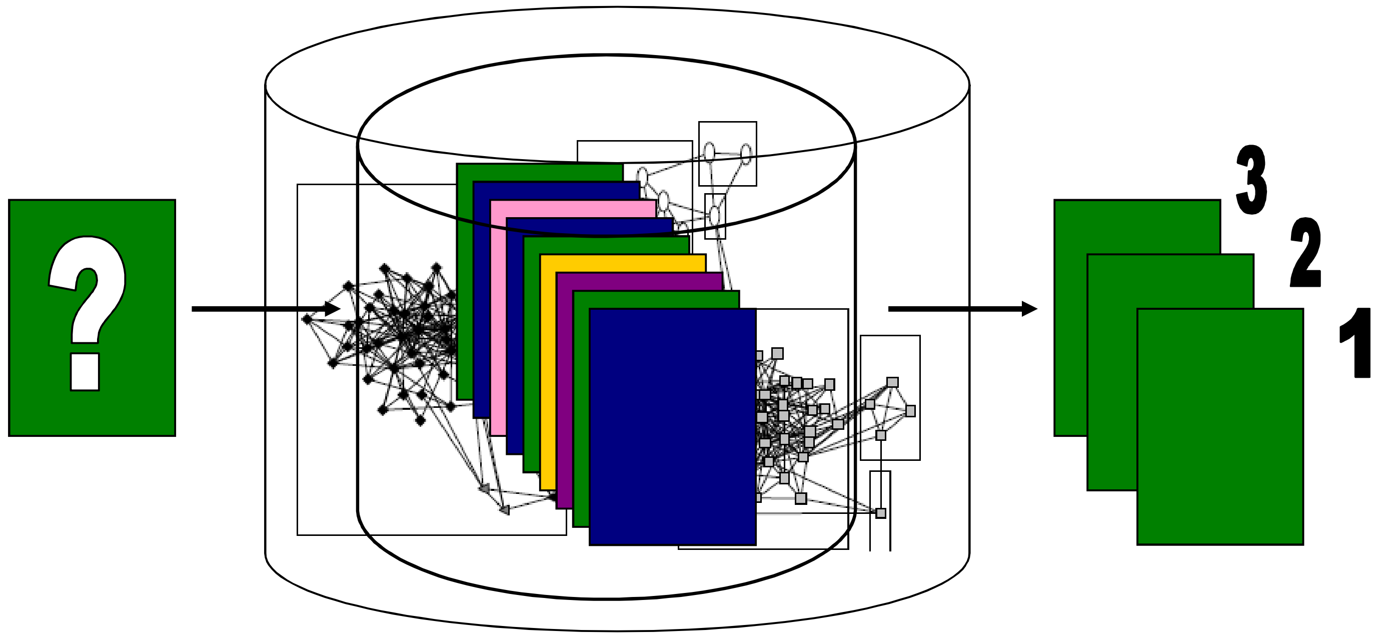
- “We are drowning in information but starved for knowledge”
  [Naisbitt 1982] John Naisbitt. Megatrends. Warner Books, Inc. (1982)

- Today, we are drowning in data and starved for information.

- Lot of data… but how much of that data will actually be useful?

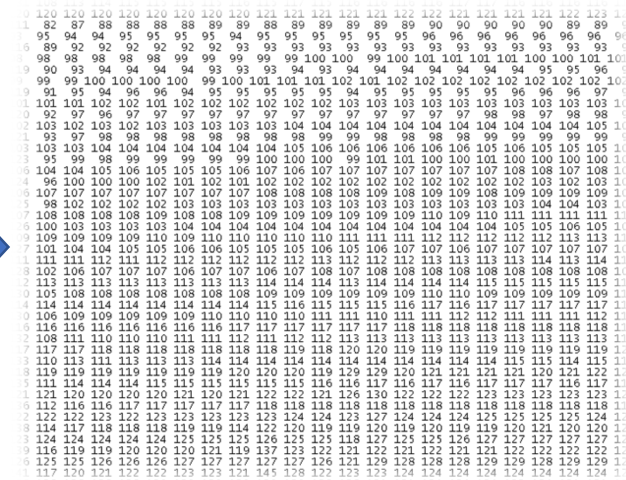
What about visual data? Much worse…

Visual search, organization



Query **Image or video archives** **Relevant content**

What the computer gets



153	156	148	152	149	147	139	146	142	150	146	144	137	125	120	119	136	146	151	164	172	175	183	188	196	200	205	208	214	214	219	217
159	151	150	148	140	138	139	129	119	104	86	82	89	97	107	115	118	130	128	132	128	144	160	168	179	188	200	208	213	220	212	214
149	146	153	147	147	146	132	99	73	78	87	96	105	120	138	151	145	157	163	171	165	161	146	126	157	184	190	201	215	212	214	214
145	150	154	148	148	126	93	67	72	78	96	107	117	127	131	134	127	154	166	167	183	194	200	195	143	140	175	190	197	203	206	207
151	153	151	147	120	85	67	75	84	83	94	92	81	78	78	91	83	117	126	144	178	200	201	203	208	175	127	159	185	196	195	206
146	144	139	123	79	66	74	83	79	69	64	62	58	50	46	54	54	66	60	80	86	108	141	191	184	200	187	123	144	175	198	199
135	130	115	87	64	77	90	79	78	85	81	63	55	57	56	53	70	62	61	68	59	58	84	105	168	194	196	183	131	151	185	197
128	116	92	71	82	94	103	101	83	101	88	66	70	90	80	42	39	53	88	73	76	82	116	87	97	144	188	195	190	166	171	203
135	120	84	83	108	127	135	115	100	92	79	49	85	74	59	0	0	0	50	69	52	79	157	141	100	84	136	187	206	204	189	200
144	103	91	115	139	147	127	91	87	80	72	44	61	84	25	0	0	0	50	181	45	69	142	164	167	113	93	130	193	199	208	203
139	102	123	143	137	131	109	85	93	84	68	47	77	86	31	0	3	0	51	156	53	75	141	169	199	151	171	108	143	181	199	208
141	135	153	142	114	104	97	97	83	98	77	42	77	96	79	21	0	23	58	46	56	77	155	199	212	161	194	193	164	187	202	205
160	172	164	141	128	112	98	95	100	96	91	73	68	86	75	73	64	65	54	69	77	115	190	212	193	181	174	188	210	194	202	207
179	189	160	140	139	116	97	97	108	103	110	99	75	80	72	83	50	55	54	95	98	174	205	185	179	188	185	190	193	217	217	224
189	183	152	130	121	105	105	117	114	108	107	115	110	81	85	85	87	81	81	124	183	202	175	180	178	171	173	204	225	215	219	225
178	161	149	135	120	115	122	129	137	145	131	121	125	115	109	91	92	111	132	159	173	170	184	176	184	190	191	217	210	226	228	223
187	159	139	127	125	115	118	121	121	131	133	134	140	137	134	139	140	152	141	154	170	163	195	194	176	198	216	209	219	224	223	226
185	164	140	122	116	110	109	108	113	118	115	116	123	127	135	148	154	162	165	170	171	160	183	198	201	210	223	216	221	222	221	226
188	175	150	130	118	117	113	110	108	115	117	123	130	132	138	150	157	158	174	182	189	186	198	221	224	221	227	221	223	218	218	222
187	179	158	141	124	127	125	127	126	129	130	135	139	141	150	165	175	172	185	195	207	210	212	226	229	222	224	224	223	218	219	221
188	184	172	159	138	135	135	143	143	143	144	146	145	147	160	174	184	191	199	207	211	213	217	224	227	223	223	221	221	218	224	223
192	191	187	174	153	139	140	147	146	149	157	162	160	159	165	174	181	198	201	210	212	216	223	224	225	225	220	215	217	215	224	224

Challenges: Complexity

- Thousands to millions of pixels in an image
- 3,000-30,000 human recognizable object categories
- 30+ degrees of freedom in the pose of articulated objects (humans)
- Billions of images indexed by Google Image Search

About half of the cerebral cortex in primates is devoted to processing visual information [Felleman and van Essen 1991]

Early years of CBIR

- QBIC (IBM) early 90
  - Image processing for retrieval by color, texture, and local geometry
- PhotoBook (MIT) mid 90
  - Semantic preserving image compression (eigen-features)
- ImageRover (BU) late 90
  - Mixing text and low-level image information
- Many others
  - Accumulative and global features, salient points, object and shape features, signs, and structural combinations
  - Similarity of pictures and objects in pictures

[Smeulders et al. 2000] Smeulders, A.W.M. Worring, M. Santini, S. Gupta, A. Jain, R.: Content-Based Image Retrieval at the End of the Early Years. IEEE Transactions on Pattern Analysis and Machine Intelligence. vol. 22, No. 12. pp. 1349-80. (2000)

The semantic gap

- The lack of coincidence between the information that one can extract from the visual data acquired from an image and the interpretation that the same data have for a user in a given situation.
  [Smeulder et al. 2000]

Difference between low-level representation of an image and high-level human perception

And then?

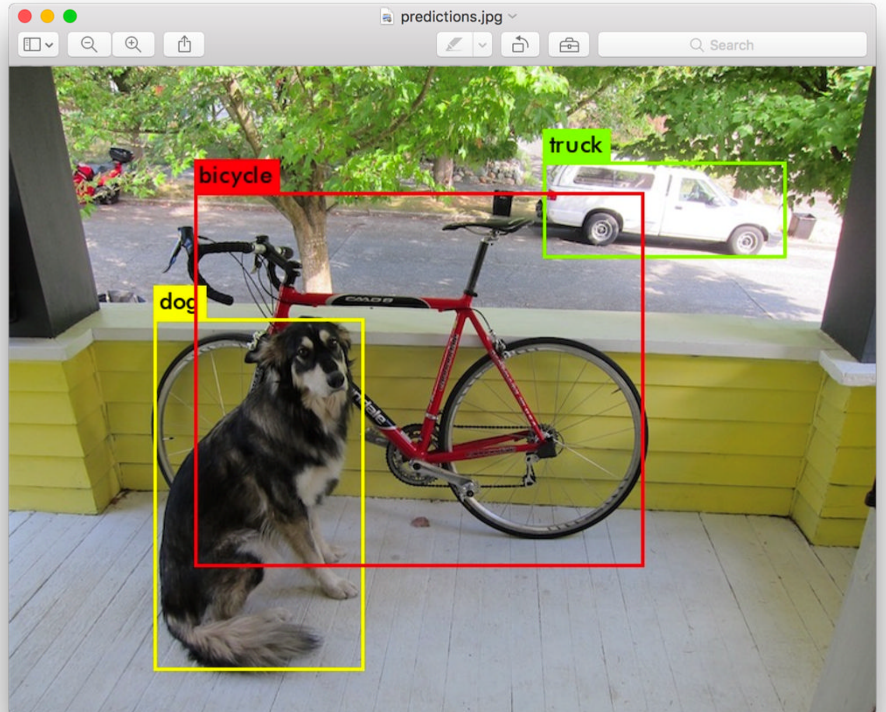
- Hundreds of papers
- More powerful visual features
- Better similarity distance
- Some significant applications in very specific domains
- …

Semantic gap? Still there

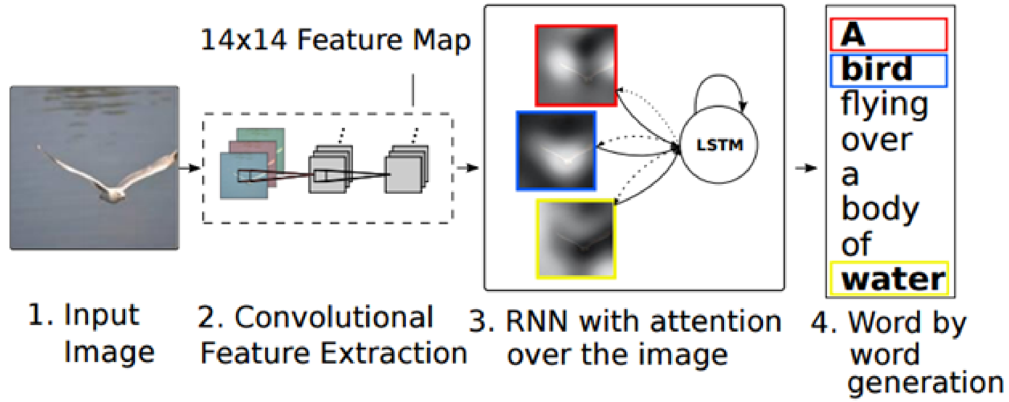
Deep learning

- CNN for image classification
- Object detection by Region based Convolutional Networks (R-CNN)
- CNN for feature extraction and representation.
  - No more human design of visual feature.
- Generative Adversarial Network (GAN) to create a variety of realistic images corresponding to the description.
- Multimodal embedding: Deep Boltzmann Machines, Restricted Boltzmann Machines, CNN to process visual data + MLP or LSTMs to embed text features
- Siamese network (two weight-sharing networks running on two images) to model similarity function

Very impressive results



Redmon,J. Divvala,S. Girshick,R. Farhadi,A.
You Only Look Once: Unified, Real-Time Object Detection.
IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2016



Xu,K. Ba,J. Kiros,R. Cho,K. Courville,A. Salakhutdinov,R. Zemel,R. Bengio,Y.
Show, Attend and Tell: Neural Image Caption Generation with Visual Attention.
International Conference on Machine Learning, 2015

Question

CBIR at the time of deep learning:
has deep learning closed the semantic gap?

ALLDATA 2019
Panel on Software and Data

Theme: Designing Complex Systems
Topic: Machine Learning and Visual Data

Tsan-sheng Hsu

徐讚昇

Institute of Information Science
Academia Sinica, Taiwan
tshsu@iis.sinica.edu.tw

http://www.iis.sinica.edu.tw/~tshsu

About me

- **Academia Sinica is THE governmental pure research laboratory in Taiwan.**
  - **Have 31 institutes working on nature science, life science and social science.**
  - **Institute of Information Science is the computer science branch.**
- **Laboratory of Massive Data Computing and Management**
  - **Since the year 2000**
  - **Efficient algorithms for processing large data**
  - **Applications**
    - ▷ *Data privacy*
    - ▷ *Classical board games*
    - ▷ *Simulations of disease spreading through contacts*
    - ▷ *Construction and understanding of human disease network*
- **Research and teaching**
  - **Areas**
    - ▷ *Algorithm design and implementation*
    - ▷ *Graph theory*
    - ▷ *Massive data computing*
  - **Teach a course once per year at National Taiwan University**

Points of interest (I)

- **Deep learning has enjoyed lots of success right now on data with patterns.**
  - **For example:**
    - ▷ *Medical images*
    - ▷ *Board games*
    - ▷ *Natural language processing*
- **Discusions:**
  - **Does deep learning actually learn the "knowledge" underneath the data, or just tags/labels assigned?**
  - **Are manually/auto-converted assigned labels inevitably biased, unfair and sometimes unethical?**
  - **Can an explanation of the "intelligence" uncovered be easily derived?**

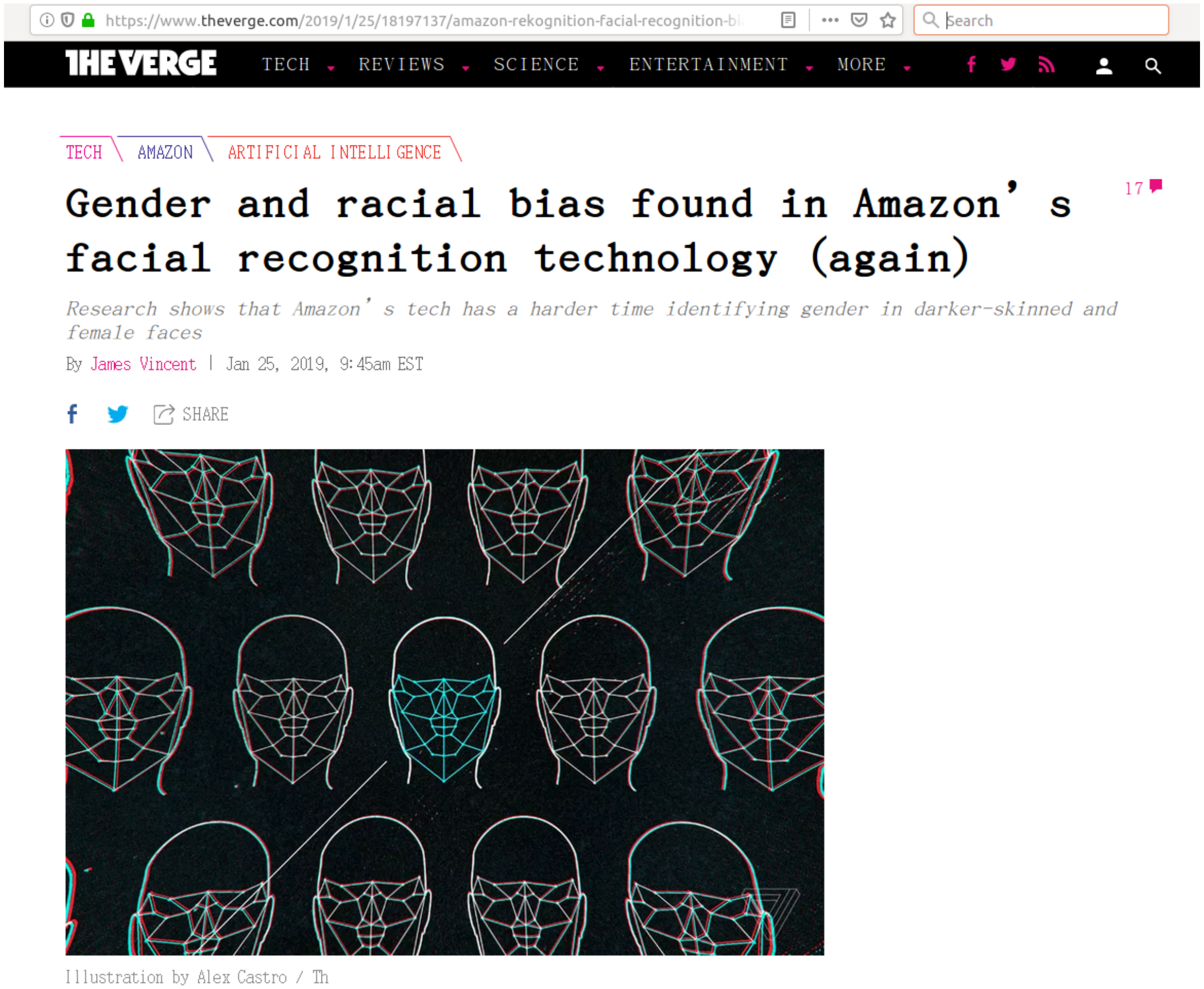
Points of interest (II)

- **Maybe to remedy the labeling problem, board game playing programs like AlphaZero use "simple" and "transparent" rules to do unsupervised learning, instead of supervised learning as in AlphaGo.**
  - **Go is a complex, not complicated game.**
  - **Go boards are "visual" data.**
  - **Supervised learning takes previous game logs with labels.**
  - **Basic rules for playing Go is simple.**
  - **Learn the basic rules, not the labels.**
  - **A great line of success for board game playing**
    AlphaGo (2016) → AlphaGo Zero (2017) → AlphaZero (2018)
- **Discussions:**
  - **Are the rules of Go really simple if a board position is complex?**
    - ▷ *Life and death*
    - ▷ *Avoid immediate loops (Ko), but allow longer loops*
  - **Are simple rules available for very complex system?**

Backup slides

-

Face recognition software bias



https://www.theverge.com/2019/1/25/18197137/amazon-rekognition-facial-recognition-bi

THE VERGE TECH REVIEWS SCIENCE ENTERTAINMENT MORE

TECH \ AMAZON \ ARTIFICIAL INTELLIGENCE

17

Gender and racial bias found in Amazon's facial recognition technology (again)

Research shows that Amazon's tech has a harder time identifying gender in darker-skinned and female faces

By James Vincent | Jan 25, 2019, 9:45am EST

SHARE

Illustration by Alex Castro / Th

James Vincent, Jan 25, 2019, the verge, https://www.theverge.com/2019/1/25/18197137/amazon-rekognition-facial-recognition-bias-race-gender

AI chat Bot bias

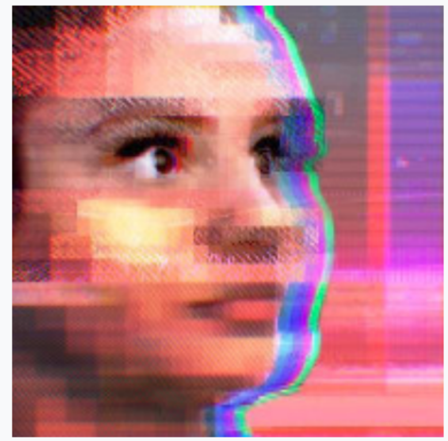
https://en.wikipedia.org/wiki/Tay_(bot)

From Wikipedia, the free encyclopedia

This article has multiple issues. Please help **improve it** or discuss these issues on the **talk page**. *(Learn how and when to remove these template messages)* [show]

Tay was an artificial intelligence chatter bot that was originally released by Microsoft Corporation via Twitter on March 23, 2016; it caused subsequent controversy when the bot began to post inflammatory and offensive tweets through its Twitter account, forcing Microsoft to shut down the service only 16 hours after its launch.[1] According to Microsoft, this was caused by trolls who "attacked" the service as the bot made replies based on its interactions with people on Twitter.[2] It was soon replaced with Zo.

Tay



The Twitter profile picture of Tay

Developer(s)	Microsoft Research, Bing
Available in	English
Type	Artificial intelligence chatterbot
License	Proprietary
Website	tay.ai

Contents [hide]

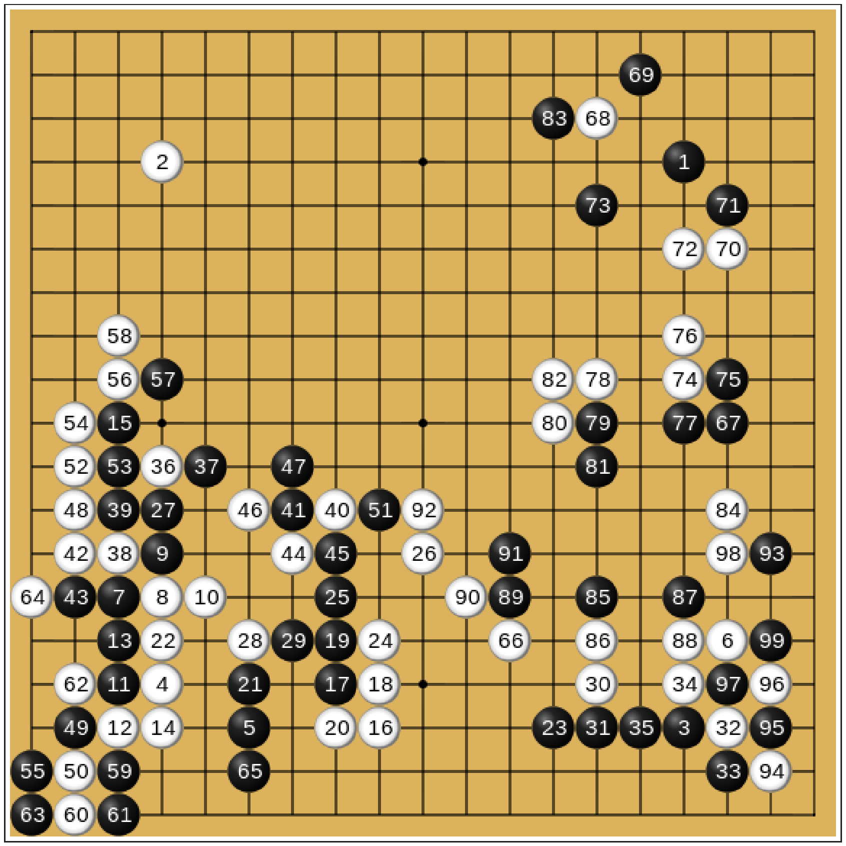
1 Background
  1.1 Creation
  1.2 Initial release
  1.3 Suspension
  1.4 Second release and shutdown
2 Legacy
3 See also
4 References
5 External links

Background [edit]

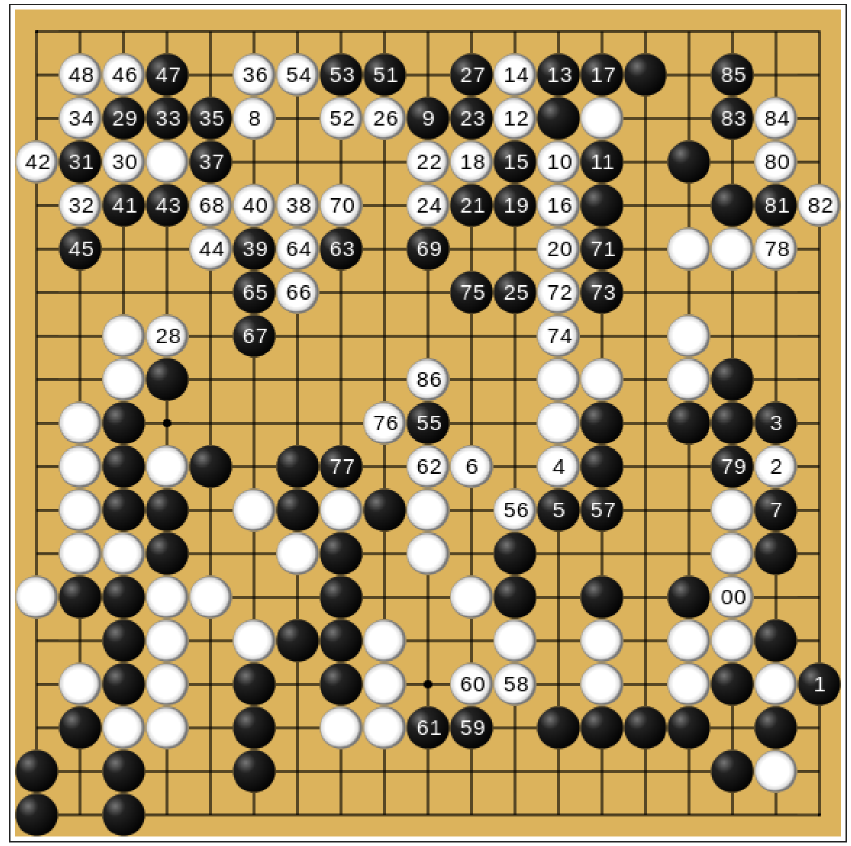
Wiki, https://en.wikipedia.org/wiki/Tay_(bot)

AlphaGo

AlphaGo Master (white) v. Tang Weixing (31 December 2016), AlphaGo won by resignation. White 36 was widely praised.



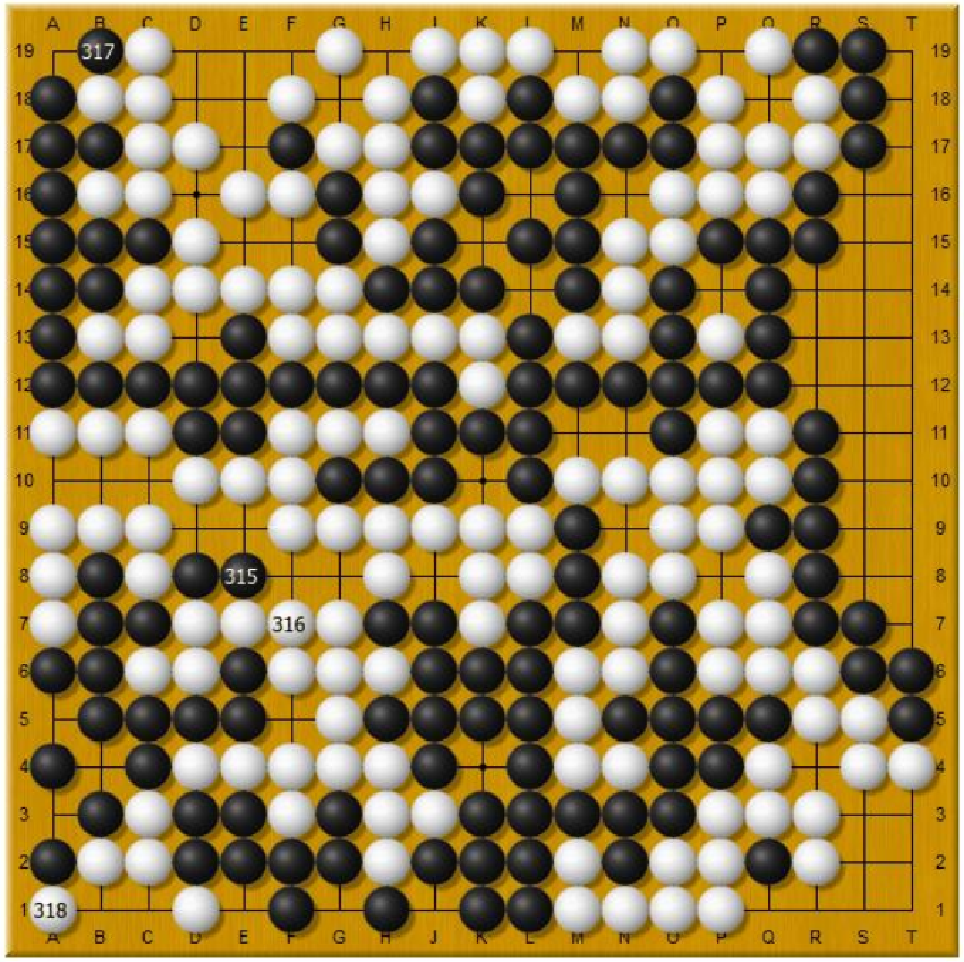
First 99 moves



Moves 100–186 (149 at 131, 150 at 130)

Wiki, https://en.wikipedia.org/wiki/AlphaGo

Complicated Go, 1/2 point win/Draw



長生劫, game played by 林海峰(black) vs 小松英樹 1993/9/2, Honinbo competition, Japan; https://kknews.cc/zh-tw/sports/6ne8oxl.html

Self-driving dilemmas

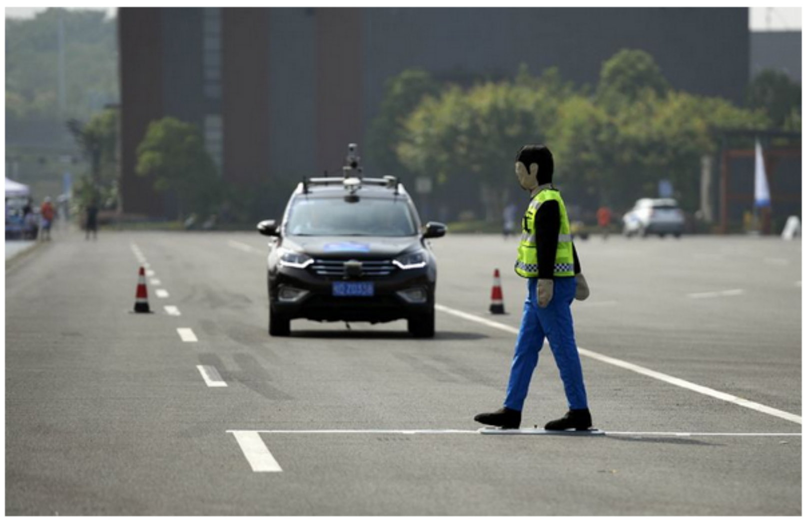
https://www.nature.com/articles/d41586-018-07135-0 80% Search

MENU nature

Self-driving car dilemmas reveal that moral choices are not universal

Survey maps global variations in ethics for programming autonomous vehicles.

Amy Maxmen



Self-driving cars are being developed by several major technology companies and carmakers. credit: VCG/Getty

PDF version

RELATED ARTICLES

Autonomous vehicles: No drivers required

Machine ethics: The robot's dilemma

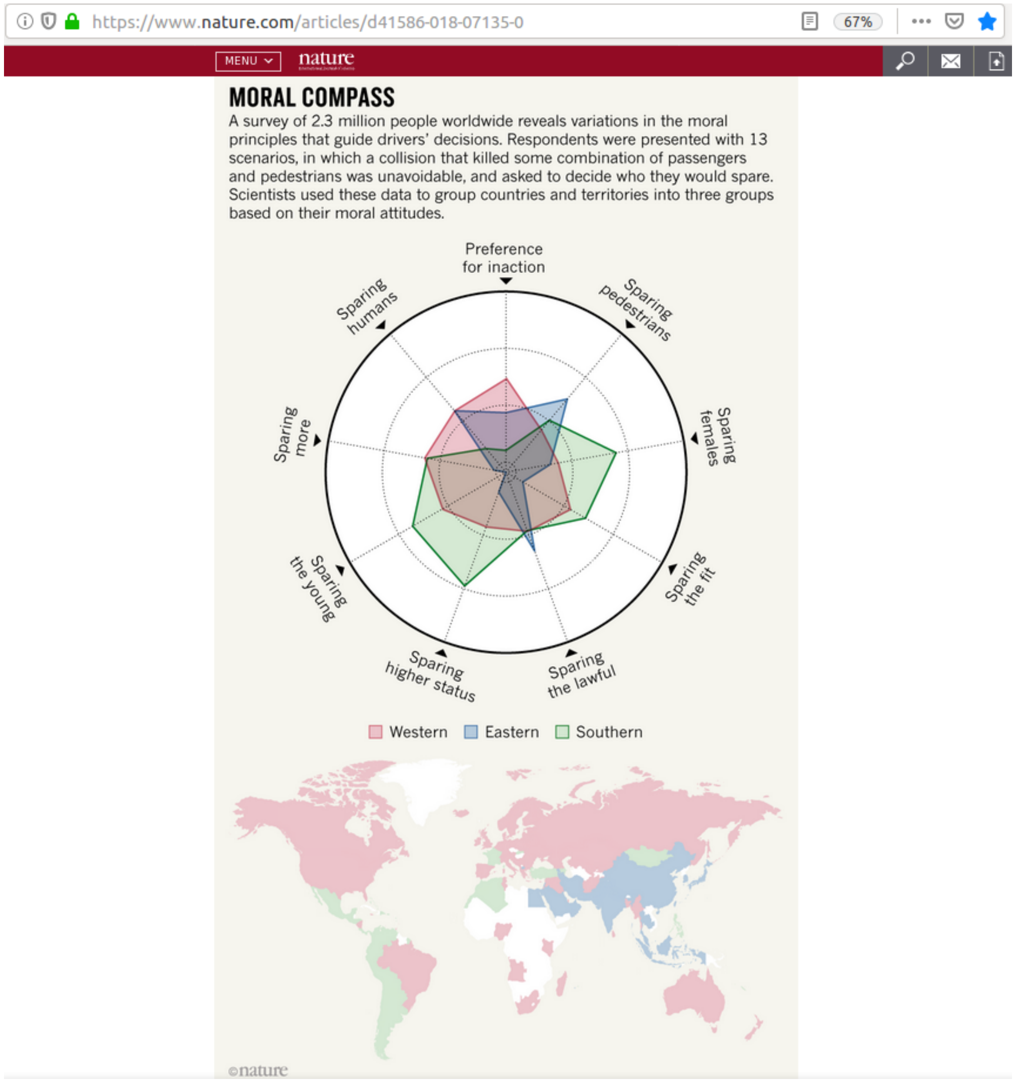
Editorial: Road test

SUBJECTS

Computer science Ethics

Amy Alexmen, 2018/10/24, Nature 562, 469-470 (2018) doi: 10.1038/d41586-018-07135-0

Moral choices are not universal



Amy Alexmen, 2018/10/24, Nature 562, 469-470 (2018) doi: 10.1038/d41586-018-07135-0