BIG DATA IN SANITA', falsi miti e opportunità dalla trusted artificial intelligence

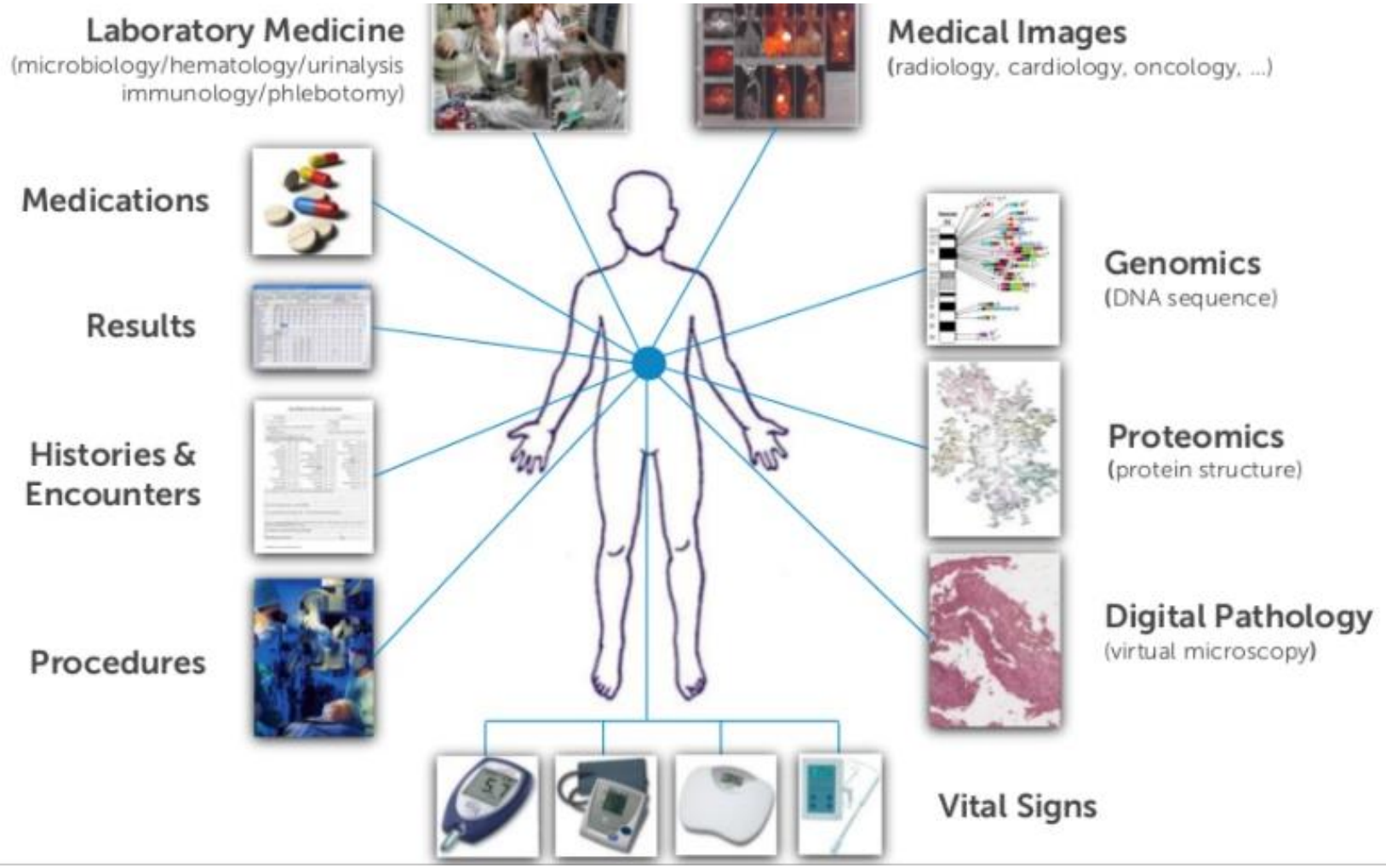
Dott. Daniela Scaramuccia, Dott. Robert Alexander

Can we talk about BIG DATA in healthcare?

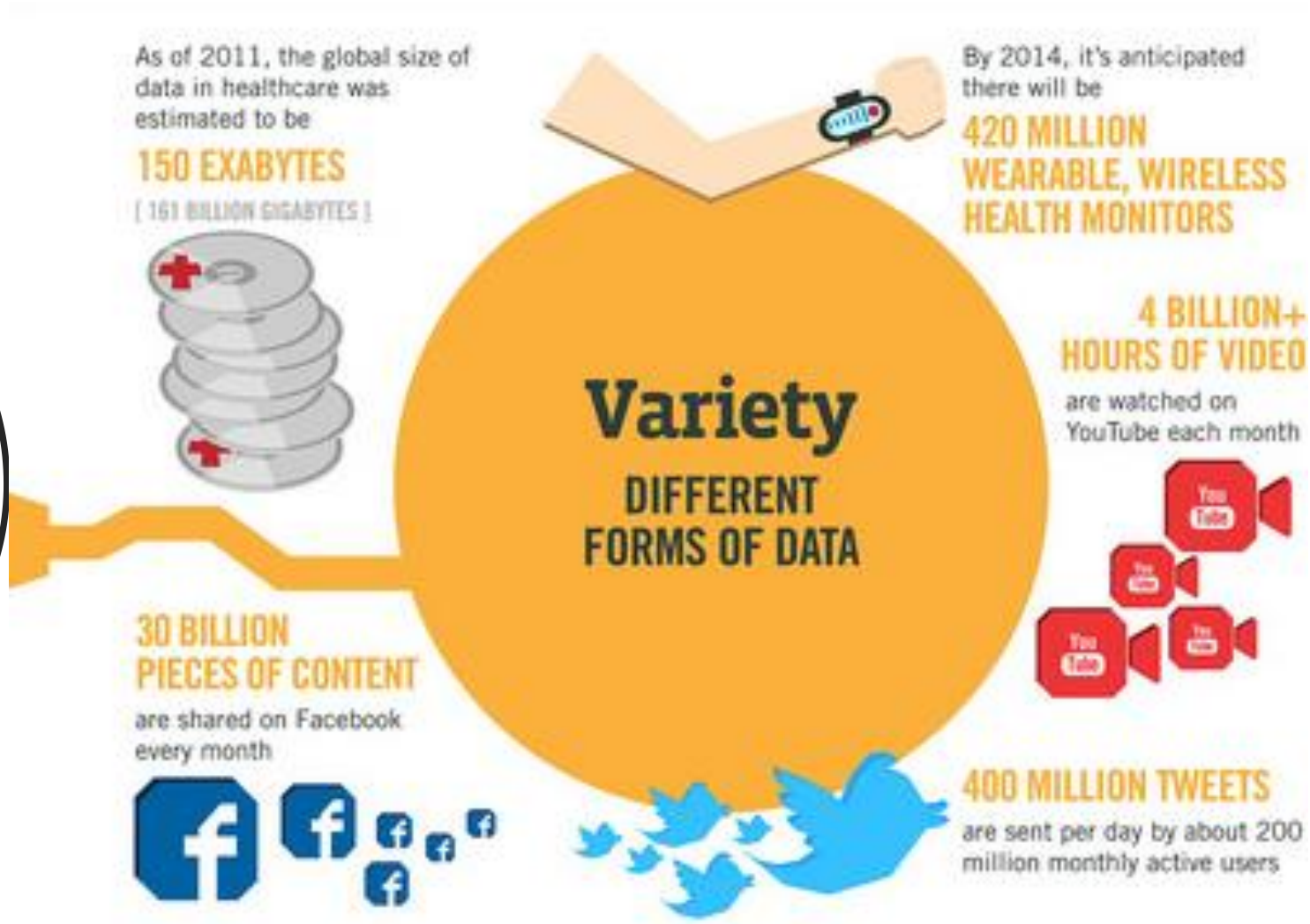
*"**BIG DATA** is **high-volume, high-velocity and/or high-variety** information assets that demand cost-effective, innovative forms of information processing that **enable enhanced insight, decision making, and process automation**"*

(Gartner)

High Volume? YES!



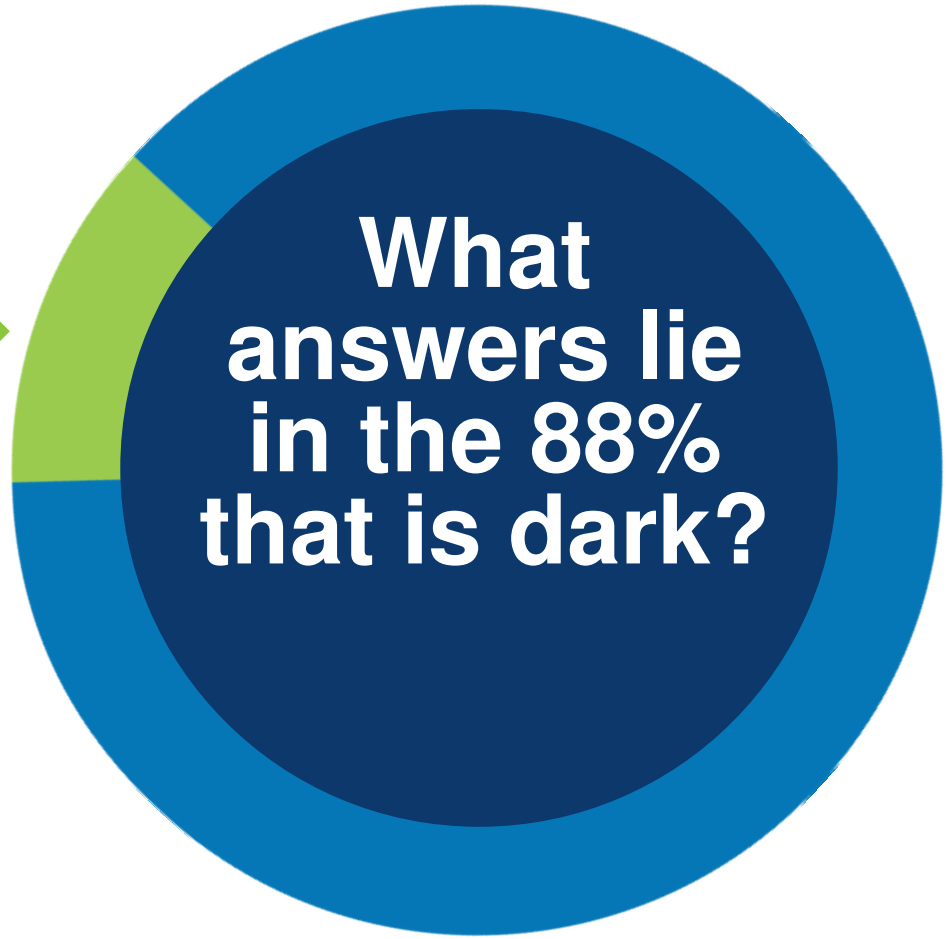
High Variety/ Velocity? YES!



"… enable enhanced insight, decision making, and process automation?
Not really…"

Cognitive can unlock DARK DATA

Think of all that has been accomplished using only **a fraction of the available data**



Unlock the possibilities.

By the year 2020, about **1.7 MB** of new information will be **created every second, for every human being** on the planet.

Desktop Medicine

RESEARCH ARTICLE

HEALTH AFFAIRS > VOL. 36, NO. 4 : MATERNITY COVERAGE, CHILDREN, DISABILITY & MORE

THE PRACTICE OF MEDICINE

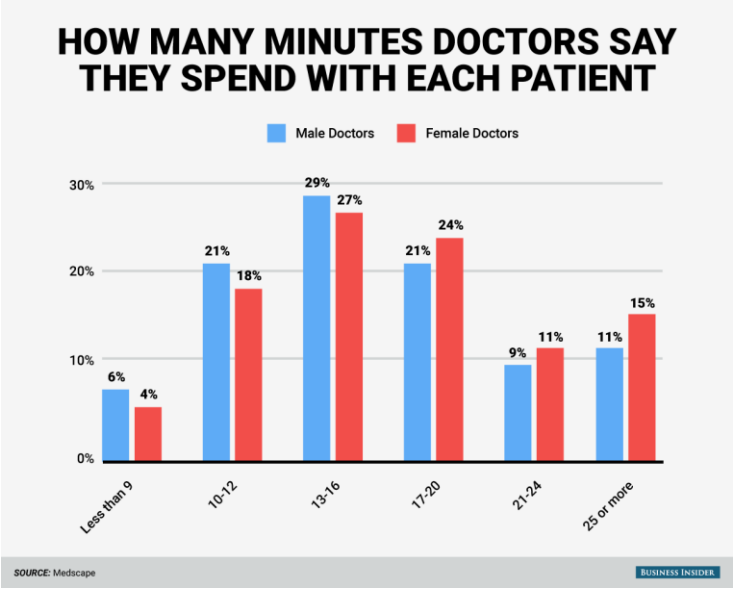
Electronic Health Record Logs Indicate That Physicians Split Time Evenly Between Seeing Patients And Desktop Medicine

Ming Tai-Seale[1], Cliff W. Olson[2], Jinnan Li[3], Albert S. Chan[4], Criss Morikawa[5], ... See all authors

AFFILIATIONS

PUBLISHED: APRIL 2017 No Access

https://doi.org/10.1377/hlthaff.2016.0811

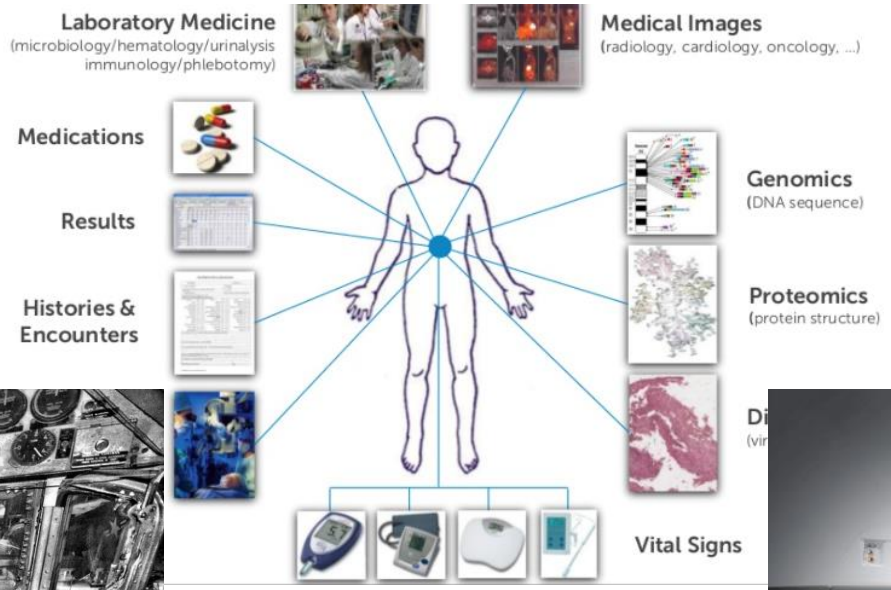


Maggiori distrattori dal rapporto terapeutico

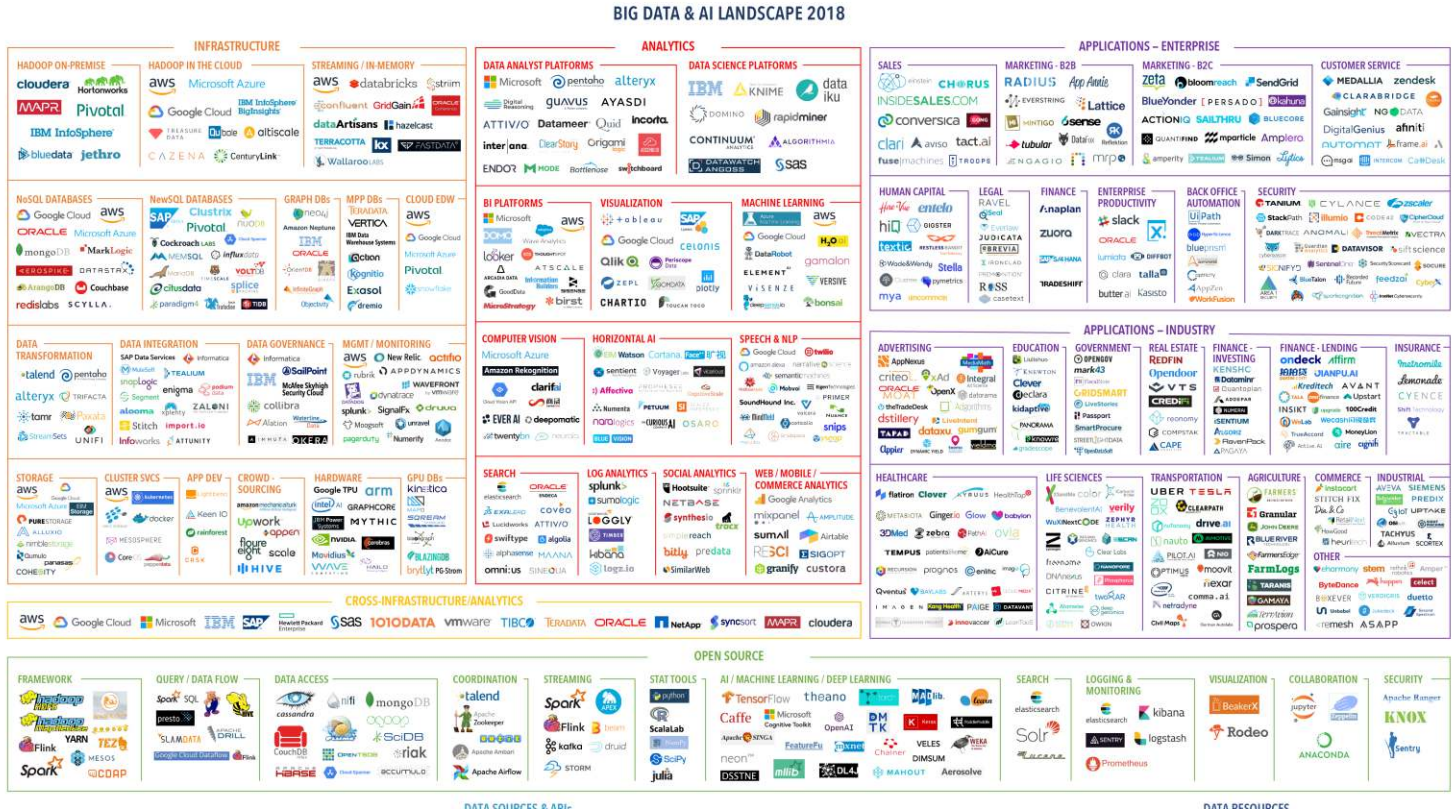
- Interazione con sistemi elettronici strutturati e molto orientati alla tracciatura clinico-amministrativa
- Ricerca di evidenze non strutturate nella documentazione del paziente
- Ricerca di evidenze non strutturate nelle linee guida e letteratura scientifica

Da EBM a medicina personalizzata e di precisione

Integrare dati eterogenei e complessi in modelli predittivi che modulino l'EBM in modo appropriato



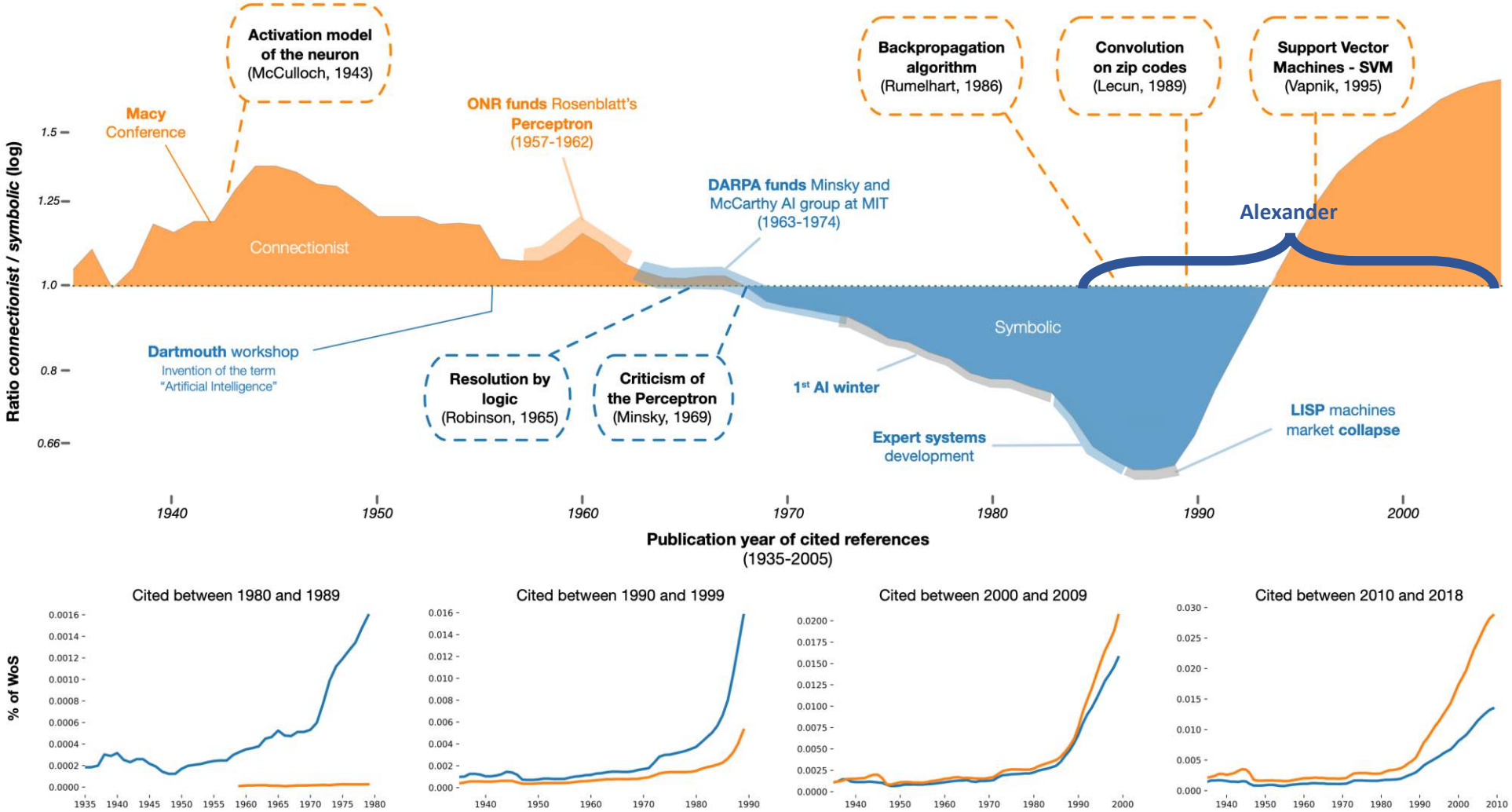
Il mercato dell'AI è in grande fermento



Punti di attenzione: narrowness, brittleness, accettabilità

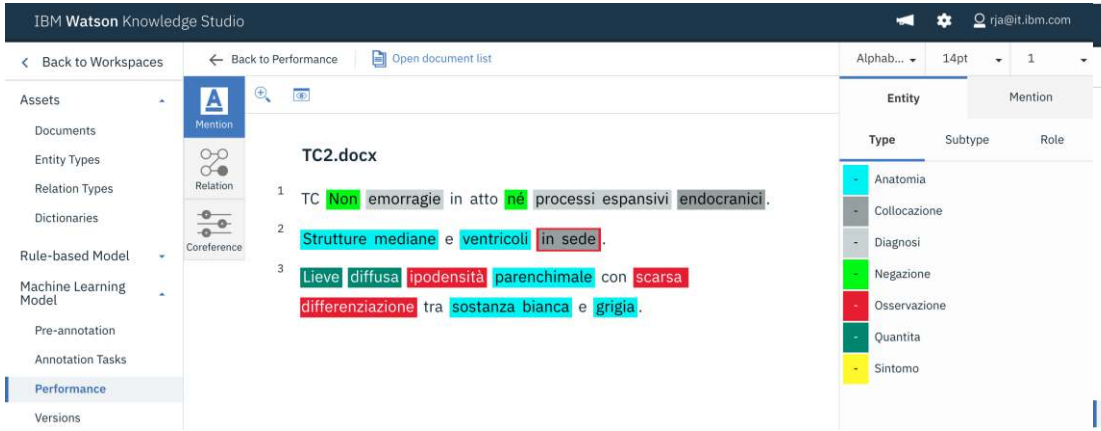
Sviluppo futuro: «common sense» (funding DARPA 2019)

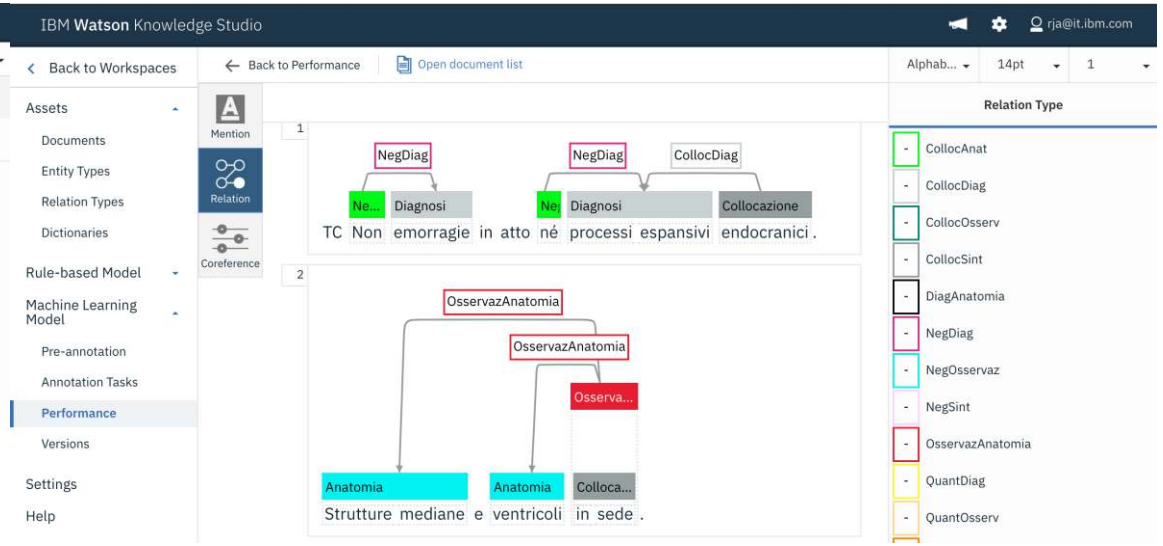
Chi non ricorda il passato è condannato a ripeterlo



NLP: estrarre concetti e dati da testo libero

- Testi «medici»
  - personali: anamnesi, cartella medica ed infermieristica, referti, consulenze etc
  - Sapere: linee guida, PDTA, articoli scientifici etc
- Altro
  - Esposizioni ambientali, comportamento, lavoro, situazione socio economica etc
- Patient Generated Data: dati non strutturati e «small big data» da app e IoT





Fun Headlines

- Hospitals are Sued by 7 Foot Doctors
- Astronaut Takes Blame for Gas in Spacecraft
- New Study of Obesity Looks for Larger Test Group
- Chef Throws His Heart into Helping Feed Needy
- Include your Children when Baking Cookies
- Iraqi Head Seeks Arms
- Juvenile Court to Try Shooting Defendant
- Kids Make Nutritious Snacks
- British Left Waffles on Falkland Islands
- Red Tape Holds Up New Bridges
- Clinton Wins on Budget, but More Lies Ahead
- Ban on Nude Dancing on Governor's Desk

Resources: lexicons, thesauri, ontologies

ICD9 – 1972
ICD10 – ICD11 already in the works
UMLS – multilingual but shallow coverage
MeSH – good for biblio not for clinical info
SNOMED-CT – no Italian
FMA – no Italian (latin, catalan, polish etc.)

Italian resources are needed

NLP: comprensione profonda richiede approcci ibridi

Trusted AI per una consapevole accettazione

What does it take to trust a decision made by a machine?
(Other than that it is “xx% accurate”)

Is it fair?

Is it easy to understand?

Did anyone tamper with it?

Is it accountable?

Trusted AI per IBM

Quattro dimensioni che coprono il ciclo di vita di una applicazione AI based

(AI360)
FAIRNESS

EXPLAINABILITY

(ART)
ROBUSTNESS

ASSURANCE

supported by an instrumented platform
AIOpenScale

AI Fairness 360

An open source extensible toolkit for detecting, understanding, and mitigating unwanted algorithmic bias

IBM Research Trusted AI

Home | Demo | Resources | Community

AI Fairness 360 Open Source Toolkit

This extensible open source toolkit can help you examine, report, and mitigate discrimination and bias in machine learning models throughout the AI application lifecycle. Containing over 30 fairness metrics and 9 state-of-the-art bias mitigation algorithms developed by the research community, it is designed to translate algorithmic research from the lab into the actual practice of domains as wide-ranging as finance, human capital management, healthcare, and education. We invite you to use it and improve it.

API Docs ↗ | Get Code ↗

Not sure what to do first? Start here!

Read More
Learn more about fairness and bias mitigation concepts, terminology, and tools before you begin.

Try a Web Demo
Step through the process of checking and remediating bias in an interactive web demo that shows a sample of capabilities available in this toolkit.

Watch a Video
Watch a video to learn more about AI Fairness 360.

Read a paper
Read a paper describing how we designed AI Fairness 360.

Use Tutorials
Step through a set of in-depth examples that introduces developers to code that checks and mitigates bias in different industry and application domains.

Ask a Question
Join our AIF360 Slack Channel to ask questions, make comments and tell stories about how you use the toolkit.

View Notebooks
Open a directory of Jupyter Notebooks in GitHub that provide working examples of bias detection and mitigation in sample datasets. Then share your own notebooks!

Contribute
You can add new metrics and algorithms in GitHub. Share Jupyter notebooks show-casing how you have examined and mitigated bias in your machine learning application.

Learn how to put this toolkit to work for your application or industry problem. Try these tutorials.

Credit Scoring
See how to detect and mitigate age bias in predictions of credit-worthiness using the German Credit dataset.

Medical Expenditure
See how to detect and mitigate racial bias in a care management scenario using Medical Expenditure Panel Survey data.

Gender Bias in Face Images
See how to detect and mitigate bias in automatic gender classification of face images.

Web experience: http://aif360.mybluemix.net/
Code: https://github.com/IBM/AIF360
Paper: https://arxiv.org/abs/1810.01943

Sicurezza e robustezza dell'AI

INFOSEC INSTITUTE

How Criminals Can Exploit AI

SecurityIntelligence

How Can Companies Defend Against Adversarial Machine Learning Attacks in the Age of AI?

Home > Security

NEWS

Hackers get around AI with flooding, poisoning and social engineering

Many defensive systems need to be tuned, or tune themselves, in order to appropriately respond to possible threats.

OPINION

The rise of artificial intelligence DDoS attacks

The leaves may change color, but the roots are the same. Are you ready for AI-based DDoS attacks?

ART - an open-source adversarial robustness toolkit

Adversarial Robustness

- Metrics
- Adversarial Sample Detection
- Input Preprocessing
- Model Hardening

Model Theft

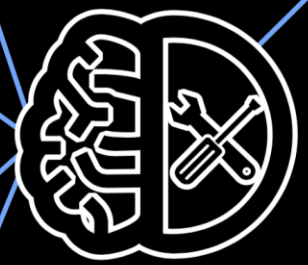
- Prevention of theft via APIs
- Detection of model theft attacks
- Deterring theft through model watermarking

Model and Data Privacy

- Provable privacy guarantees for training data (local differential privacy)
- Secure federated learning

Poisoning Attacks

- Detect poisoned training and models
- Poison can degrade performance or insert backdoors



IBM **ART**
Adversarial Robustness Toolbox
a.k.a. Nemesis

Model Robustness for AI DevOps

- Develop ART as platform agnostic library
- Modular framework to evaluate robustness, generate adversarial samples, and harden models
- Integration into IBM offerings to build secure model building pipelines

https://adversarial-robustness-toolbox.readthedocs.io/en/latest/

Interpretabilità è fondamentale specie in Sanità

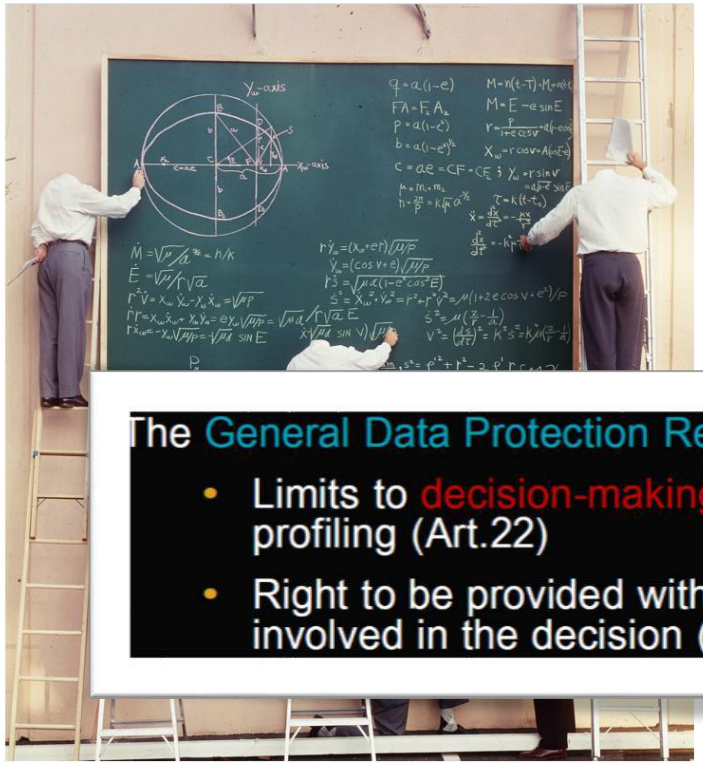
The quest for "explainable AI"

CIO JOURNAL.
Companies Grapple With AI's Opaque Decision-Making Process
THE WALL STREET JOURNAL.

Why Explainable AI Will Be the Next Big Disruptive Trend in Business
AlleyWatch

When a Computer Program Keeps You in Jail

Don't Trust Artificial Intelligence? Time To Open The AI 'Black Box'



The General Data Protection Regulation (GDPR)

- Limits to decision-making based solely on automated processing and profiling (Art.22)
- Right to be provided with meaningful information about the logic involved in the decision (Art.13 (2) f and 15 (1) h)

"meaningful" ???

Interpretabilità: ricerca attiva multidisciplinare

One explanation does not fit all

Different stakeholders require explanations for different purposes and with different objectives, and explanations will have to be tailored to their needs

End users

"Why did you recommend this treatment?"

Who: Physicians, judges, loan officers, teacher evaluators
Why: trust/confidence, insights(?)

Affected users

"Why was my loan denied? How can I be approved?"

Who: Patients, accused, loan applicants, teachers
Why: understanding of factors

Regulatory bodies

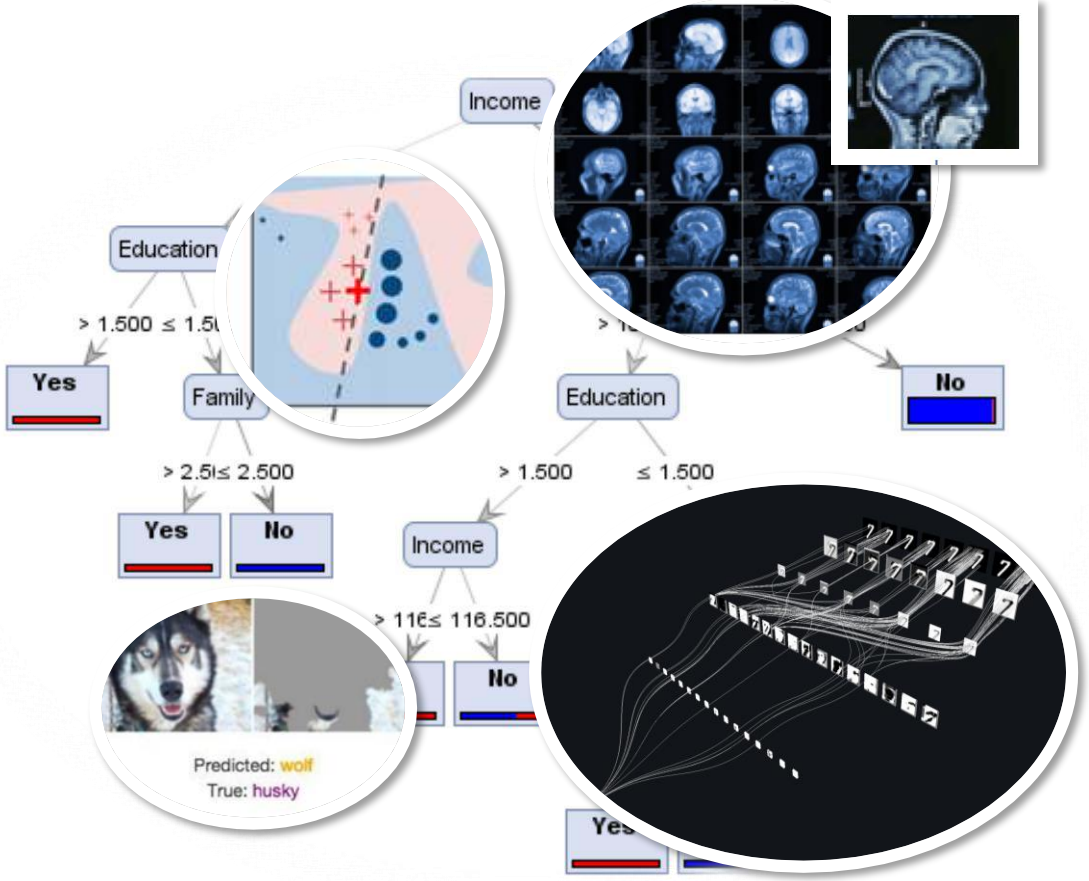
"Prove that your system didn't discriminate."

Who: EU (GDPR), NYC Council, US Gov't, etc.
Why: ensure fairness for constituents

AI system builders/stakeholders

"Is the system performing well? How can it be improved?"

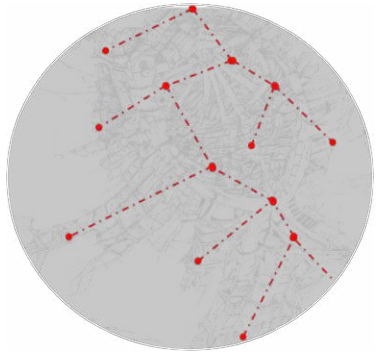
Who: EU (GDPR), NYC Council, US Gov't, etc.
Why: ensure or improve performance



Interpretabilità: alcuni riferimenti di letteratura

Our research focuses on tackling different ways to explain

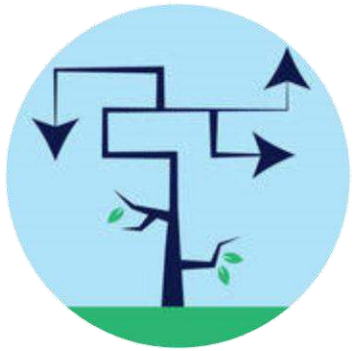
Selected 2018 explainability innovations from IBM Research



GLOBAL, POST-HOC

Improving Simple Models with Confidence Profiles

NIPS 2018
https://arxiv.org/abs/1807.07506



GLOBAL, DIRECTLY INTERPRETABLE

Boolean Decision Rules via Column Generation

NIPS 2018
https://arxiv.org/abs/1805.09901

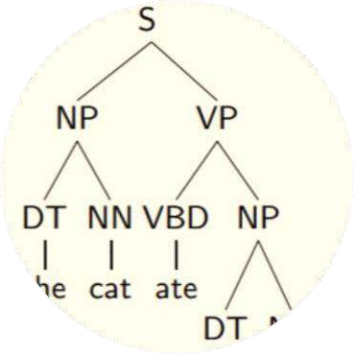
Variational Inference of Disentangled Latent Concepts from Unlabeled Observations

ICLR 2018
https://arxiv.org/abs/1711.00848

LOCAL, POST-HOC

Explanations Based on the Missing: Towards Contrastive Explanations with Pertinent Negatives

NIPS 2018
https://arxiv.org/abs/1802.07623



INTERACTIVE MODEL VISUALIZATION

Seq2Seq-Vis: A Visual Debugging Tool for Sequence-to-Sequence Models

IEEE VAST 2018
http://seq2seq-vis.io/

Riflessioni conclusive

- L'AI fornisce strumenti ed approcci utili per governare la complessità dei dati relativi alla salute.
- La varietà di approcci, tecniche, algoritmi, ed i limiti inerenti dell'AI, compresi correttamente aggiungono strumenti potenti di interpretazione della realtà a supporto della ricerca e dell'assistenza.
- Tecniche AI applicate ai dati non strutturati farebbero molto per dischiudere l'80% dei «dark data» sia per la ricerca che per l'assistenza.
- L'utilizzo di AI ha il potenziale di liberare risorse dalla «burocrazia» al rapporto umano e terapeutico, aumentando la qualità dell'assistenza e del lavoro degli addetti.
- La validazione «in vivo» deve essere condotta con il rigore ed i tempi del processo scientifico. C'è tanto da fare, partiamo subito assieme.