Harnessing big data for evaluation and regulation of medicines

Digital respiratory medicine – realism v futurism
A digital health summit of the ERS

Luis Pinheiro
Data Analytics and Methods – European Medicines Agency

Disclaimer

The views expressed in this presentation are the personal views of the speaker and may not be understood or quoted as being made on behalf of or reflecting the position of the European Medicines Agency or one of its committees or working parties.

Artificial Intelligence in Healthcare| The Promise of AI

Greater efficiency

Automate processes, address scalability issues

Reduce human error

Facilitate access, reduce human cognitive load

Structure & expose data

Transform text data into structured data, reduce dimensionality of data

Expand insights on data

Diagnostic models, predictive modelling, confounding adjustment

AI & Big Data | Big data facilitates AI

- AI not new but has gone through phases of dormant development (i.e. AI Winters)

- The advent of Big data and the increase in *computing power* have been considered the reason why it has gone from hypothetical applications to real use.

  - But also:

    - *internet* (information explosion and content sharing), and

    - *open-source technologies* (democratizing of techniques, accessible to a wider set of people)
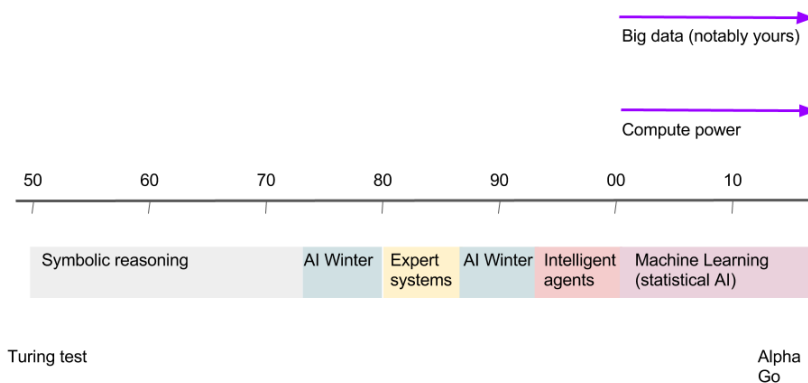
AI & Big Data | Big Data Definition

A formal definition of Big Data based on its essential features

Andrea De Mauro , Marco Greco , Michele Grimaldi ▾

Library Review
ISSN: 0024-2535
Article publication date: 4 April 2016   Reprints & Permissions

ALTMETRICS
12

Beyond the hype: Big data concepts, methods, and analytics

Amir Gandomi*, Murtaza Haider

Ted Rogers School of Management, Ryerson University, Toronto, Ontario M5B 2K3, Canada

RESEARCH ARTICLE

What is your definition of Big Data?
Researchers' understanding of the
phenomenon of the decade

Maddalena Favaretto *, Eva De Clercq, Christophe Olivier Schneble, Bernice
Simone Elger

- The V(s) approach:
  - 3 Vs (Velocity, variety, volume)
  - 5 Vs (+ Veracity and value)

- Heads of Medicines Agency (HMA)/EMA Big Data Taskforce definition:
  - '[E]xtremely large datasets which may be complex, multi-dimensional, unstructured and heterogeneous, which are accumulating rapidly, and which may be analysed computationally to reveal patterns, trends, and associations.

AI & Big Data | Big data in Healthcare

Observational data

Pharmacovigilance data

Clinical Trial Data

Social media & mHealth

Genomics & other omics

Digital respiratory medicine – realism v futurism • A digital health summit of the ERS

AI & Big Data | Use cases in Healthcare

- Clinical prediction models
  - Response to treatment, dose, adherence, adverse reactions, etc.
- Heterogeneity of treatment effects
  - Who will benefit most from treatment (and its counterfactual)
- Confounding adjustment
  - To estimate propensity scores
- Phenotype identification
  - Probabilistic phenotyping to identify who had a disease or specific condition (e.g. pregnancy)

- Impute missing data
  - Making use of non-parametric modelling, ensembles, etc.
- Expose additional data
  - Structure text information
- Decision support or automated systems
  - Automated drug delivery systems
- Monitoring and Endpoint adjudication
  - Clinical trials applications

Probabilistic phenotyping | An example

- A particular genotype can increase the probability of causing a combination of a set of ill-defined adverse reactions (i.e., a syndrome) to a class of medicines

- There are tens of thousands of reactions for that class of medicines in a global pharmacovigilance database

- Defining public health impact:

  - How can we tell *how many* had the susceptible genotype?

    - Consider that prior knowledge of that genotype will affect the probability of treatment.

An Application of Machine Learning in Pharmacovigilance: Estimating Likely Patient Genotype From Phenotypical Manifestations of Fluoropyrimidine Toxicity

Luis Correia Pinheiro[1,*], Julie Durand[1] and Jean-Michel Dogné[2,3]

Dihydropyrimidine dehydrogenase (DPD)-deficient patients might only become aware of their genotype after exposure to dihydropyrimidines, if testing is performed. Case reports to pharmacovigilance databases might only contain phenotypical manifestations of DPD, without information on the genotype. This poses a difficulty in estimating the cases due to DPD. Auto machine learning models were developed to train patterns of phenotypical manifestations of toxicity, which were then used as a surrogate to estimate the number of cases of DPD-related toxicity. Results indicate that between 8,878 (7.0%) and 16,549 (13.1%) patients have a profile similar to DPD deficient status. Results of the analysis of variable importance match the known end-organ damage of DPD-related toxicity, however, accuracies in the range of 90% suggest presence of overfitting, thus, results need to be interpreted carefully. This study shows the potential for use of machine learning in the regulatory context but additional studies are required to better understand regulatory applicability.

AI in the EU network | Strategic initiatives

- Three strategic initiatives provide recommendations on Big Data and the use of AI:
  - HMA/EMA Big Data Taskforce
  - Regulatory Science Strategy 2025
  - European Medicines Agencies Network Strategy to 2025

Digital respiratory medicine – realism v futurism • A digital health summit of the ERS

AI in the EU network | Highlights of actions

- Data Analysis Real World Interrogation Network - DARWIN EU®
  - Deliver a sustainable platform to access and analyse healthcare data (RWD) from across the EU
  - When mature will evolve to a node in the European Health Data Space (EHDS)
- Joint HMA/EMA workshop 19-20 April 2021
  - Stakeholders reflected on their priorities from the several recommendations on AI:
    - Framework to validate AI and Framework to develop guidance on AI
    - Supported by training of EU network, Academic and Research Partnerships and International collaborations

Internal EMA AI initiatives | Governance and cross-agency collaboration

- EMA AI Coordination Group
  - Catalyse the ongoing work on AI across the Agency
  - Promote oversight and a harmonised approach
  - Map expertise and define clearer roles and responsibilities
  - Foster sharing of knowledge and best practices

- EMA AI Technical Group
  - Forum for knowledge sharing and capacity building in practical AI
  - Forum for researching, debating and developing recommendations and best practices in technical aspects of AI
  - Safe environment for guided practical learning on AI methods and technical aspects (sandbox)
 - To support the monitoring of innovation in data science and AI

Future-proofing EMA | AI and novel digital technologies drivers

- **Internal drivers**
  - Increase efficiency:
    - Structure unstructured (text) data (e.g. NLP)
    - Automate menial activities (e.g. RPA)
    - Improve access to data and reduce data harvesting / collection time (e.g. implementation of data models / NLP)
  - Extract deeper insights from data
    - Advanced analytics (machine learning, deep learning)

- **External or client-side drivers**
  - Regulate products with AI elements
    - Digital endpoints
    - Probabilistic phenotyping / automated event adjudication
    - Heterogeneity of treatment effects
    - Patient-level predictions
    - Automated ICSR coding
    - New signal detection methods
    - Automated literature monitoring

Vision

Innovate to turn data into decisions for a healthier world
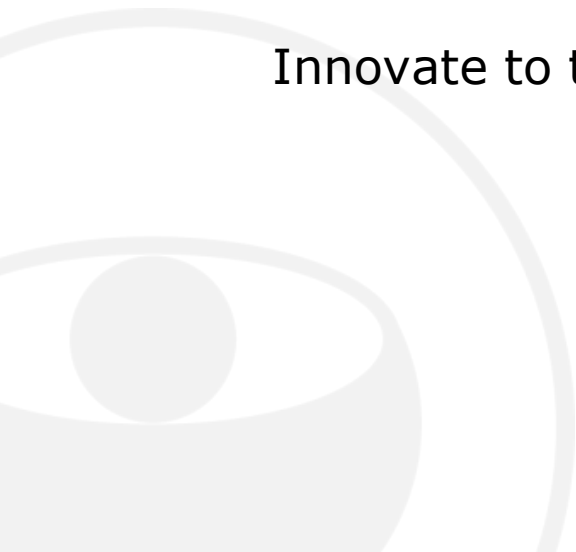
Any questions?

Further information

luis.pinheiro@ema.europa.eu

Official address Domenico Scarlattilaan 6 ● 1083 HS Amsterdam ● The Netherlands
Send us a question Go to www.ema.europa.eu/contact
Telephone +31 (0)88 781 6000

Follow us on 🐦 **@EMA_News**