

Machine Learning, Big Data, AI How does it apply to my project?

Guillaume Labilloy
Center for Data Solutions
03.03.2020

Agenda

- ML/AI in research/clinical
- Brief history of AI and ML
- Learning techniques and Algorithms
- Data science and full fledged ML/AI
- Current utilization
- Examples of projects

ML/AI in research/clinical

Short term risk prediction:

Wijnberge, M., et al., *Effect of a Machine Learning-Derived Early Warning System for Intraoperative Hypotension vs Standard Care on Depth and Duration of Intraoperative Hypotension During Elective Noncardiac Surgery*. JAMA, 2020.

Tumor detection:

McKinney, S.M., et al., *International evaluation of an AI system for breast cancer screening*. Nature, 2020. 577(7788): p. 89-94.

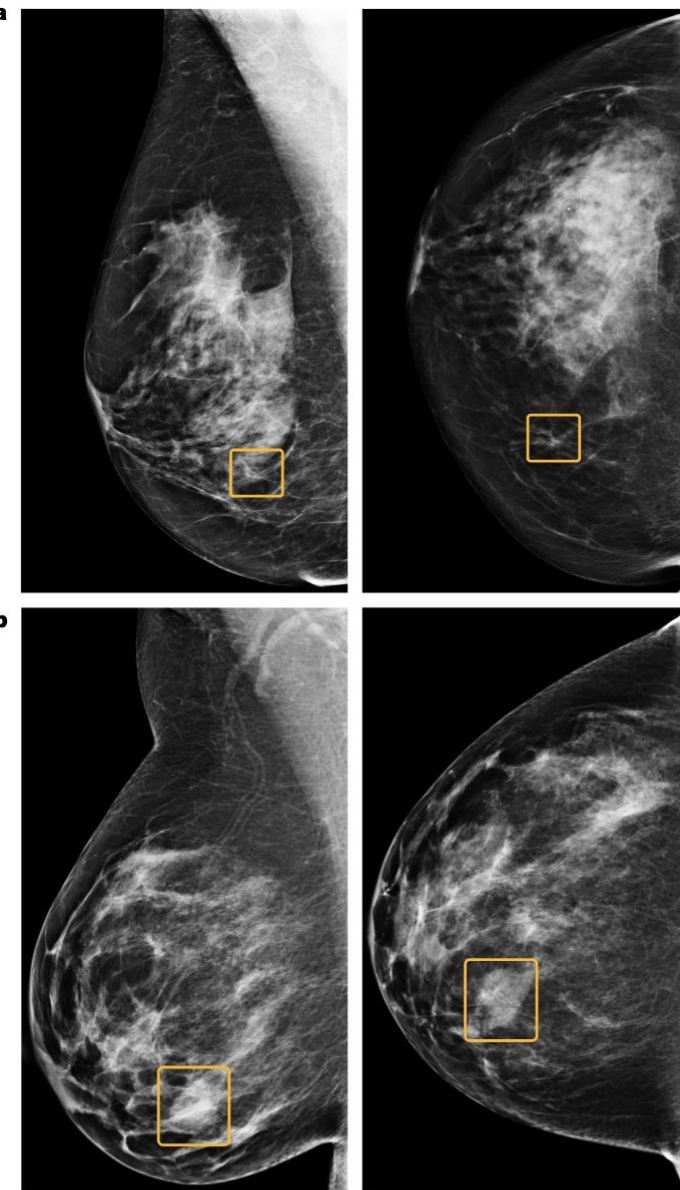


Fig. 4 | Discrepancies between the AI system and human readers. a, A sample cancer case that was missed by all six readers in the US reader study, but correctly identified by the AI system. The malignancy, outlined in yellow, is a small, irregular mass with associated microcalcifications in the lower inner right breast. b, A sample cancer case that was caught by all six readers in the US reader study, but missed by the AI system. The malignancy is a dense mass in the lower inner right breast. Left, mediolateral oblique view; right, craniocaudal view.

ML/AI Brief History (1/2)

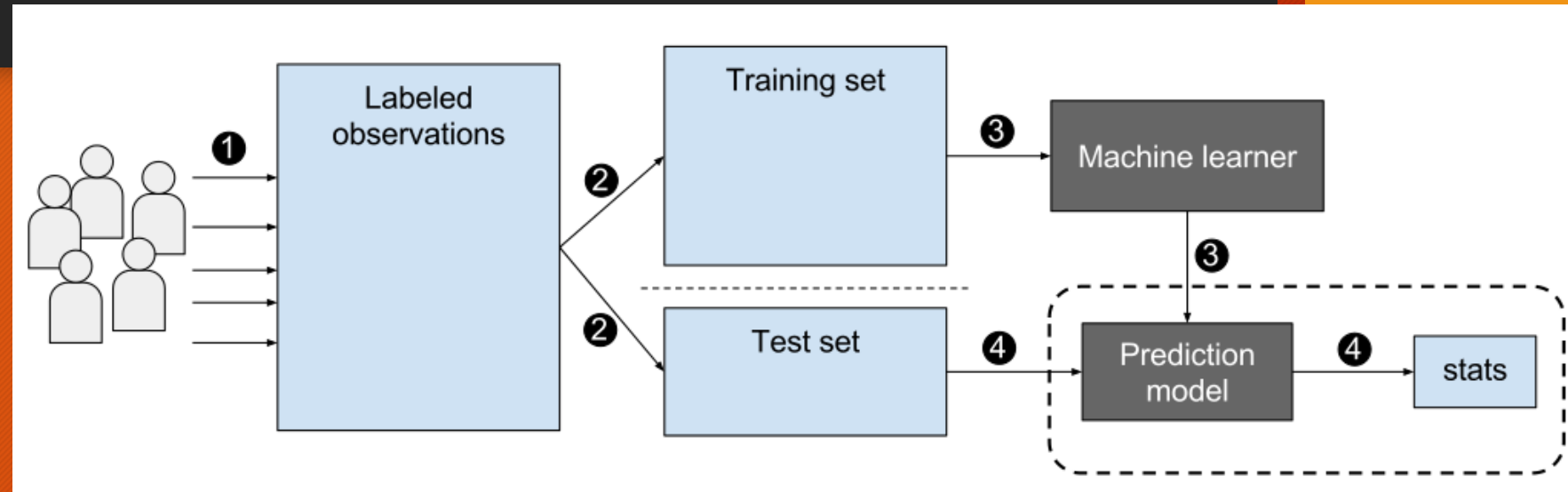
- 1949 - Donald Hebb, book titled *The Organization of Behavior*
- 1950 - Alan Turing propose the Turing test in "Computing Machinery and Intelligence" paper
- 1952 - Arthur Samuel (IBM) first came up with the phrase "Machine Learning", while working on a game of checkers program.
- 1957 - The perceptron by Frank Rosenblatt (one layer network)
- 1967 - The Nearest Neighbor Algorithm (the traveling salesperson's problem)

ML/AI Brief History (2/2)

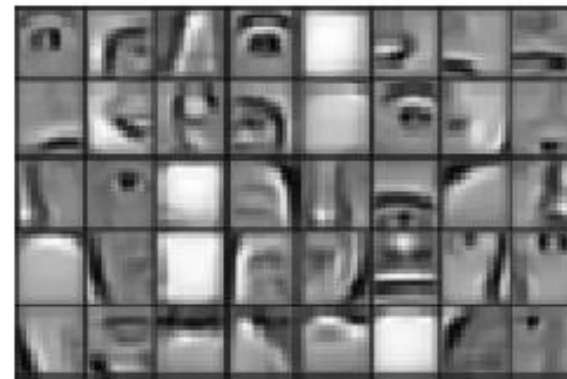
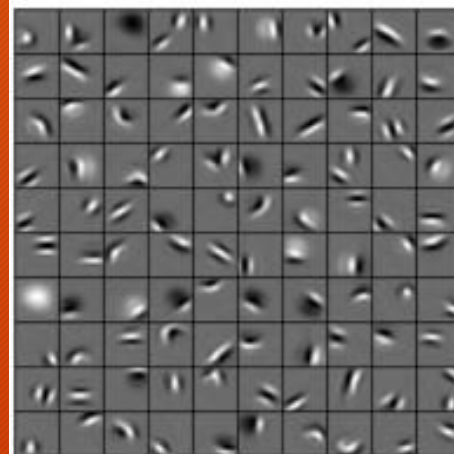
- 60s-70s - Multilayer neural networks (feedforward and backpropagation)
- 70s-80s - Differentiation between AI and ML
 - AI struggled
 - ML shifted from training program for AI to solving practical problems
- 90s
 - ML flourishes, due to the Internet and availability of data
 - Boosting: “A set of weak learners can create a single strong learner.” - Robert Schapire
- 1997 - Speech recognition using NN Long Short-Term Memory (LSTM)
- 2012 - Google’s X Lab can autonomously browse and find videos containing cats

Learning techniques (1/2)

- Supervised

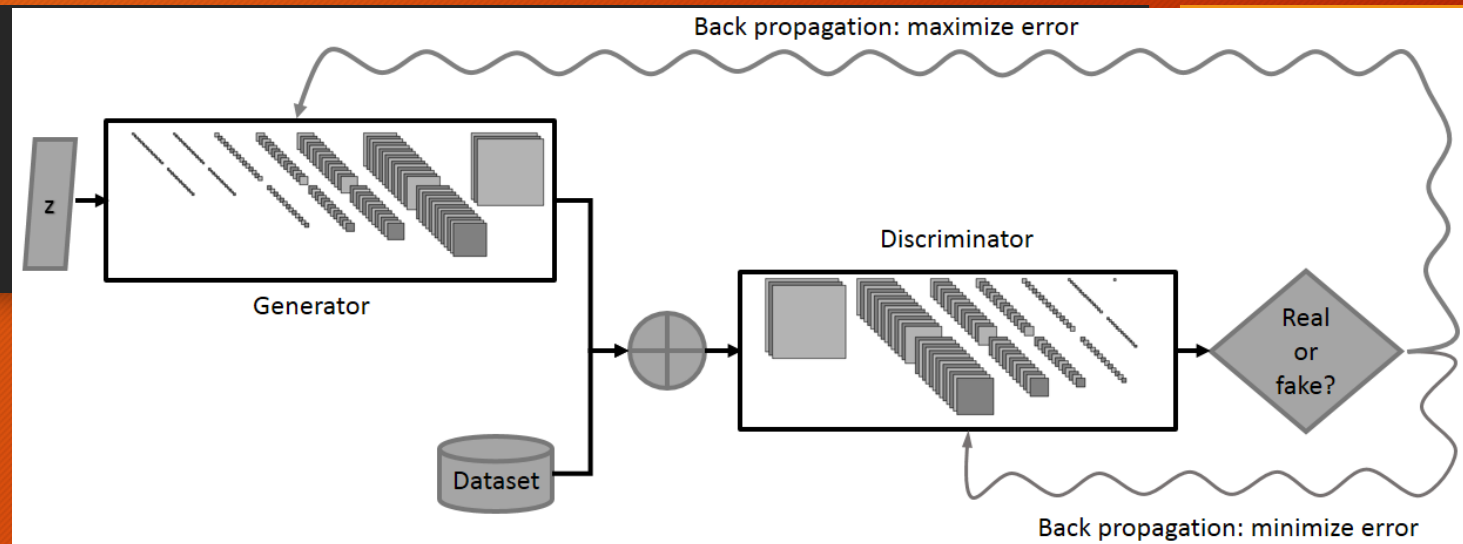


- Unsupervised

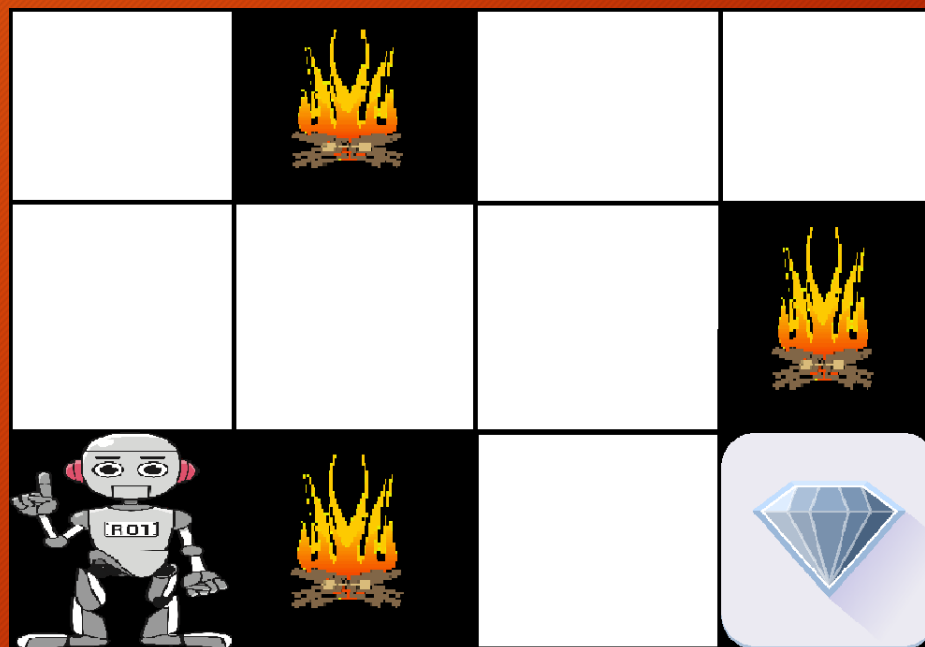


Learning techniques (2/2)

- Semi-Supervised (General adversarial network)



- Reinforcement



Types of algorithms to classify and/or predict

- Regression Algorithms (Logistic Regression, Stepwise Regression, ...)
- Instance-based Algorithms (k-Nearest Neighbor (kNN), SVM, ...)
- Regularization Algorithms (Least Absolute Shrinkage and Selection Operator LASSO, ...)
- Decision Tree Algorithms (Chi-squared Automatic Interaction Detection CHAID, ...)
- Bayesian Algorithms (Naive Bayes , ...)
- Clustering Algorithms (Hierarchical Clustering, ...)
- Association Rule Learning Algorithms (Apriori algorithm, ...)
- Artificial Neural Network Algorithms (Multilayer Perceptrons (MLP), ...)
- Deep Learning Algorithms (Convolutional Neural Network (CNN), ...)
- Dimensionality Reduction Algorithms (Principal Component Analysis (PCA), ...)
- Ensemble Algorithms (Boosting, Weighted Average (Blending), ...)
- And more such as NLP, Computer vision, etc.

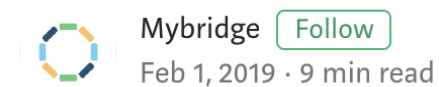
Rich Open Source Ecosystem

Lime:



Amazing Machine Learning Open Source of the Year (v.2019)

Top 49 Tools & Projects. Average Github ★: 3,566.



Feb 1, 2019 · 9 min read



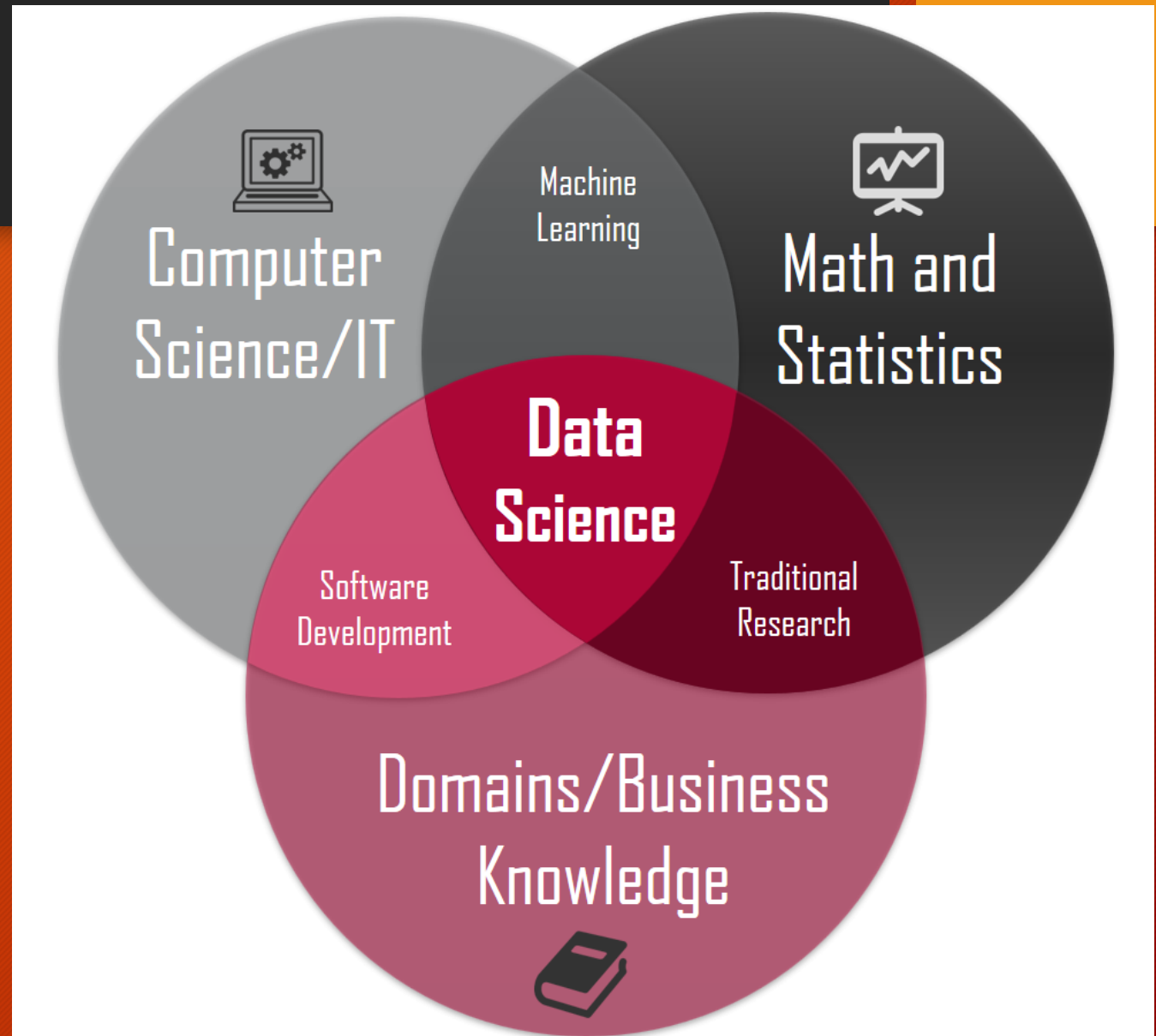
For the past year, we've compared nearly 22,000 Machine Learning open source tools and projects to pick Top 49 (0.22% chance).

The tools and projects are broken down by 6 categories

- Computer Vision (1~5)
- Reinforcement Learning (6~13)
- NLP (14~20)
- GAN (21~26)
- Neural Network (27~35)
- Toolkit (36~49)

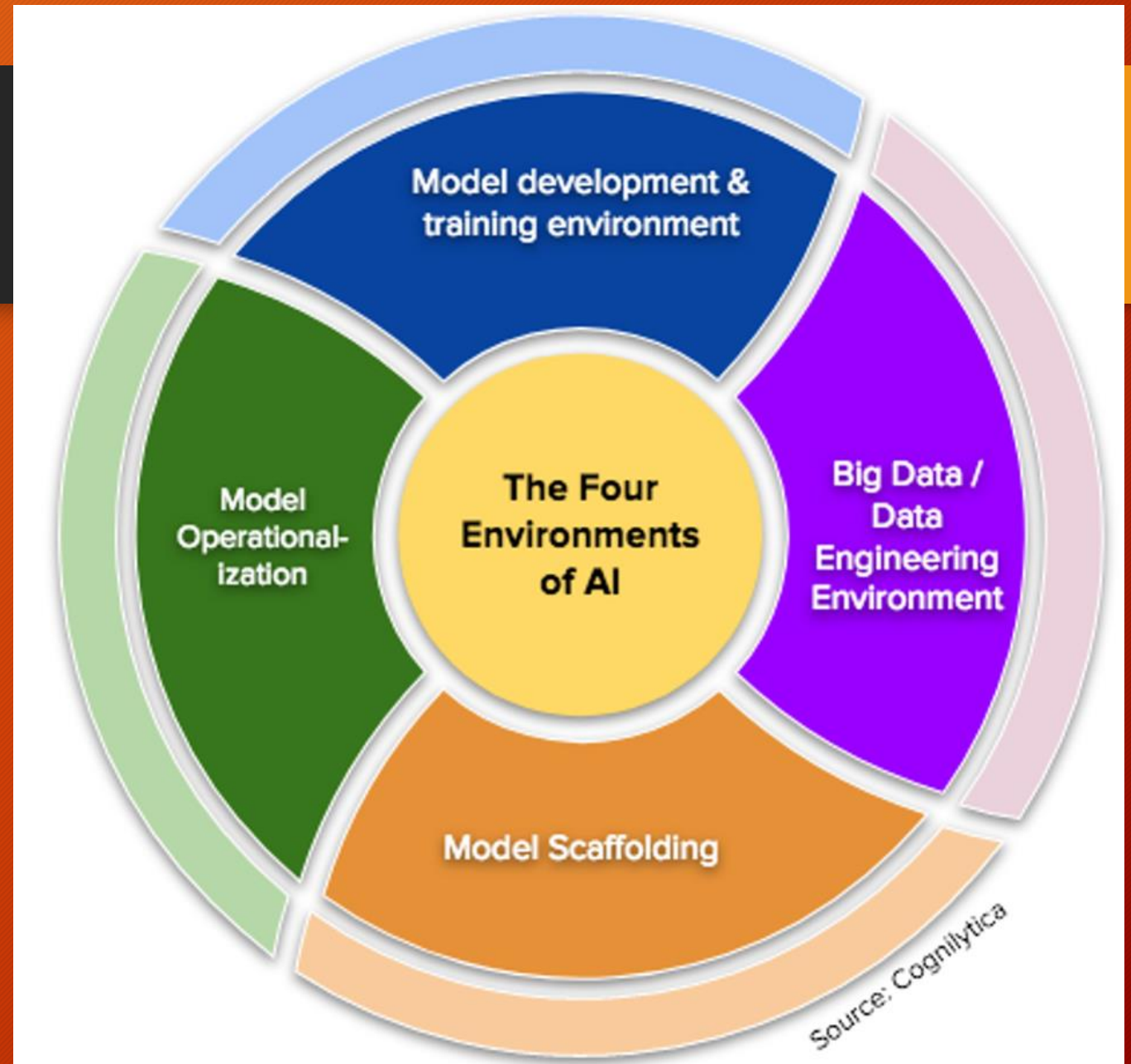
Data Science

- Computer Science/IT
- Math and Statistics
- Domains/Business Knowledge



Full Scale Of ML/AI

- Software
- Infrastructure/Big Data
- Integration
- Operation



ML/AI is mature and widely used

- **Analyzing Sales Data:** Streamlining the data
- **Real-Time Mobile Personalization:** Promoting the experience
- **Fraud Detection:** Detecting pattern changes
- **Product Recommendations:** Customer personalization
- **Learning Management Systems:** Decision-making programs
- **Dynamic Pricing:** Flexible pricing based on a need or demand
- **Natural Language Processing:** Speaking with humans

ML and Stats (1/2)

What's required to create good machine learning systems?

- Data preparation capabilities.
- Algorithms – basic and advanced.
- Automation and iterative processes.
- Scalability.
- Ensemble modeling.



Did you know?

- In machine learning, a target is called a label.
- In statistics, a target is called a dependent variable.
- A variable in statistics is called a feature in machine learning.
- A transformation in statistics is called feature creation in machine learning.

ML and Stats (2/2)

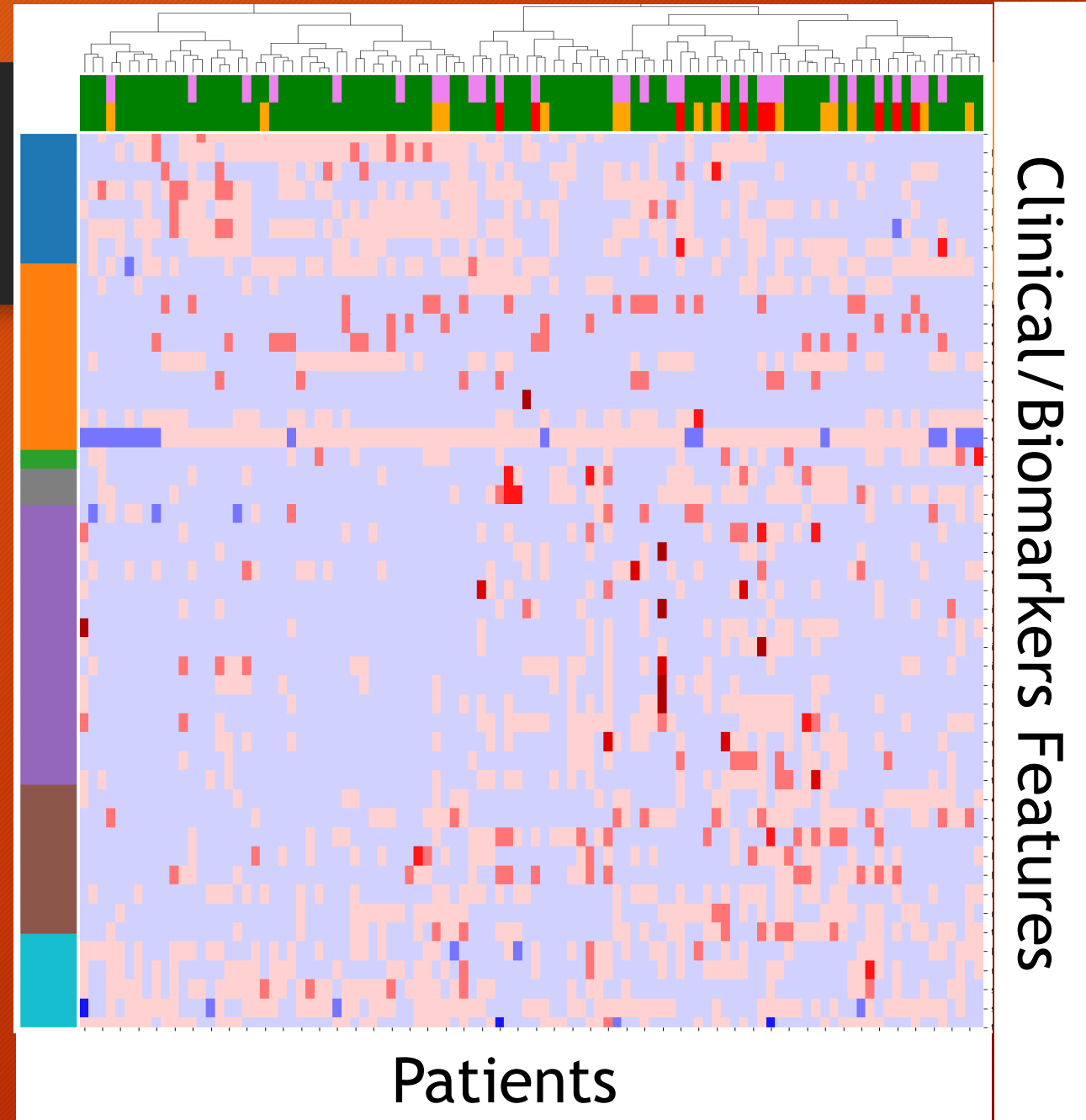
Statistics	Machine learning
Covariates	Features
Outcome variable	Target
Model	Network, graphs
Parameters	Weights
Model for discrete var.	Classifier
Model for continuous var.	Regression
Log-likelihood	Loss
Multinomial regression	Softmax
Measurement error	Noise
Subject/observation	Sample/instance
Dummy coding	One-hot encoding
Measurement invariance	Concept drift

Statistics	Machine learning
Prediction	Supervised learning
Latent variable modeling	Unsupervised learning
Fitting	Learning
Prediction error	Error
Sensitivity	Recall
Positive predictive value	Precision
Contingency table	Confusion matrix
Measurement error model	Noise-aware ML
Structural equation model	Gaussian Bayesian network
Gold standard	Ground truth
Derivation-validation	Training-test
Experiment	A/B test

ML implementation

Sepsis project:

Can we predict outcome based on current clinical data and specific biomarkers?



Thank you!

Any questions?



DATA SOLUTIONS

<https://centerfordatasolutions.org/>