Artificial Intelligence, the Future of Life Sciences

董伟 (Wei Dong)

SAS Institute Inc.

THE POWER TO KNOW.

Agenda

- AI and Relevant Terms
- Applications in Life Sciences
- SAS'S AI Capability
- Successful Stories
- Q & A

AI and Relevant Terms

Artificial Intelligence

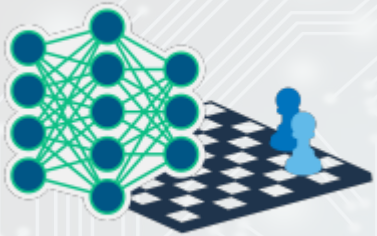
is the science of training systems to emulate human tasks through **Learning** and **Automation**

Understand Context

Learn Patterns

Recognize Objects

Evolution of Artificial Intelligence



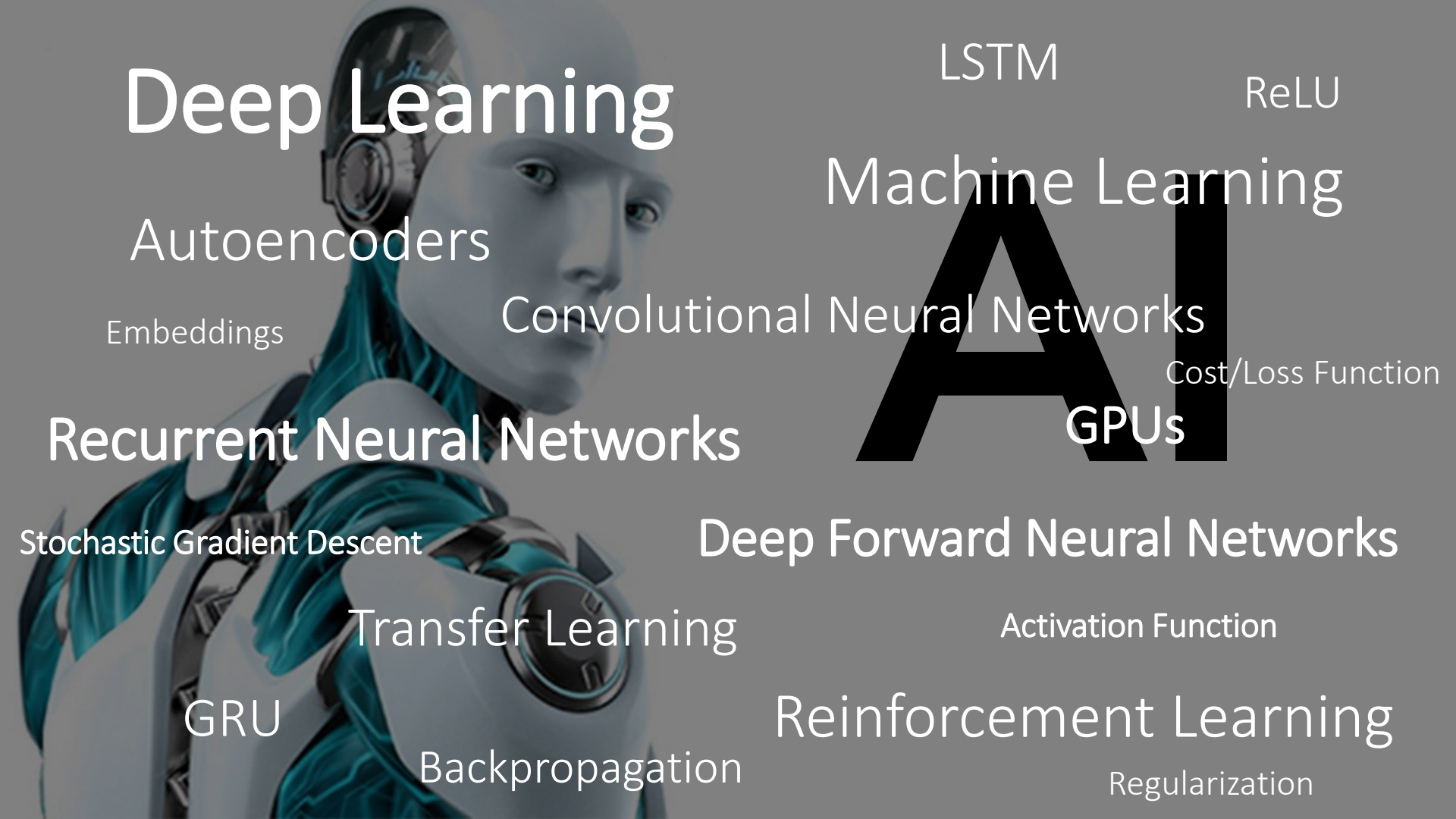
1950s-1970s
Neural Networks

1980s-2010s
Machine Learning

Present Day
Deep Learning and Cognitive Systems

SAS

1976


Deep Learning
LSTM
ReLU
Machine Learning
Autoencoders
AI
Convolutional Neural Networks
Embeddings
Cost/Loss Function
GPUs
Recurrent Neural Networks
Stochastic Gradient Descent
Deep Forward Neural Networks
Transfer Learning
Activation Function
GRU
Reinforcement Learning
Backpropagation
Regularization

Evolution

The term "Artificial Intelligence"

Neural Network

Machine Learning

Classical

Modern

Deep Learning

Applied Artificial Intelligence

Data Size , Variety & Compute Power

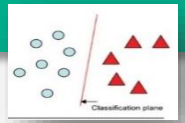
Past to Present

1950s | 1990s | 2010's | Present

Types of Machine Learning Methods

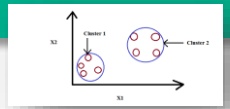
SUPERVISED LEARNING

- Labeled data
- Classification, Prediction
- Algorithms: Logistic Regression, Gradient Boosting etc.



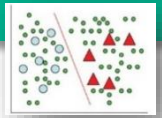
UNSUPERVISED LEARNING

- Unlabeled data
- Clustering, Feature Extraction
- Algorithms: K-means clustering, PCA, etc.



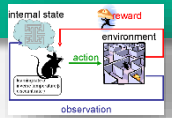
SEMI-SUPERVISED LEARNING

- Labeled and unlabeled data
- Classification, Prediction
- Algorithms: Autoencoders, TSVM etc.



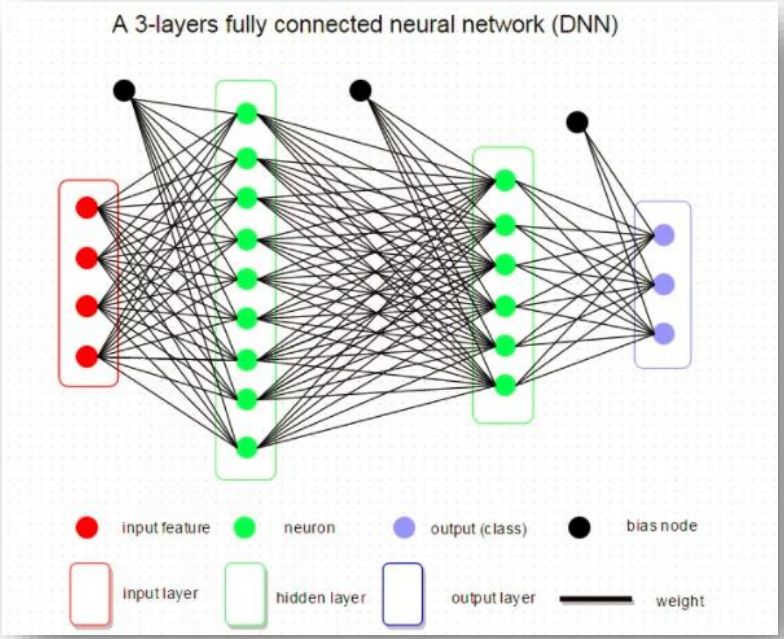
REINFORCEMENT LEARNING

- Agent, environment and actions
- Robotics, Gaming and Navigation
- Monte-Carlo methods etc.



Deep Learning

A specialization of machine learning



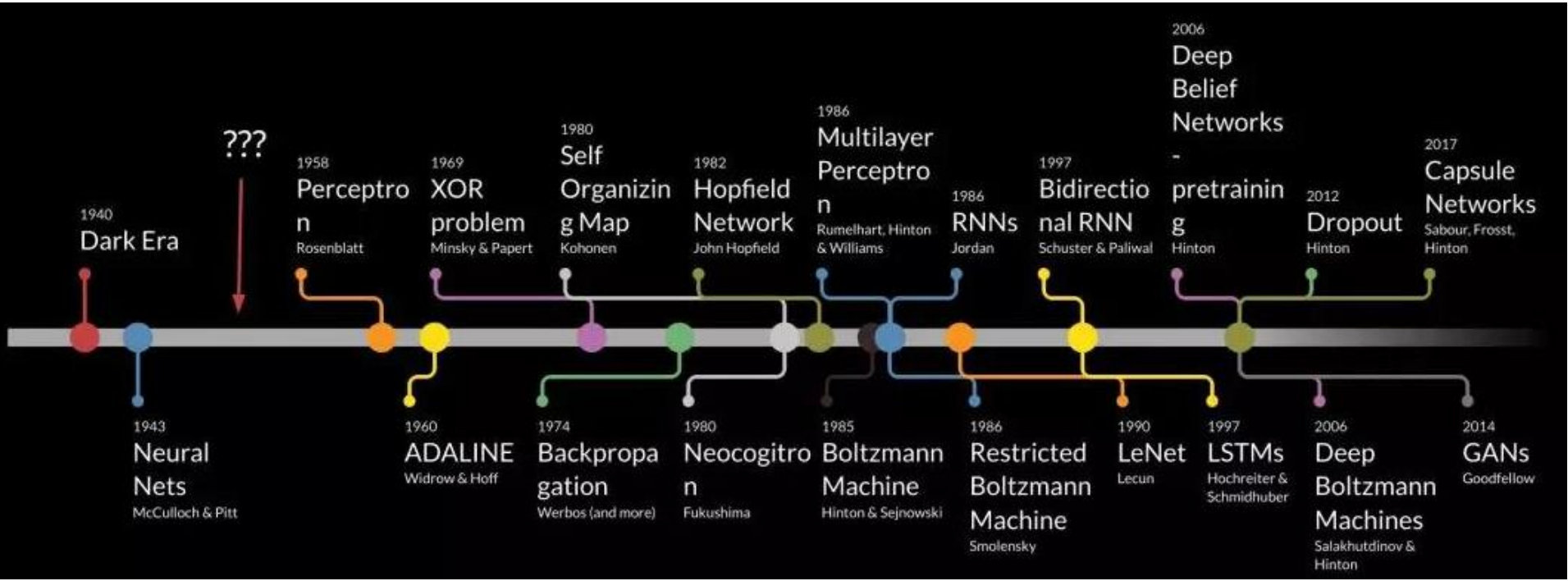
- Neural networks with many layers and different types of …
  - Activation functions
  - Network architectures
  - Sophisticated optimization routines

Each layer represents an optimally weighted, non-linear combination of the inputs.

- Automatic feature generation
- Extremely accurate results if well-trained; use for classification, prediction, or pattern recognition, especially in unstructured data

Image from: https://www.gabormelli.com/RKB/Multi_Hidden-Layer_Neural_Network

History of Deep Learning



What about transfer learning?

Portability of learned features is a key benefit

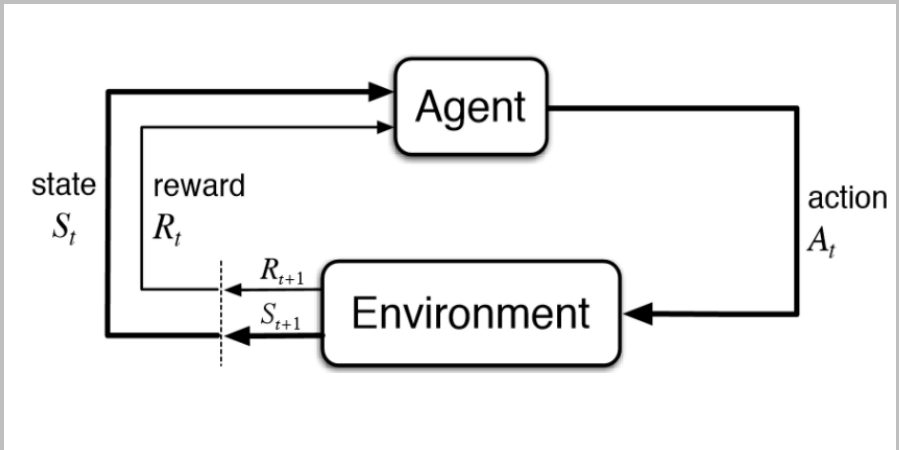
- A common and highly effective approach to deep learning on smaller datasets is to use a ***pre-trained network*** – that is, a saved network that was previously trained on a large dataset, typically on a large-scale task.

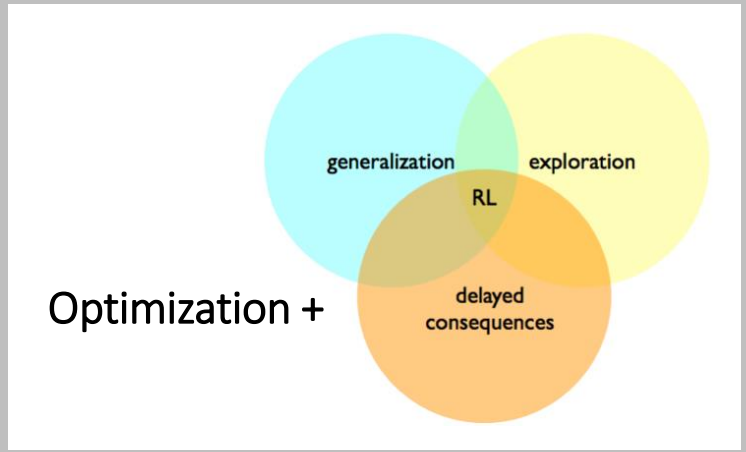
  If this original dataset is large enough and general enough (e.g. ImageNet for images, Wikipedia for text), then the features learned by the pre-trained network can effectively act as a generic model that can be applied to new data, even if it's a slightly different context.

- Users can import SAS pre-packaged networks such as LeNet, ResNet, and VGG. They can also add a custom network on top of an already-trained base network* and jointly train both layers. In the 18w25 release, users will also be able to import architectures, specify a layer name, and freeze this layer as well as all previous layers.

What is reinforcement learning?

A machine learning technique where the goal is to learn a behavior strategy that maximizes the long term sum of rewards in an unknown and stochastic environment.





Unlike standard supervised learning, correct input/output pairs need not be presented, and sub-optimal actions need not be explicitly corrected. Instead the focus is on performance, which involves finding a balance between exploration (of uncharted territory) and exploitation (of current knowledge).

Applications in Life Sciences

- Diagnosis and Disease Identification
- Advancing Research Of New Products
- Personalized Medicine
- Drug Discovery and Manufacture
- Clinical Trial Design and Data Interrogation in Drug Discovery
- Image Recognition
- Electronic Health Records
- Real World Evidence

Diagnosis and Disease Identification

- The biggest challenge in medicine is correct diagnosis and identification of diseases, which makes it priority one in machine learning development.
- Histopathology image analysis and automated diagnosis were ripe for AI, given the technological progress in digitalization of complete histology slides, which permit all microscopic magnifications. AI and pattern recognition, combined with complex algorithms and automated immunohistochemical measurement systems, have advanced pathologists' ability to oversee the analysis and concentrate on more-difficult cases.

Advancing Research Of New Products

- Using sophisticated learning algorithms to mine real-world structured and unstructured data to uncover insights can lead to the identification of new mechanisms of disease, potential new line extension, and design for preclinical experiments.
- Knowledge gaps of how candidates act on proteins to aid design of new drugs can be filled.
- Knowledge can be extracted in real time from commercial, scientific, and regulatory literature, allowing researchers to identify competitive white space, eliminate blind spots in research, and discover disease similarities.

Personalized Medicine

- There is much research going on regarding the use of machine learning and predictive analytics in customizing treatment to a person's unique health history. If successful, this can result in optimized diagnoses and treatment protocols. Currently, the focus is on supervised learning where doctors can use genetic information and symptoms to narrow down diagnostic options or make an educated guess about a patient's risk. This can lead to better preventive measures.
- The predicted surge in the use of advanced health-measuring mobile apps as well as micro biosensors and devices in the next 10 years will provide a wealth of data that can help point the way for effective research and development and better treatment protocols. Aside from better health management, personalized medicine also means lower costs overall, up and down the chain.

Drug Discovery and Manufacture

- Machine learning plays many roles in early-stage drug discovery, such as the development of new drug compounds, and in discovery technologies, such as next-generation sequencing. One of the first in this field is precision medicine, which makes identification of complex diseases and possible treatment modalities more efficient. The research uses unsupervised learning, which seeks data patterns without predicting outcomes.
- Across the industry, product development timelines range from seven to 10 years from discovery to launch, with sights set on reducing them to five to seven years. Advancements in AI and machine learning to reduce the time it takes to develop, manufacture, and launch new patient therapies support the goal of reducing overall product development timelines. Scientists are integrating research data, lab data, and clinical data, in combination with new information sources (e.g., social media and wearables) across the drug development spectrum, creating a holistic picture of the drug development candidate. Improving ways to acquire and mine data in real time allows scientists to use AI and machine learning to make improved decisions faster, which will accelerate the product development and scale-up process.

Clinical Trial Design and Data Interrogation in Drug Discovery

- Clinical trial research is a long and arduous progress. Machine learning can help make it less in various ways. One is by using advanced predictive analytics on a wide range of data to identify candidates for clinical trials for target populations much more quickly.
- Compliance is often a burden on companies and requires an approach to mitigate costs while meeting regulation. New applications are emerging utilizing advanced algorithms based on customized NLP technologies incorporating scientific-specific taxonomies and text-mining models. Using these advanced models, it is possible to identify keywords, phrases, and data patterns (such as adverse event dates) that may require redaction or anonymization. These new applications provide the higher level of accuracy required to meet the policy requirements while also automating manual activities.
- For regulatory reasons, late stage clinical trials have been conducted on large, diverse patient populations over many sites. AI is playing an increasingly important role in the design of clinical trials and the interpretation of the data they produce. This is enabling patient enrolment to be much more effective, trials to be conducted on smaller patient populations and for those patients to participate remotely from a wider geographical area, reducing cost while increasing the likelihood of obtaining accurate and relevant data.

Image Recognition

- The use of AI to interrogate medical images to identify disease is already under way and is likely to be one of the early success stories in the use of AI in this sector. An example that received a lot of publicity last year was a model jointly developed by Google DeepMind and Moorfields Eye Hospital in London. The model was trained on approximately 15,000 images after which it identified eye disease in approximately 1000 images more accurately than a team of retinal specialists. It will be able to complete its analysis more quickly and accurately the more images it reviews. A further benefit is that the algorithm may be adaptable for use in reviewing radiotherapy and mammogram images.

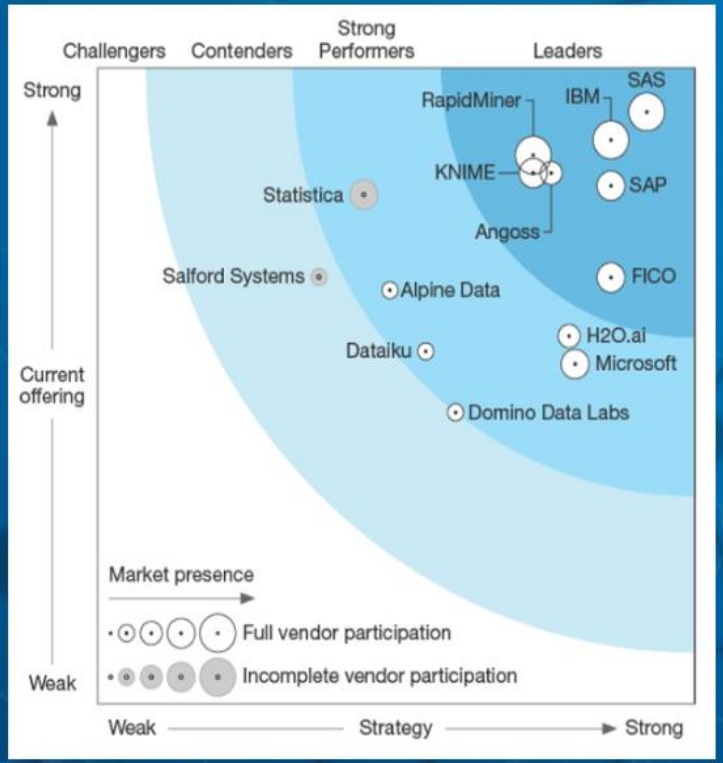
Electronic Health Records

- There is little standardization of patient health records, even where they are digitized. Consequently, it is difficult to extract relevant information or to make connections that may allow meaningful insights into the underlying causes of ill health. AI can overcome some of these limitations. For example, natural language processing tools can ensure that information is captured in a more standardized way, making it more accessible to search tools. Other free text search programs are able to extract key terms from less structured data. Diagnostic algorithms are helping predict (and therefore track and manage) risk of future illness on the basis of historic health data.

Real World Evidence

- The increasing availability of real world data as a way to assess performance of drugs in the real world gives rise to a number of benefits. First, it is allowing regulators to approve new drugs sooner – and therefore more cheaply – on the basis that they will be monitored on an ongoing basis for effectiveness and side effects. Secondly, it is enabling healthcare systems to push manufacturers into payment by results charging models where therapies are paid for on the basis of outcomes. It is expected that this will reduce waste in healthcare systems considerably.

SAS’S AI Capability

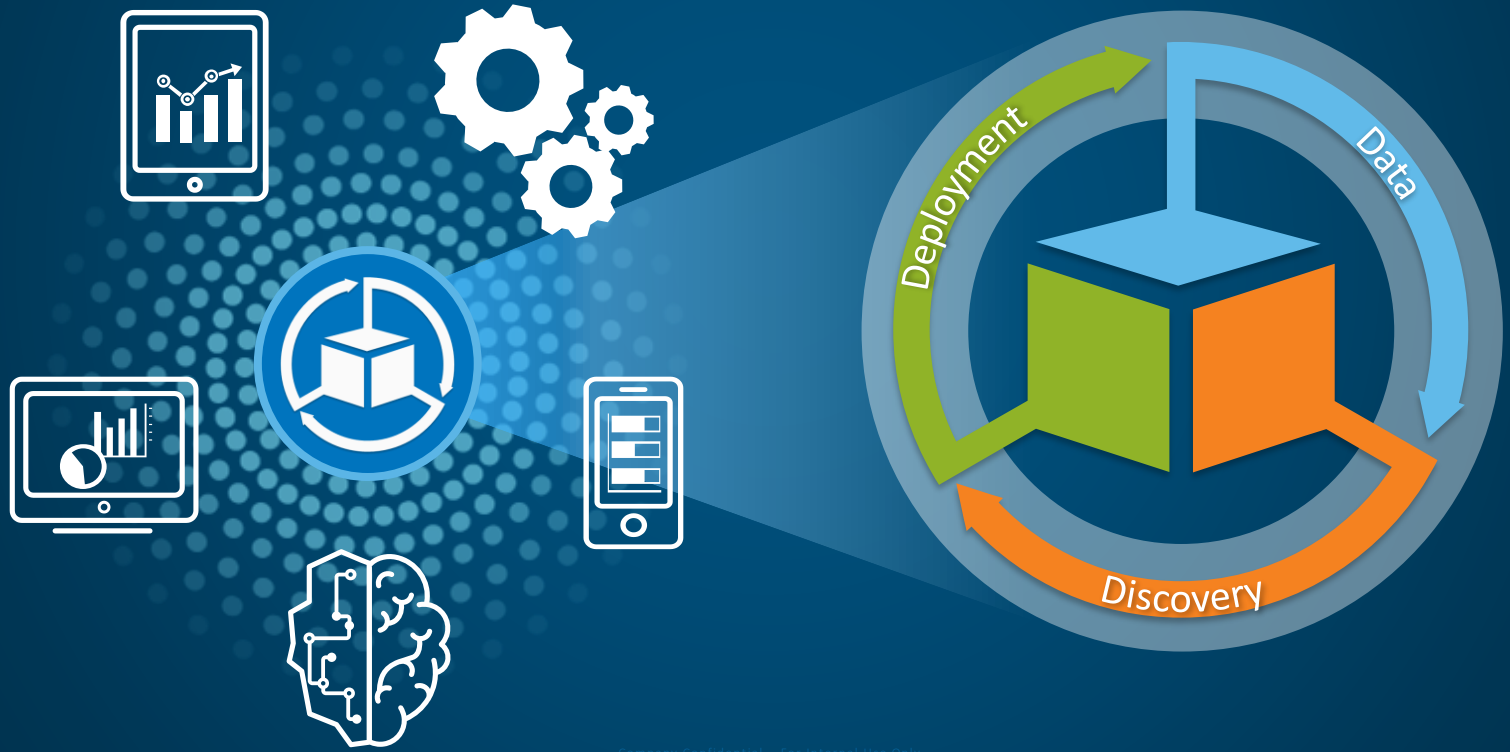


SAS' vision for data science is not limited to innovation in tools. It has been quick to jump on any new promising analytical methods across multiple disciplines, such as statistics, econometrics, optimization, machine learning, deep learning and natural language interaction.

The Forrester Wave™: Predictive Analytics and Machine Learning Solutions, Q1 2017

The SAS Platform

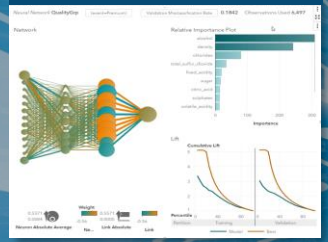
… is a software foundation that's engineered specifically to generate insights from data, in any computing environment.



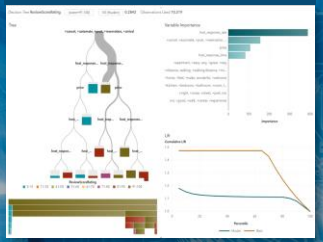
Artificial Intelligence at SAS

Core Capabilities

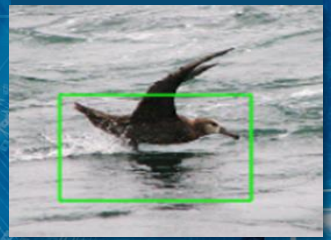
Machine Learning



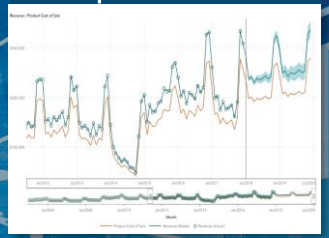
Natural Language



Computer Vision

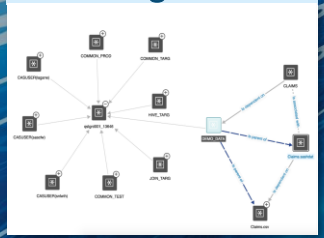


Forecasting and Optimization



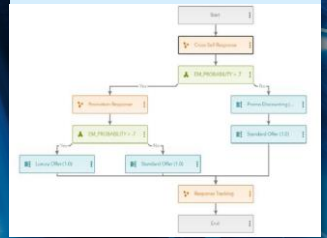
Supporting Technologies

Data Management

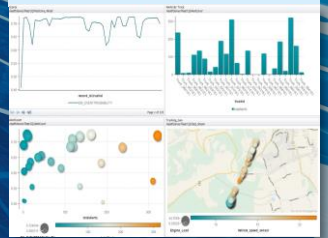


Visualization

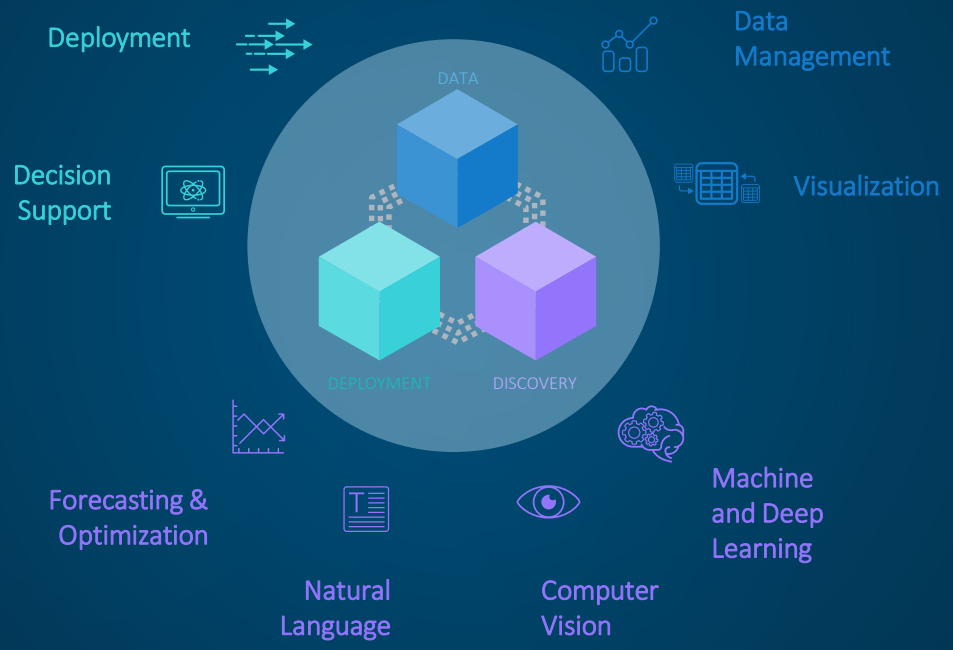
Decision Support



Deployment



SAS embeds AI capabilities into entire Analytics Lifecycle



DIVERSITY

Range of analytics, from exploration to AI

Embraces and extends open source technologies

SCALE

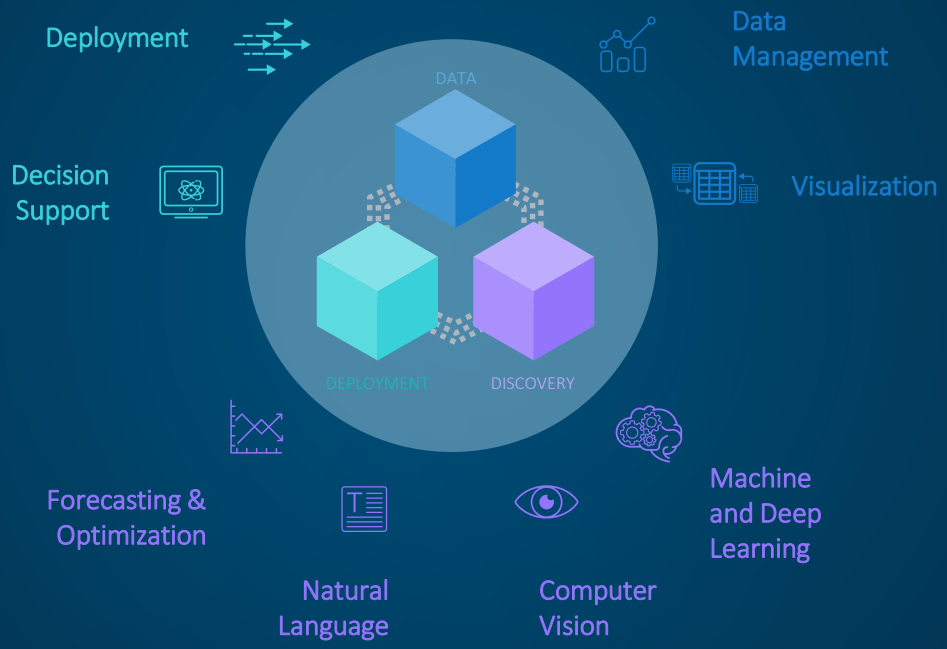
Data – Speed and Volume

Performance of insights – from data to action

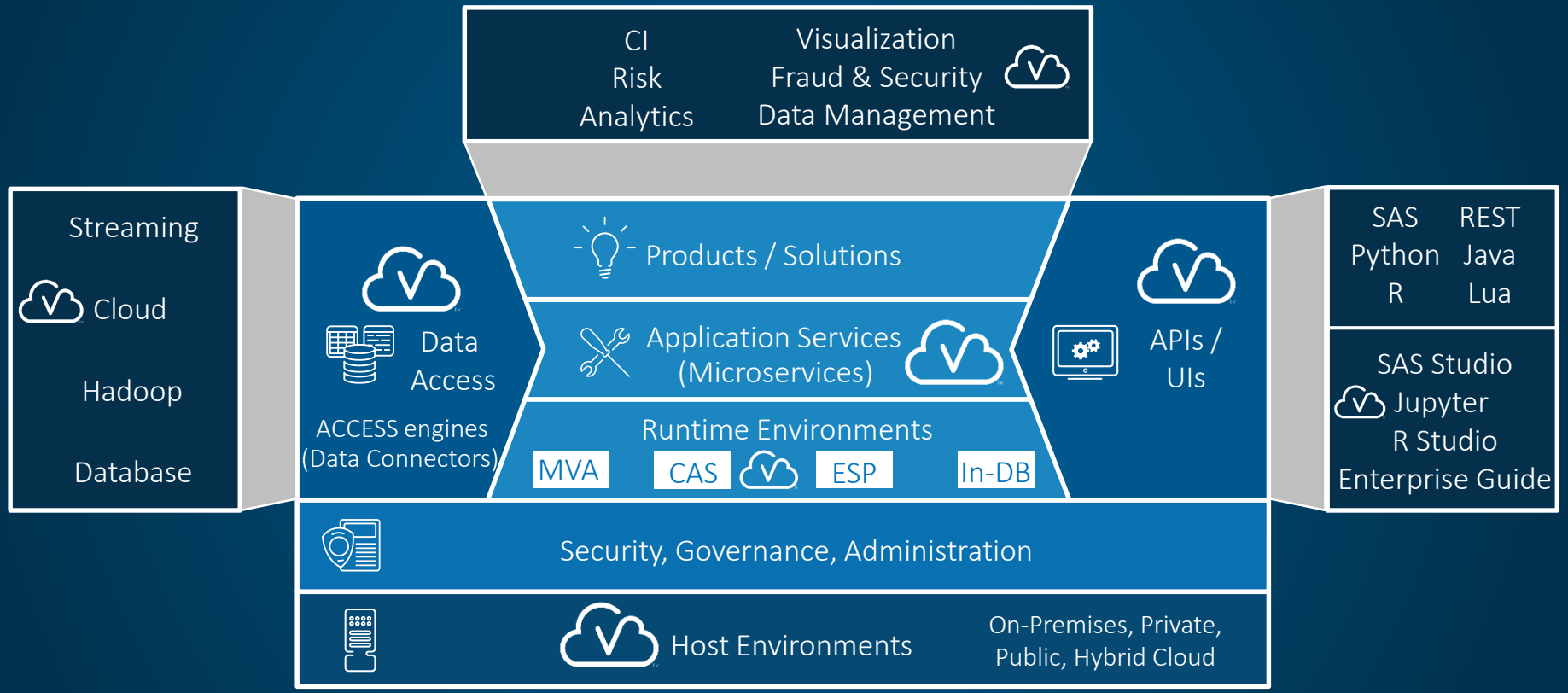
TRUST

Experience & Leadership

Resiliency, Security & Governance



The SAS Platform Conceptual Architecture



- The promise of AI in healthcare and life sciences is profound. It can help physicians and researchers prevent disease, speed recovery and save lives, by unlocking complex and varied data sets to develop new insights. AI can speed genomics processing and make medical image analysis faster and more accurate for personalized treatment. It can also be used to detect and correct massive waste, fraud and abuse in healthcare spending.

Successful Stories

Cancer Detection

1,735,350

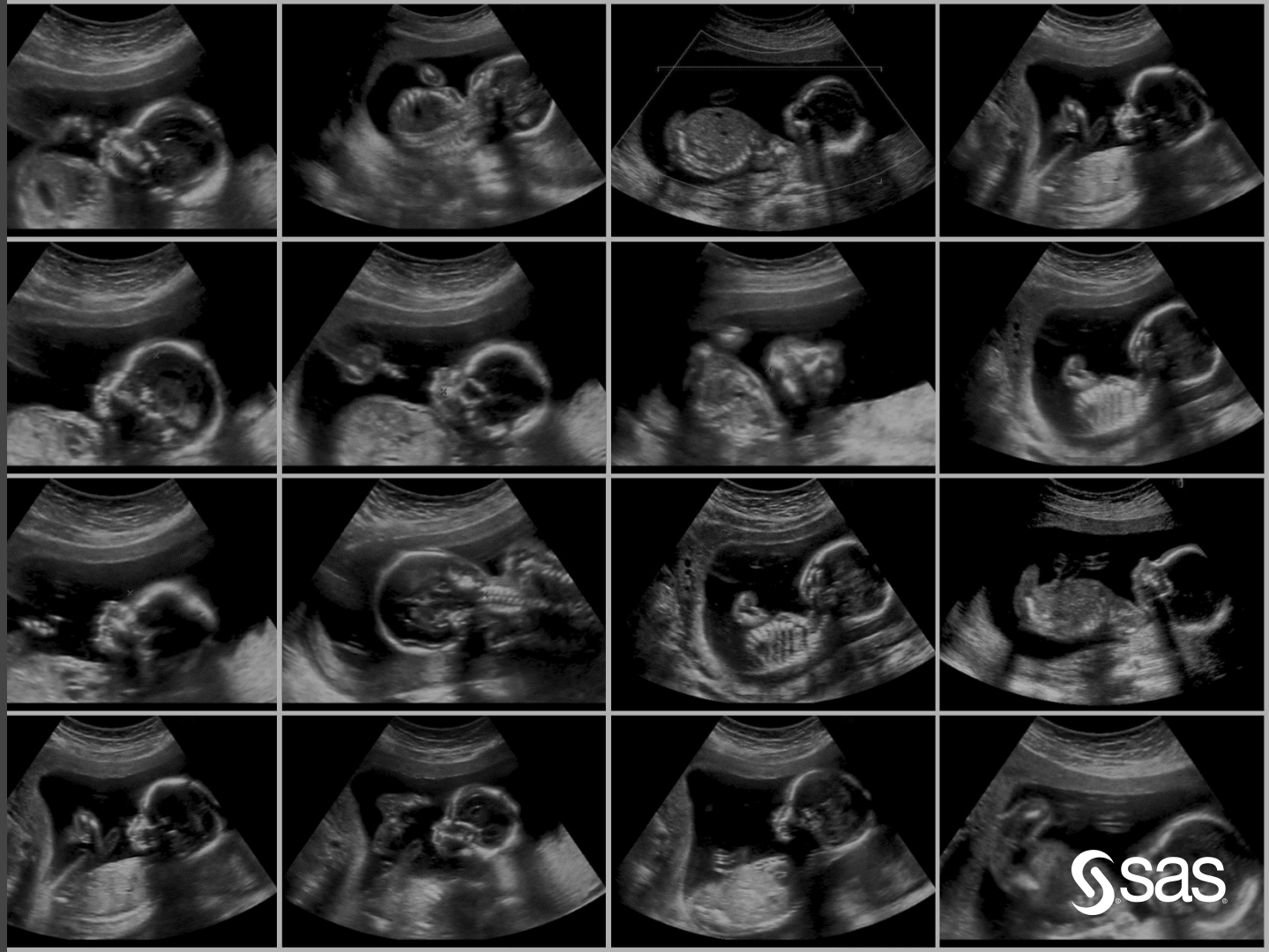
estimated new cases of cancer diagnosed in the US in 2018

SAS Computer Vision helps identify areas of concerns in liver and brain of cancer patients

Proper oxygenation

of the placenta is essential for a successful birth

Computer Vision makes it possible to monitor oxygenation through the umbilical cord



Thank You!

sas.com

THE POWER TO KNOW.