Deep Learning & SNOMED CT

Dr Chris Bates

Director of Research & Analytics, TPP

Visiting Research Fellow, University of Leeds, UK

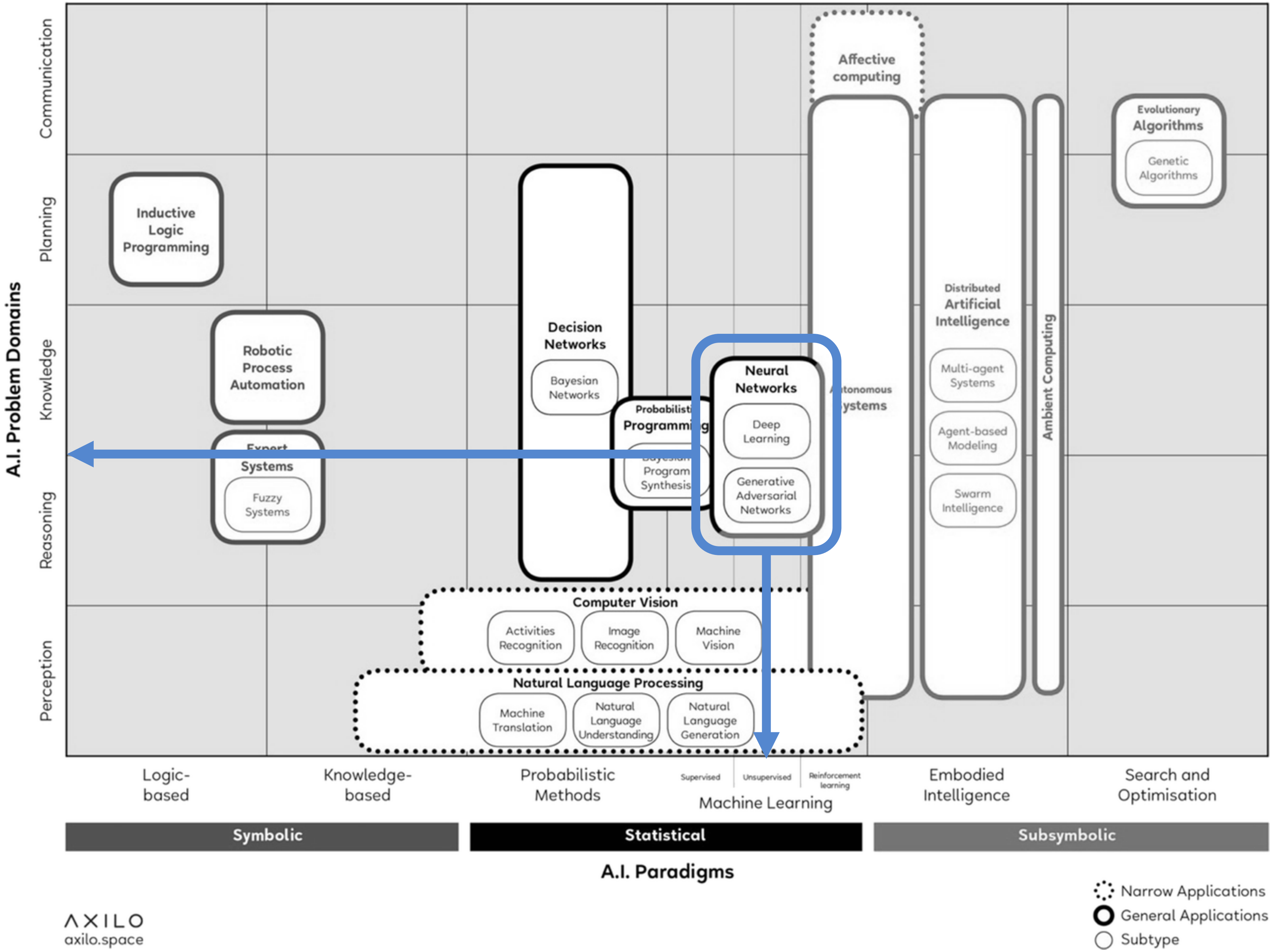
1. AI & Deep Learning

2. Role of EHRs & SNOMED CT

3. Challenges & Opportunities

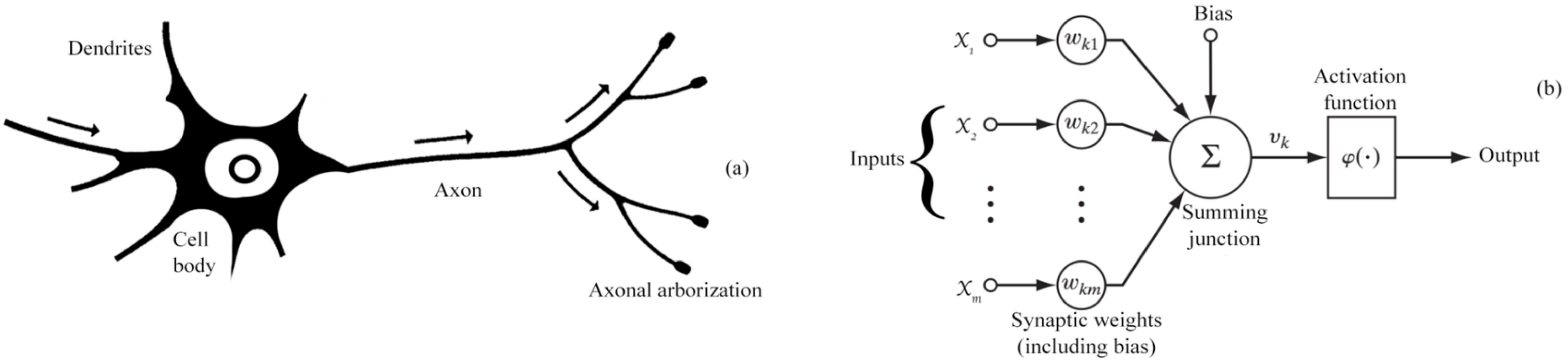
“For the present purpose, the artificial intelligence problem is taken to be that of making a machine behave in ways that would be called **intelligent if a human were so behaving**.”

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A proposal for the Dartmouth summer research project on artificial intelligence, August 31, 1955. *AI Magazine, 27*(4), 12.

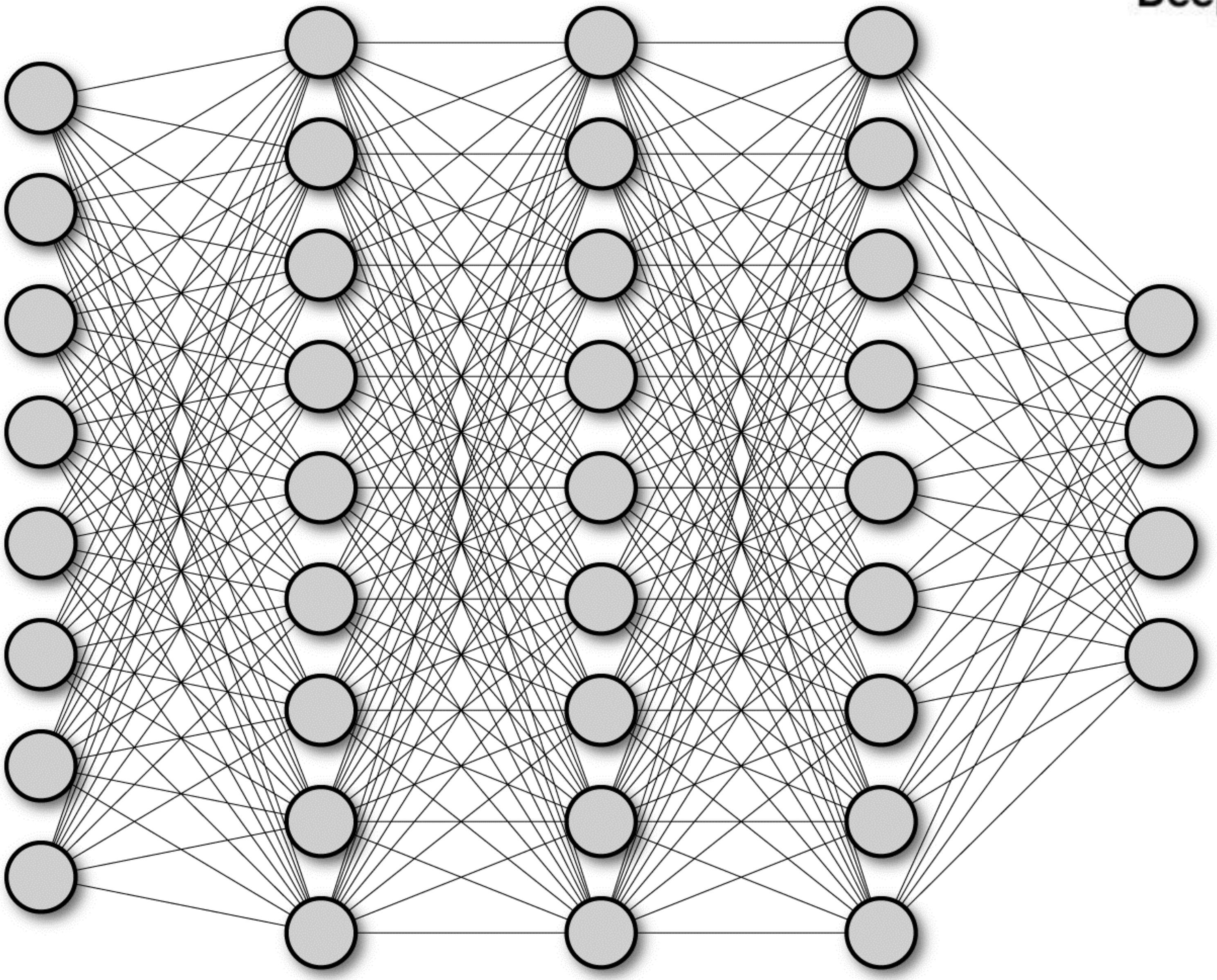


AI knowledge map (AIKM), source: (Corea Aug. 29, Corea 2018)

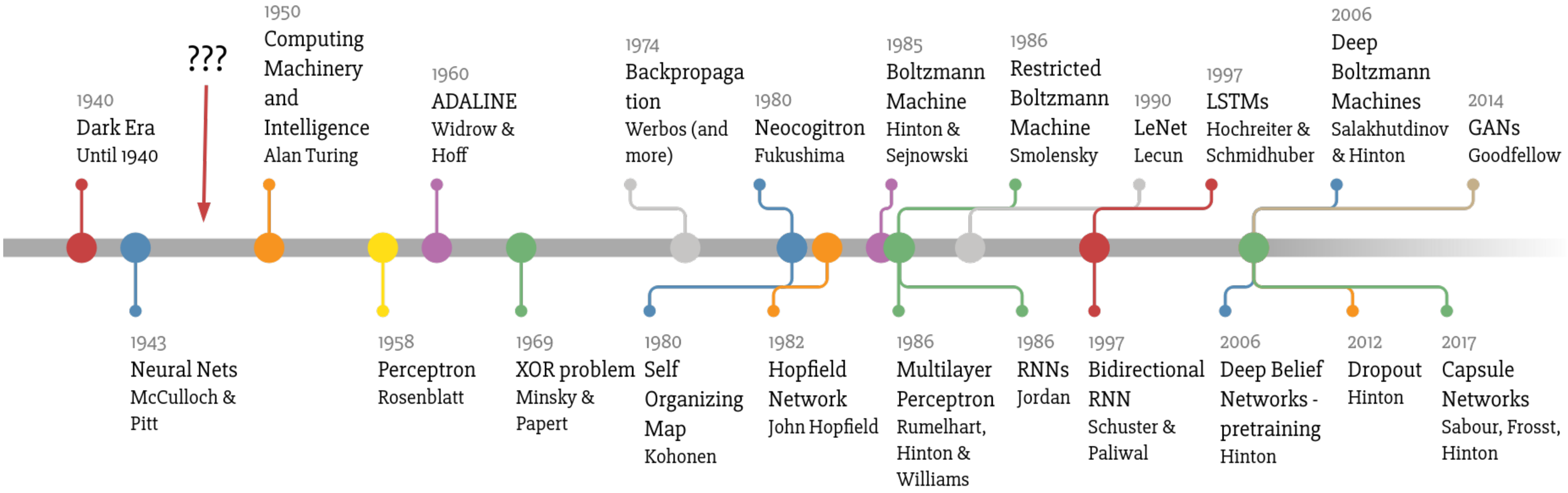
Biological Neuron versus Artificial Neural Network

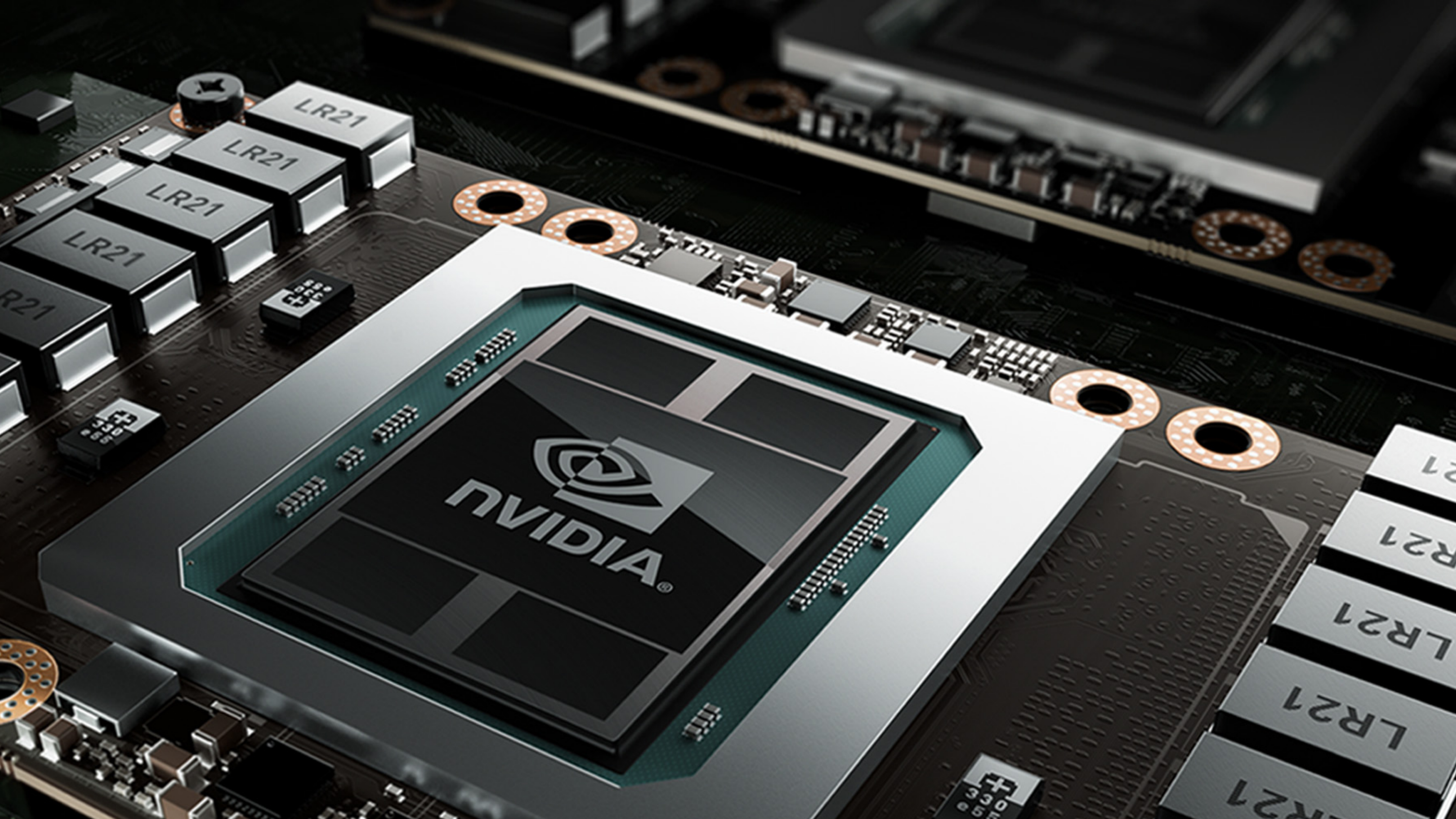


Deep Neural Network



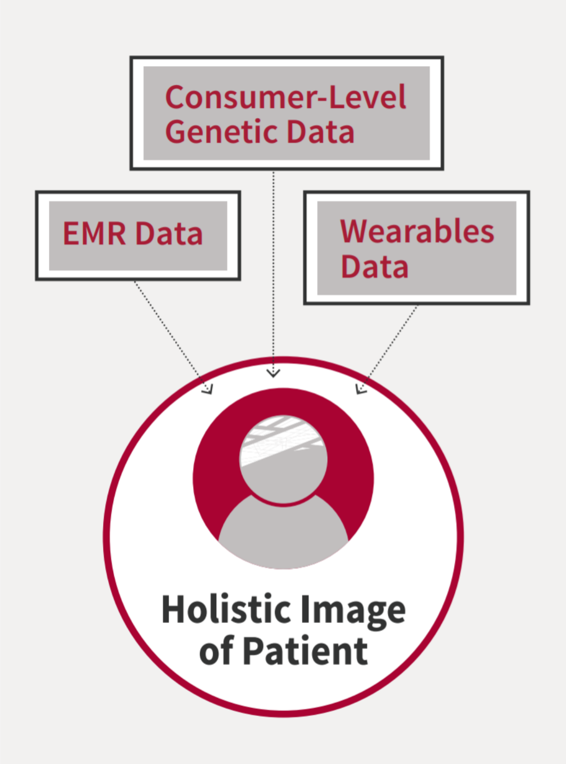
Deep Learning Timeline

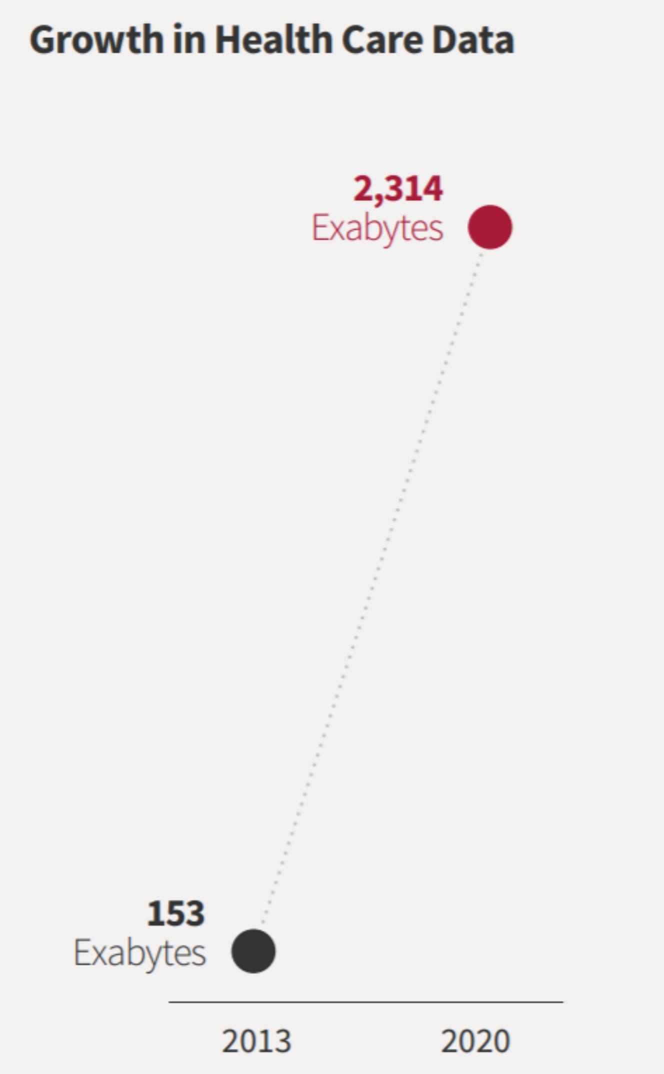



LR21
LR21
LR21
LR21
NVIDIA®

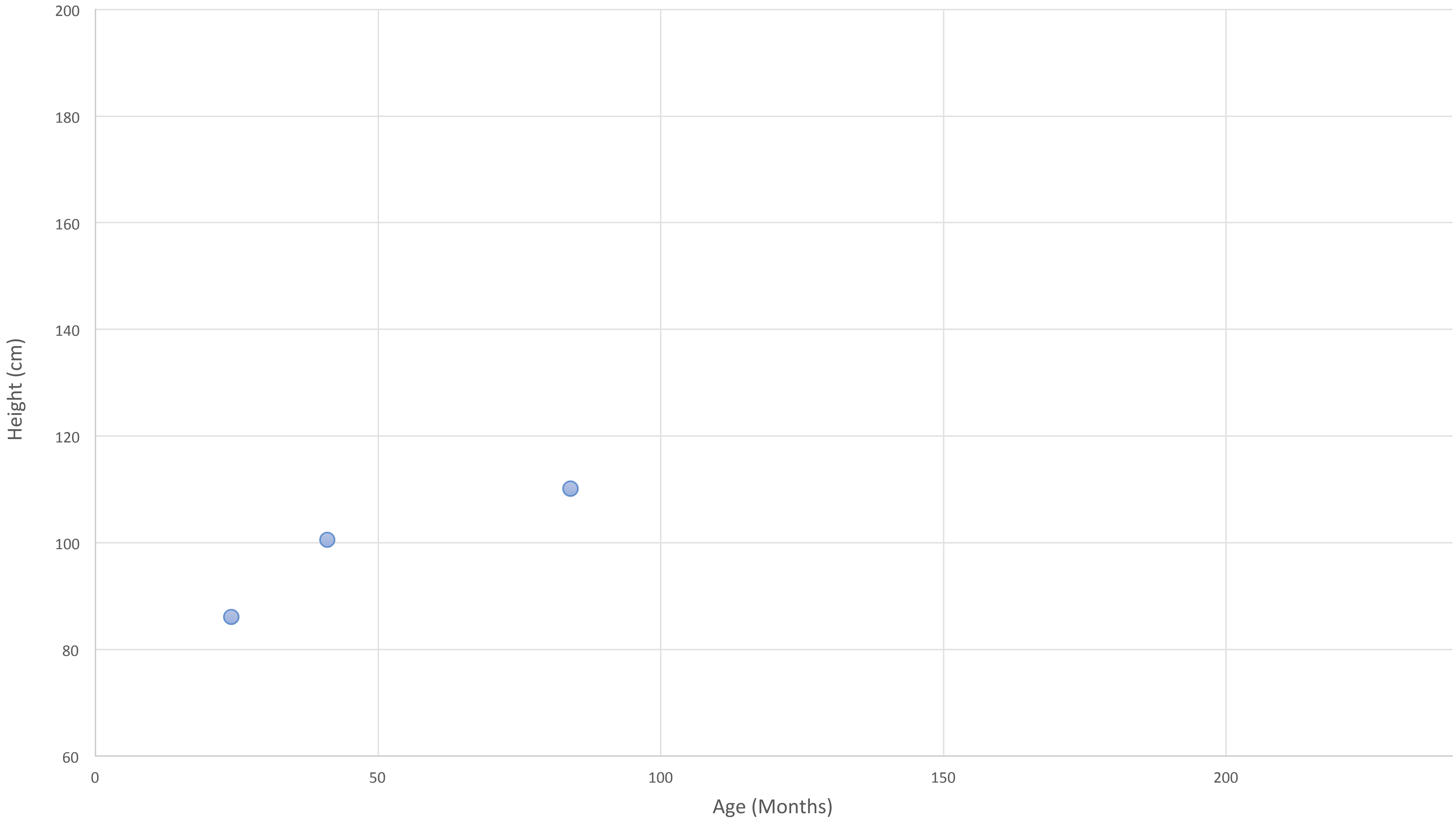
Data's Impact on Health Care

2,314 exabytes will be produced in 2020 translating to an overall rate of increase at least 48 percent annually.[8]

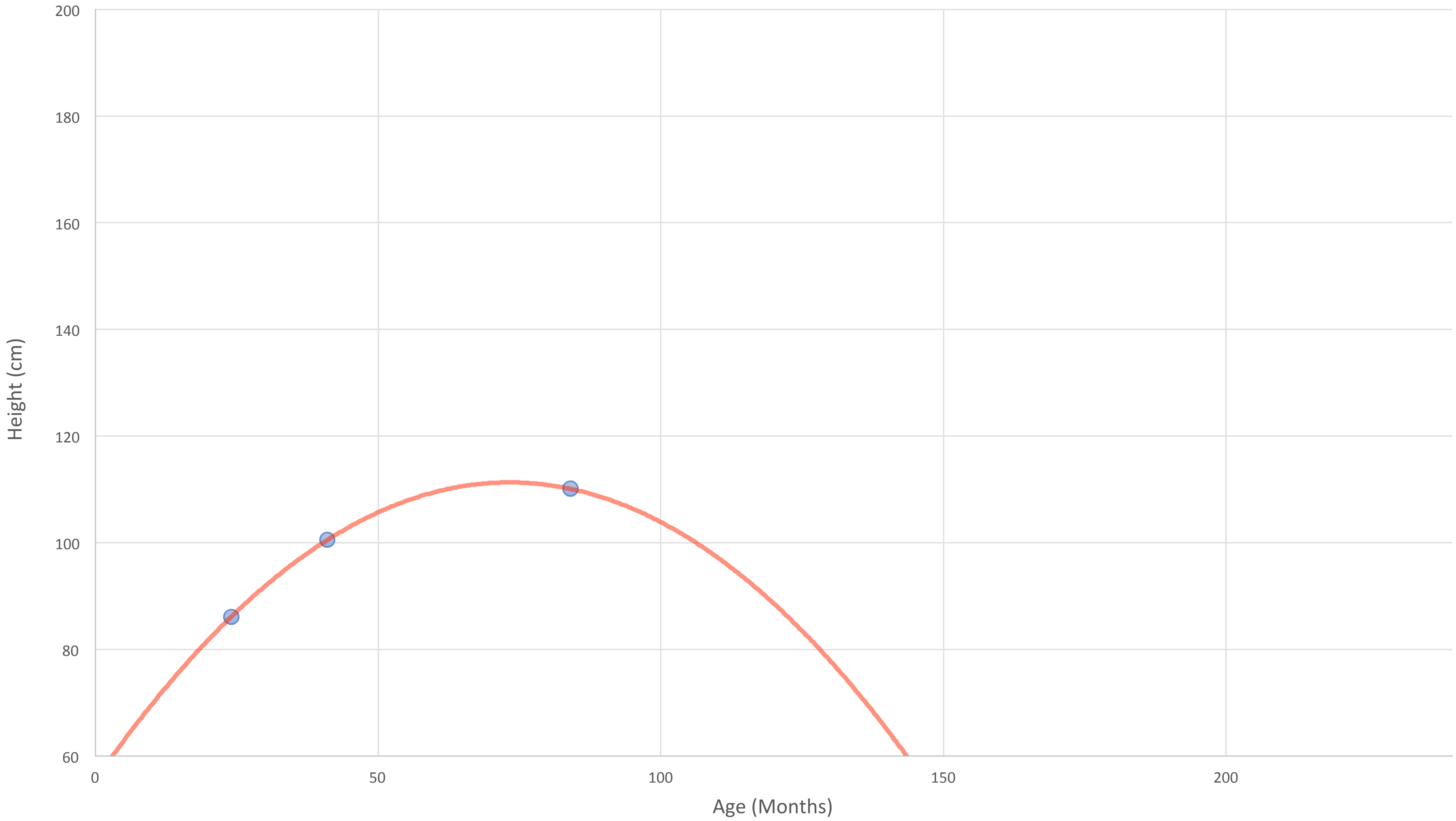




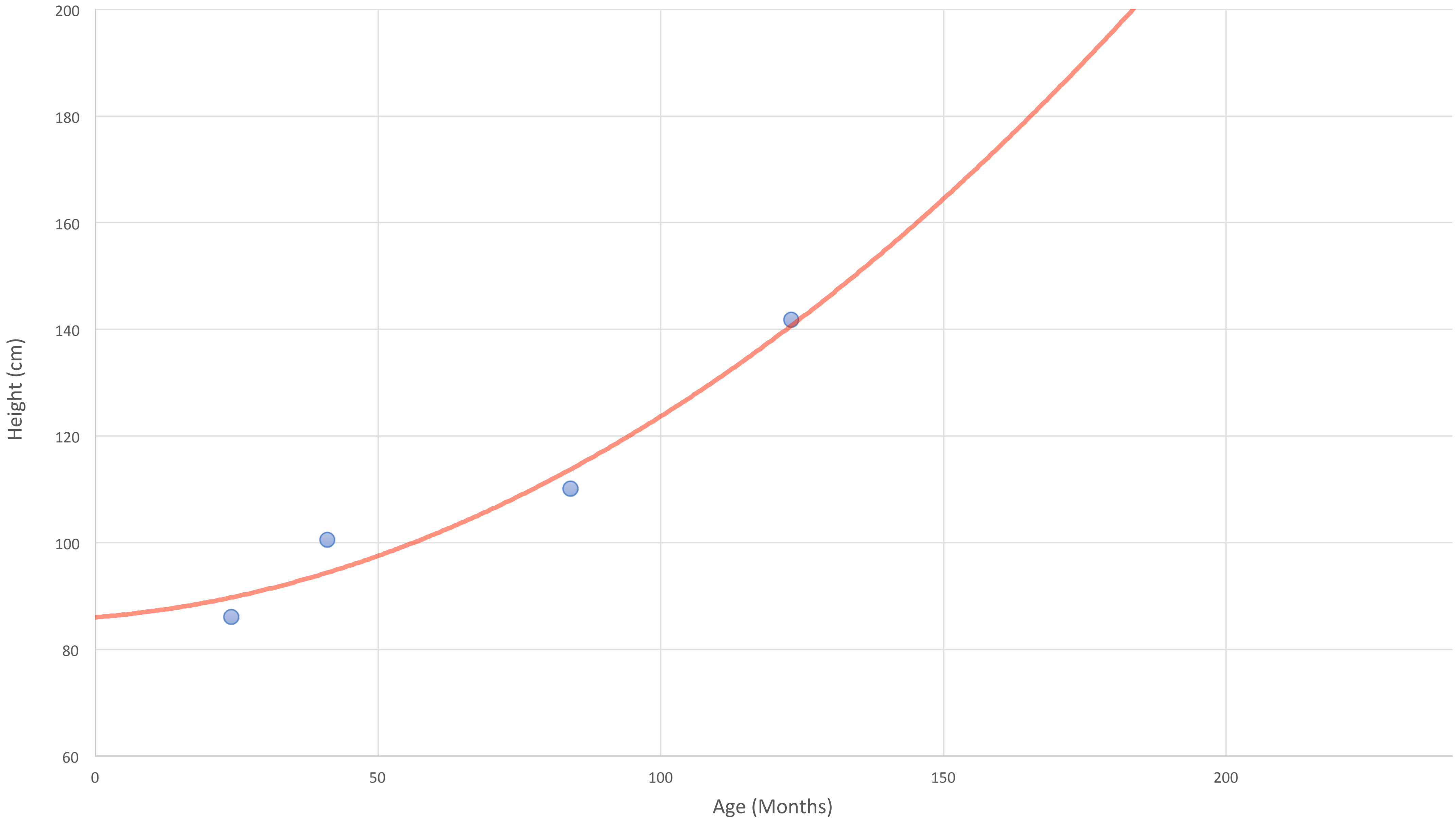
Source: International Data Corporation (IDC)

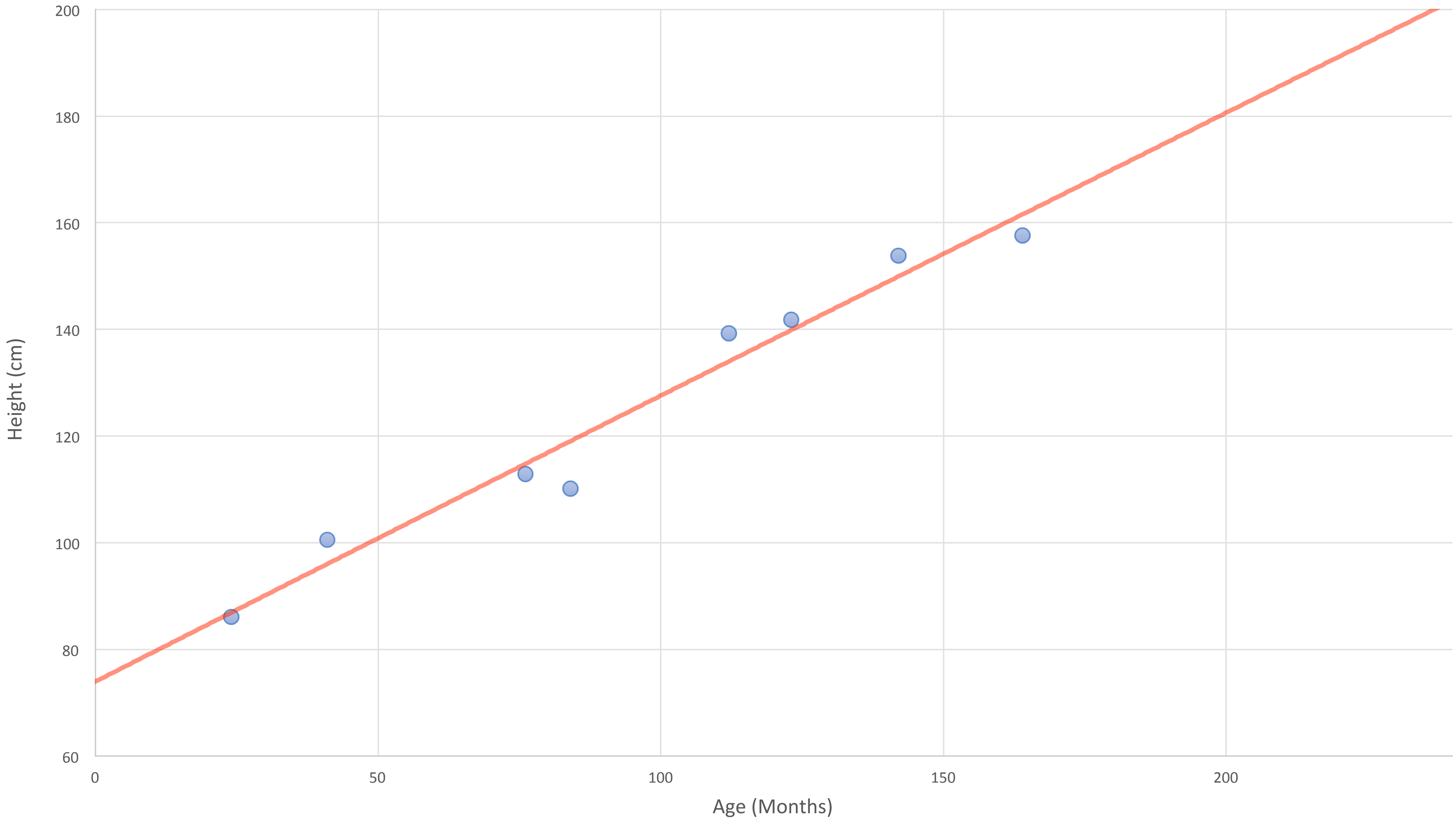


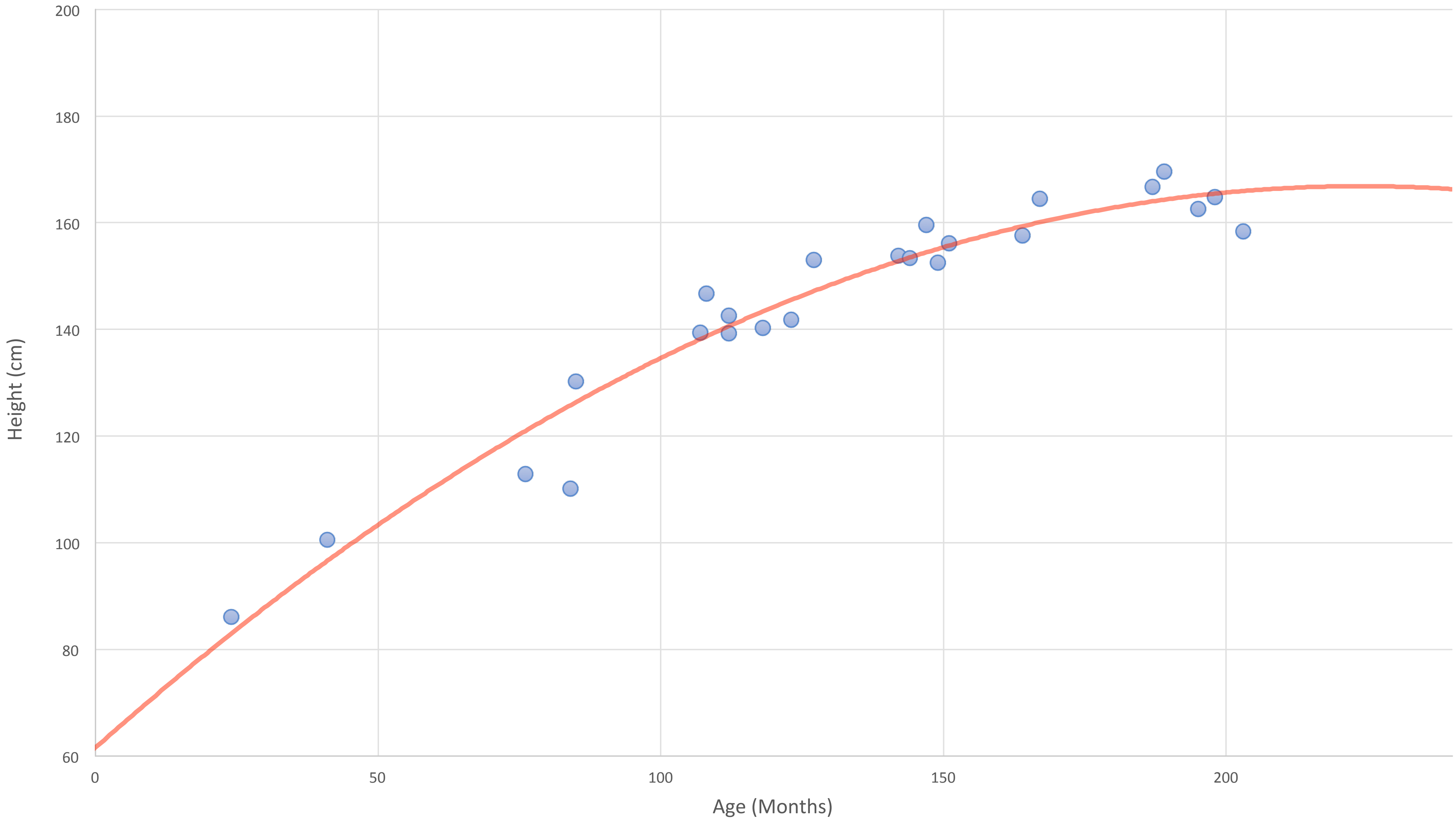
Height (cm)
200
180
160
140
120
100
80
60
0
50
100
150
200
Age (Months)



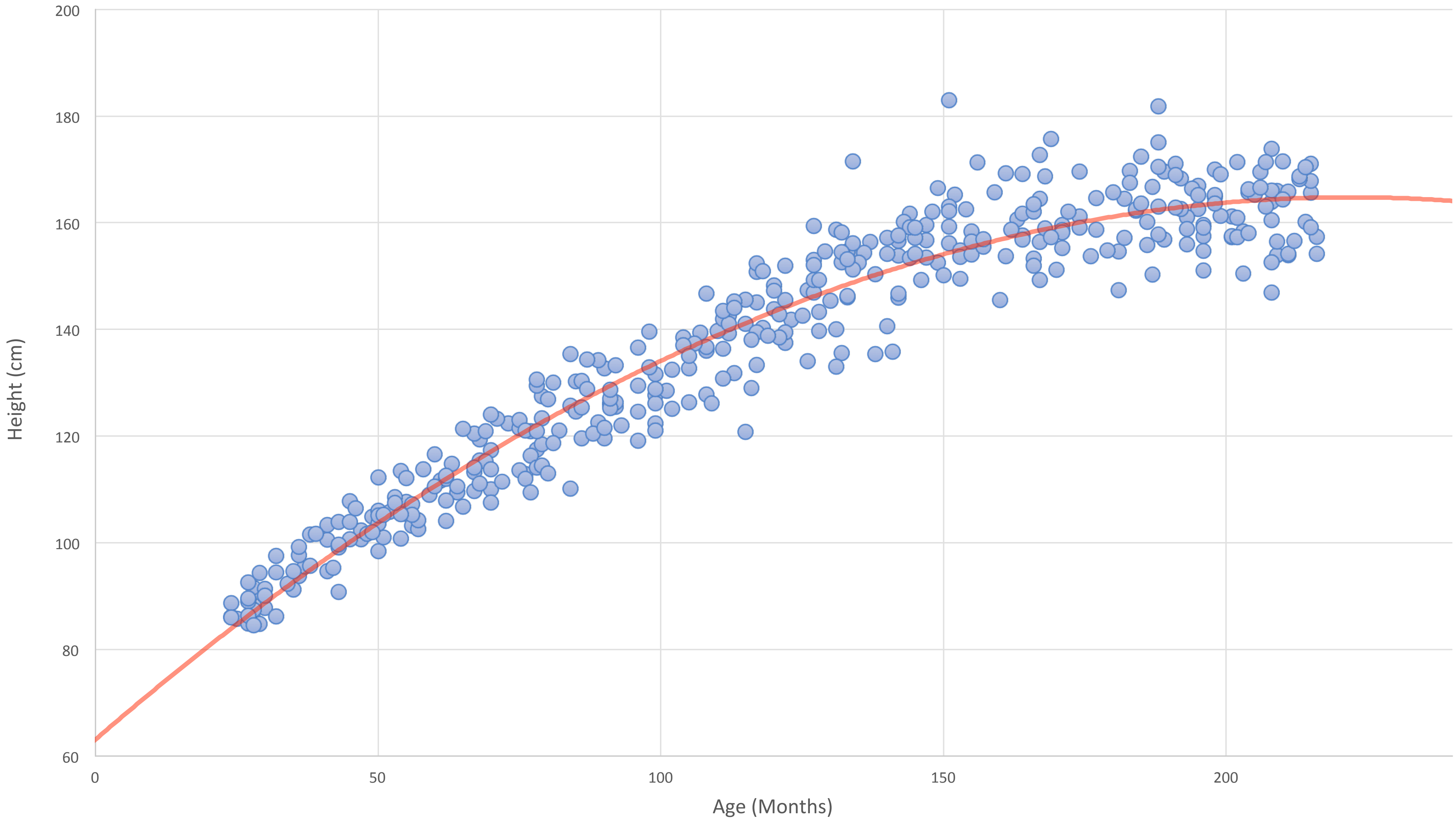
Height (cm)
200
180
160
140
120
100
80
60
0
50
100
150
200
Age (Months)


200
180
160
140
120
100
80
60
Height (cm)
0
50
100
150
200
Age (Months)

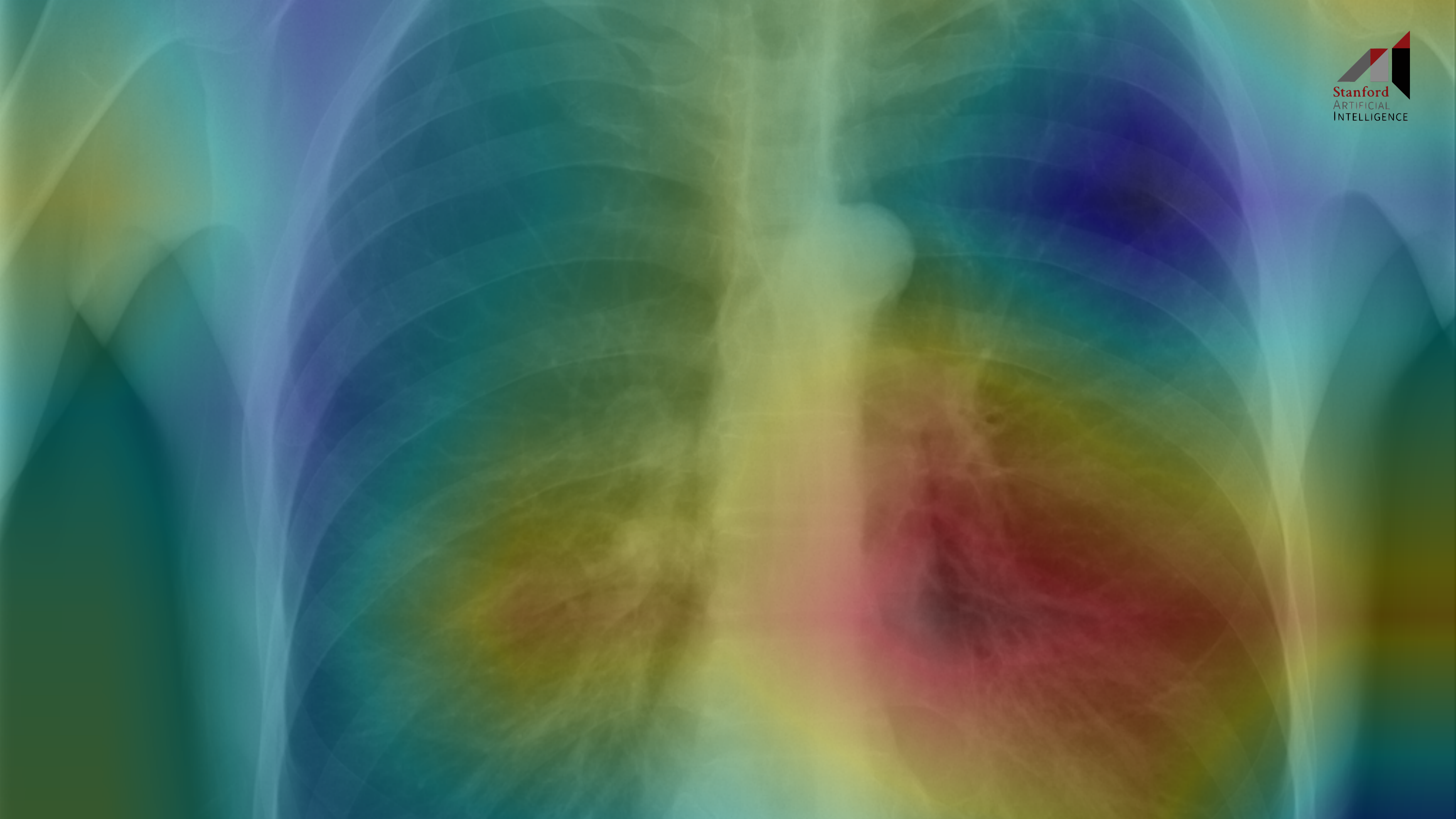

Height (cm)
200
180
160
140
120
100
80
60
0
50
100
150
200
Age (Months)



Height (cm)
200
180
160
140
120
100
80
60
0
50
100
150
200
Age (Months)


200
180
160
140
120
100
80
60
0
50
100
150
200
Height (cm)
Age (Months)


Stanford
ARTIFICIAL
INTELLIGENCE


Stanford
Artificial
Intelligence

nature biomedical engineering

ARTICLES
https://doi.org/10.1038/s41551-018-0195-0

Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning

Ryan Poplin[1,4], Avinash V. Varadarajan[1,4], Katy Blumer[1], Yun Liu[1], Michael V. McConnell[2,3], Greg S. Corrado[1], Lily Peng[1,4]* and Dale R. Webster[1,4]

Traditionally, medical discoveries are made by observing associations, making hypotheses from them and then designing and running experiments to test the hypotheses. However, with medical images, observing and quantifying associations can often be difficult because of the wide variety of features, patterns, colours, values and shapes that are present in real data. Here, we show that deep learning can extract new knowledge from retinal fundus images. Using deep-learning models trained on data from 284,335 patients and validated on two independent datasets of 12,026 and 999 patients, we predicted cardiovascular risk factors not previously thought to be present or quantifiable in retinal images, such as age (mean absolute error within 3.26 years), gender (area under the receiver operating characteristic curve (AUC) = 0.97), smoking status (AUC = 0.71), systolic blood pressure (mean absolute error within 11.23 mmHg) and major adverse cardiac events (AUC = 0.70). We also show that the trained deep-learning models used anatomical features, such as the optic disc or blood vessels, to generate each prediction.

The Telegraph

3 SEPTEMBER 2019 • 7:45PM

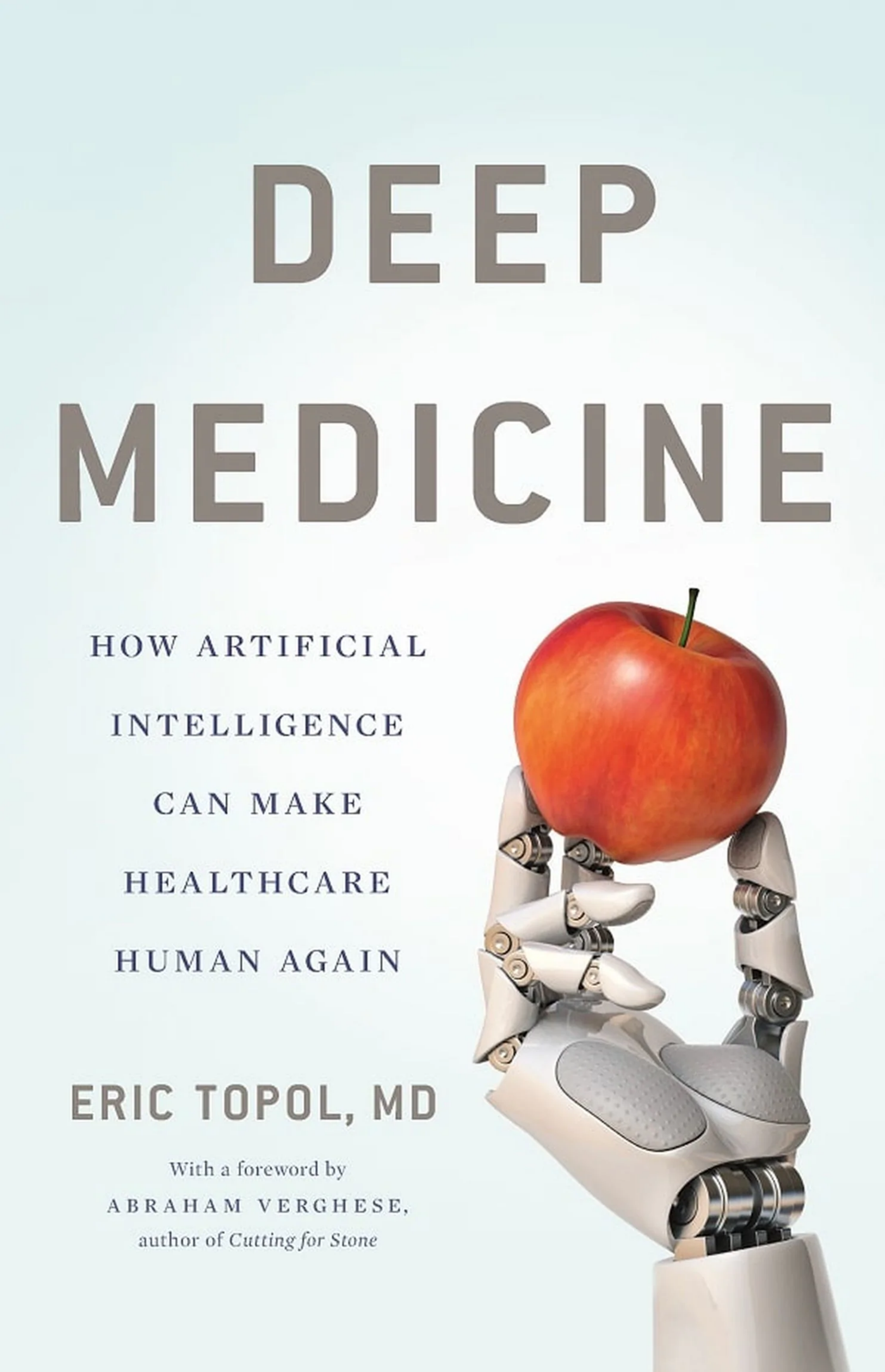
Scientists can predict risk of heart attack with 90 per cent accuracy

Scientists have found a way to predict the risk of heart attack with 90 per cent accuracy - almost a decade in advance.

The breakthrough by Oxford University uses artificial intelligence to look "beneath the surface" of routine CT scans and spot biomarkers which can give early red flags.

Currently, patients experiencing chest pains are sent for CT scans.

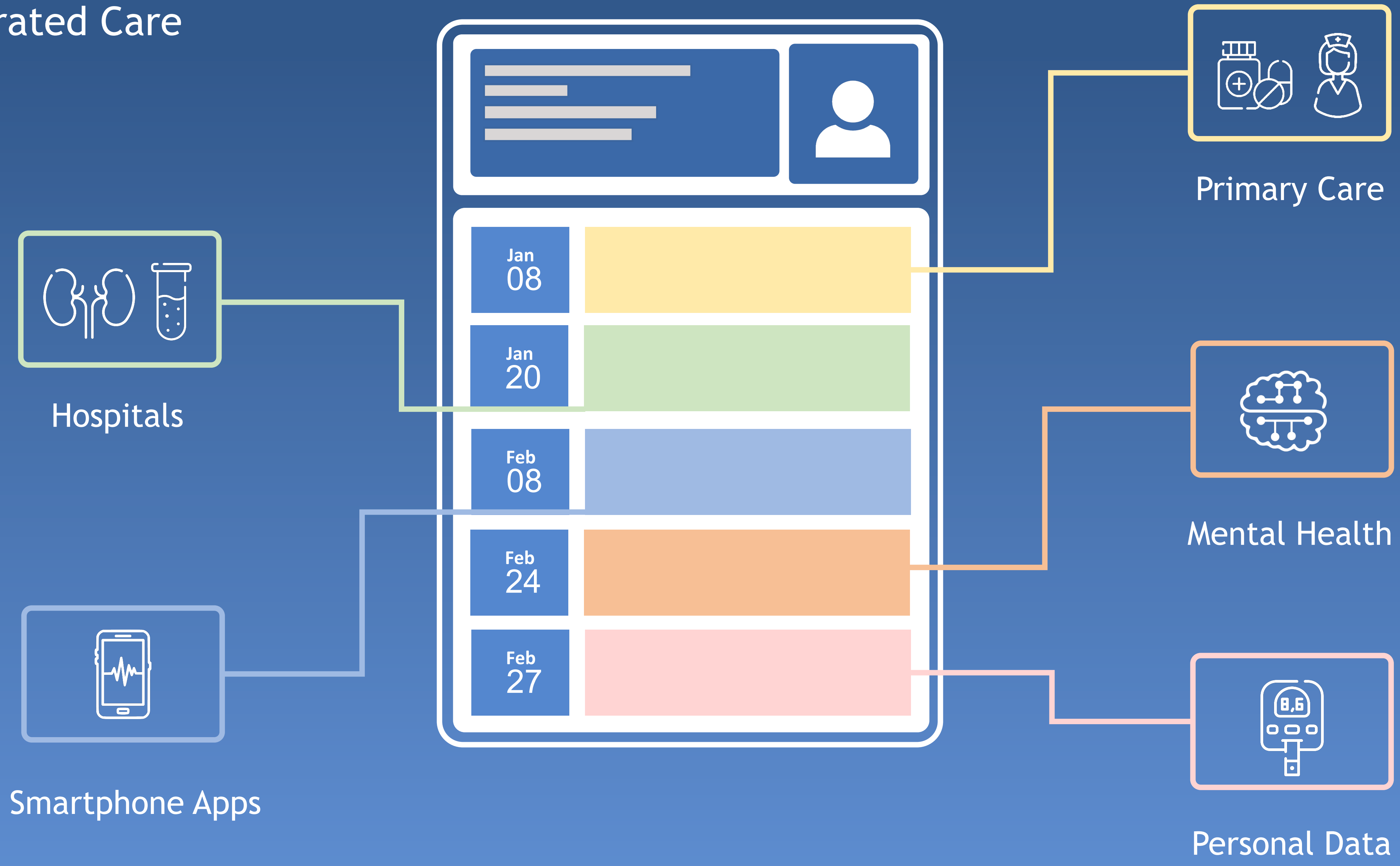
In around one quarter of cases, these show blockages which can be treated with surgery.



"To help bring this point home, for every **one hundred Medicare recipients** age sixty-five or older, each year there are more than **50 CT scans, fifty ultrasounds, fifteen MRIs, and ten PET scans**. It's estimated that 30 to 50 percent of the 80 million CT scans in the United States are unnecessary".

Shallow Medicine, Chapter 2, Page 27

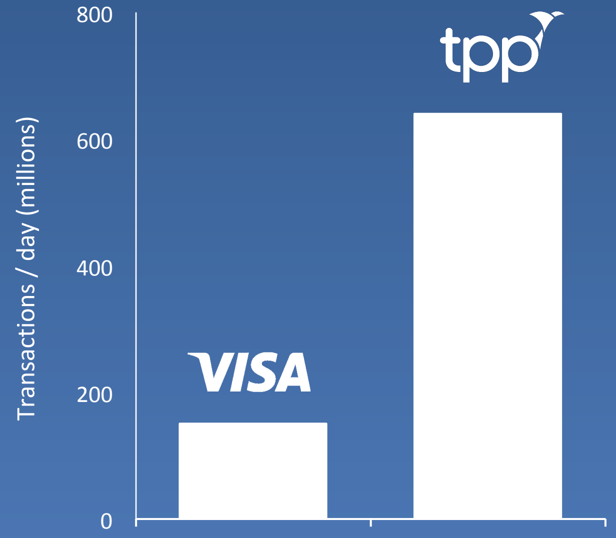
Digital Health & Integrated Care



50,000,000+
Electronic Patient Records

7,000+
NHS Organisations

200,000+
Doctors & Nurses



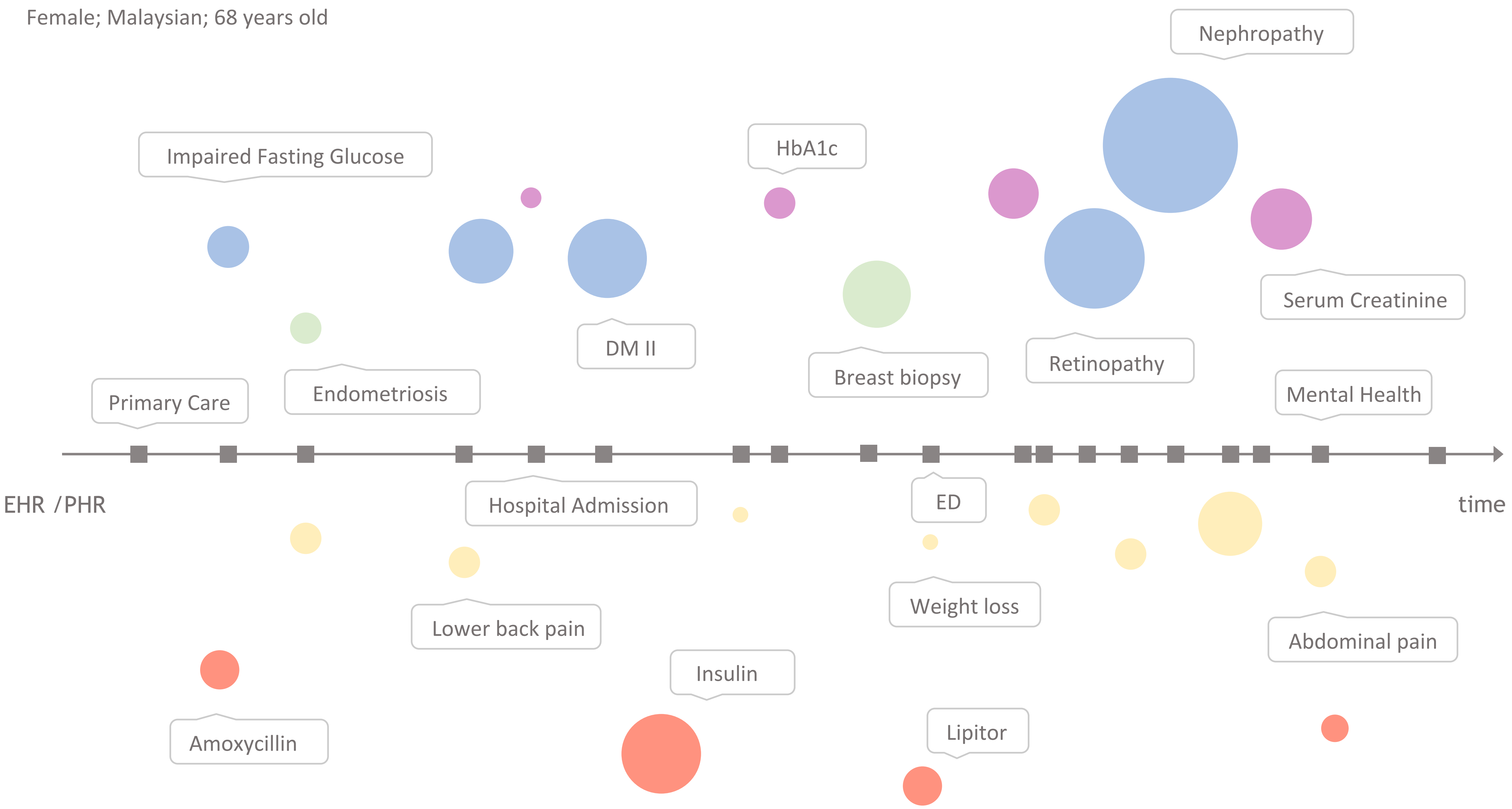
10,000
functions / second

114,000,000
Clinical events per month

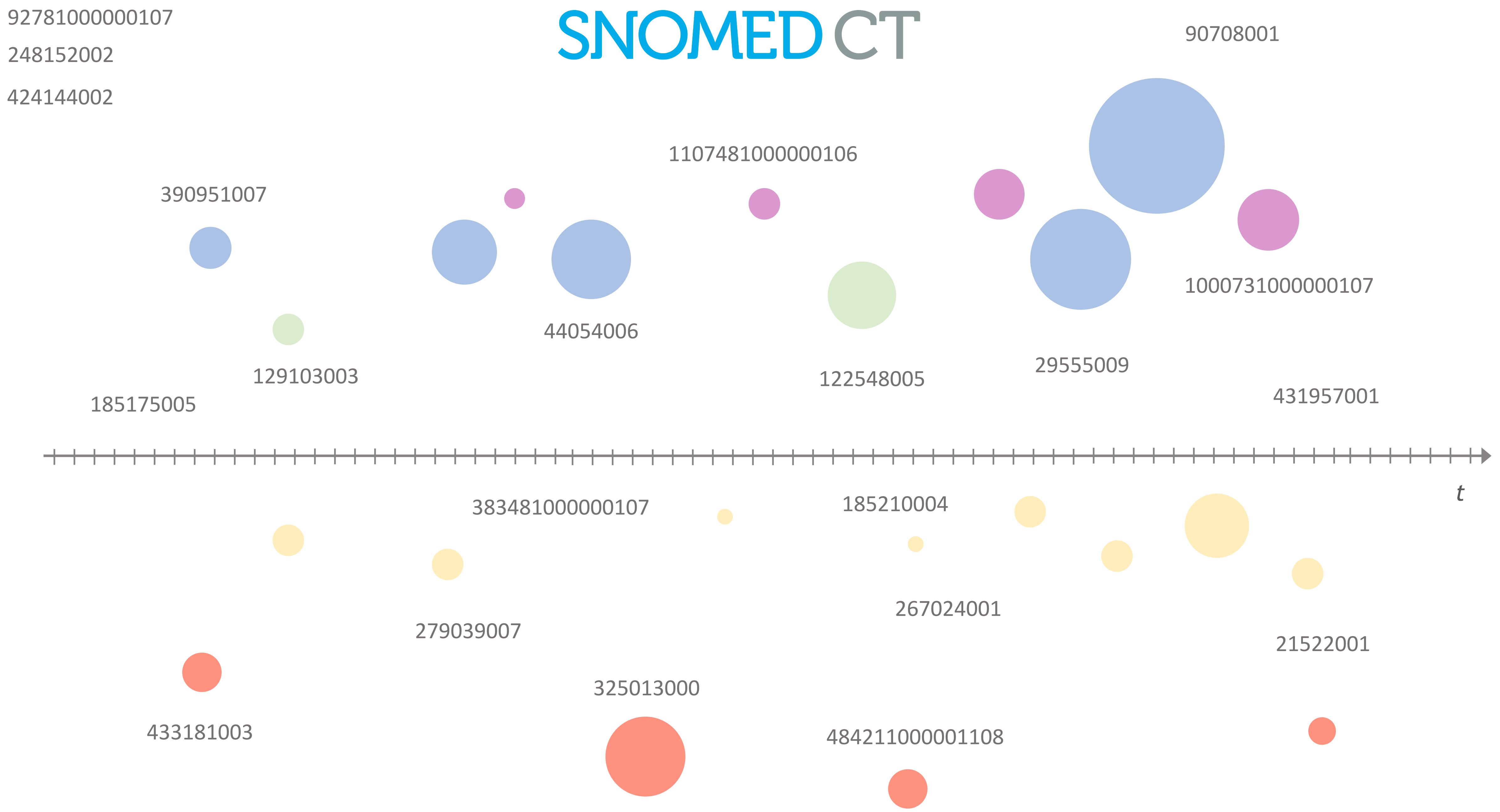
16,600,000,000
Total Coded Entries

SNOMED CT

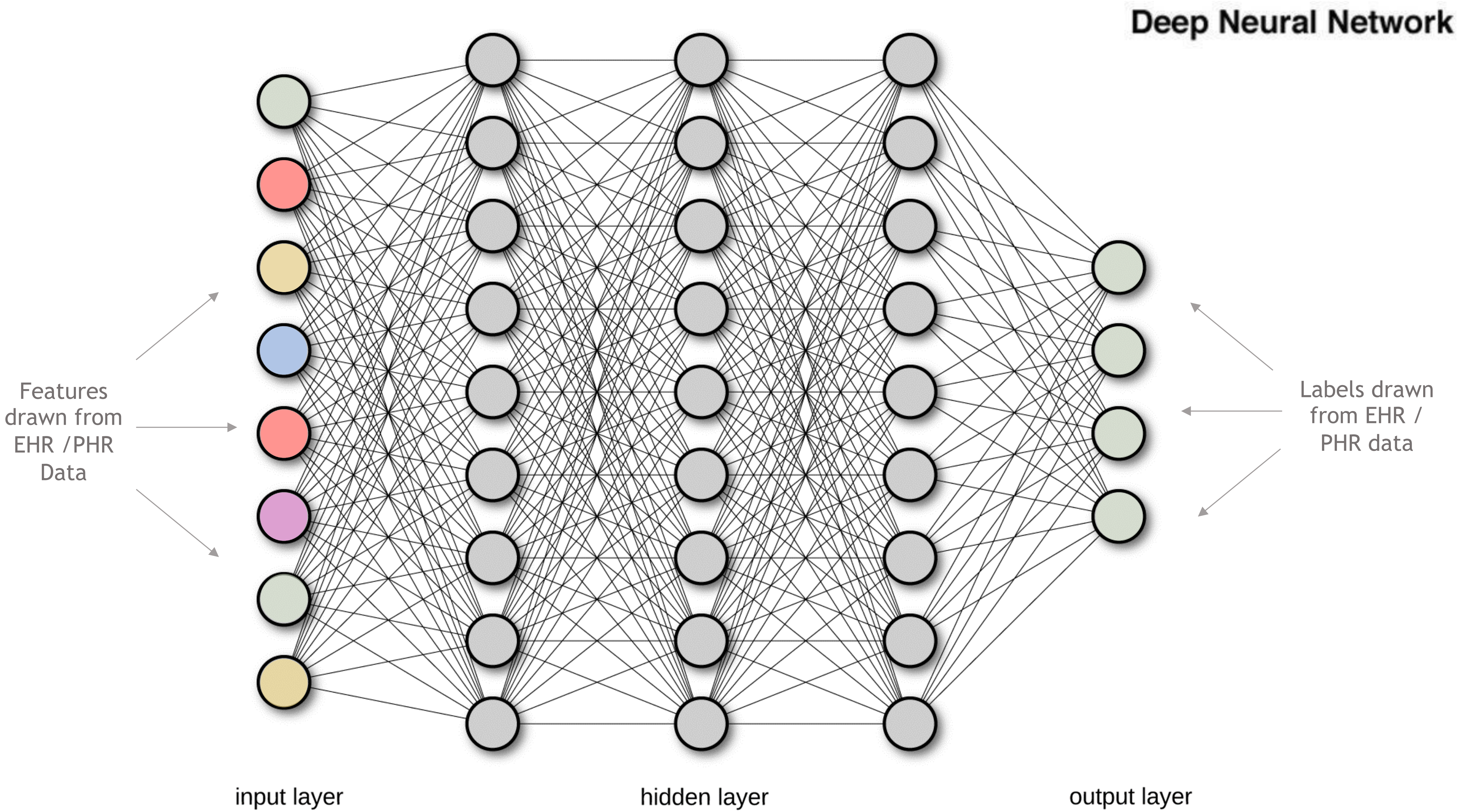
https://usa.visa.com/run-your-business/small-business-tools/retail.html



Female; Malaysian; 68 years old
Nephropathy
Impaired Fasting Glucose
HbA1c
Serum Creatinine
DM II
Breast biopsy
Retinopathy
Primary Care
Endometriosis
Mental Health
EHR /PHR
Hospital Admission
ED
time
Lower back pain
Weight loss
Abdominal pain
Insulin
Lipitor
Amoxycillin



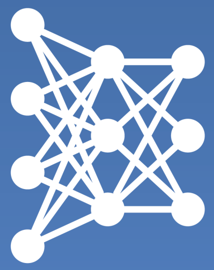
92781000000107
248152002
424144002
SNOMED CT
90708001
1107481000000106
390951007
44054006
1000731000000107
129103003
122548005
29555009
185175005
431957001
t
383481000000107
185210004
267024001
279039007
21522001
325013000
433181003
484211000001108


Deep Neural Network
Features drawn from EHR /PHR Data
Labels drawn from EHR / PHR data
input layer
hidden layer
output layer

DNNs: Early cancer detection

23% of all cancer diagnosis are made in urgent care settings

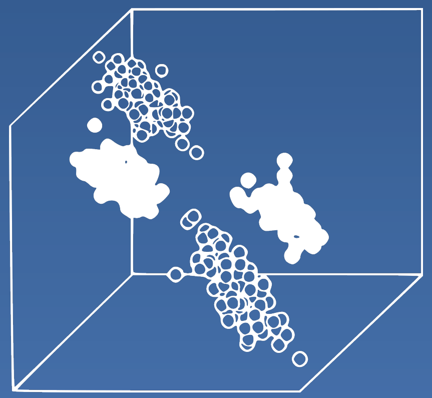
60% of ovarian cancer diagnoses at stage III or IV



Clinical assistance / Safety-netting

50% Detection before GP

Unsupervised Learning: Diabetes II



Clusters of patients, based on HbA1c history & co-morbidities

Different odd-ratios for complications in each cluster

Personalised targets & care plans

Improving Elderly Care

SNOMED CT

Electronic Frailty Index

An algorithm to detect frailty in elderly individuals based on comprehensive electronic health record data

NHS England · UNIVERSITY OF BIRMINGHAM · UNIVERSITY OF LEEDS

NICE National Institute for Health and Care Excellence

70% Frail elderly citizens with a targeted medication review

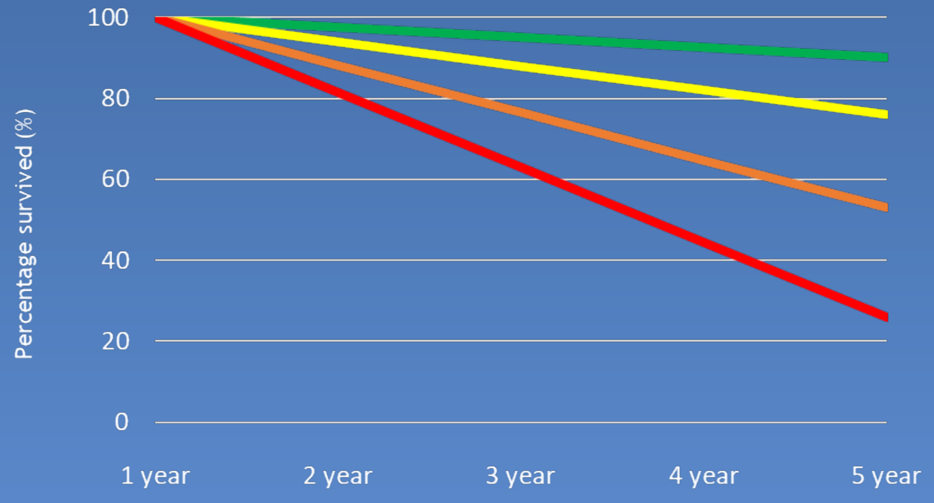
30% Geriatric assessment based on eFI screening programme

Population Health

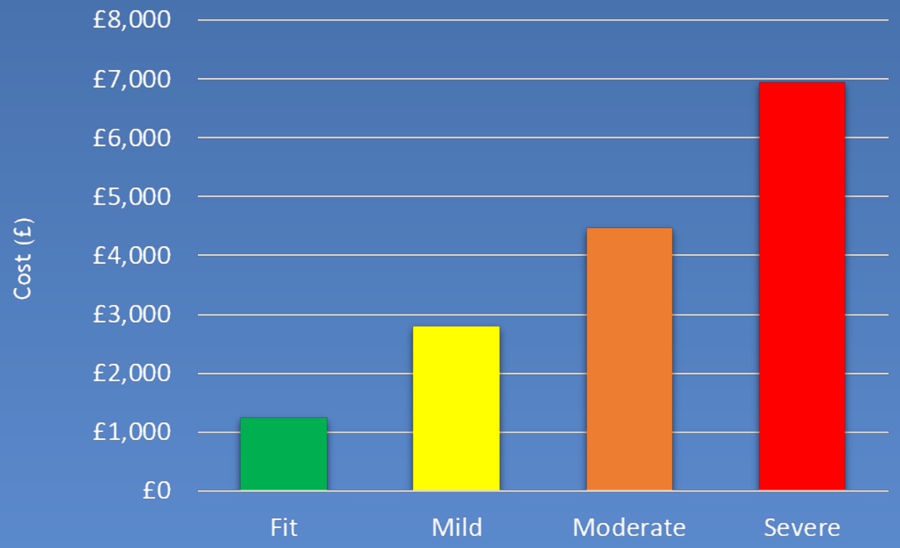
Surgery

Oncology

Five-year Kaplan-Meier survival curve



Average healthcare cost for elderly citizens



Bates, Clegg et al, Development and validation of an electronic frailty index using routine primary care electronic health record data, Age and Ageing, Volume 45, Issue 3, May 2016

tpp

Rare Diseases

Ataxia Telangiectasia

0.002%

Paget's Disease

0.31%

Rett Syndrome

0.007%

All Rare Diseases

10%

Unsupervised learning approaches to case identification

HL7 FHIR

Process & Optimisation

Appointment non-attendance

Appointment length

Delayed discharges

Intelligent triage

Medication errors

SNOMED CT

https://www.the -scientist.com/infographics/infographic-rare-disease-by-the-numbers-33339

Challenges / Opportunities

SNOMED CT

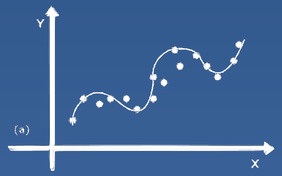
Bias

Iatrogenic risk

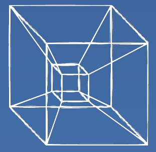
Local epidemiology

User interface / Context

Overfitting



High dimensionality



Validation

Data sets / Validation

Example: Kawasaki's Disease

Morbilliform eruption (disorder)

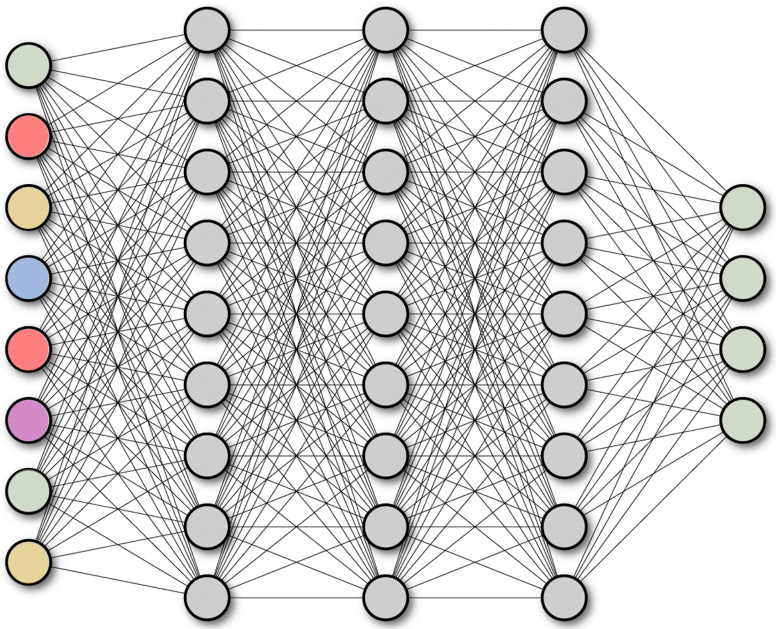
SCTID: 247470007

247470007 | Morbilliform eruption (disorder) |

- Morbilliform eruption
- Morbilliform rash
- Morbilliform eruption (disorder)

Associated morphology → Maculopapular rash
Finding site → Skin structure

Associated morphology → Cutaneous eruption
Finding site → Skin structure



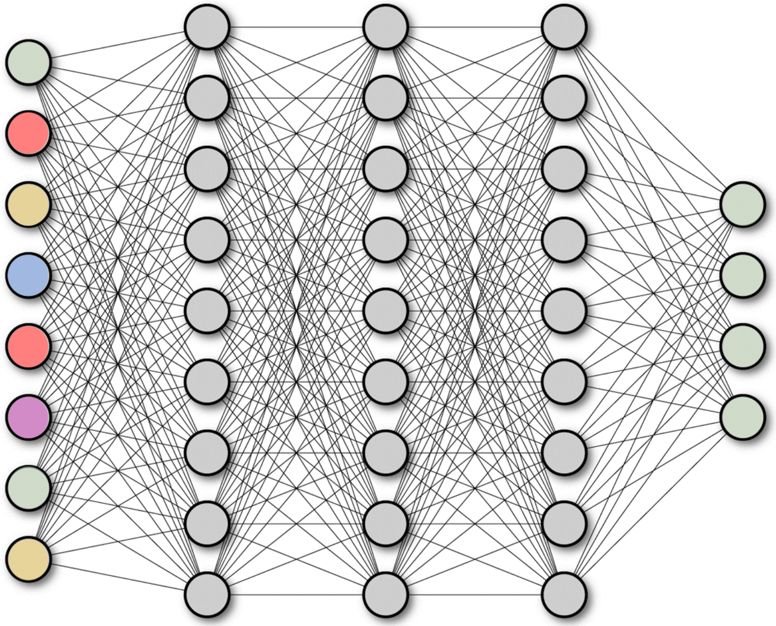
Acute febrile mucocutaneous lymph node syndrome (disorder)

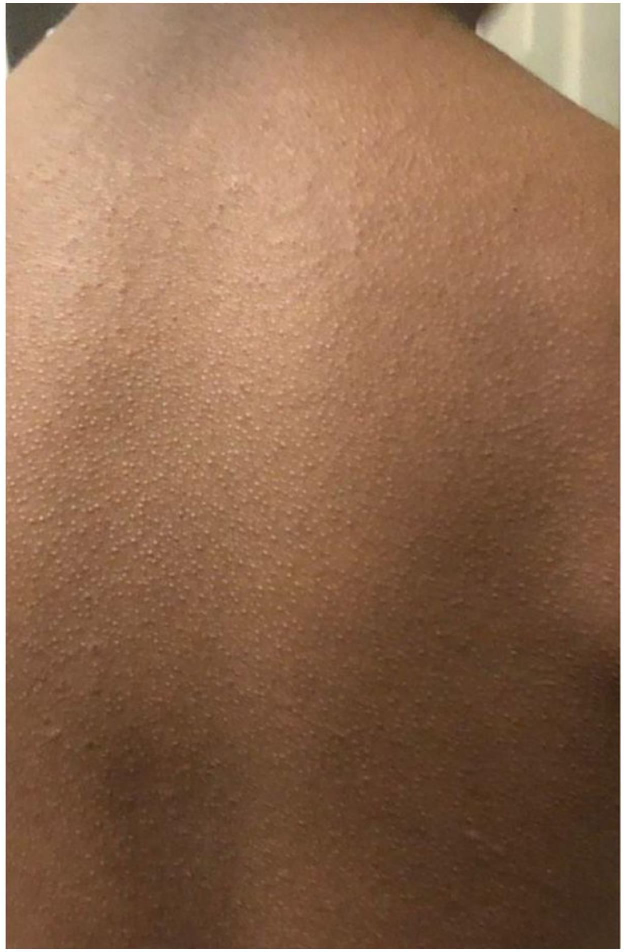
SCTID: 75053002

75053002 | Acute febrile mucocutaneous lymph node syndrome (disorder) |

- Acute febrile mucocutaneous lymph node syndrome
- MCLS
- Kawasaki's disease
- Kawasaki disease
- Mucocutaneous lymph node syndrome
- MLNS
- Kawasaki's syndrome
- Acute febrile mucocutaneous lymph node syndrome (disorder)
- Kawasaki syndrome
- Kawasakis mucocutaneous lymph node syndrome

Example: Kawasaki's Disease





@brownskinmatters

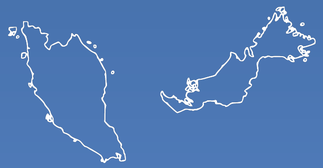
Challenges / Opportunities

SNOMED CT

Bias

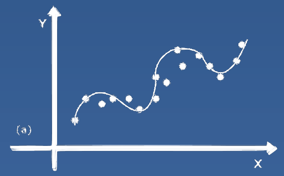
Iatrogenic risk

Local epidemiology

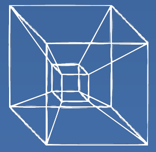


User interface / Context

Overfitting



High dimensionality



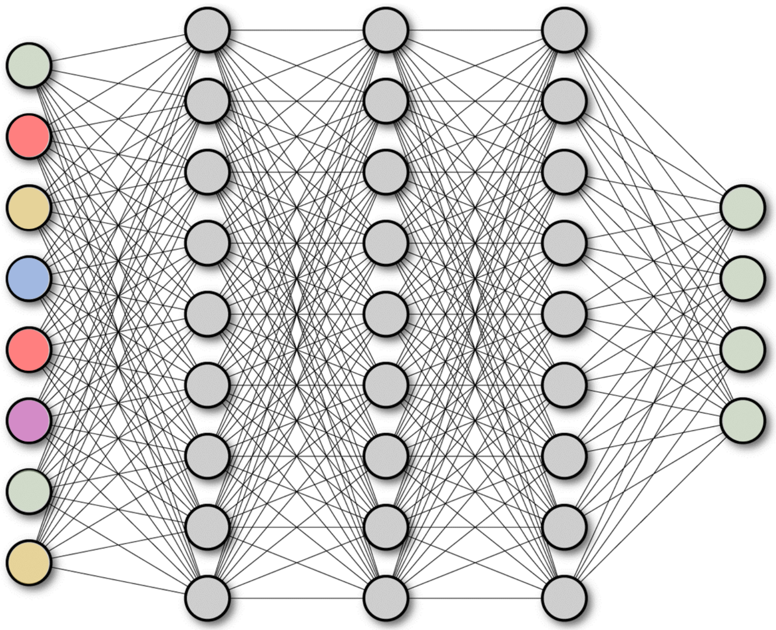
Validation

Data sets / Validation

Dimensionality

SNOMED CT

35,645 - findings



SNOMED CT

75,991 disorders

$2^{35,645}$ possibilities

$2^{33} = 8,589,934,592$

Challenges / Opportunities

SNOMED CT

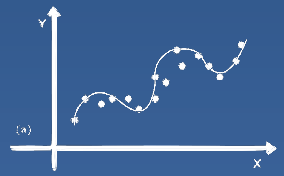
Bias

Iatrogenic risk

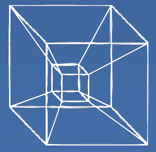
Local epidemiology

User interface / Context

Overfitting



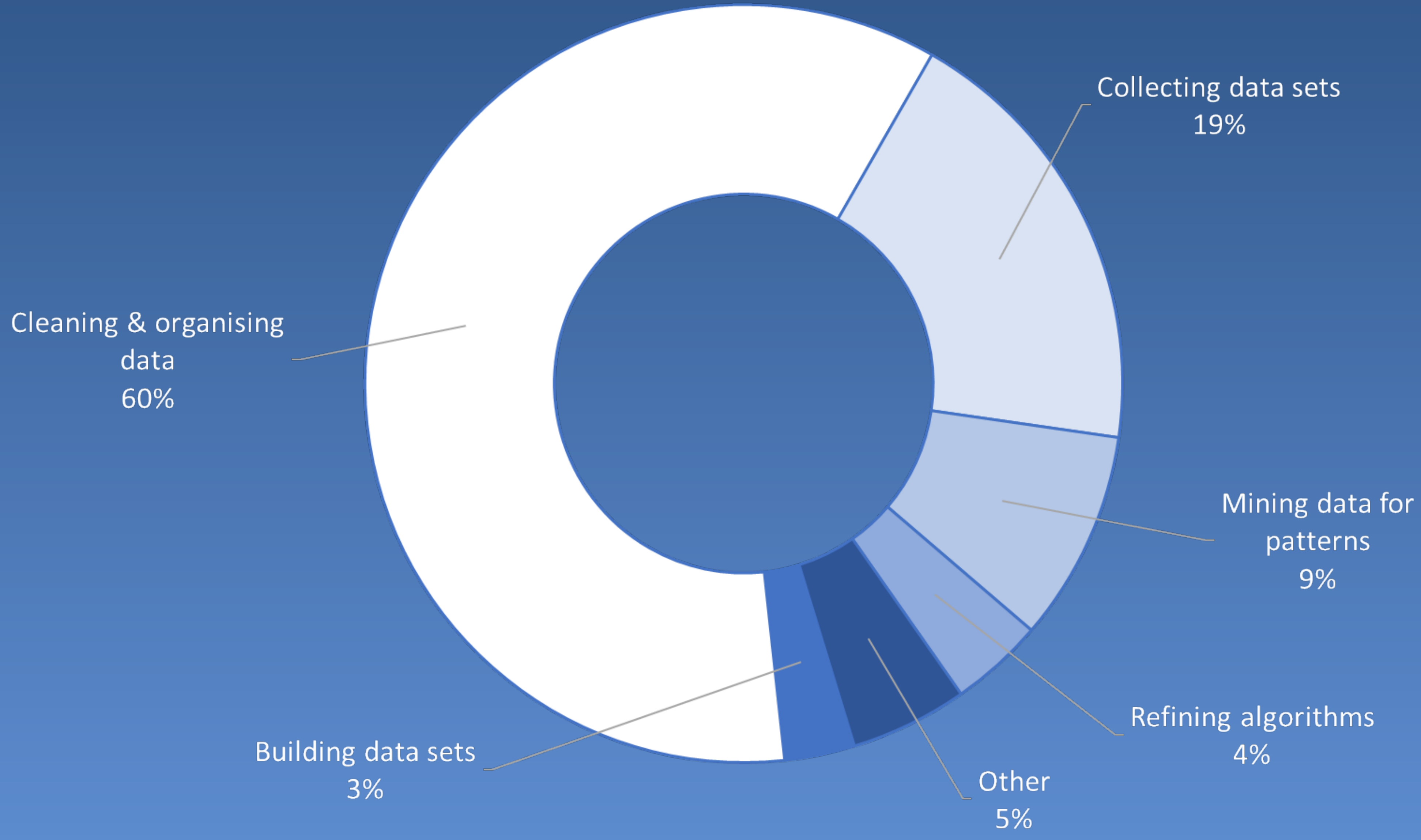
High dimensionality



Validation

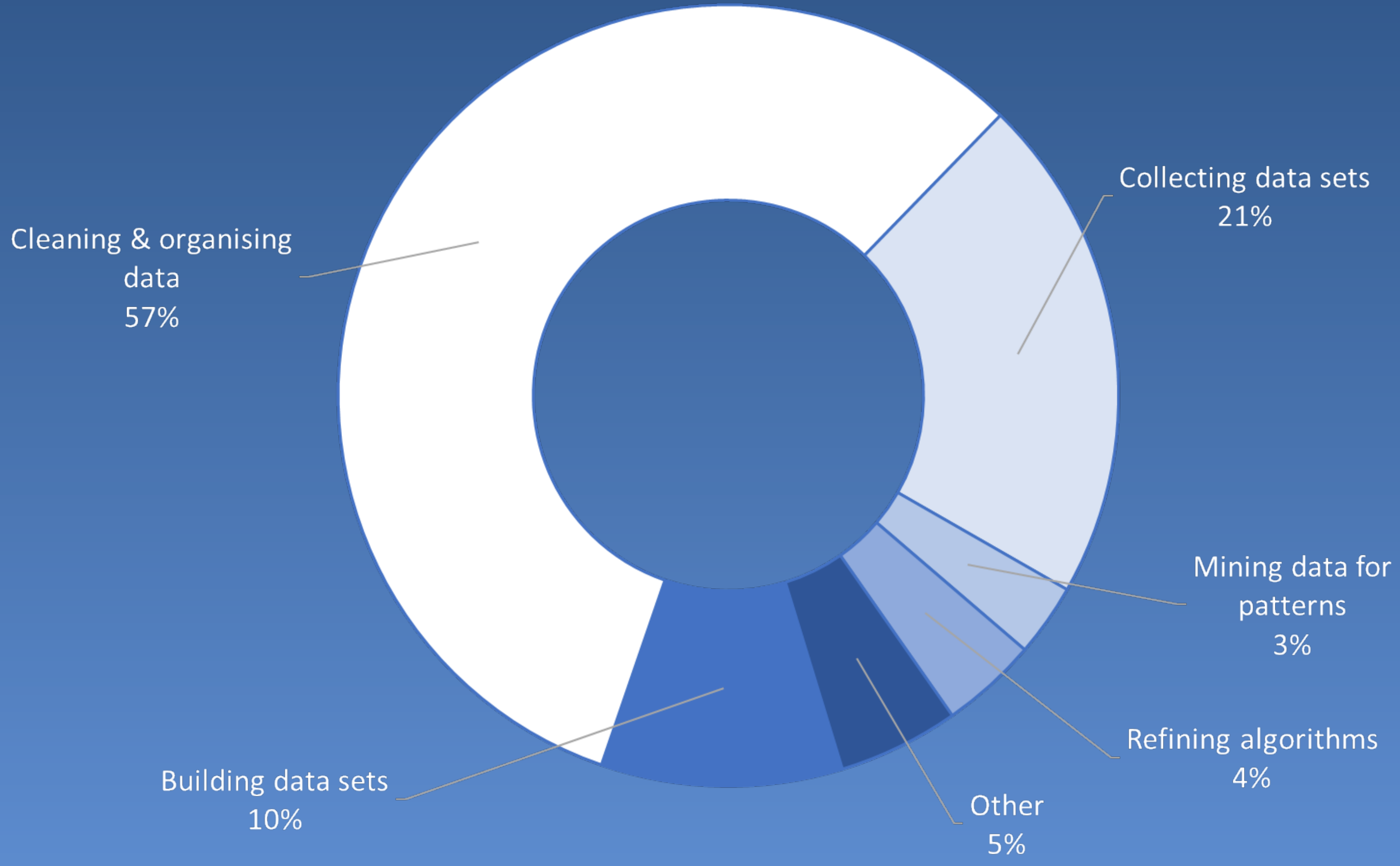
Data sets / Validation

What data scientists spend the most time doing



CrowdFlower 2016 Data Science Report

What's the least enjoyable part of data science?



CrowdFlower 2016 Data Science Report

Challenges / Opportunities

SNOMED CT

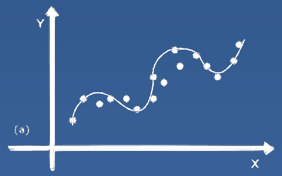
Bias

Iatrogenic risk

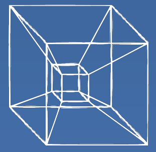
Local epidemiology

User interface / Context

Overfitting



High dimensionality



Validation

Data sets

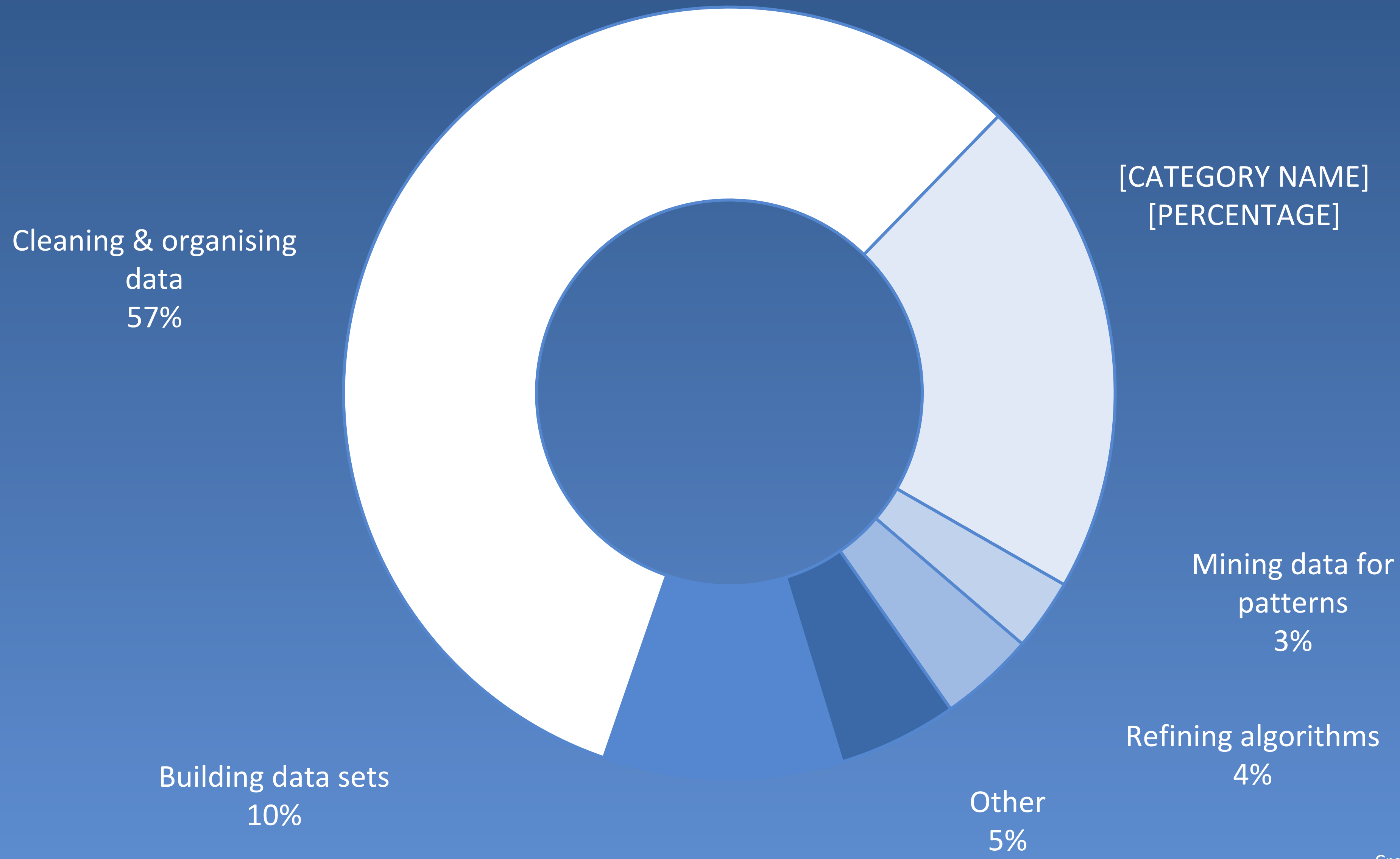
Thank you

tpp

Thank you

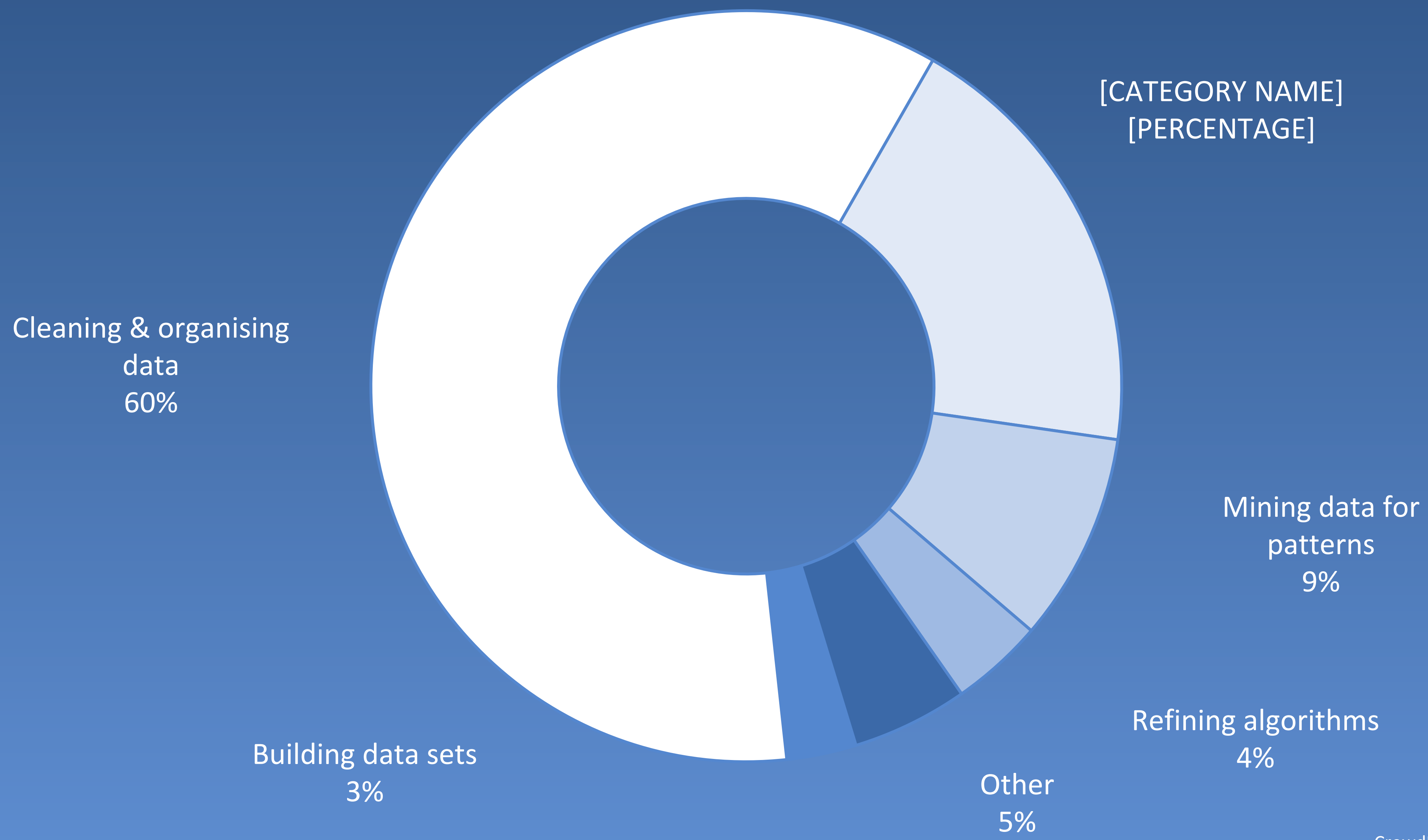
Thank you

What’s the least enjoyable part of data science?



CrowdFlower 2016 Data Science Report

What data scientists spend the most time doing



CrowdFlower 2016 Data Science Report