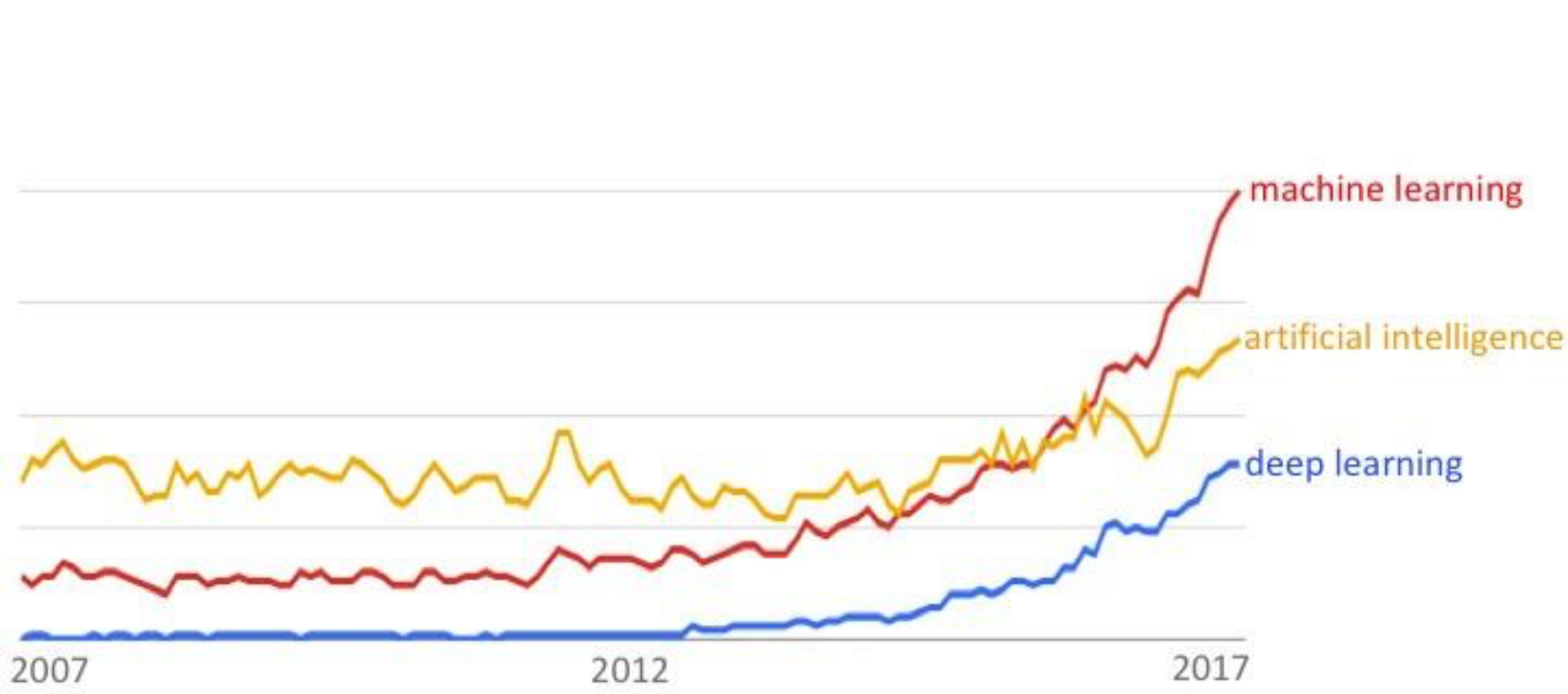
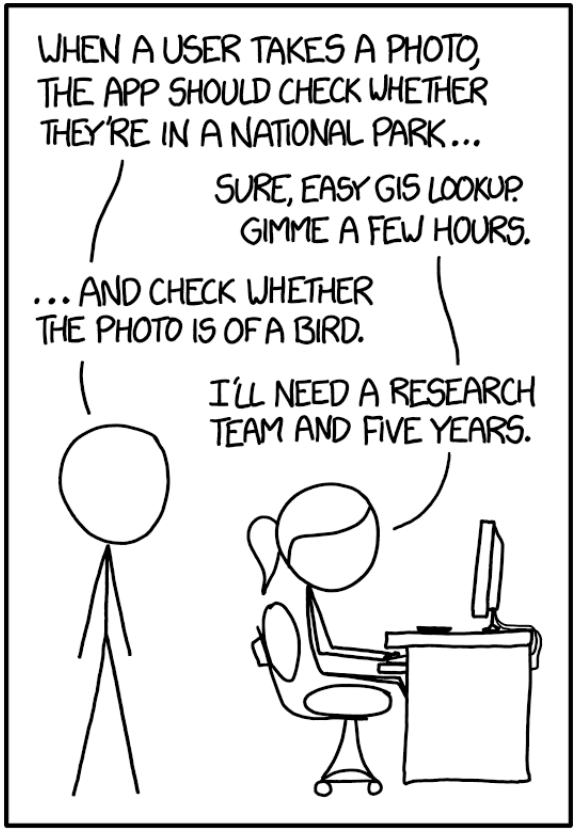
Applications of Deep Learning (Beyond Text & Images)

Brian Mac Namee

APPLICATIONS OF MACHINE LEARNING

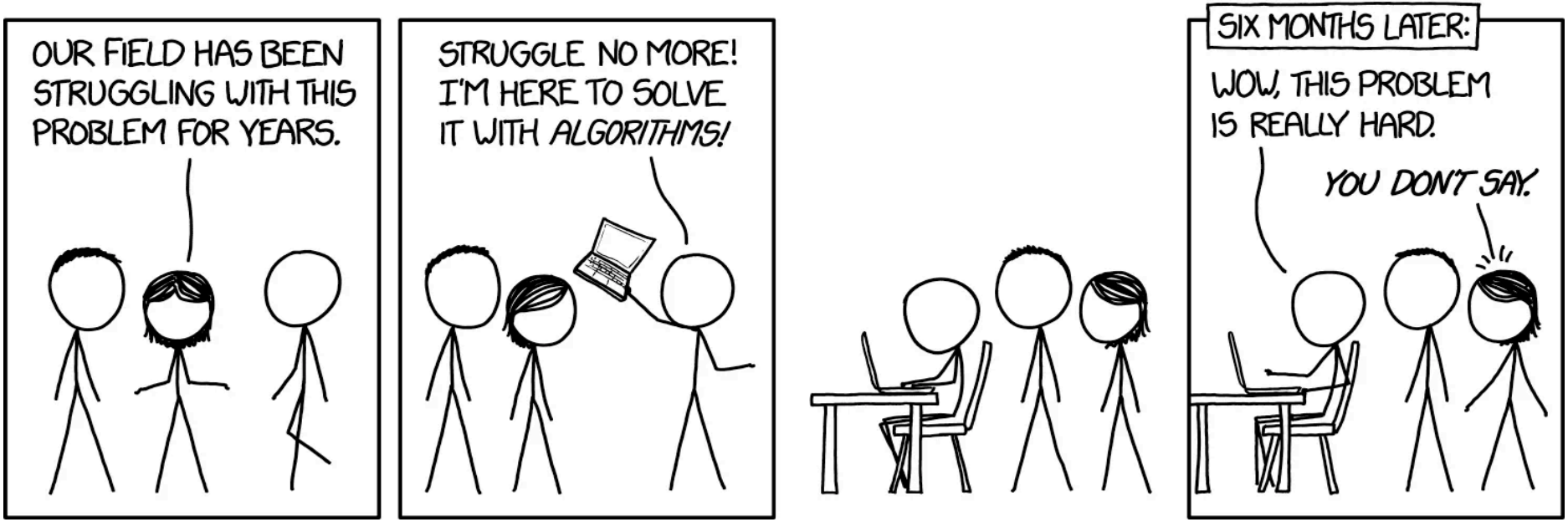


https://trends.google.com/trends/

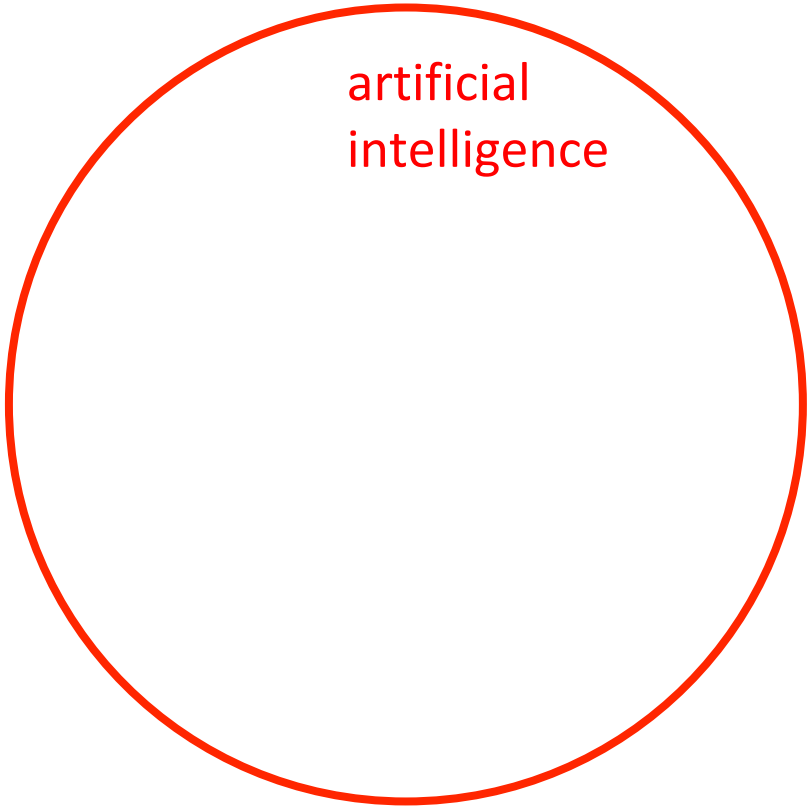


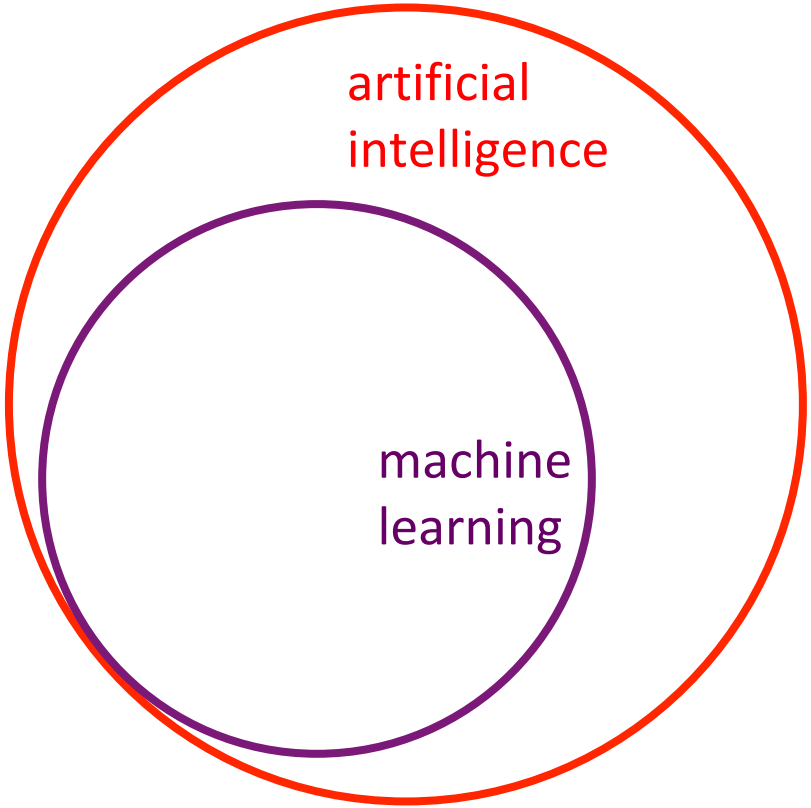
IN CS, IT CAN BE HARD TO EXPLAIN THE DIFFERENCE BETWEEN THE EASY AND THE VIRTUALLY IMPOSSIBLE.

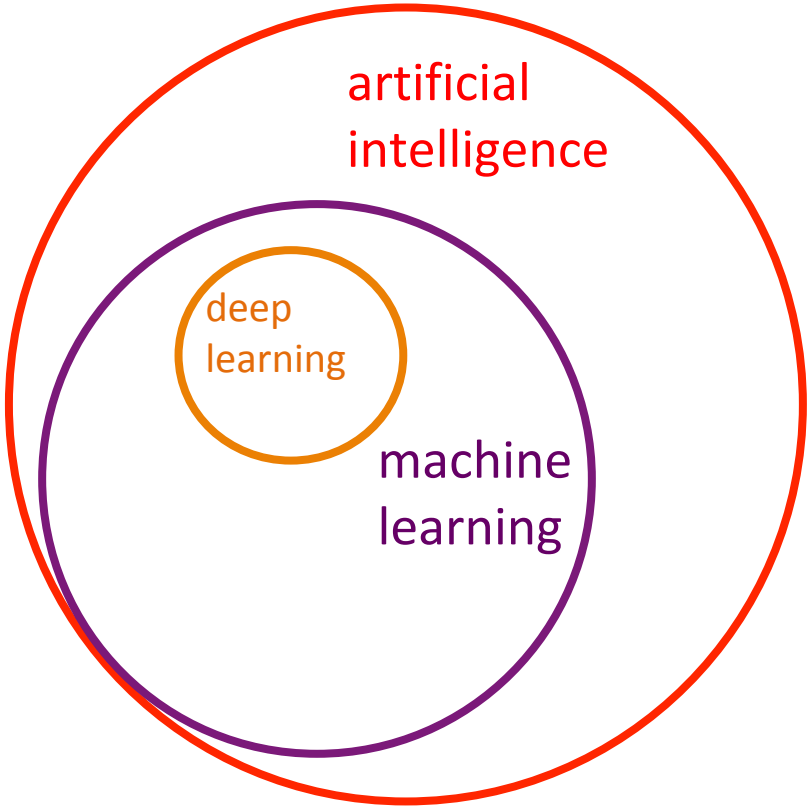
https://xkcd.com/1425/

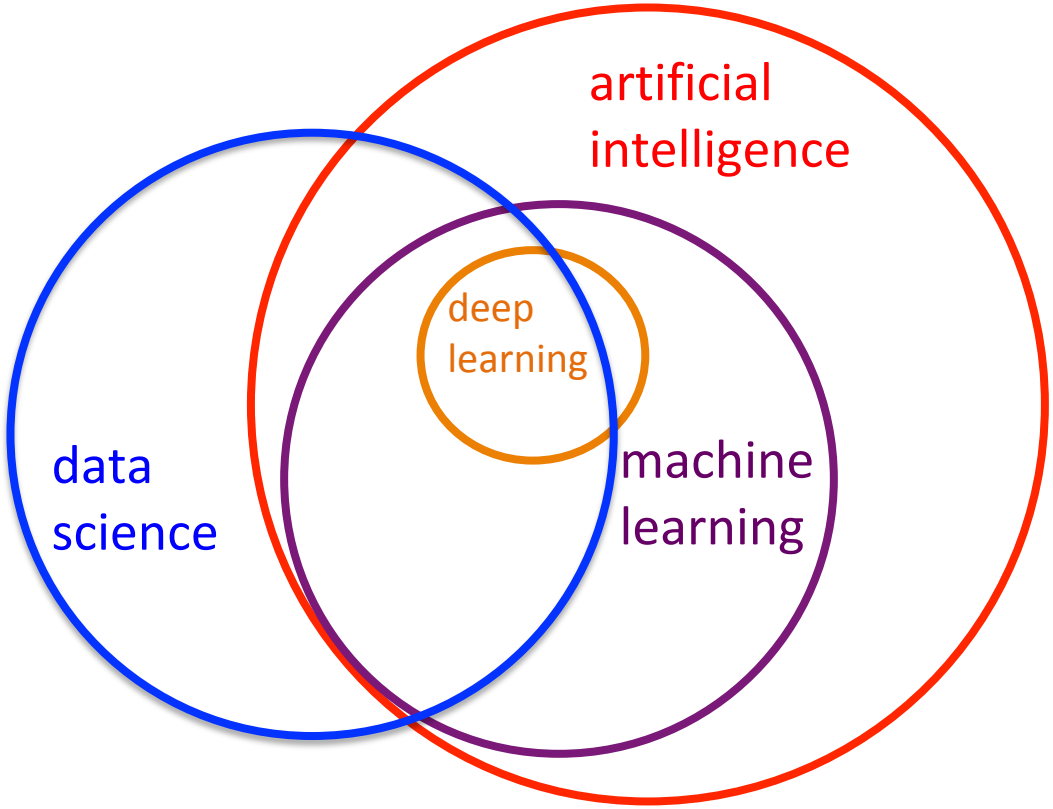


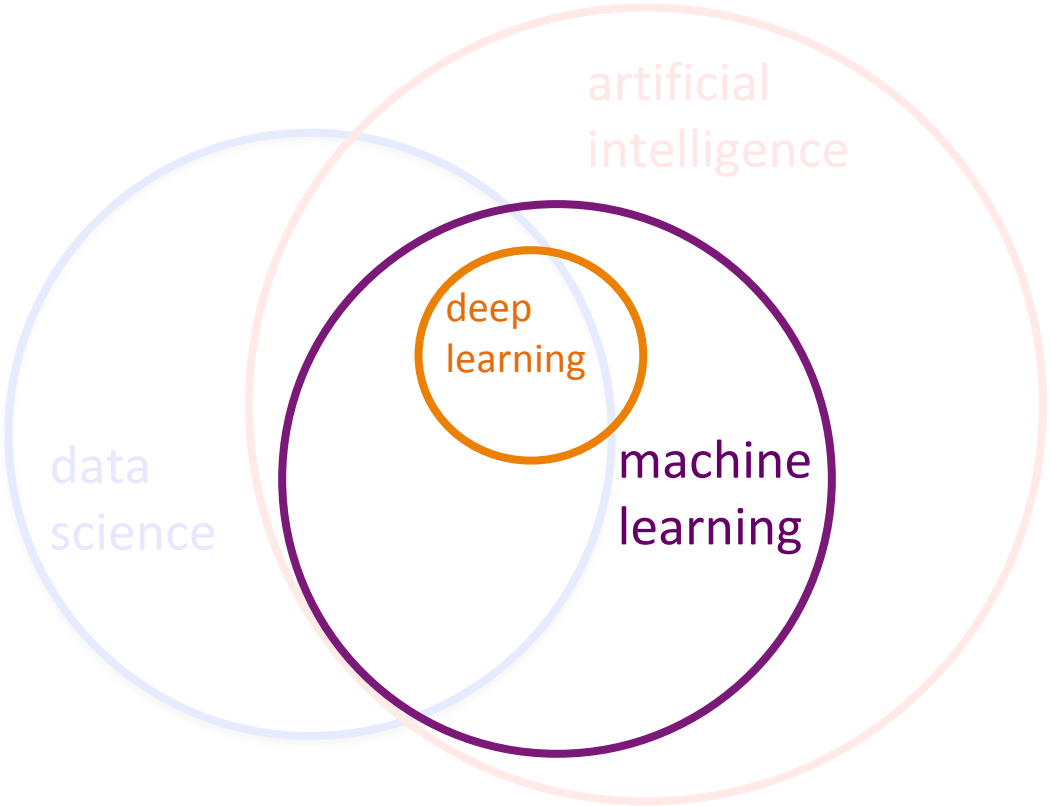
https://xkcd.com/1831/

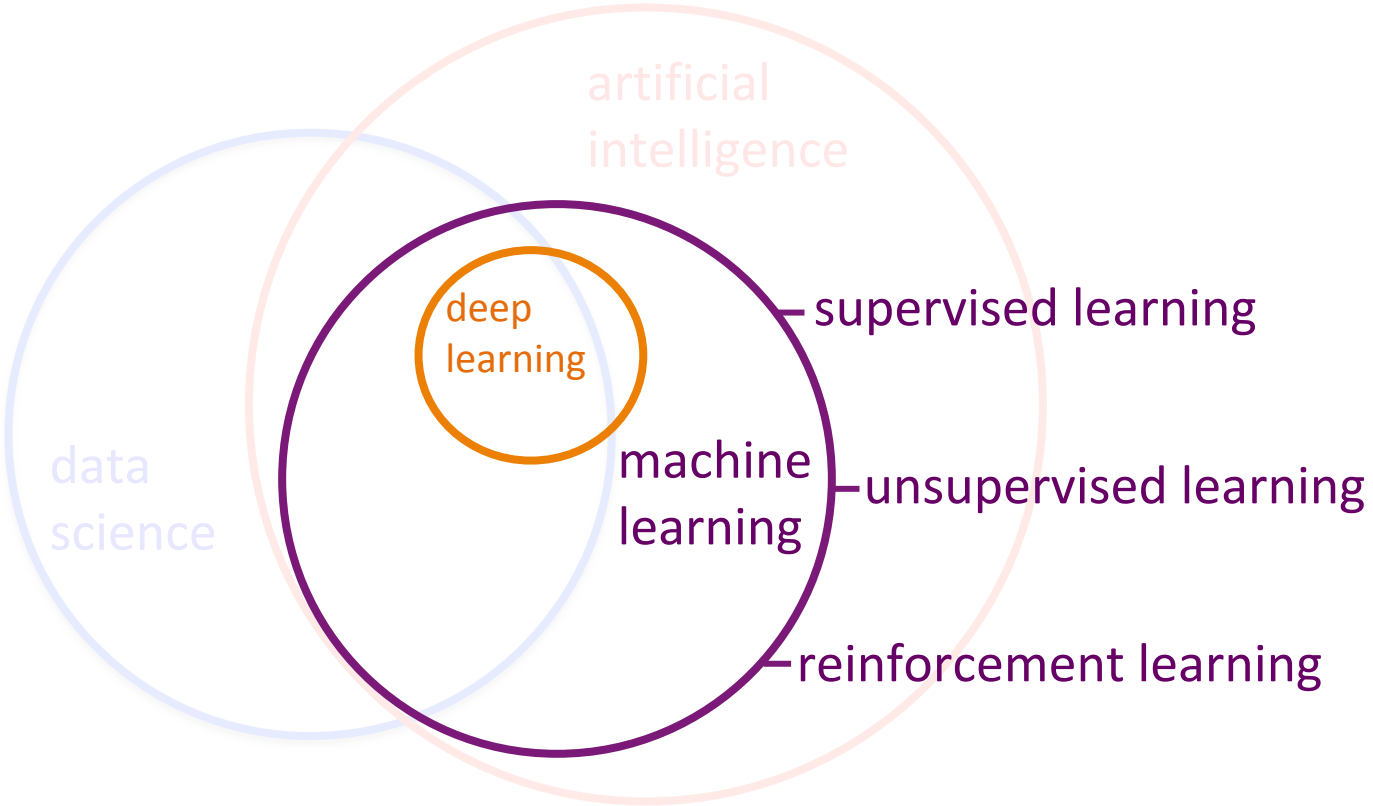

artificial
intelligence

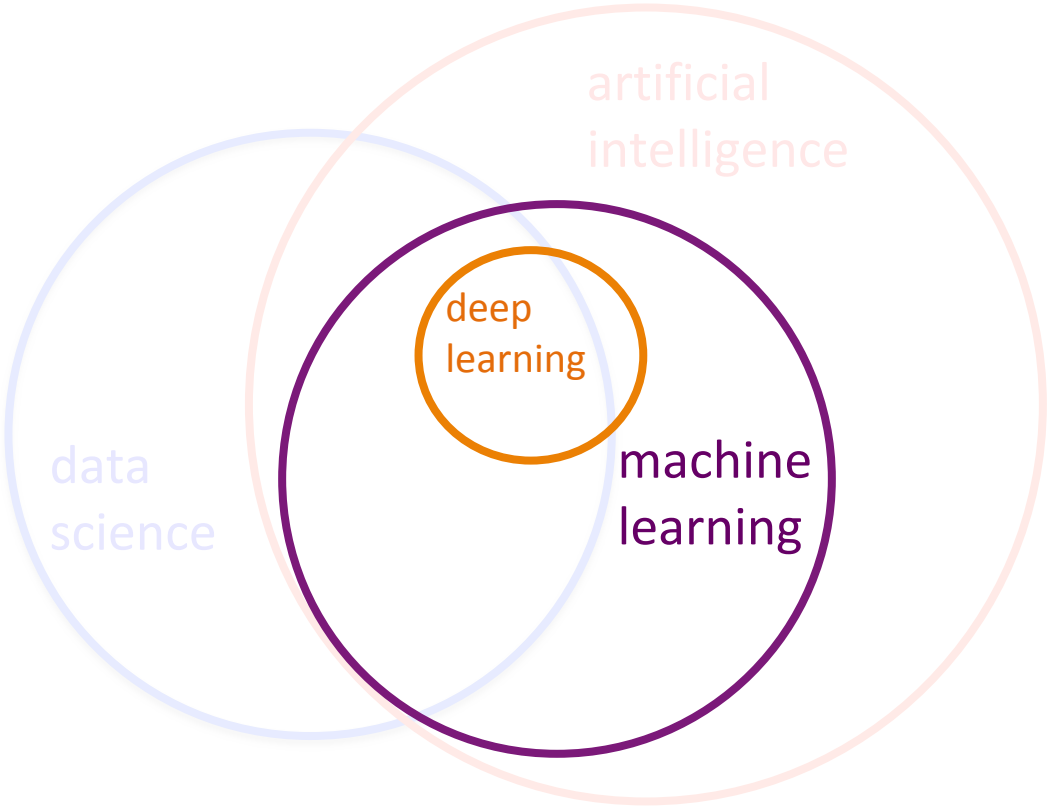

artificial
intelligence
machine
learning

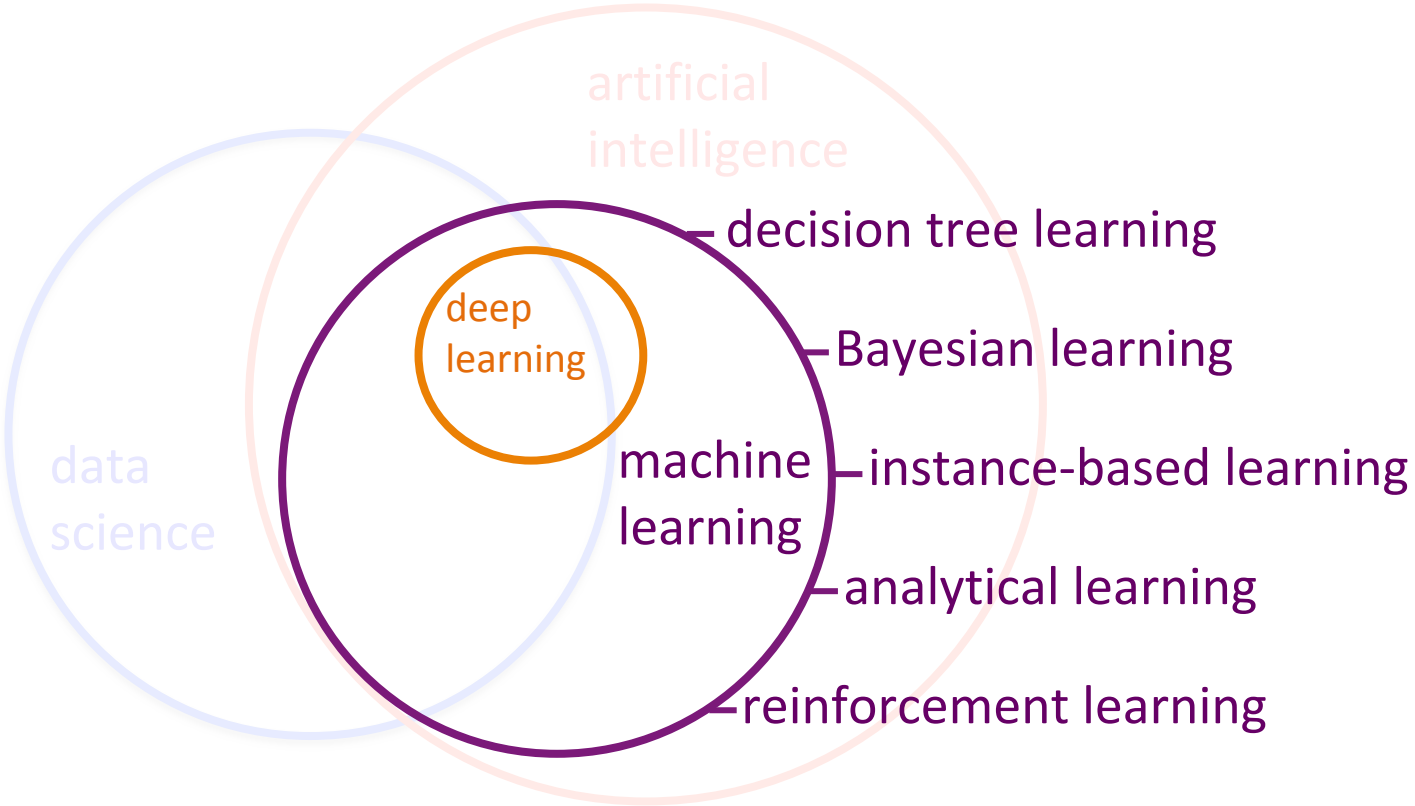

artificial
intelligence
deep
learning
machine
learning

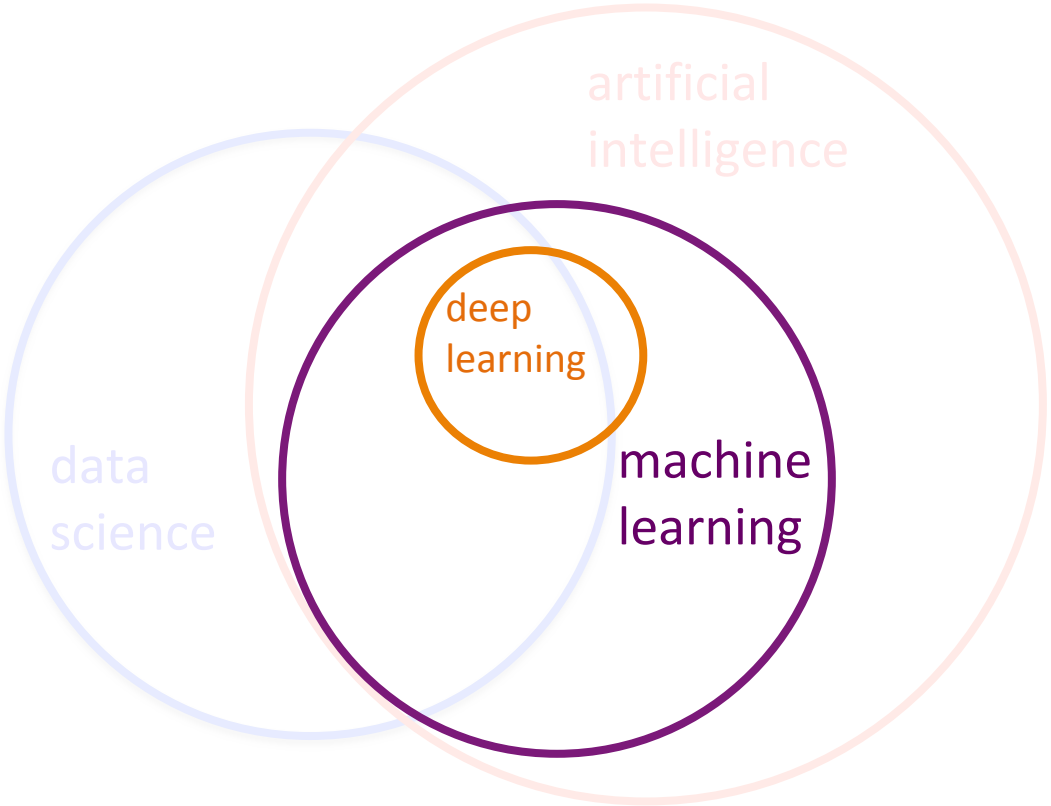

artificial
intelligence
deep
learning
data
science
machine
learning

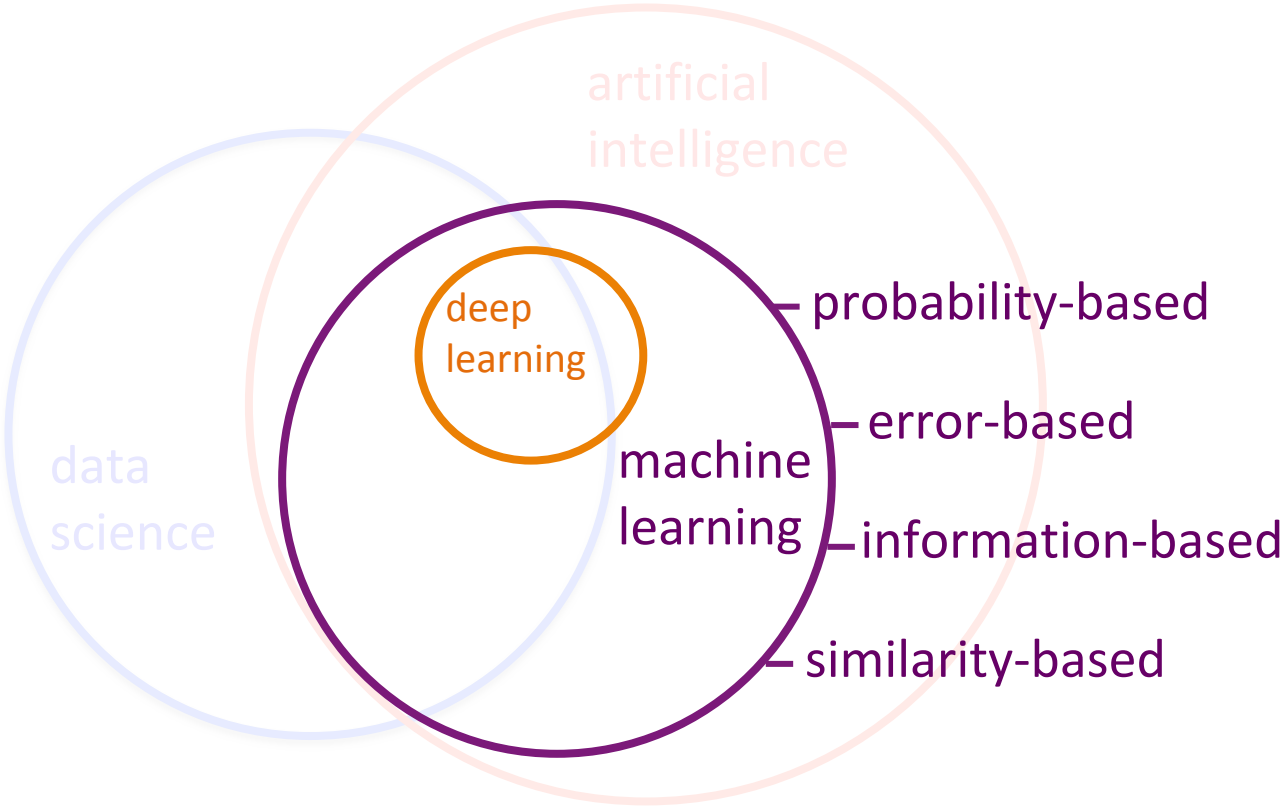

artificial
intelligence
data
science
deep
learning
machine
learning

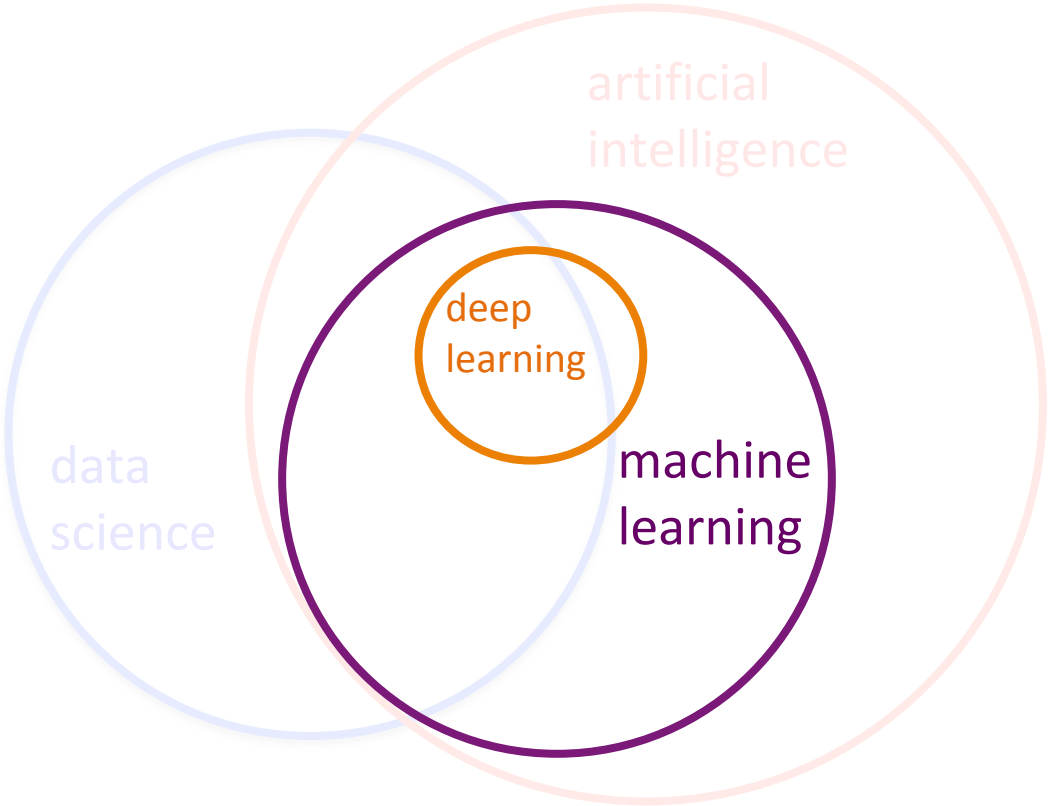

artificial
intelligence
data
science
deep
learning
machine
learning
supervised learning
unsupervised learning
reinforcement learning

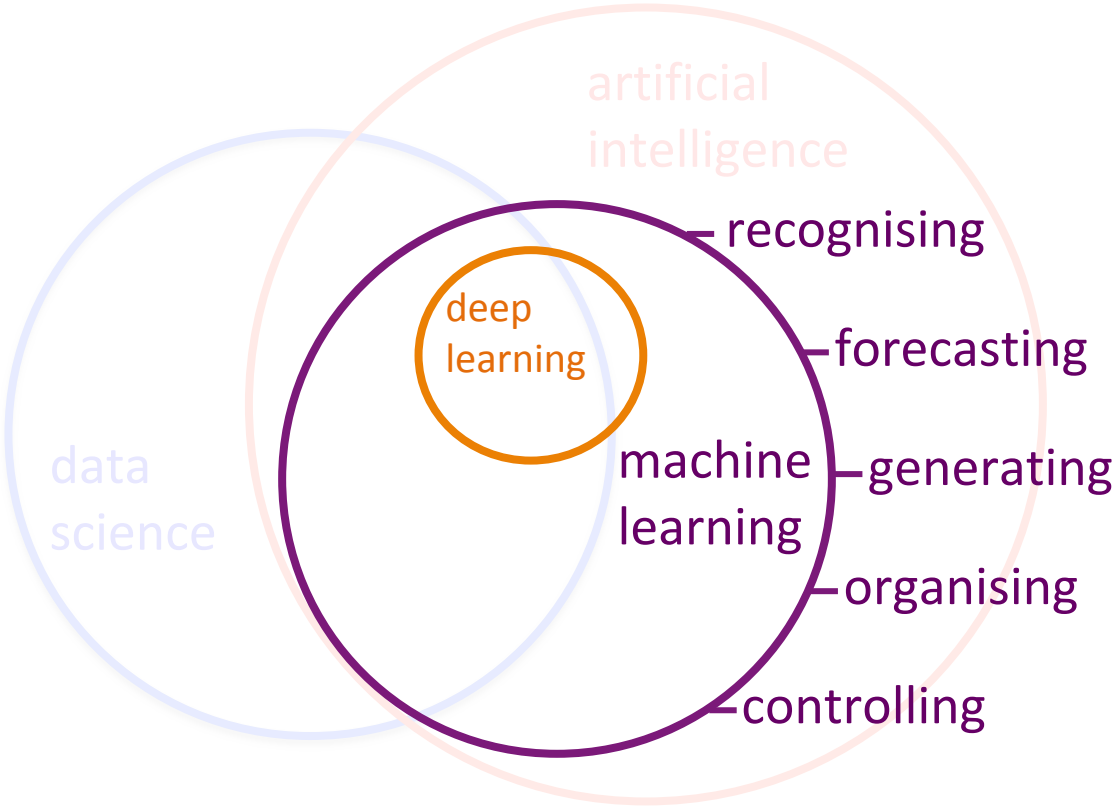

artificial
intelligence
data
science
deep
learning
machine
learning

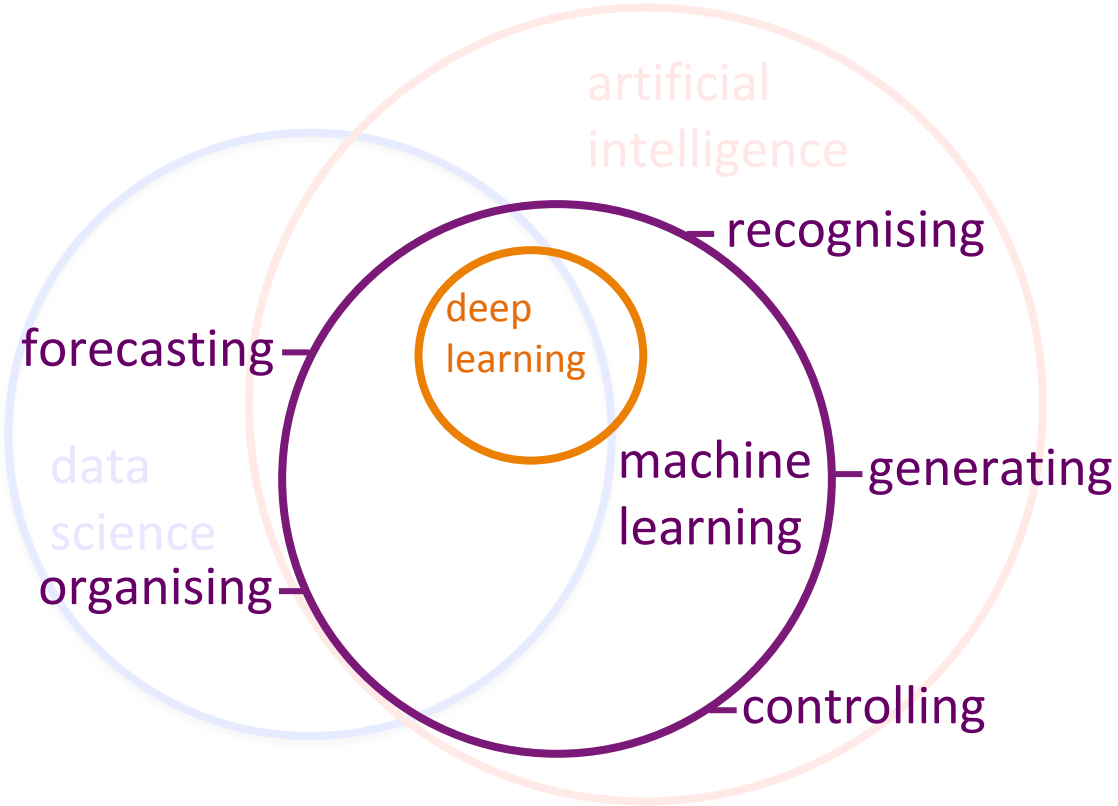

artificial
intelligence
data
science
deep
learning
machine
learning
decision tree learning
Bayesian learning
instance-based learning
analytical learning
reinforcement learning

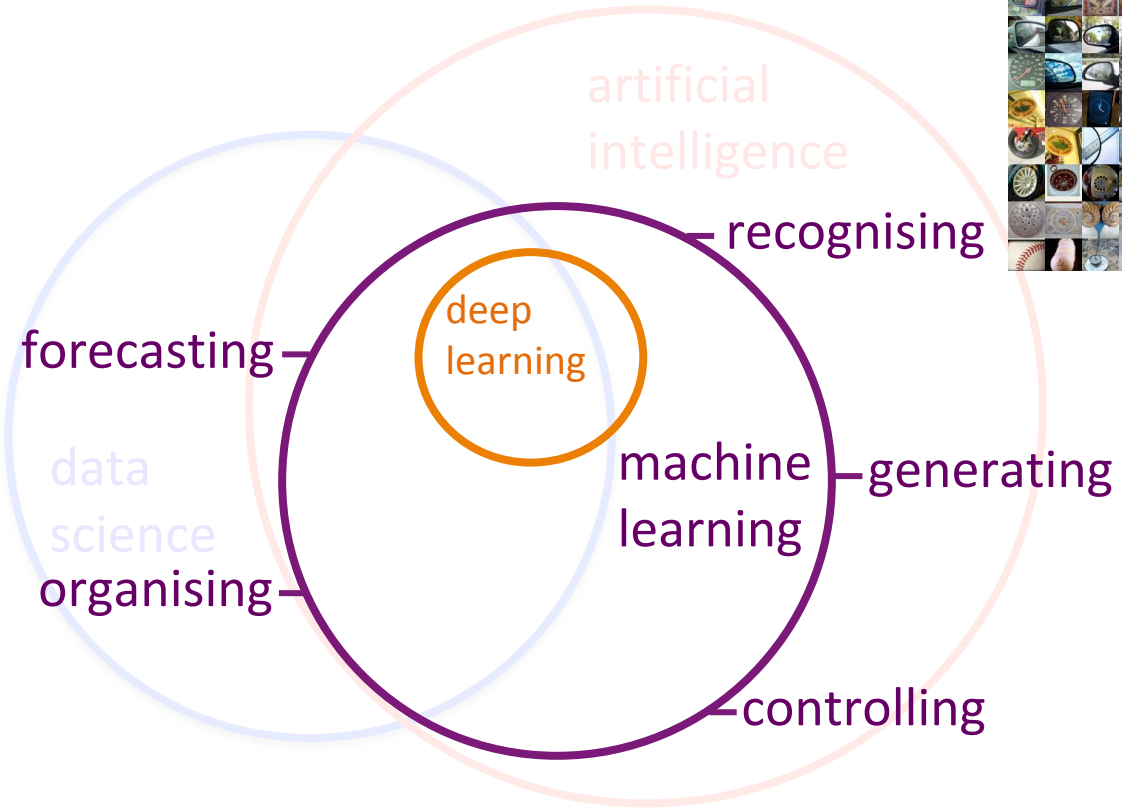

artificial
intelligence
data
science
deep
learning
machine
learning

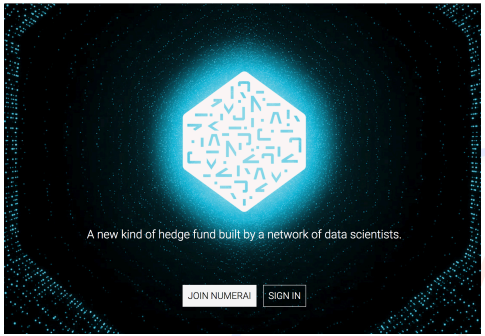

artificial
intelligence
data
science
deep
learning
machine
learning
probability-based
error-based
information-based
similarity-based


artificial
intelligence
data
science
deep
learning
machine
learning


artificial
intelligence
data
science
deep
learning
machine
learning
recognising
forecasting
generating
organising
controlling


artificial
intelligence
recognising
deep
learning
forecasting
data
science
machine
learning
generating
organising
controlling


artificial
intelligence
recognising
deep
learning
forecasting
data
science
machine
learning
generating
organising
controlling



artificial intelligence

recognising

deep learning

forecasting

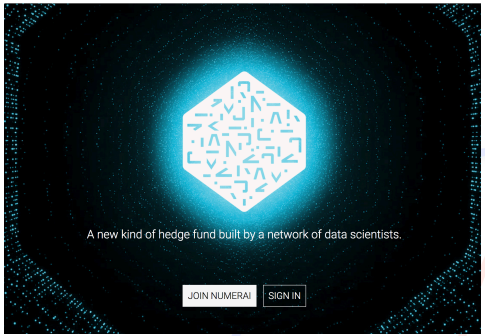
data science

machine learning

generating

organising

controlling



artificial intelligence

recognising

forecasting

deep learning

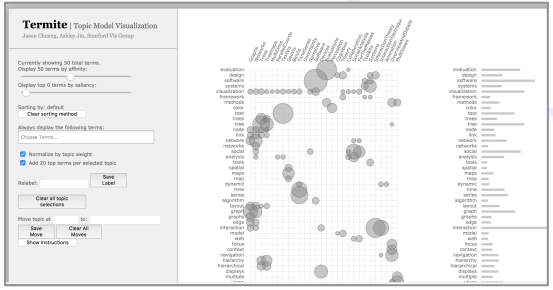
data science

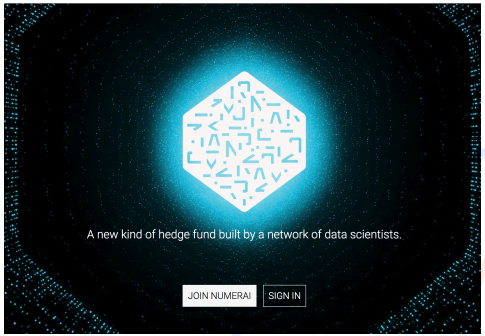
machine learning

generating

organising

controlling





artificial intelligence

recognising

forecasting

deep learning

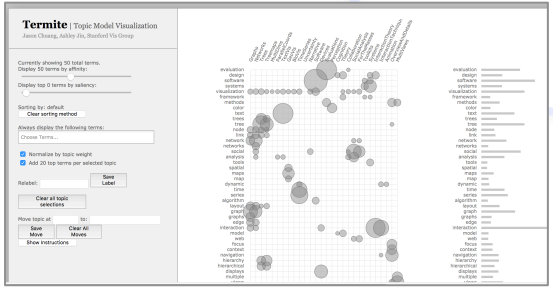
data science

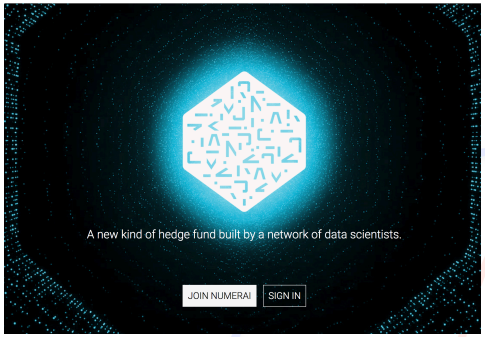
machine learning

generating

organising

controlling





artificial intelligence

recognising

deep learning

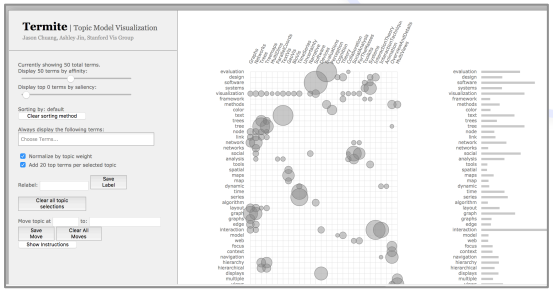
forecasting

data science

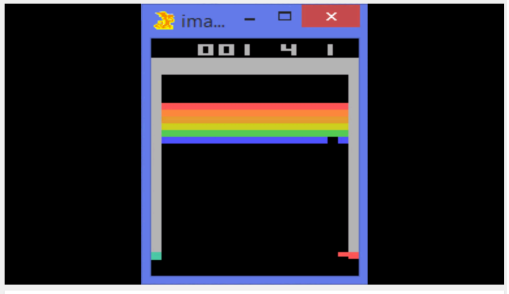
machine learning

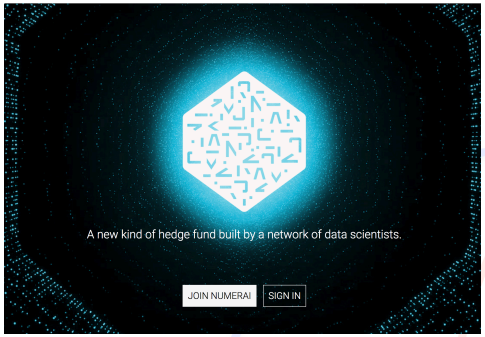
generating

organising



controlling





artificial intelligence

recognising

deep learning

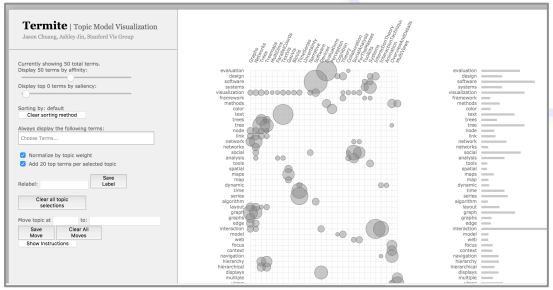
forecasting

data science

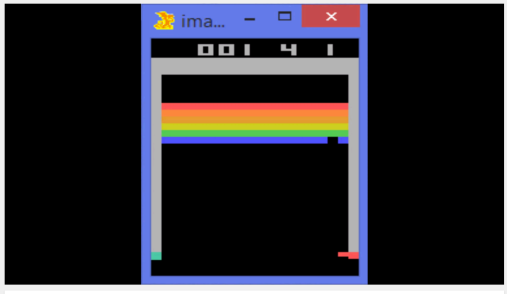
machine learning

generating

organising



controlling

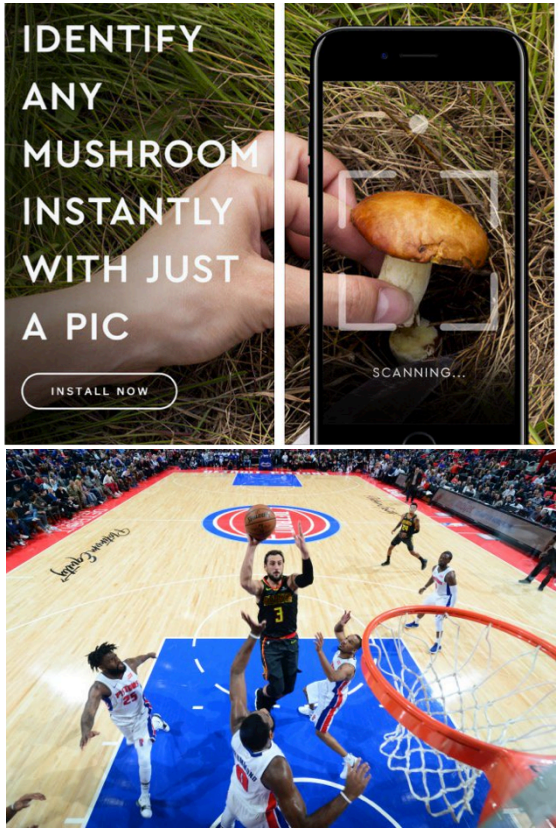


Domains Ripe for Application of Machine Learning

- Involve repetitive tasks with defined outcomes
- Massive collections of historical examples of the task with solutions already exist
- Involve simple decisions rather than complex recommendations
- The domain does not change too rapidly
- The opportunity to augment human performance rather than replace it exists

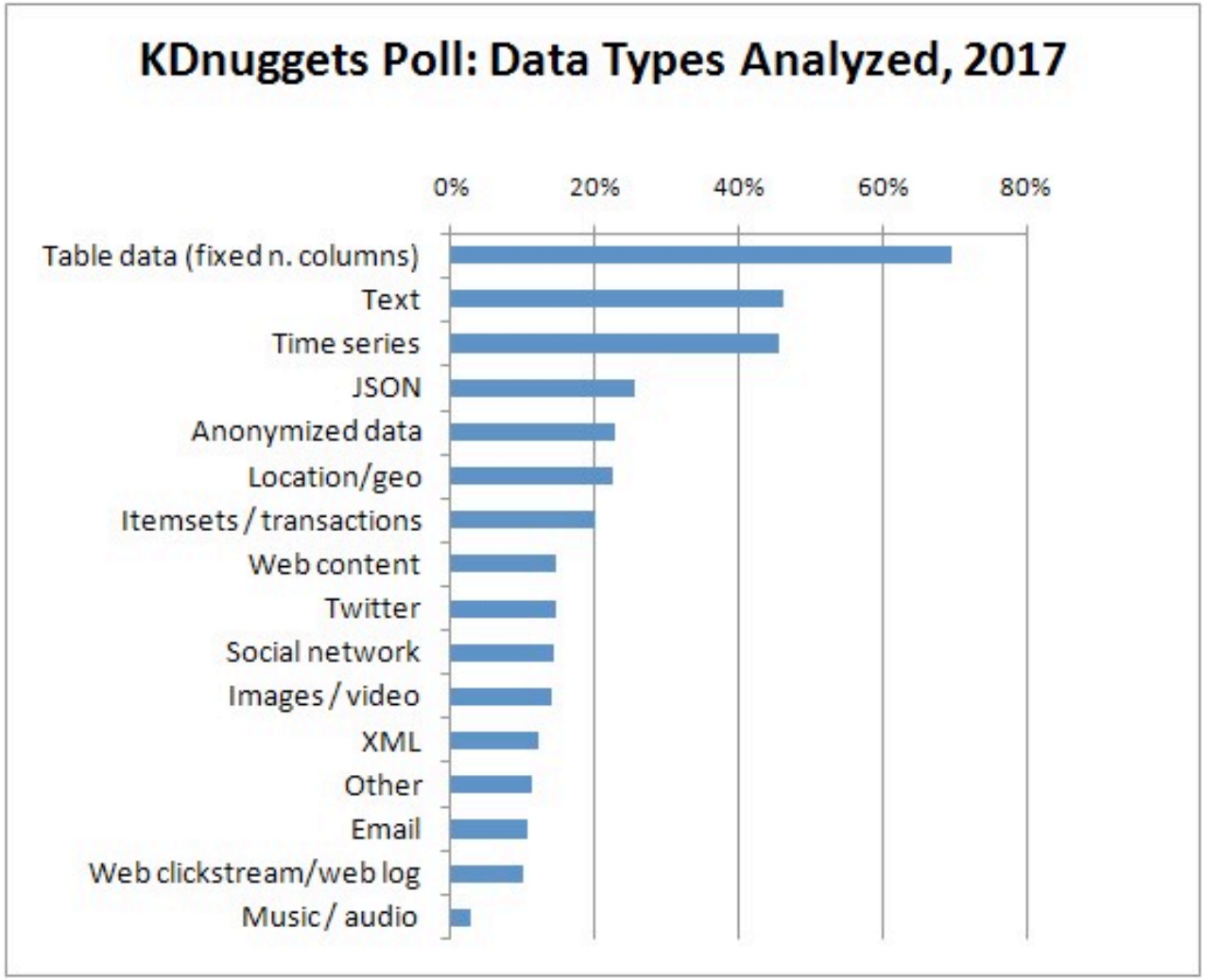
Limitations of Machine Learning

- Still best for one-level questions
- Struggles to deal with subtle context
- Encode biases that exist in datasets
- Making machine learning models that continuously learn is still difficult
- Explanation of models (in domains where trust is required) remains challenging



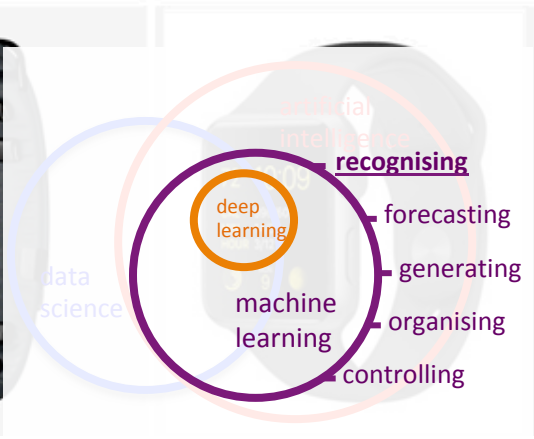
(BEYOND TEXT & IMAGES)

There’s All Kinds Of Data Out There!



What Data You Analyzed – KDnuggets Poll Results and Trends
https://www.kdnuggets.com/2017/04/poll-results-data-analyzed.html

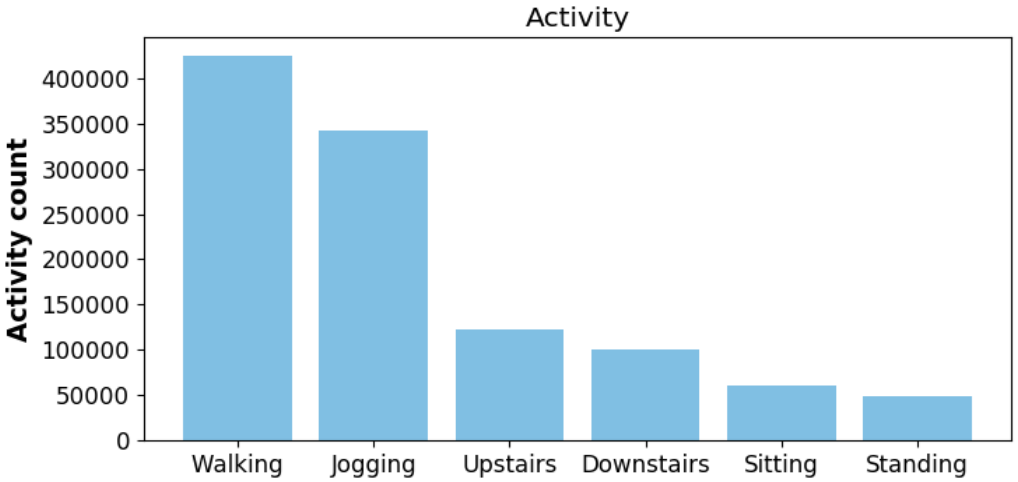
Activity Tracking



WISDM v1.1 Activity Recognition Data

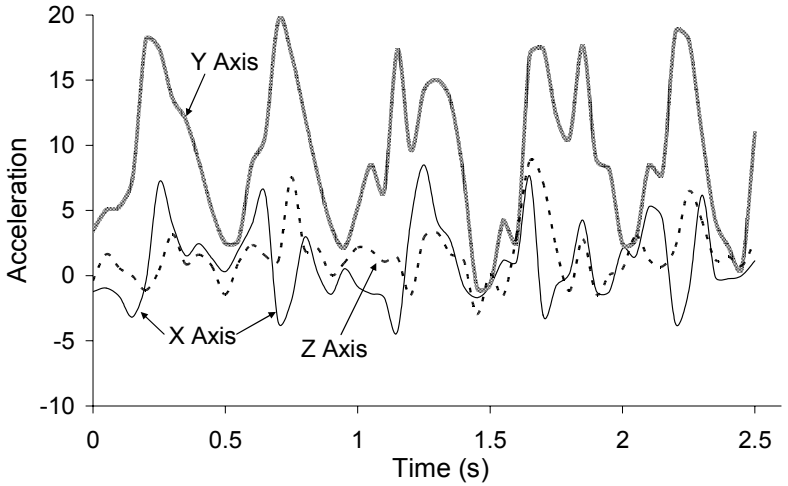
Accelerometer data recorded in controlled conditions for activity recognition

- 1,098,207 instances
- 3 attributes
- 6 activity classes

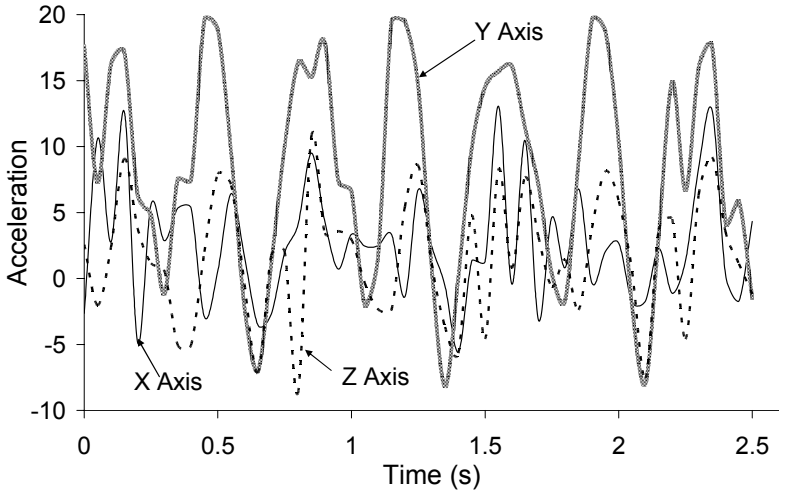


Assume signals contain both spatial and temporal structure

WISDM v1.1 Activity Recognition Data



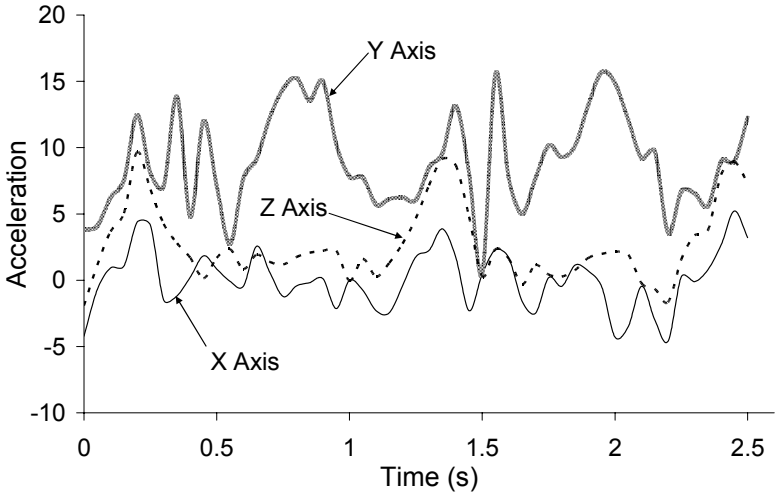
(a) Walking



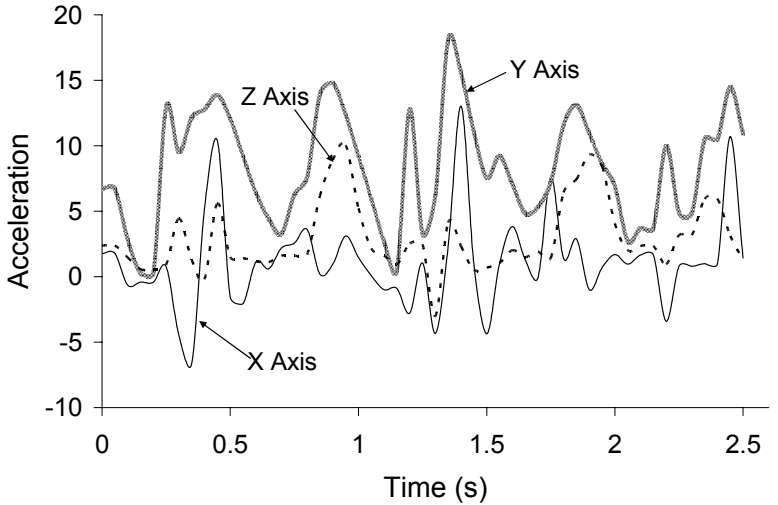
(b) Jogging

Jennifer R. Kwapisz, Gary M. Weiss and Samuel A. Moore (2010). Activity Recognition using Cell Phone Accelerometers, Proceedings of the Fourth International Workshop on Knowledge Discovery from Sensor Data (at KDD-10)
http://www.cis.fordham.edu/wisdm/dataset.php

WISDM v1.1 Activity Recognition Data



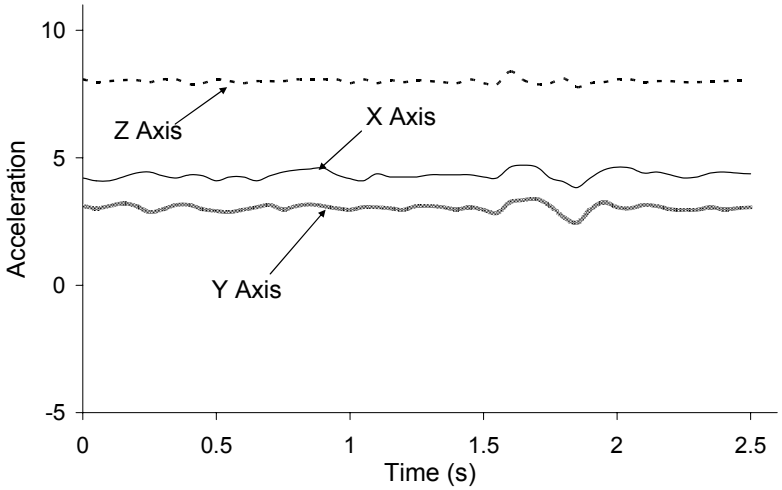
(c) Ascending Stairs



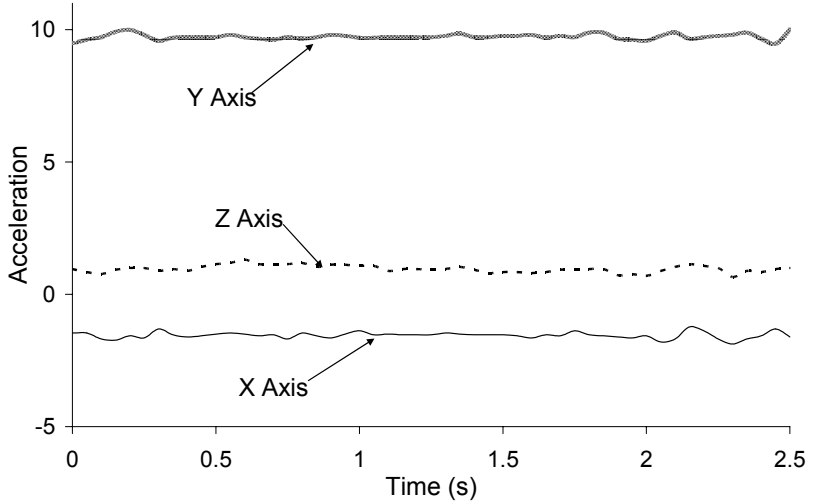
(d) Descending Stairs

Jennifer R. Kwapisz, Gary M. Weiss and Samuel A. Moore (2010). Activity Recognition using Cell Phone Accelerometers, Proceedings of the Fourth International Workshop on Knowledge Discovery from Sensor Data (at KDD-10)
http://www.cis.fordham.edu/wisdm/dataset.php

WISDM v1.1 Activity Recognition Data



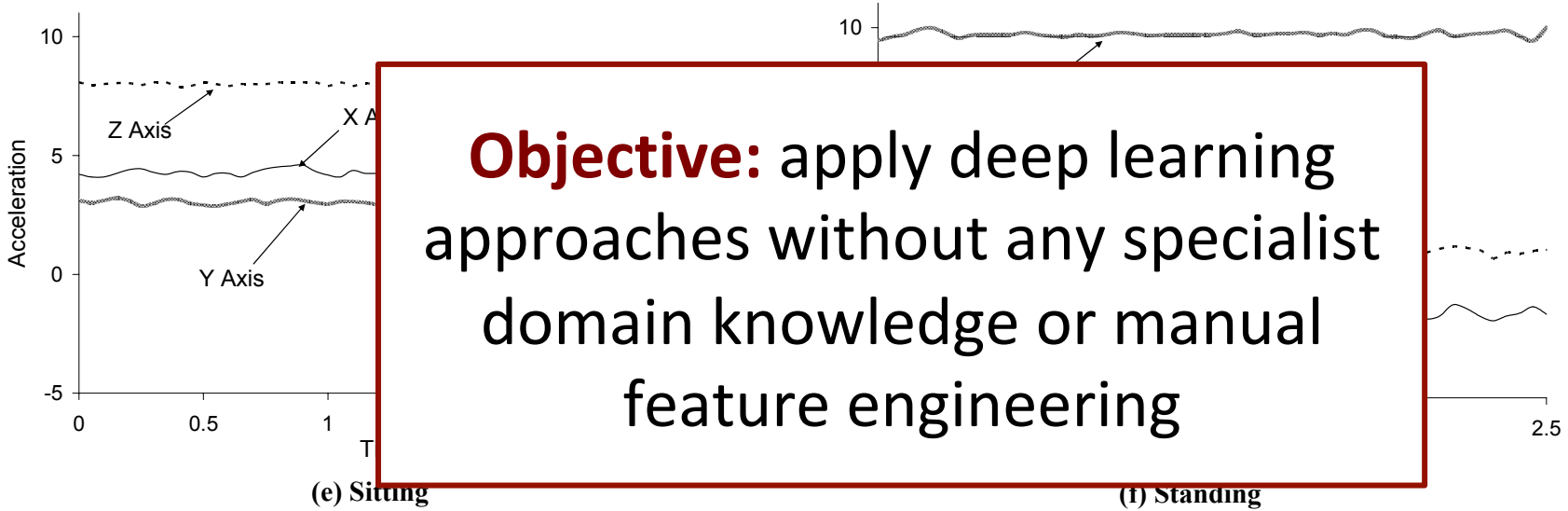
(e) Sitting



(f) Standing

Jennifer R. Kwapisz, Gary M. Weiss and Samuel A. Moore (2010). Activity Recognition using Cell Phone Accelerometers, Proceedings of the Fourth International Workshop on Knowledge Discovery from Sensor Data (at KDD-10)
http://www.cis.fordham.edu/wisdm/dataset.php

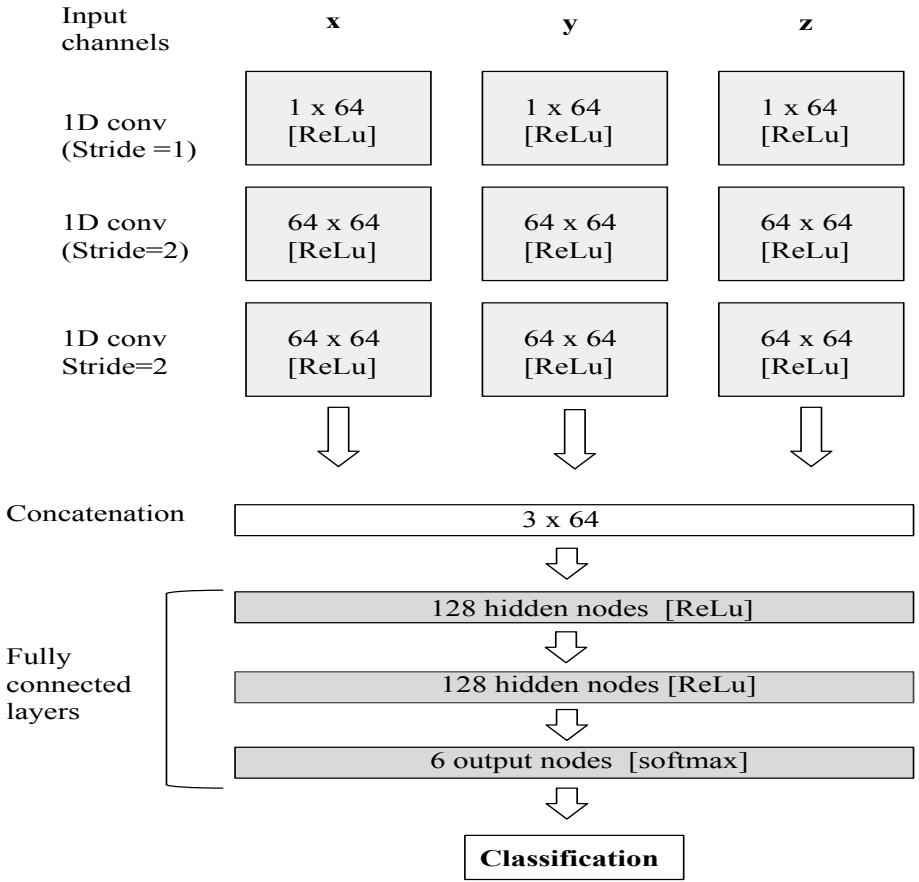
WISDM v1.1 Activity Recognition Data



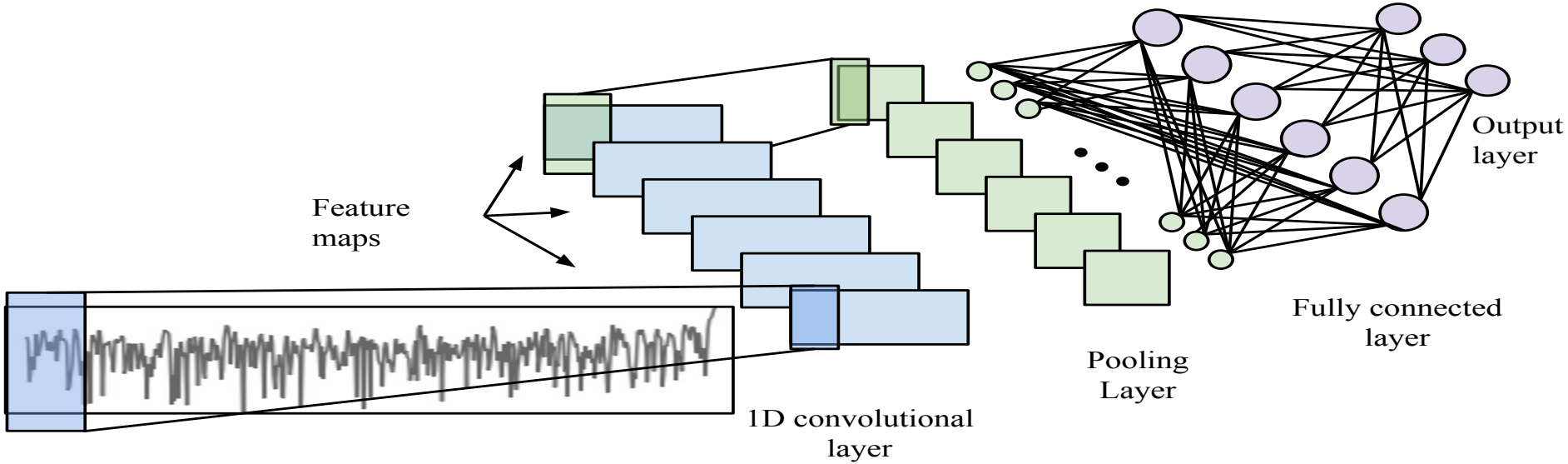
Objective: apply deep learning approaches without any specialist domain knowledge or manual feature engineering

Jennifer R. Kwapisz, Gary M. Weiss and Samuel A. Moore (2010). Activity Recognition using Cell Phone Accelerometers, Proceedings of the Fourth International Workshop on Knowledge Discovery from Sensor Data (at KDD-10)
http://www.cis.fordham.edu/wisdm/dataset.php

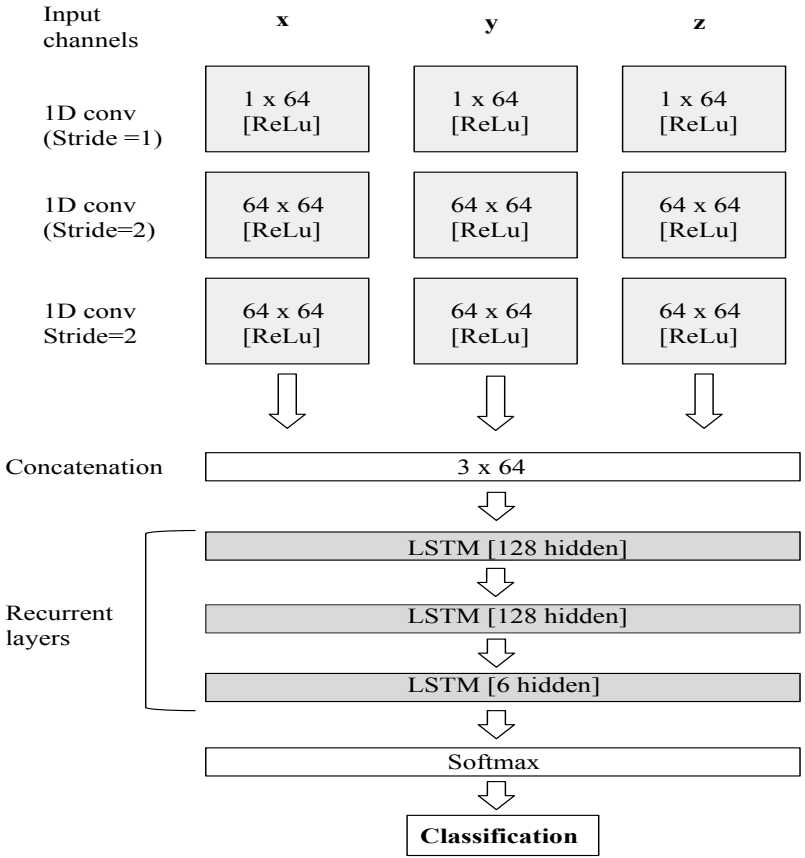
CNN Based Architecture



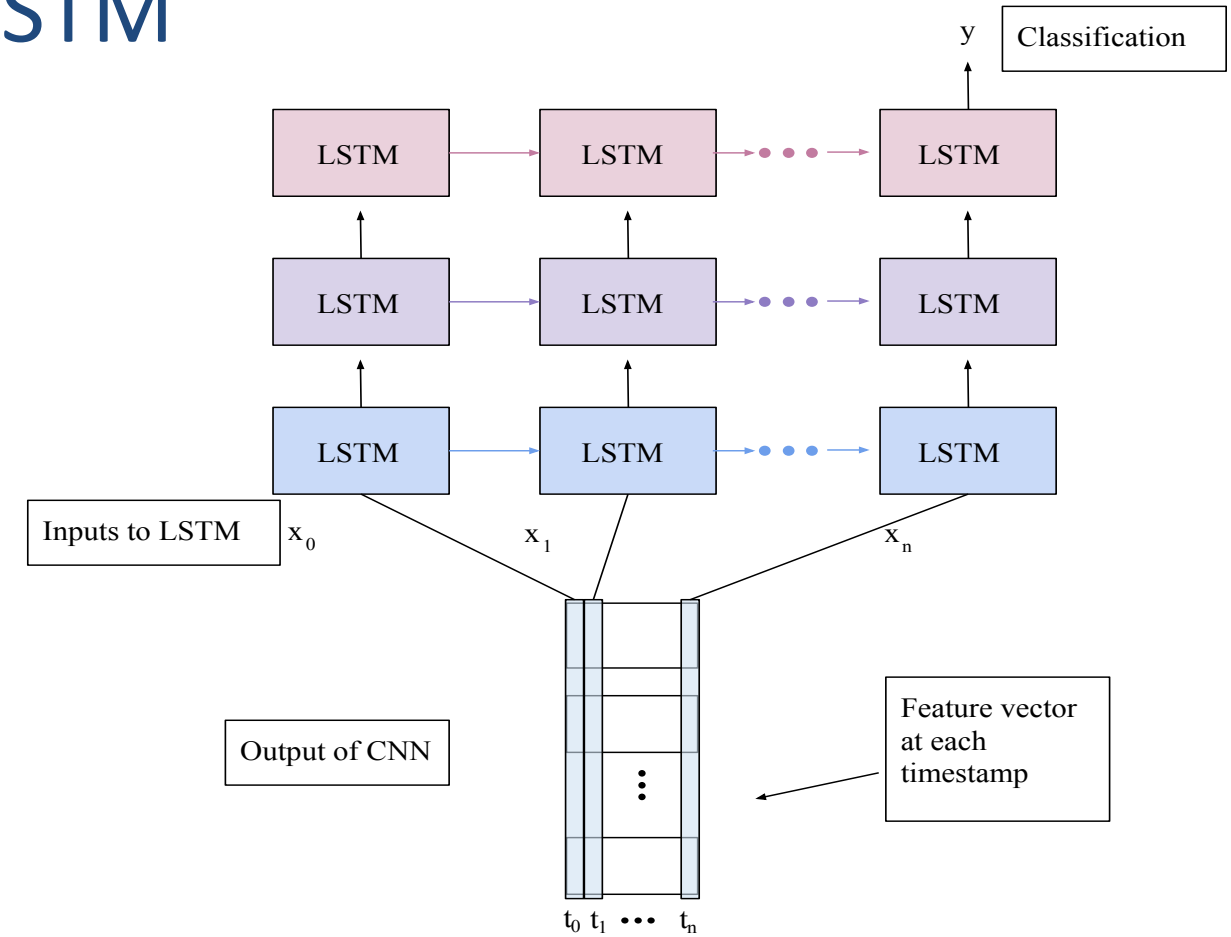
CNN on 1-D Time Series Channel



CNN-LSTM based architecture



CNN to LSTM


y
Classification
LSTM
LSTM
LSTM
LSTM
LSTM
LSTM
LSTM
LSTM
LSTM
Inputs to LSTM
x0
x1
xn
Output of CNN
Feature vector at each timestamp
t0 t1 tn

Results

Table 1: F1-scores for hybrid models.

Fold	CNN	CNN-LSTM
1	**0.944**	0.911
2	**0.929**	0.924
3	0.918	**0.922**
4	0.922	**0.930**
5	**0.919**	0.914
Avg.	**0.926**	0.920

Table 2: F1-score for impersonal models.

Fold	CNN	CNN-LSTM
1	**0.839**	0.819
2	0.803	**0.857**
3	0.858	**0.882**
4	**0.813**	**0.813**
5	0.843	**0.882**
Avg.	0.831	**0.850**

User Centric Problem

Impersonal Data

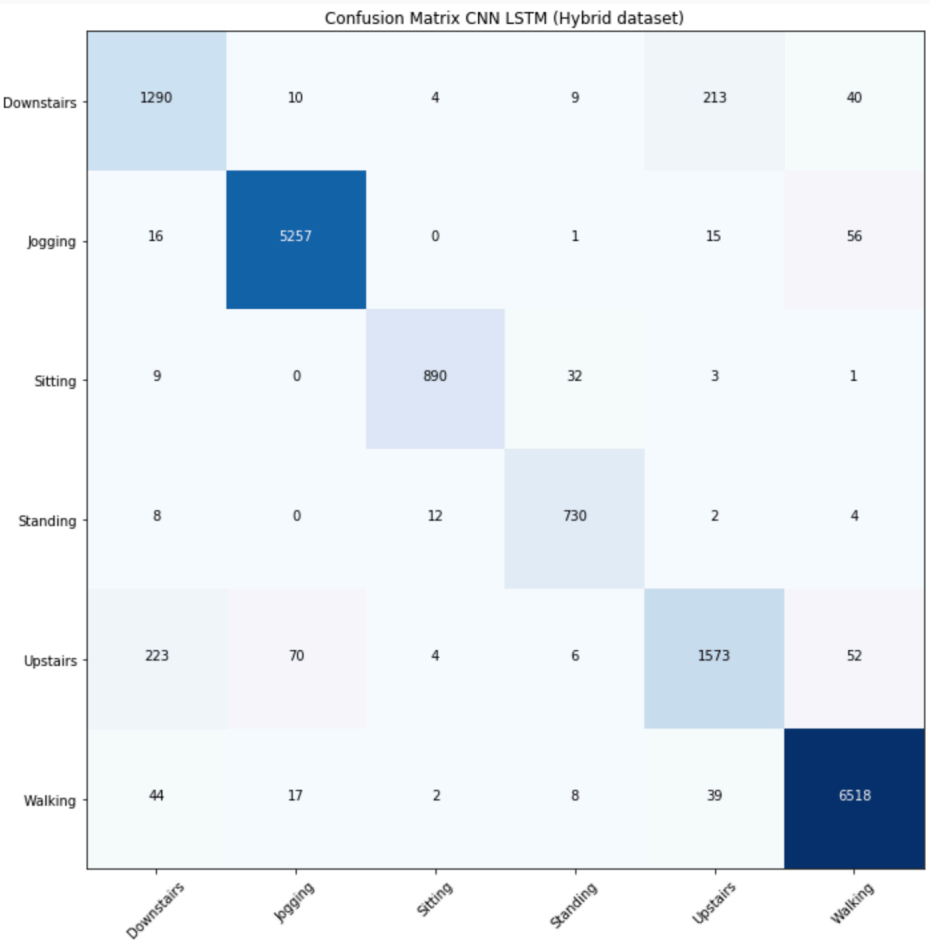
- Model trained on data from only users outside the test set.
- Don’t require user-specific data but are less accurate

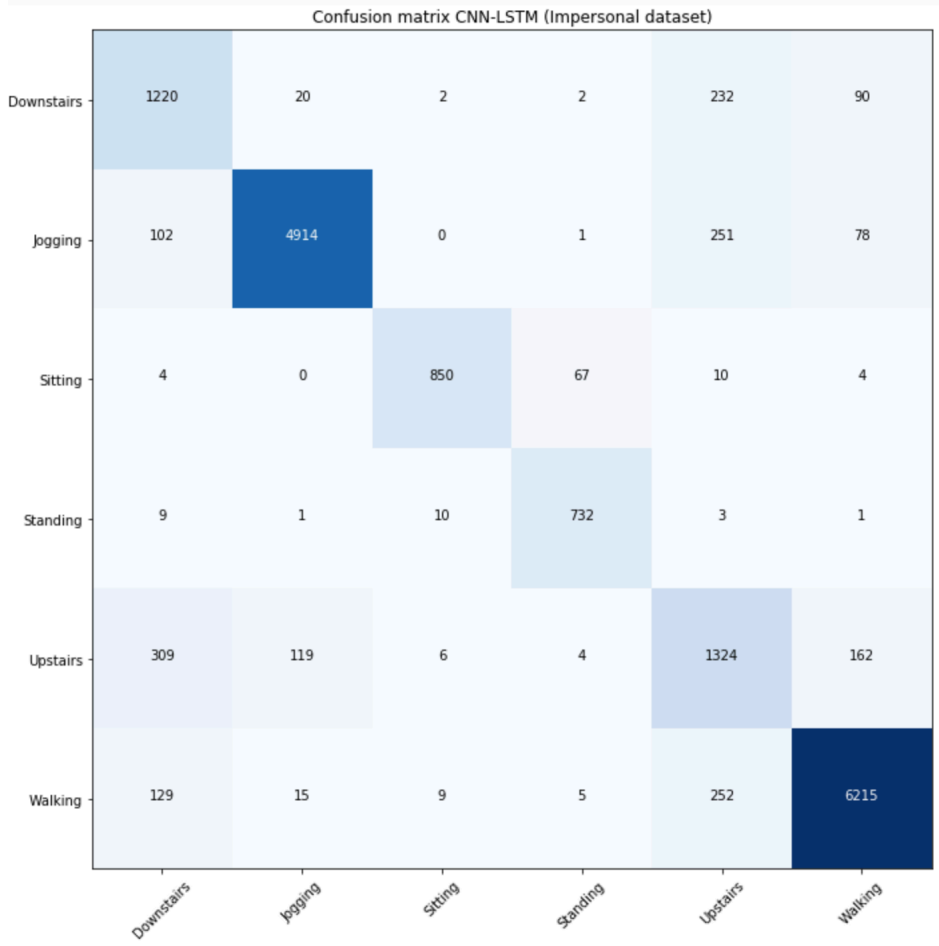
Personal Data

- Model trained on data only from the test user.
- Require user-specific data but tend to be accurate

Hybrid Data

- Model trained on data from both the test users and users outside the test set.


Confusion Matrix CNN LSTM (Hybrid dataset)
Downstairs
Jogging
Sitting
Standing
Upstairs
Walking
1290 10 4 9 213 40
16 5257 0 1 15 56
9 0 890 32 3 1
8 0 12 730 2 4
223 70 4 6 1573 52
44 17 2 8 39 6518
Downstairs
Jogging
Sitting
Standing
Upstairs
Walking


Confusion matrix CNN-LSTM (Impersonal dataset)
Downstairs
Jogging
Sitting
Standing
Upstairs
Walking
1220 20 2 2 232 90
102 4914 0 1 251 78
4 0 850 67 10 4
9 1 10 732 3 1
309 119 6 4 1324 162
129 15 9 5 252 6215
Downstairs
Jogging
Sitting
Standing
Upstairs
Walking

Malware Detection

Kaggle Microsoft Malware Classification Challenge

Malware is malicious code which is often encountered as compiled executable byte code

Kaggle Microsoft malware classification challenge

- Over 400 GB uncompressed data
- 9 labelled malware classes
- 10,868 malware files as raw byte code (plus disassembled machine code) in training set

Malware Class	Instances
Ramnit	1541
Lollipop	2478
Kelihos_v3	2942
Vundo	475
Simda	42
Tracur	751
Kelihos_v1	398
Obfuscator.ACY	1228
Gatak	1013

Kaggle Microsoft Malware Classification Challenge

```
.text:00401000 56              push   esi
.text:00401001 8D 44 2
.text:00401005 50
.text:00401006 8B F1
.text:0040100D C7 06 0

.text:00401013 8B C6
.text:00401015 5E
.text:00401016 C2 04 00        retn   4
.text:00401019 CC CC CC        align  10h
.text:00401020 C7 01 08        mov    dword ptr [ecx],
                               offset off_42BB08
.text:00401026 E9 26 1C        jmp    sub_402C51
```

```
00401000 56 8D 44 24 08 50 8B F1
                                5E C2 04
                                00 E9 26
                                08 BB 42
                                56 E8 6C
1E 00 00 83 C4 04 8B C6
00401050 5E C2 04 00 CC CC CC CC
CC CC CC CC CC CC CC CC
00401060 8B 44 24 08 8A 08 8B 54
24 04 88 0A C3 CC CC CC
```

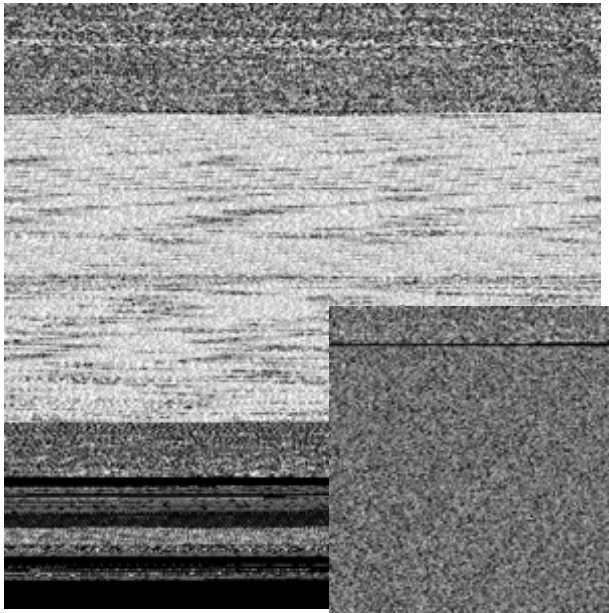
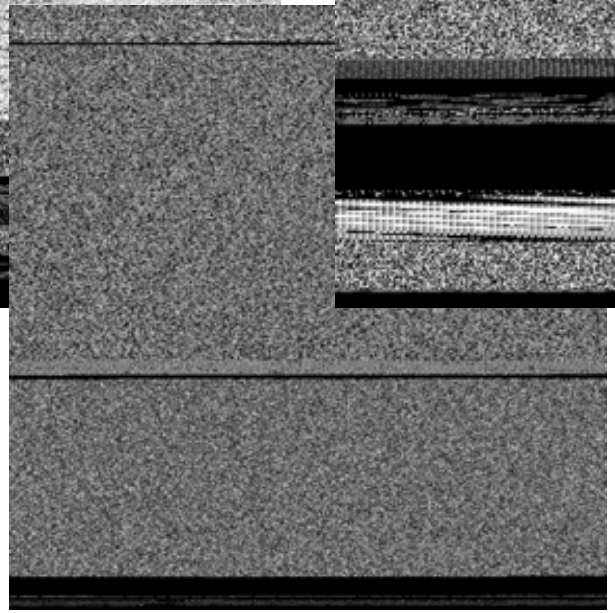
Objective: apply deep learning approaches without any specialist domain knowledge or manual feature engineering

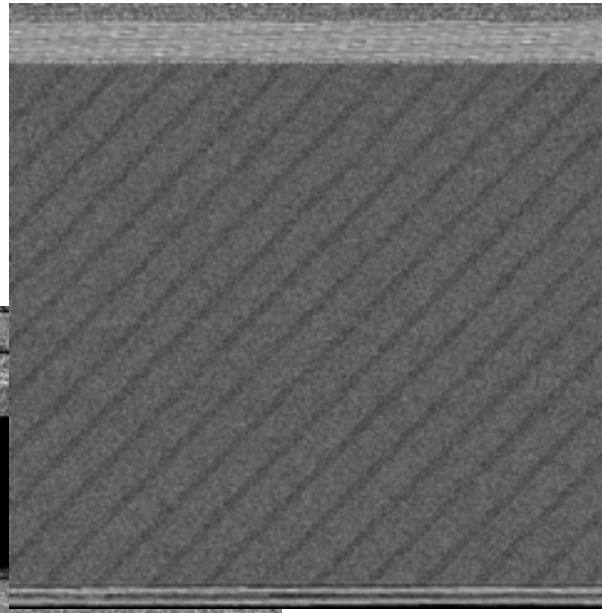
Kaggle Microsoft Malware Classification Challenge

```
.text:00401000 56              push   esi
.text:00401001 8D 44 24 08     lea    eax, [esp+8]
.text:00401005 50              push   eax
.text:00401006 8B F1           mov    esi, ecx
.text:0040100D C7 06 08        mov    dword ptr [esi]
                               offset off_42BB08
.text:00401013 8B C6           mov    eax, esi
.text:00401015 5E              pop    esi
.text:00401016 C2 04 00        retn   4
.text:00401019 CC CC CC        align  10h
.text:00401020 C7 01 08        mov     dword ptr [ecx],
                               offset off_42BB08
.text:00401026 E9 26 1C        jmp    sub_402C51
```

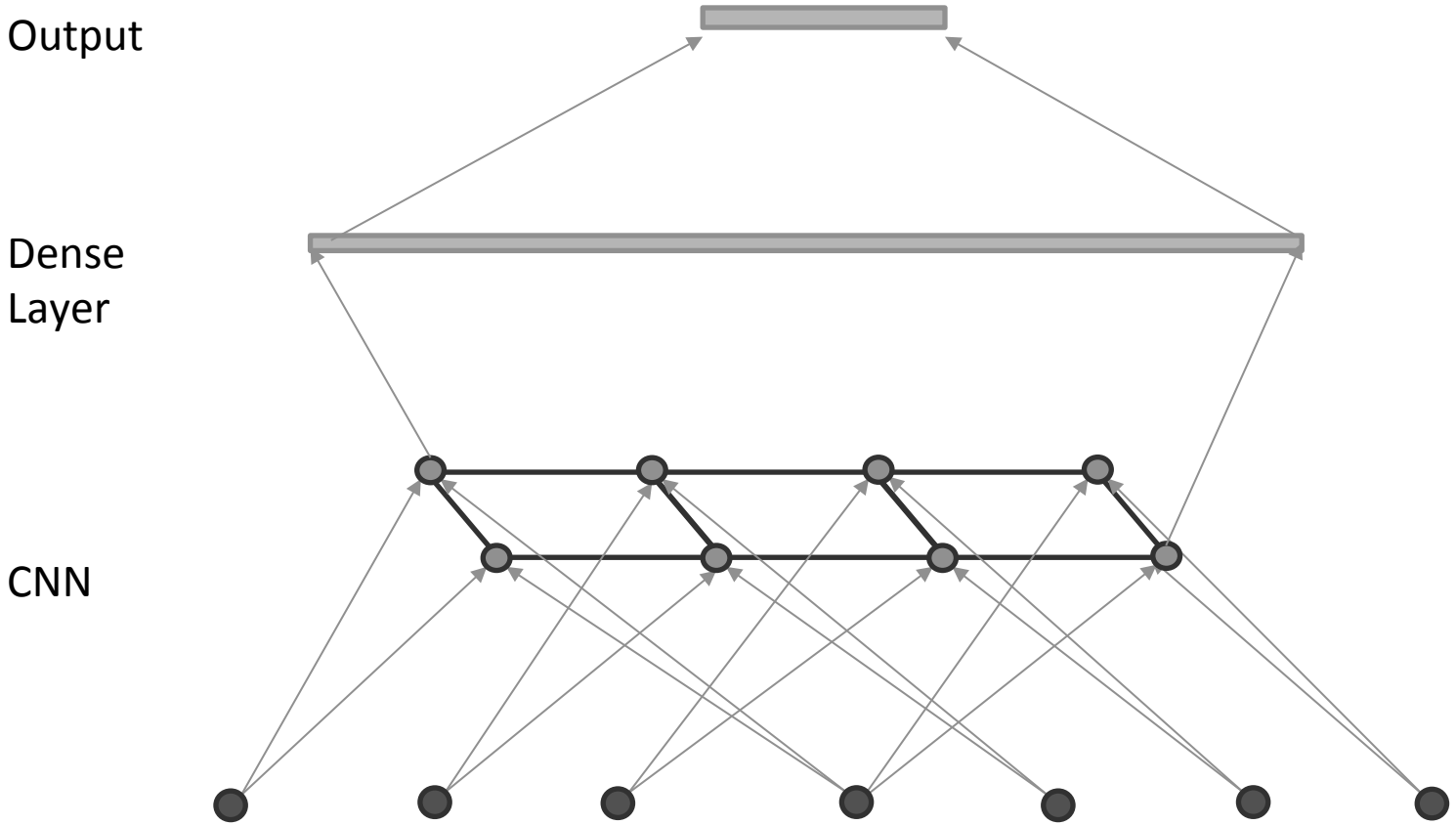
```
00401000 56 8D 44 24 08 50 8B F1
E8 1C 1B 00 00 C7 06 08
00401010 BB 42 00 8B C6 5E C2 04
00 CC CC CC CC CC CC CC
00401020 C7 01 08 BB 42 00 E9 26
1C 00 00 CC CC CC CC CC
00401030 56 8B F1 C7 06 08 BB 42
00 E8 13 1C 00 00 F6 44
00401040 24 08 01 74 09 56 E8 6C
1E 00 00 83 C4 04 8B C6
00401050 5E C2 04 00 CC CC CC CC
CC CC CC CC CC CC CC CC
00401060 8B 44 24 08 8A 08 8B 54
24 04 88 0A C3 CC CC CC
```

Kaggle Microsoft Malware Classification Challenge https://www.kaggle.com/c/malware-classification

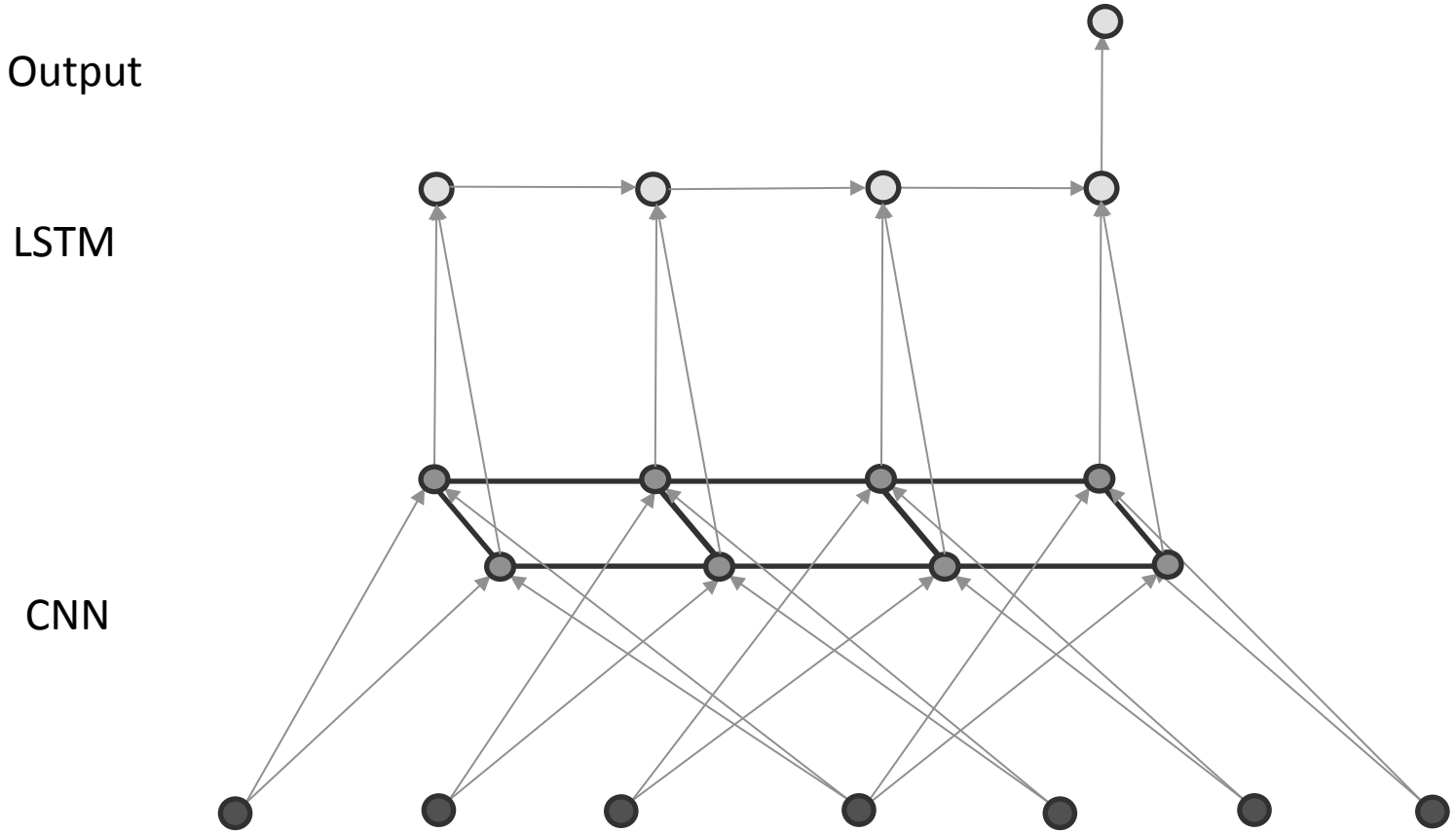



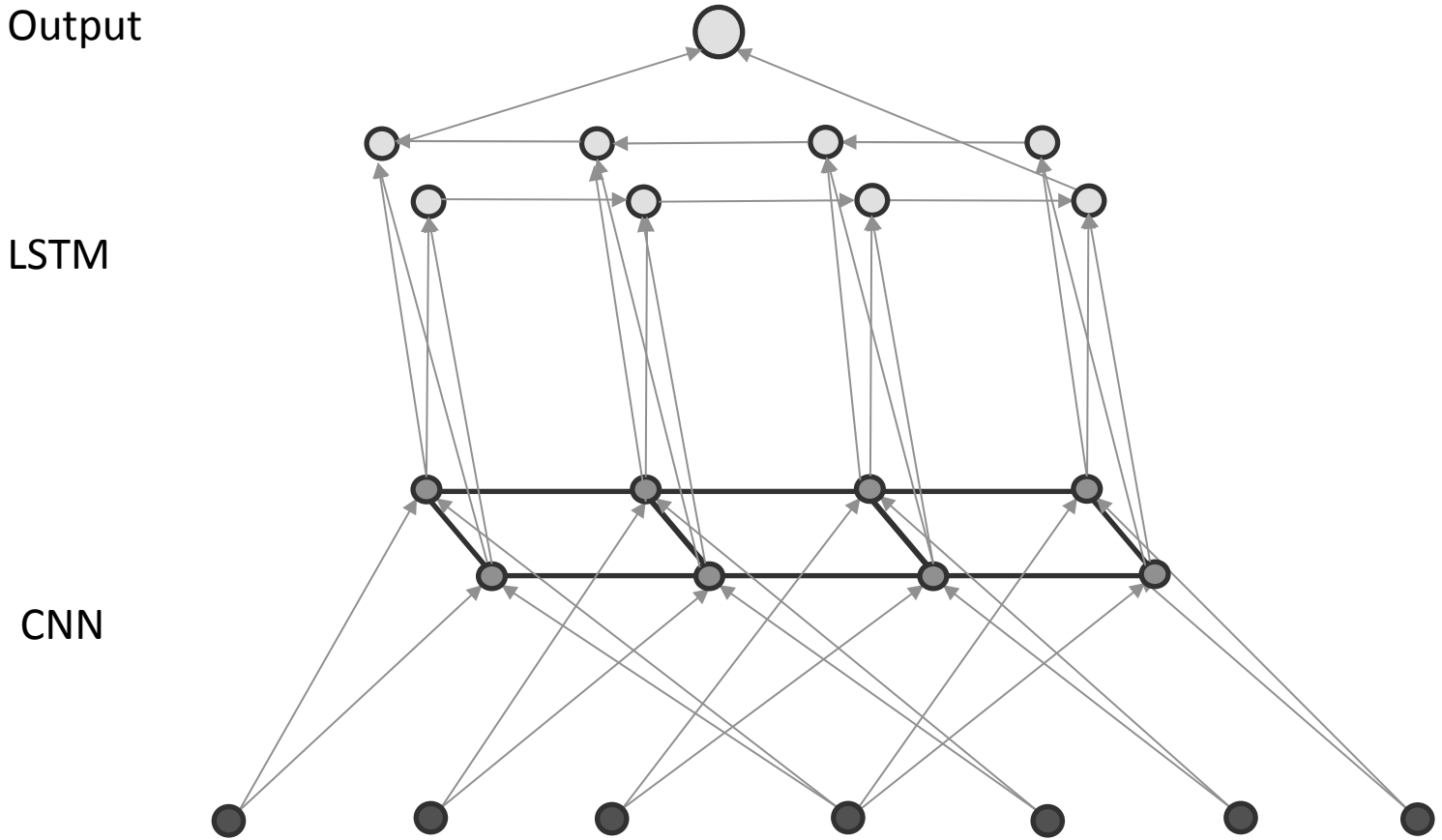
CNN Model



CNN – UniLSTM Model



CNN – BiLSTM Model



Results

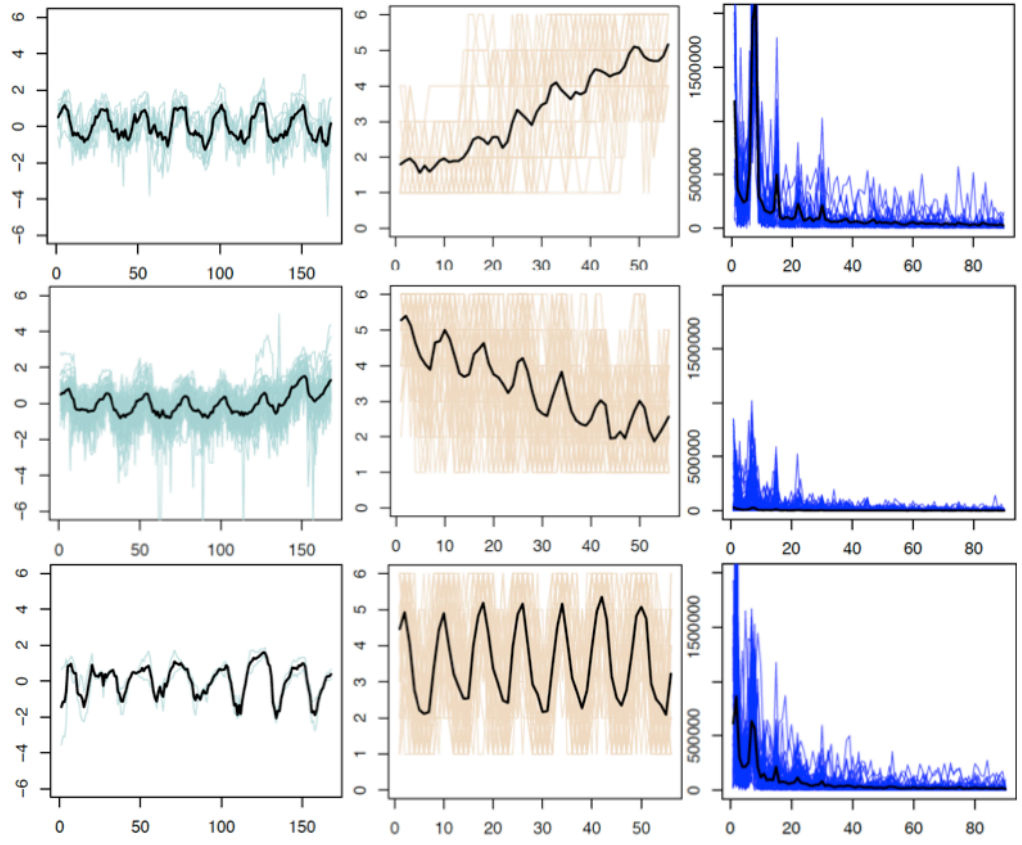
Deep Learning Configuration	Accuracy (%)	F1-score (%)
CNN (Default Sample)	95.10	92.14
CNN (Rebalanced Sample)	95.80	92.14
CNN UniLSTM (Default Sample)	97.64	94.15
CNN UniLSTM (Rebalanced Sample)	98.12	95.92
CNN BiLSTM (Default Sample)	97.91	95.52
CNN BiLSTM (Rebalanced Sample)	**98.20**	**96.05**

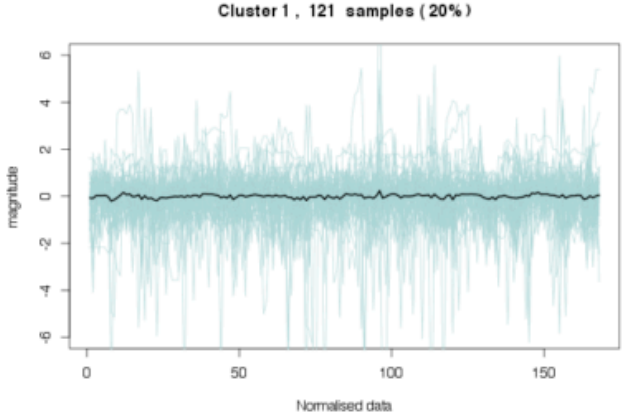
5 Fold Cross-Validation Experiment

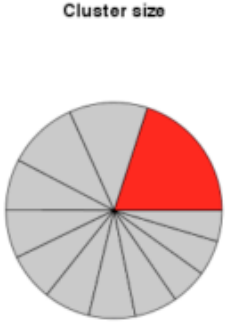
Predictive Maintenance

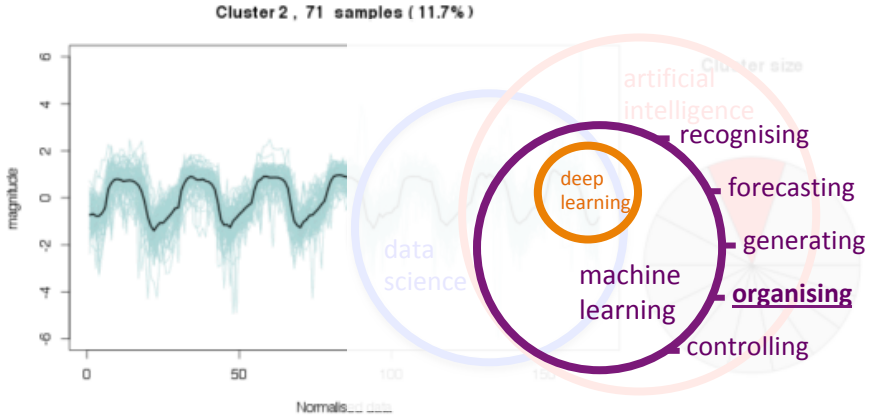
Seizure Detection

Generic Time Series Clustering









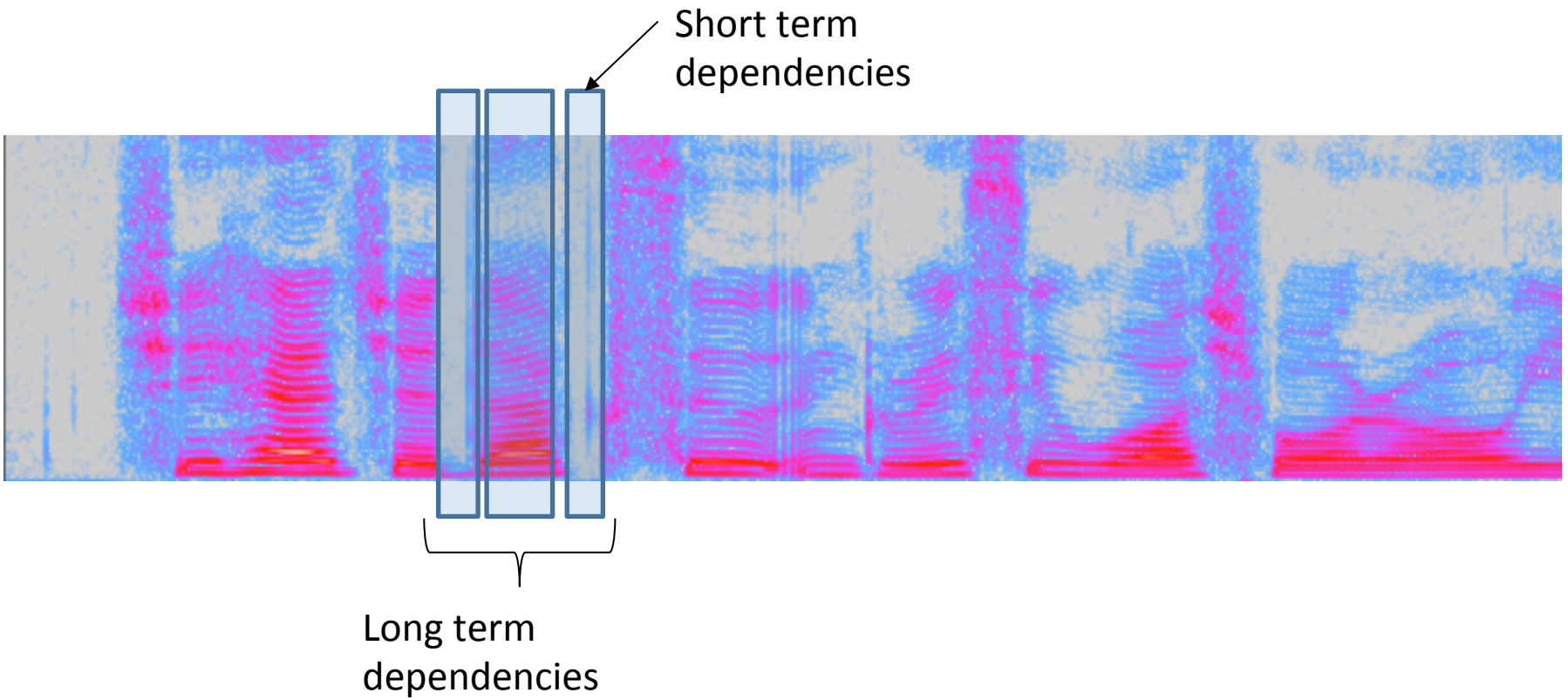
FLIRTING WITH AUTOML

Flirting With AutoML

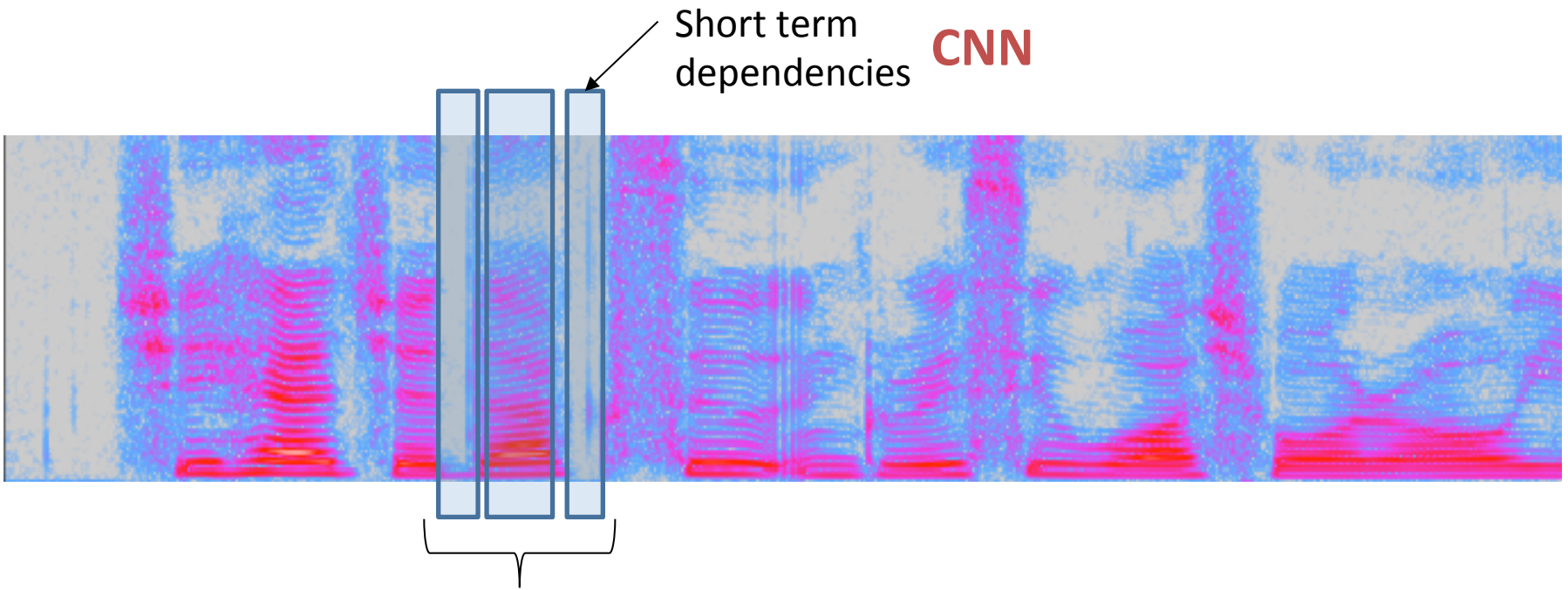
Opaque data is raw data when domain expertise is not avilable, where feature engineering has not been studied, or from newly released products and new domains

Can we build a generic solution that will work X% of the time with minimal tuning?

What Features To Model?



What Features To Model?



Collaborators

Ellen Rushe

Oisin Boydell

Quan Le

Luis Pechaun

Atif Qureshi

Jing Su

CeADAR Centre for Applied Data Analytics Research

Insight

Brian Mac Namee

@brianmacnamee

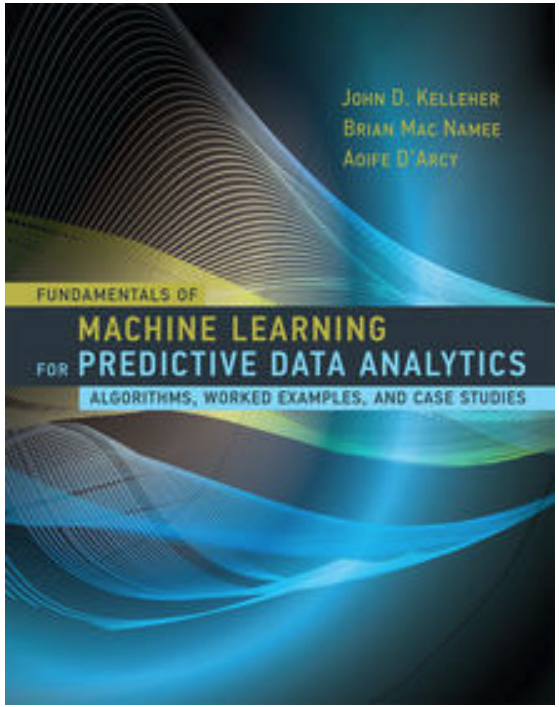
brian.macnamee@ucd.ie

University College Dublin
School of Computer Science

Insight
www.insight-centre.org

CeADAR
Centre for Applied Data Analytics Research
www.ceadar.ie

The Analytics Store
www.theanalyticsstore.ie



www.machinelearningbook.com