Introduction to machine learning

What is machine learning?

Herman Kamper

http://www.kamperh.com/

Let's write a spam filter

Simple first attempt in Python:

```
def check_spam(email):
    spam = False
    for word in email:
        if word in ["inheritance", "win", "uncle", ...]:
            spam = True
    return spam
```

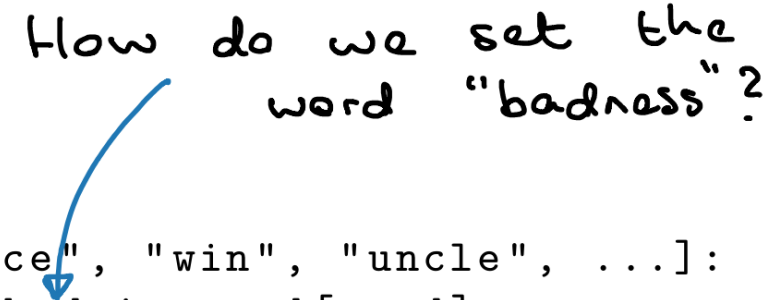
Let's write a spam filter

Simple first attempt in Python:

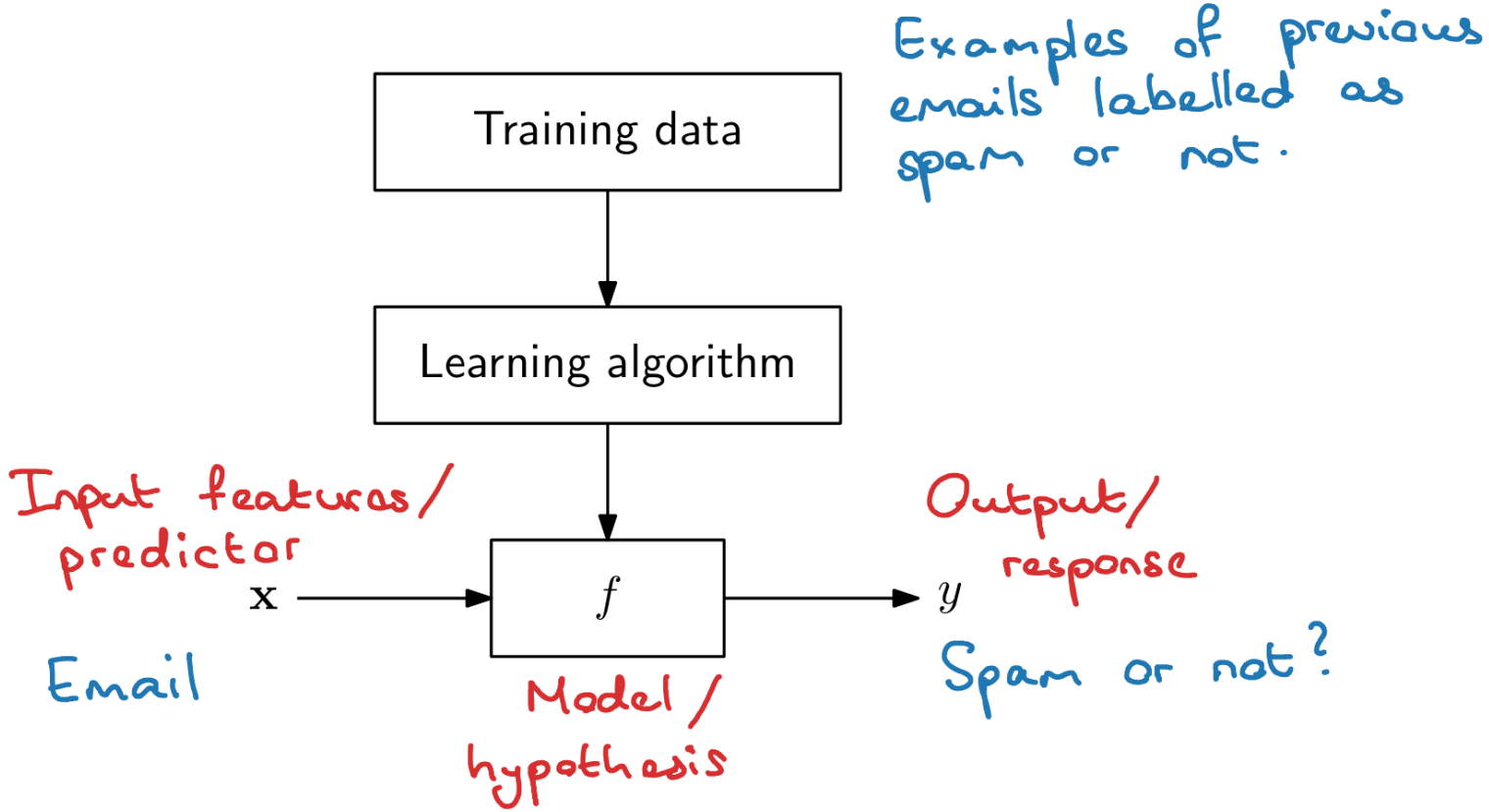
```
def check_spam(email):
    spam = False
    for word in email:
        if word in ["inheritance", "win", "uncle", ...]:
            spam = True
    return spam
```

More advanced attempt:

```
def check_spam(email):
    spam_score = 0
    for word in email:
        if word in ["inheritance", "win", "uncle", ...]:
            spam_score += how_bad_is_word[word]
    if spam_score > threshold:
        spam = True
    else:
        spam = False
    return spam
```



Machine learning



Two types of learning

- **Supervised learning:** Learn from labelled input-output $(\mathbf{x}, y)$ pairs.
- **Unsupervised learning:** Learn from unlabelled data, i.e. only have inputs $\mathbf{x}$.

Other types of learning: `https://en.wikipedia.org/wiki/Machine_learning#Approaches`

Machine learning, statistics, data science/analytics

Machine learning:

- "Field of study that gives computers the ability to learn without being explicitly programmed." [Arthur Samuel, 1959]
- "The goal of machine learning is to develop methods that can automatically detect patterns in data, and then to use the uncovered patterns to predict future data or other outcomes of interest." [Kevin P. Murphy]
- Often focusses on prediction.

Statistics:

- "Analysis, interpretation and presentation of data." [Wikipedia]
- More emphasis on explaining than prediction: "What does the data say?"

Data science/analytics:

- Combination of machine learning and statistics.
- Also engineering infrustructure for acquiring, storing and (pre-)processing data.

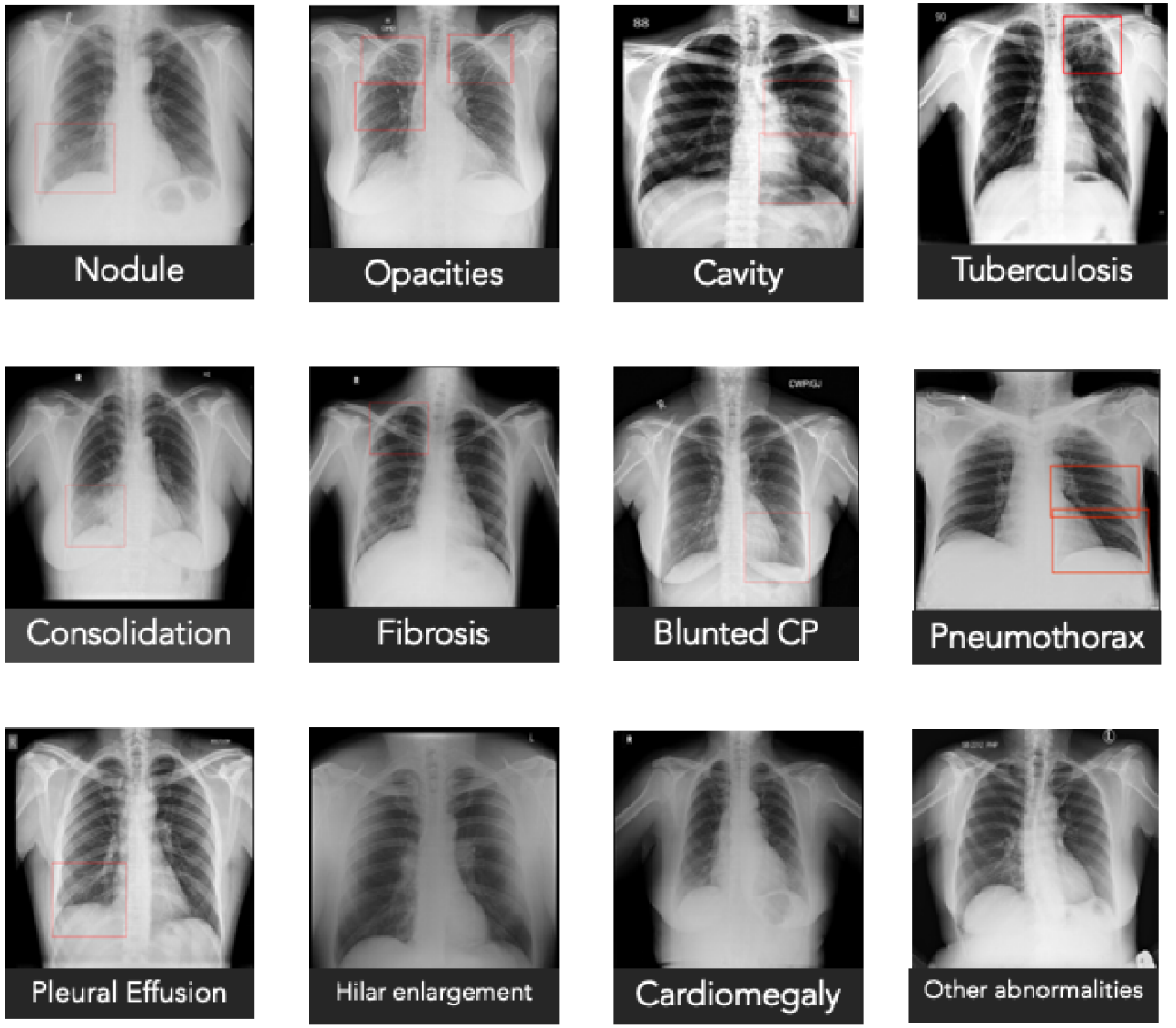
Introduction to machine learning

Supervised and unsupervised learning

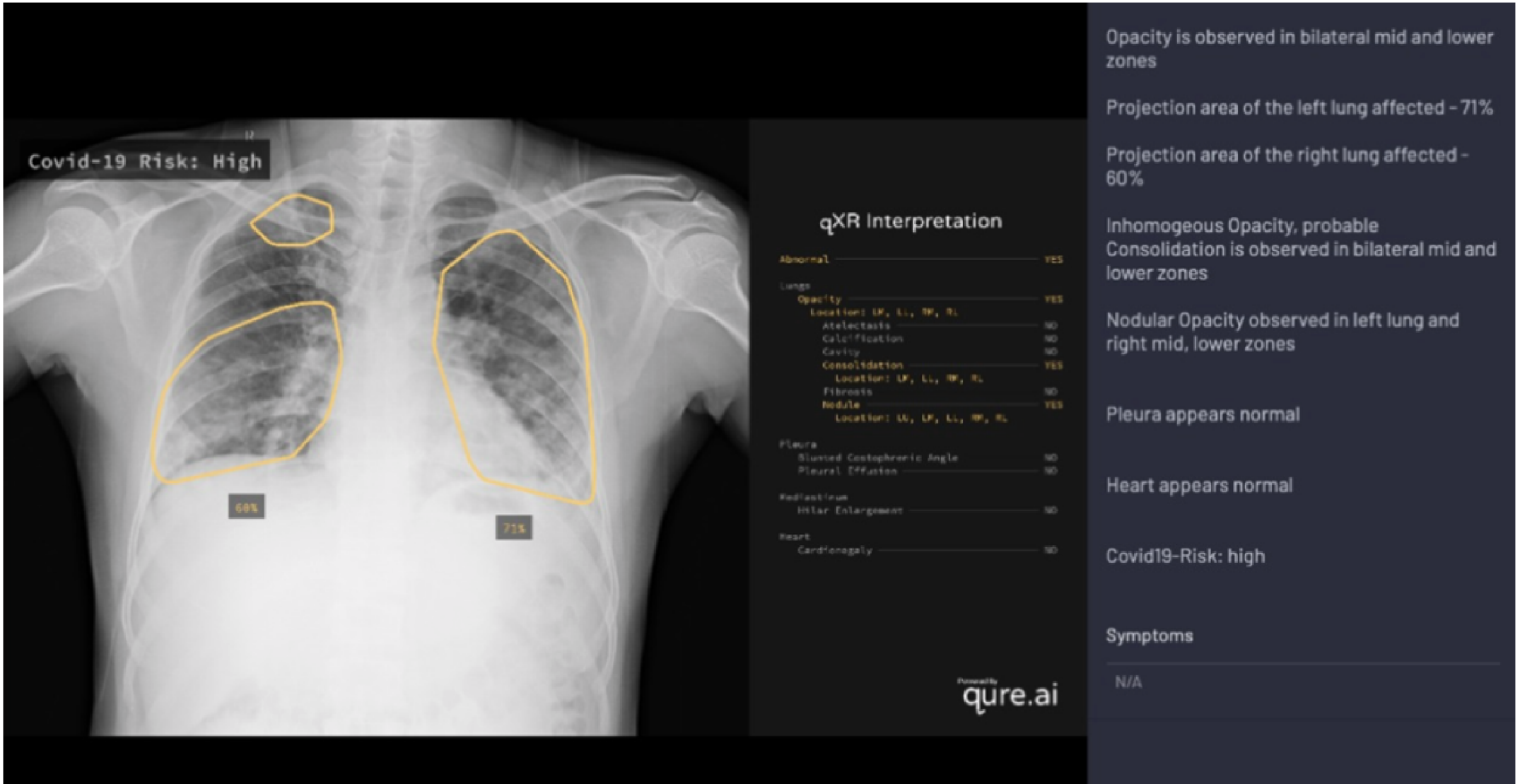
Herman Kamper

http://www.kamperh.com/

Supervised learning



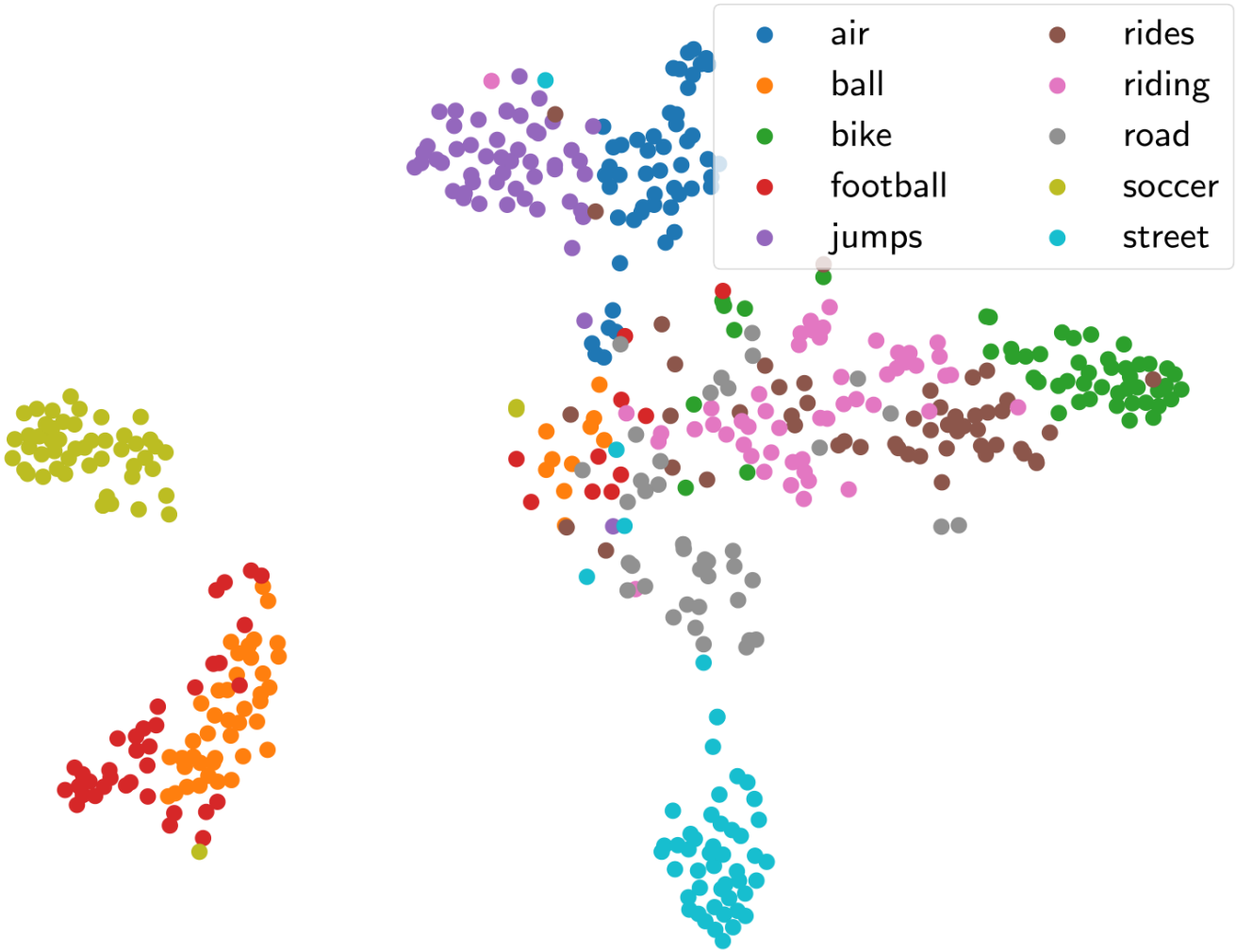
Source: https://www.technologyreview.com/2020/04/23/1000410/ai-triage-covid-19-patients-health-care/



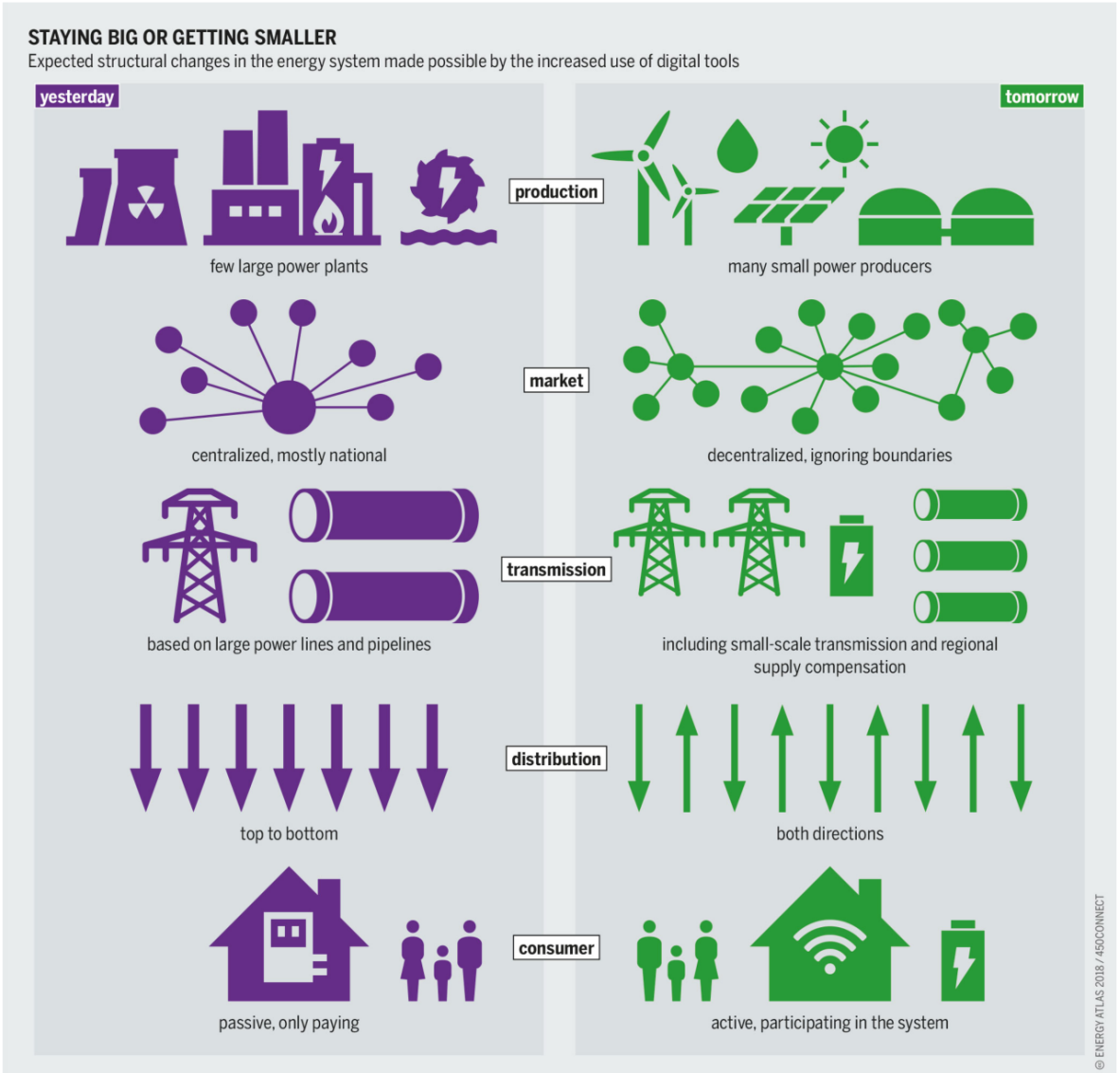
Source: https://www.technologyreview.com/2020/04/23/1000410/ai-triage-covid-19-patients-health-care/

Unsupervised learning

[Kamper et al., TASLP'19; https://arxiv.org/abs/1710.01949]



[Kamper et al., TASLP'19; https://arxiv.org/abs/1710.01949]



Source: https://en.wikipedia.org/wiki/Smart_grid