

CLASSIFYING PATENTS USING MACHINE LEARNING CIPHER CLASSIFICATION

THE BIG DATA PROBLEM IN PATENTS

Data is growing.

IBM says that
90% of the world's data has been
created in the last 2 years

There are
89m patents in the world

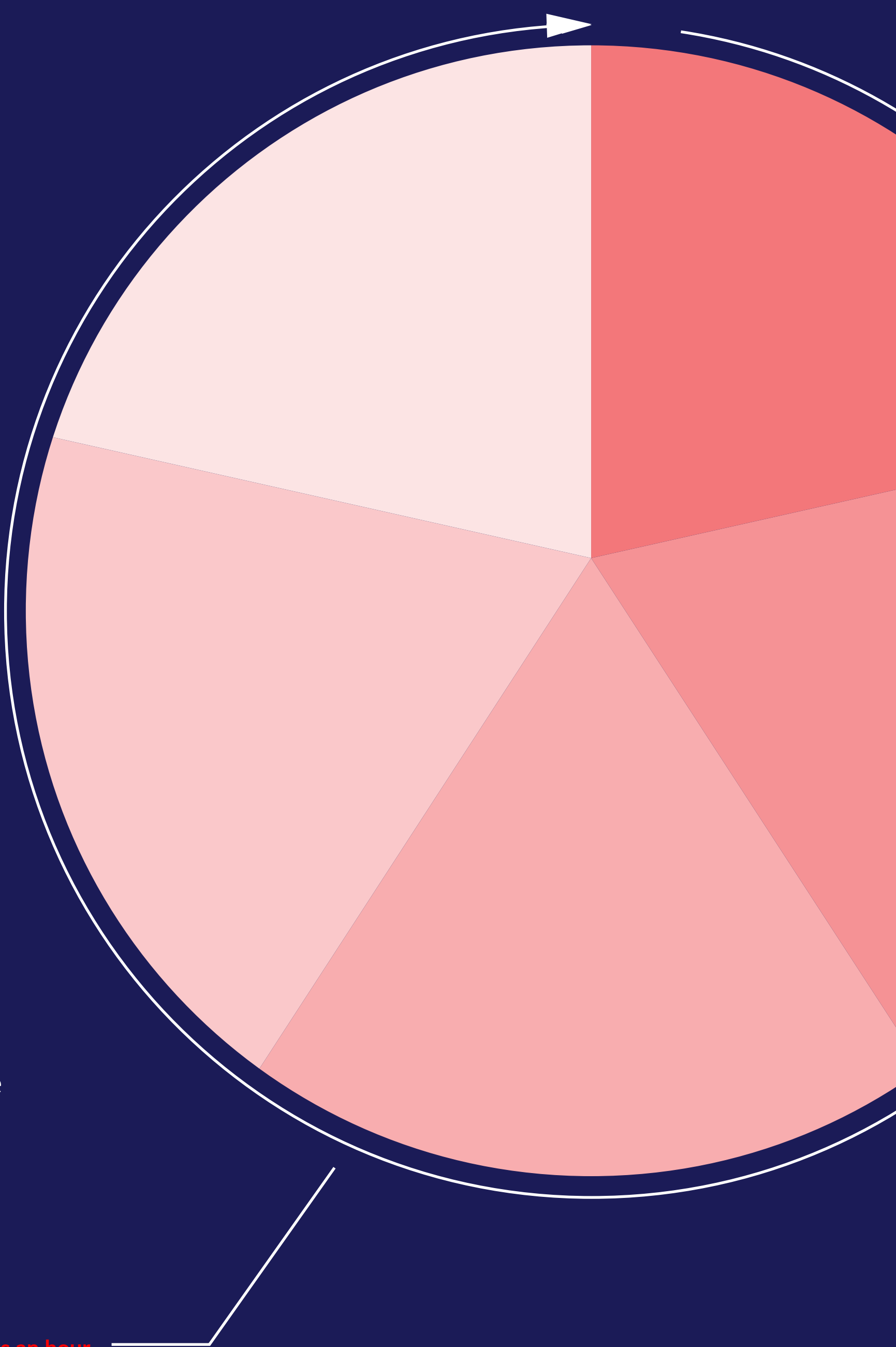
In 2019
3.8m new patents were filed,
up
35% from 2016

“
How can you know what patents
are relevant to you?”

Whether you are trying to
understand the competitive
landscape, benchmark against it,
review your portfolio, explore
monetisation options or perform
due diligence. It's unrealistic to be
able to read them all to find out.

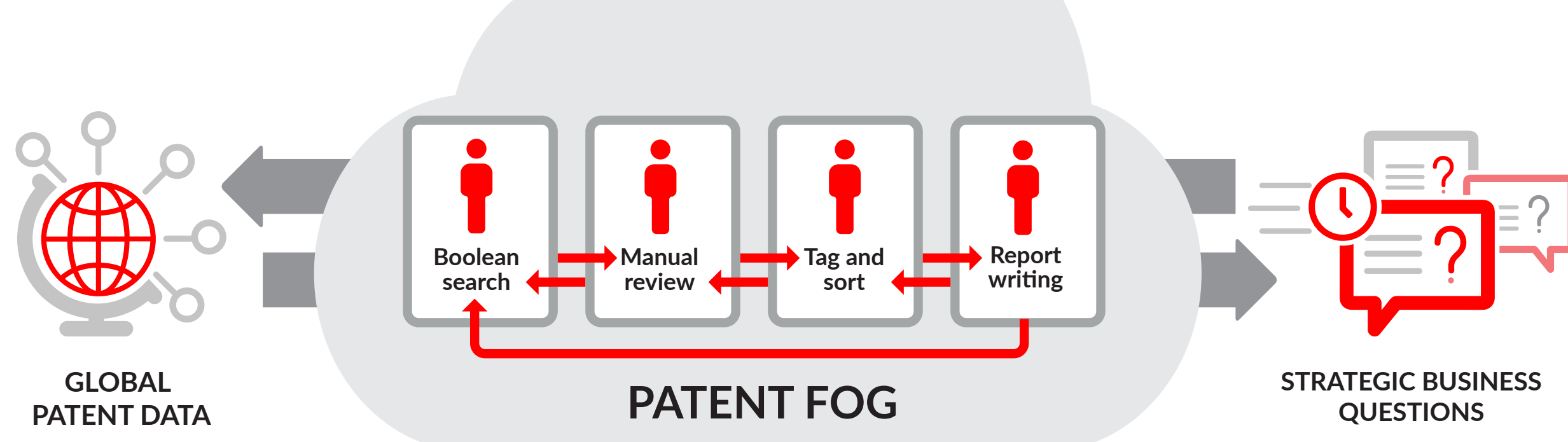
Cipher can help.

A human can read **5** patents an hour
Cipher can read **61 million** patents an hour



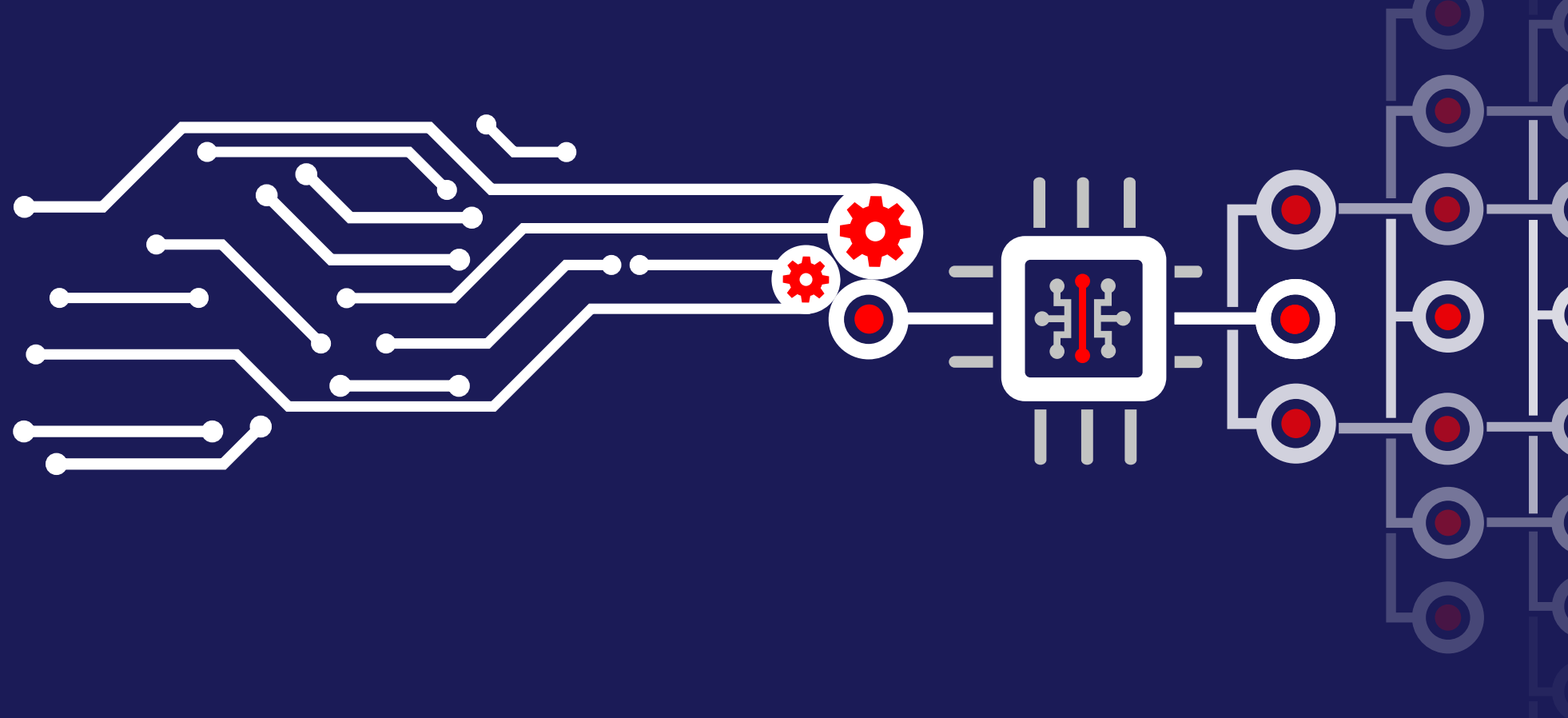
WHY CIPHER?

Manually classifying patents is too painful.



Cipher was created to automate the manual
process of sorting, tagging and classifying
patents to your view of the world.

CIPHER CAN CLASSIFY 61 MILLION PATENTS IN AN HOUR BECAUSE IT USES MACHINE LEARNING

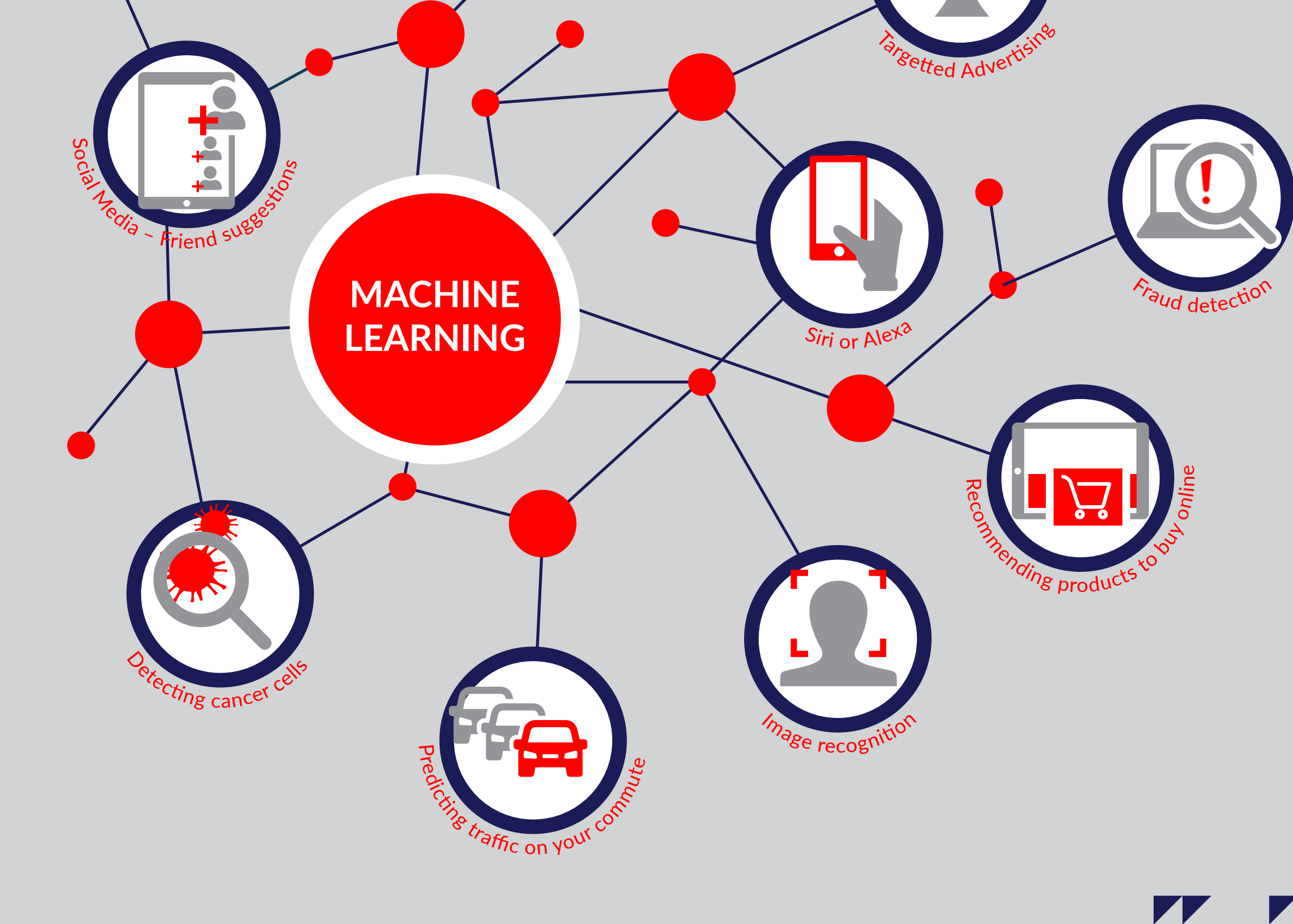


WHAT IS MACHINE LEARNING?

Machine learning is a way to
achieve human-like results,
by training algorithms how to
perform complex tasks instead
of explicitly coding them.

Machine Learning is
everywhere today

Gmail have stopped
99% of spam
through Machine Learning



HOW CAN YOU BE SURE MACHINE LEARNING WORKS TO CLASSIFY PATENTS?

“
How can we **trust** the machine?”

How can we see if the Cipher
Machine Learning algorithm
performs when classifying patents?

... We need to test it.



HOW TO ROBUSTLY TEST THE CIPHER MACHINE LEARNING ALGORITHM

This is what we did to test the Cipher
algorithm:

1. The patents from the data sets were
manually sorted
2. A small section of the data sets were used
to train the Cipher algorithm, then
3. The results delivered by the Cipher
algorithm were compared against the
manual results.

The following criteria had to be met to
deliver a robust test:

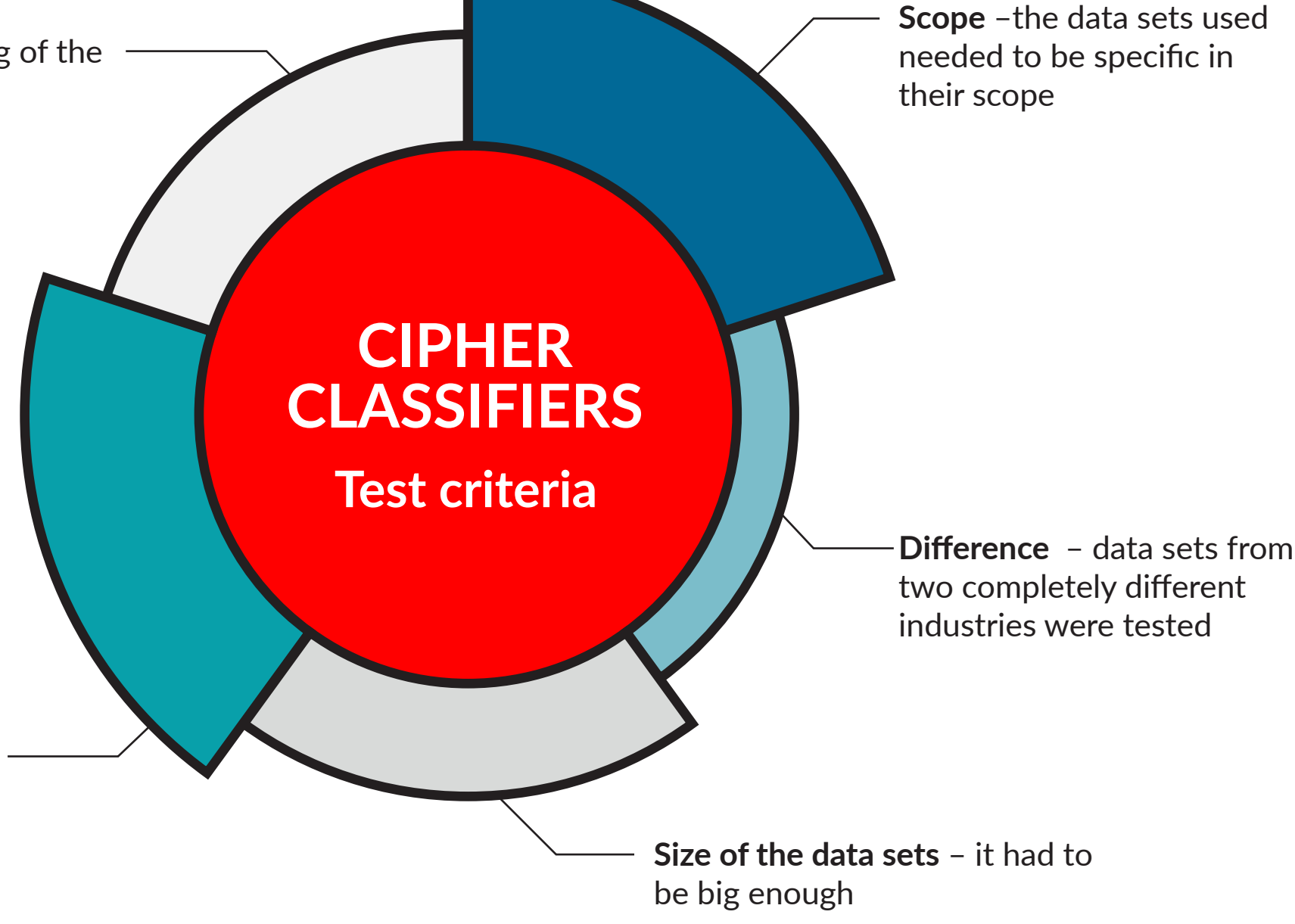
Independent - the manual sorting of the
patent data sets was done by an
independent consultant

Challenging - the data matter
had to be hard to classify

Scope - the data sets used
needed to be specific in
their scope

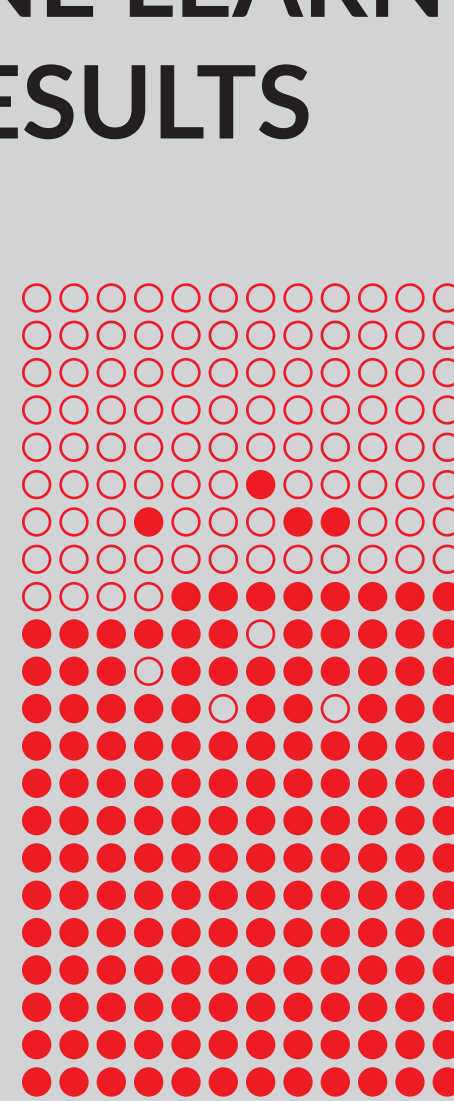
Difference - data sets from
two completely different
industries were tested

Size of the data sets - it had to
be big enough



THE CIPHER MACHINE LEARNING ALGORITHM TEST RESULTS

How did Cipher perform in the test to
classify patents?



Out of a data set of 300,
100 of the patents were relevant ○
and 200 were not related ●

Cipher sorted and classified these
patents and got it **96%** right

Under robust testing the Cipher algorithm
to classify patents **performs**

