



**SEARCH  
PARTY**

ASX Announcement  
ASX code: ALA  
February 11, 2016

## **SEARCH PARTY DATA SCIENTIST TO DELIVER KEYNOTE AT CHIEF DATA OFFICER FORUM**

Australian-based online recruitment marketplace [Search Party](#), announced that Head Data Scientist, Dylan Hogg will speak at the [Chief Data Officer Forum](#) in Sydney today, 11 February 2016, at the Amora Jamison Hotel.

Leading data and analytics executives from around the globe will share their insights on efficient data management and governance as well as effective data monetisation strategies. Search Party's Hogg will deliver a keynote presentation on the significance of leveraging machine learning for competitive advantage in the recruitment industry.

"Data science has become a key competitive advantage in many industries," said Search Party Data Scientist, Dylan Hogg.

"The case study I will present focuses on how Search Party's data science team built a scalable data processing platform that takes prototype machine learning algorithms to production quickly, thereby adding automation to the recruitment process for our customers.

"This represents a cutting edge approach to modern data science. Today, both the algorithms and underlying technology are advancing at a rapid pace. Adopting new methods and tools can lead to a competitive advantage for many businesses," said Hogg.

Leveraging machine learning is key in scaling Search Party's recruitment marketplace and positioning it for further global expansion. Since its beta launch in October 2014, Search Party has helped define a new way of recruiting across Australia, the UK and Canada. Search Party's marketplace of over 15 million agency-represented professional candidates is empowering businesses to hire candidates six times faster at 60 per cent less cost than traditional methods.



**SEARCH  
PARTY**

It is also supporting recruiters by giving them access to new customers at no cost, using their existing database of candidates.

According to CIGNA's Global Head of Enterprise Data Management, Don Gray on the CIGNA website: "The CDO Forum advances thought leadership on how to most effectively blend executive leadership and data science in today's information-driven business environment."

The emergence of big data and analytics has demonstrated the importance of the Chief Data Officer's role in modern business. Whilst the organisational structure is variable, the CDO serves as a bridge between the CIO and the business, often moving around the business to extract the most value from the company's data. This three day event will address some of the most pressing issues CDOs face today.

### **About Applabs Technologies Ltd**

Applabs Technologies Ltd (ASX:ALA) is the first ASX listed app development and technology financier. The company offers a unique business model of both an app developer and a technology financier, enabling the company to produce apps for customers as well as the ability to fund, develop and market exciting new leading edge apps internally.

### **About The Search Party**

[Search Party](#) is an online recruitment marketplace. For employers it means hiring 6X faster at 60% less cost than traditional methods without sacrificing candidate quality. For recruiters, it means getting new customers at no cost, using their database of candidates. The company was founded in Sydney with offices in London and Toronto. Search Party also owns JobAdvisor, the Australian employer review site. For more information, visit [thesearchparty.com](http://thesearchparty.com) & [jobadvisor.com.au](http://jobadvisor.com.au).

### **For further Information, please contact:**

Director  
Charles Thomas  
[Charles@applabs.com.au](mailto:Charles@applabs.com.au)

CEO – Search Party  
Ben Hutt  
[Ben@thesearchparty.com](mailto:Ben@thesearchparty.com)



**SEARCH  
PARTY**



[facebook.com/TheSearchPartyPlatform](https://facebook.com/TheSearchPartyPlatform)



[twitter.com/TheSearchParty](https://twitter.com/TheSearchParty)



[linkedin.com/company/the-search-party](https://linkedin.com/company/the-search-party)



[jobadvisor.com.au/the-search-party](https://jobadvisor.com.au/the-search-party)



**SEARCH  
PARTY**

# **LEVERAGING MACHINE LEARNING FOR COMPETITIVE ADVANTAGE**

Dylan Hogg - Head of Data Science  
February 2016

# OUTLINE

- About me and Search Party
- Scalable Data Platform: Hadoop and Spark
- Machine Learning: What and Why?
- Two Different Applications of Machine Learning

# ABOUT ME

- Head of Data Science at Search Party (3 years)
- Software Engineering, Database and BI experience (15 years)
- Postgrad study in Machine Learning from UNSW (2012) & B.Comm from Auckland University (1999)

# ABOUT SEARCH PARTY

- A marketplace that connects employers, recruiters and candidates
- Employers search millions of anonymised candidates and engage with selected recruiters to fill a vacancy



- Offices in Australia, UK and Canada

# SEARCH PARTY MISSION

Leverage cutting edge technology to evolve recruitment into an efficient digital process that provides a better outcome for all three parties involved - recruiters, employers and job seekers.



SEARCH  
PARTY

# SCALABLE DATA PLATFORM

# SEARCH PARTY DATA PLATFORM

15 million candidate  
resumes

From over 800  
recruiters

With over 80 million  
role descriptions

## Joe Bloggs

[Address, City, ST ZIP Code] | [Telephone] | [Email]

### Objective

- Check out the quick tips below to help you get started. To replace tip text with your own, just click it and start typing.

### Education

[DEGREE] | [DATE EARNED] | [SCHOOL]

- Major: [Click here to enter text]
- Minor: [Click here to enter text]
- Related coursework: [Click here to enter text]

[DEGREE] | [DATE EARNED] | [SCHOOL]

- Major: [Click here to enter text]
- Minor: [Click here to enter text]
- Related coursework: [Click here to enter text]

### Skills & Abilities

#### MANAGEMENT

- Need another experience or education entry? You got it. Just click in the second sample entry for either and then click the plus sign that appears.

#### SALES

- On the Design tab of the ribbon, check out the Themes, Colors, and Fonts galleries to get a custom look with just a click.
- Looking for a matching cover letter? All you had to do was ask! On the Insert tab, select Cover Page.

#### COMMUNICATION

- You delivered that big presentation to rave reviews. Don't be shy about it now! This is the place to show how well you work and play with others.

#### LEADERSHIP

- Are you president of your fraternity, head of the condo board, or a team lead for your favorite charity? You're a natural leader—tell it like it is!

### Experience

[JOB TITLE] | [COMPANY] | [DATES FROM - TO]

- This is the place for a brief summary of your key responsibilities and most stellar accomplishments.

[JOB TITLE] | [COMPANY] | [DATES FROM - TO]

- This is the place for a brief summary of your key responsibilities and most stellar accomplishments.

Job titles

Job descriptions

Skills

Company names

Universities

Executive summaries

Qualifications

Industries

# SCALABLE DATA PLATFORM

- Cloudera distribution of Hadoop hosted with Rackspace
- Adopted Spark in 2014 for fast data processing
- One platform for prototype and production algorithms
- Spark has Scala, Java, Python and R language support





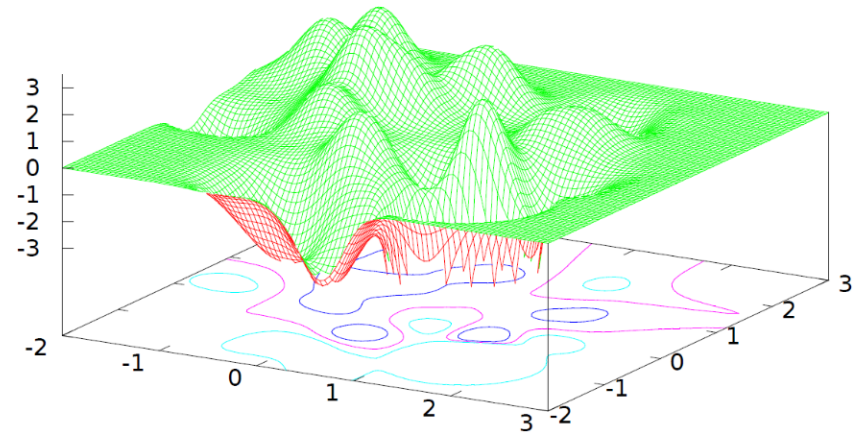
**SEARCH  
PARTY**

# **MACHINE LEARNING**

**(a branch of artificial intelligence)**

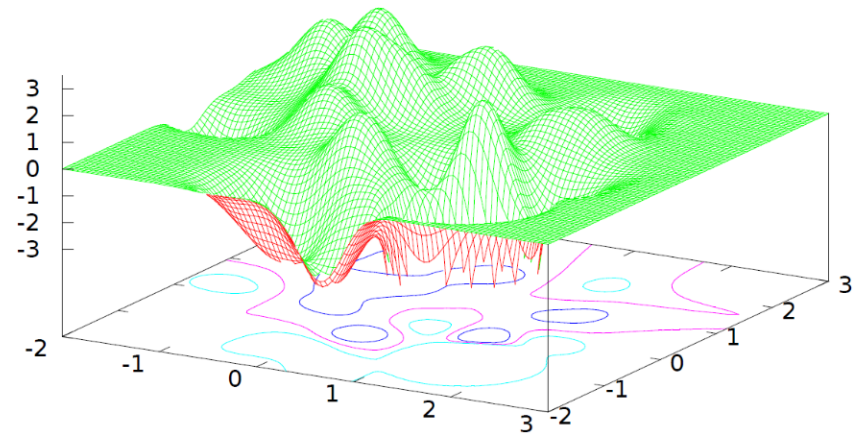
# WHAT IS MACHINE LEARNING?

- Origins in computer science and statistics
- Algorithms that can learn a specific task from data without being explicitly programmed
- Supervised learning to make predictions
- Unsupervised learning to discover patterns



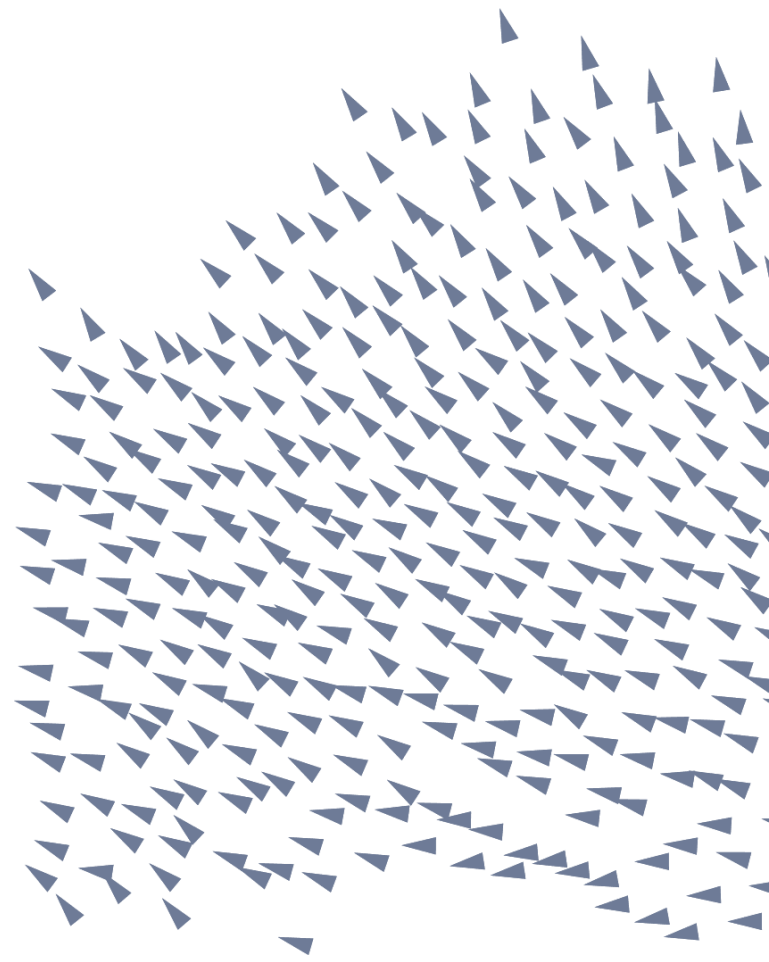
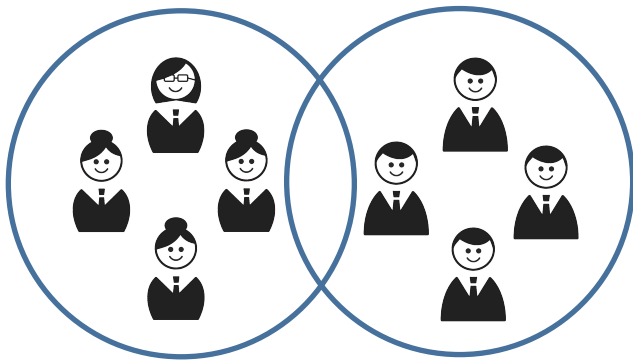
# WHY MACHINE LEARNING?

- Machine learning provides a set of tools to extract value and insights from data
- Creates a barrier to entry with data driven products and services
- With more data, machine learning models can become more accurate



# MACHINE LEARNING

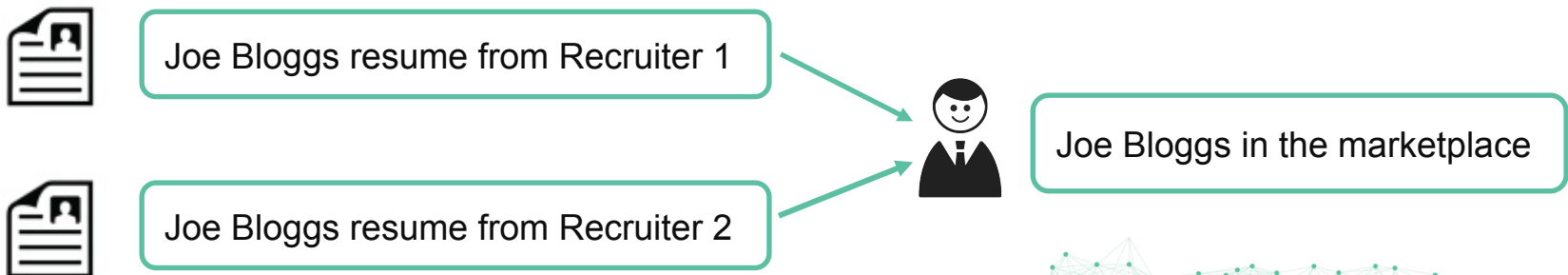
## Algorithm 1: Detecting Duplicate candidates



# DETECTING DUPLICATE CANDIDATES

## - THE PROBLEM

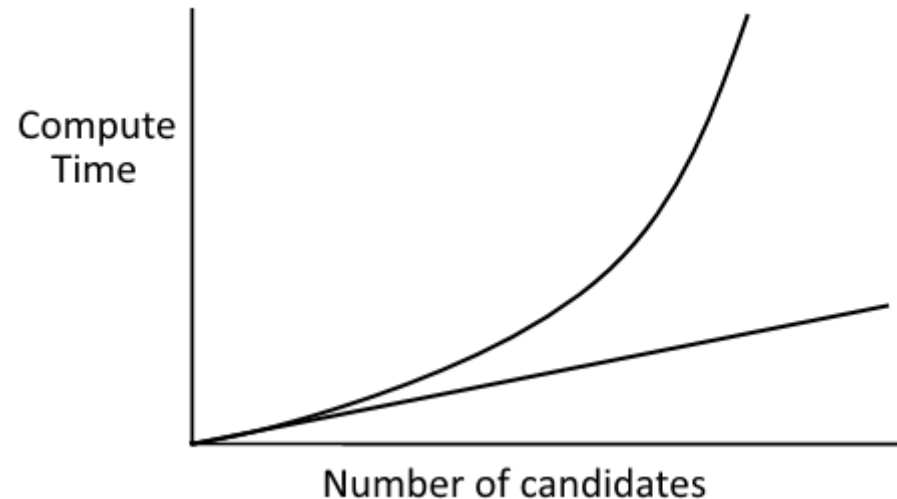
- Why does Search Party need to detect duplicate candidates?
  - Candidates may be represented by multiple recruiters
  - We resolve duplicate candidate resumes to a single candidate in our marketplace



# DETECTING DUPLICATE CANDIDATES

## - APPROACH

- The naïve implementation scales poorly with candidate number
- Several versions of “smart” implementations
- Final version: custom clustering algorithm in Spark
  - De-duplicate 15M candidates in 9 hours on 8 node cluster
  - Algorithm considers full name, email, phone, skills and list of employers



# DETECTING DUPLICATE CANDIDATES

## - EXAMPLE

Email	Full Name	Phone	Skills	Employers
j.blogs@gmail.com	Joe Bloggs	0409 123 456	Budgeting External Audit Forecasting General Ledger Intuit Quicken	Wrigley Company Pty Ltd PepsiCo Australia Holdings Austbrokers Holdings Pty
jblogs@hotmail.com	Joseph Bloggs	0409 555 555	Accruals External Audit General Ledger	PepsiCo Australia Holdings Austbrokers Holdings Pty
jblogs@hotmail.com	Joe Bloggs	+61 0409555555	Accruals General Ledger	AUSTBROKERS HOLDINGS PTY LTD

# DETECTING DUPLICATE CANDIDATES

## - ALGORITHM

1. Tokenise candidate fields e.g. “Dylan” -> dy, yl, la, an
2. Compute vectors from tokens for each candidate field (using TD-IDF)
3. Fast Canopy Clustering<sup>1</sup> on candidate full name only
4. Slower Correlation Clustering<sup>2</sup> on each canopy using all candidate fields
5. Main platform uses results to resolve duplicates

<sup>1</sup> McCallum et al (2000) "Efficient Clustering of High Dimensional Data Sets with Application to Reference Matching"

<sup>2</sup> Elsner et al (2009) "Bounding and Comparing Methods for Correlation Clustering Beyond ILP"

# HOW DOES CANDIDATE “DEDUPE” GIVE SEARCH PARTY A COMPETITIVE ADVANTAGE?

- ✓ Solves an important business problem of grouping duplicate candidates in marketplace search results
- ✓ This ability is a barrier to entry for potential competitors
- ✓ Improves candidate reporting and analysis
- ✓ Coming soon: Allowing recruiters to de-duplicate their CRM database



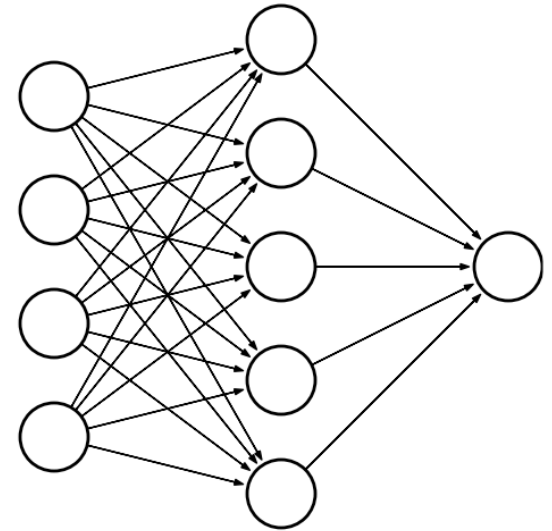
SEARCH  
PARTY

# DEEP LEARNING

(a branch of machine learning)

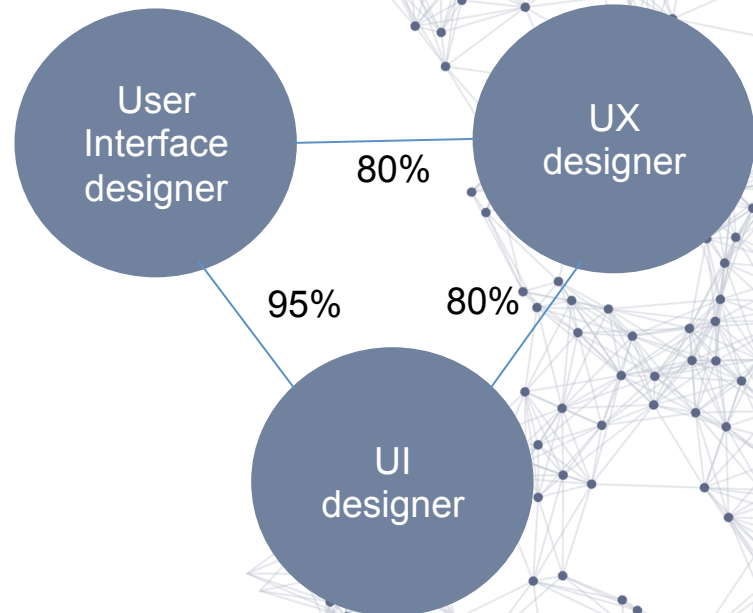
# WHAT IS DEEP LEARNING?

- Neural networks with many hidden layers
- Use of GPU processors for performance
- Requires lots of training data
- Slow to train, fast to make predictions



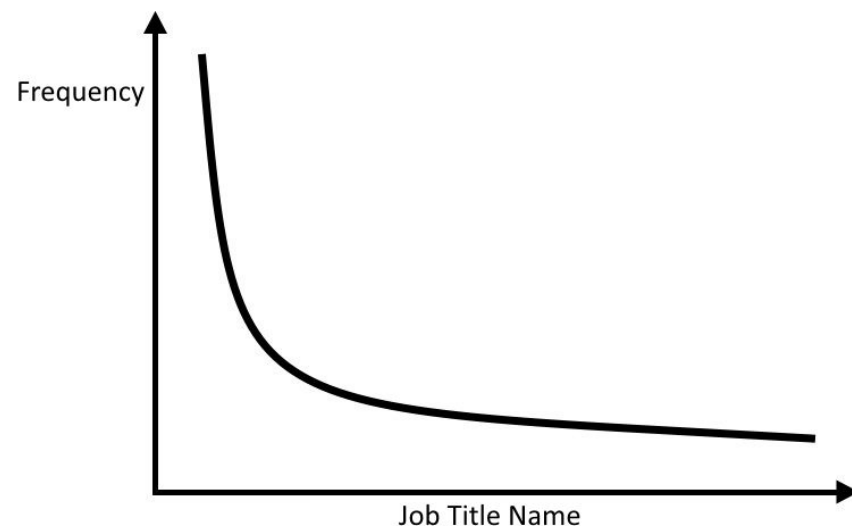
# DEEP LEARNING

## Algorithm 2: Job Title Resolution



# JOB TITLE RESOLUTION - THE PROBLEM

- 6 million distinct job titles extracted from resumes
- Most common 25 job titles cover 10% of records
- Want to resolve distinct titles that make up the long tail
- E.g. “Chief Data Officer - Europe” with “Interim CDO” with “CDO - Financial Services”



# JOB TITLE RESOLUTION - TRAINING DATA

- Training data from a partner who has millions of public vacancies
- Training example: Raw text vacancy description with corresponding job title

## **Vacancy Description**

We are seeking a Chief Data Officer who can manage standard data operating procedures, data quality standards as well as being able to understand how to combine different data sources within the organisation with each other.



**Chief Data Officer**

# JOB TITLE RESOLUTION - PREDICTION

Prediction Output: Probabilities for a list of job titles

Chief Data Officer - Asia

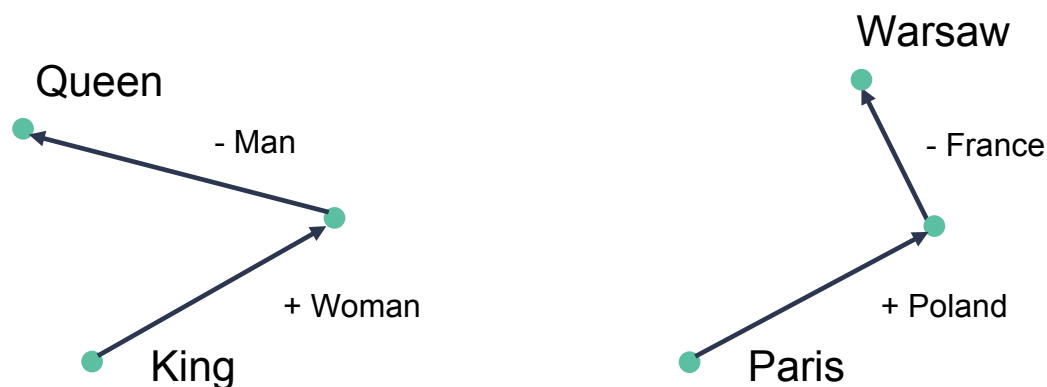


Probability	Suggested Job Title
80%	Chief Data Officer
75%	CDO
60%	Data Officer
50%	Data Quality Officer
40%	Senior Data Analyst

# JOB TITLE RESOLUTION

## - PREPARING INPUTS

- Neural network input must be numerical vectors representing words
- Create word vectors using Google's word2vec<sup>1</sup> algorithm
- Enables vector arithmetic on words:



<sup>1</sup> Mikolov et al (2013) "Efficient Estimation of Word Representations in Vector Space"

# JOB TITLE RESOLUTION

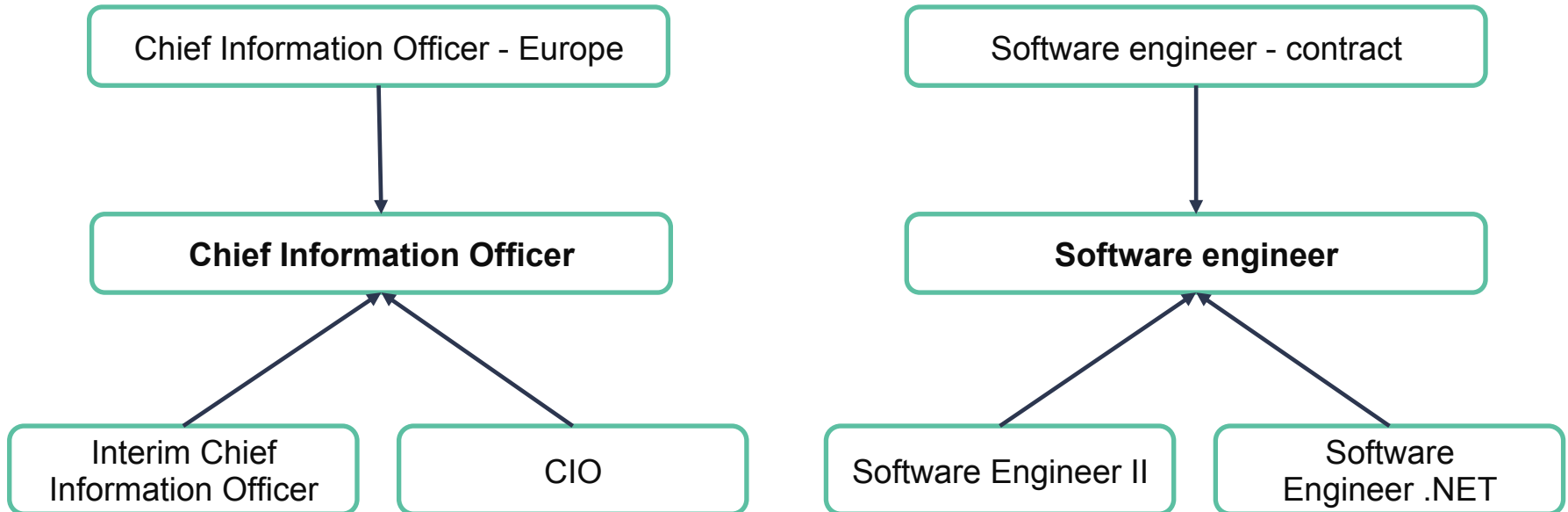
## - DEEP LEARNING NETWORK

- Training the network learns parameters that allows it to to correctly make predictions
- Training uses a vacancy description and correct job title as well as an incorrect job title
- Adapted into a ranking problem to try and rank the correct job title above the incorrect one

# JOB TITLE RESOLUTION

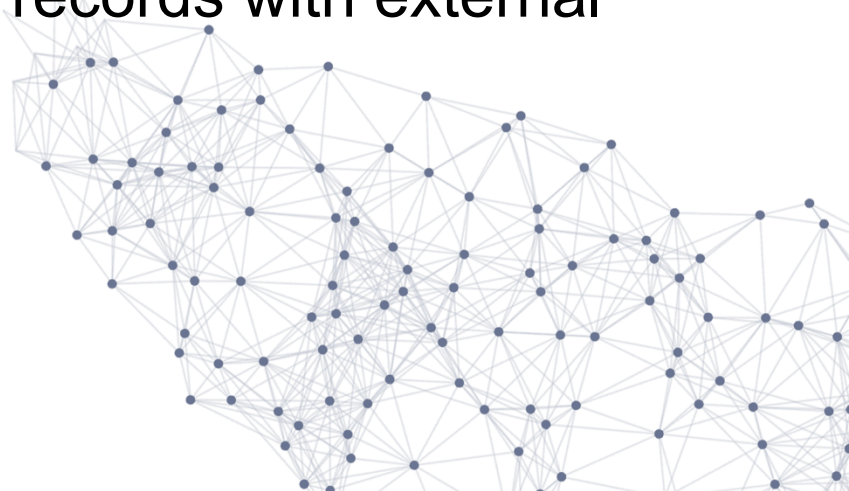
## - EXAMPLES

Resolution handles abbreviations, re-orderings, synonyms and misspellings.



# HOW DOES JOB TITLE RESOLUTION GIVE SEARCH PARTY A COMPETITIVE ADVANTAGE?

- ✓ Improvements to our candidate search engine via improved understanding of user search queries
- ✓ Increased chance of a successful candidate placement for a vacancy
- ✓ Data cleansing enables linking records with external data sources



# SUMMARY

- Machine learning algorithms give Search Party the ability to extract value from our data and build great data products
- These products provide a better outcome for all parties involved - recruiters, employers and job seekers
- Our scalable machine learning algorithms increase the barrier to entry for any potential competitors
- As our data grows, supervised machine learning algorithms improve and this directly benefits our platform

# KEY TAKEAWAY

Algorithms are as important as Data  
and are a key competitive advantage for  
your organisation



# SEARCH PARTY

Questions?  
Connect with me...



[www.linkedin.com/in/dylanhogg](http://www.linkedin.com/in/dylanhogg)



[@dylanhogg](https://twitter.com/dylanhogg)



[www.thesearchparty.com](http://www.thesearchparty.com)