MANCHESTER 1824
The University of Manchester

Choosing a Data Science PhD: is it even for you?

Professor Christopher Yau

University of Manchester
Health Data Research UK

About Me

Professor of Artificial Intelligence at the University of Manchester.

I trained as an Engineer but then started working in Statistics as a PhD student.

My core research is in methodological development:

1. Bayesian Statistics.

2. Adaptations of generic machine learning approaches to real world applications.

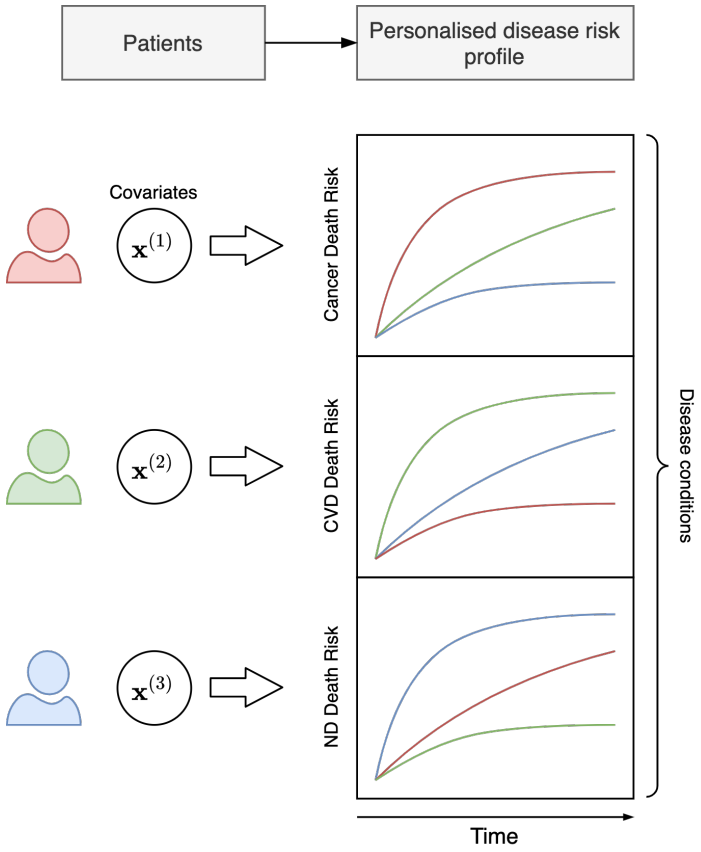
Survival analysis

Time-to-event modelling.

Understand relationship between time-to-event distributions and patient-specific characteristics.

- Learning patient-specific (competing) risk profiles.
- Relax (strong) classical assumptions (e.g. Cox models).
- Function across a range of small and large data.

The fundamental problem is **probability density estimation**.



Nonparametric density estimation

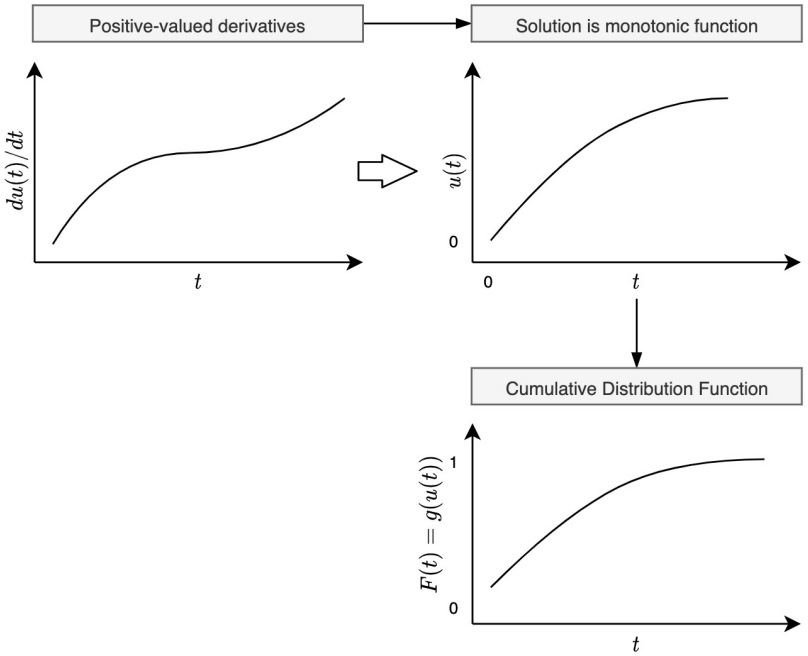
A valid **cumulative distribution function (CDF)** needs to satisfy three conditions:

1. Starts at zero.
2. Monotonicity.
3. Does not exceed one.

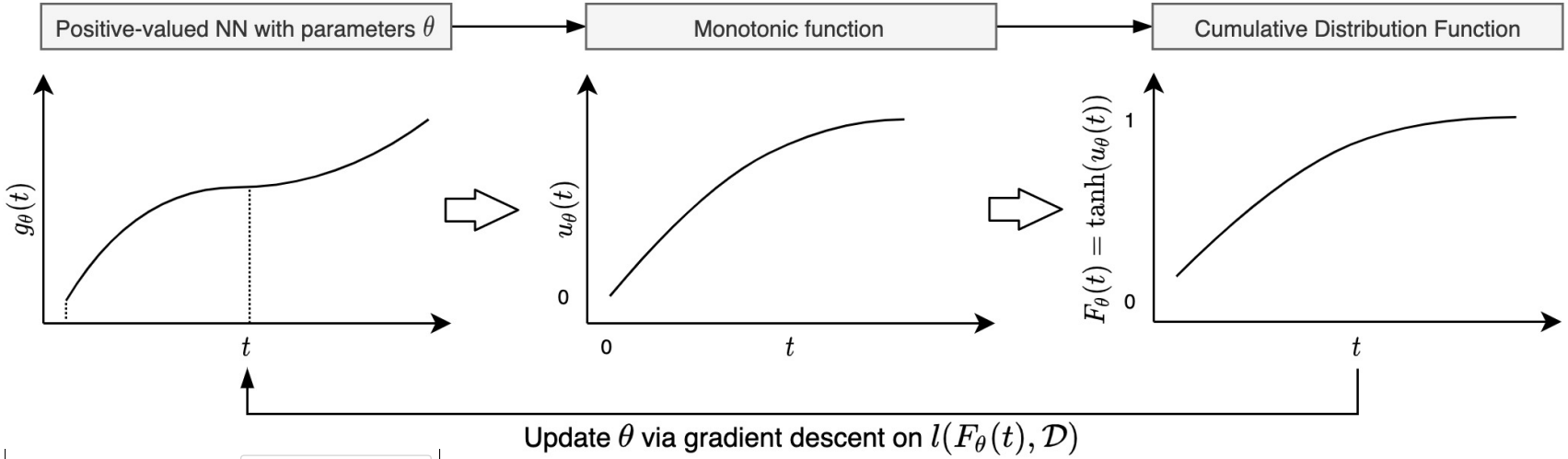
We want to learn a CDF so we need a model that has **large support** over valid CDFs with minimal additional constraints.

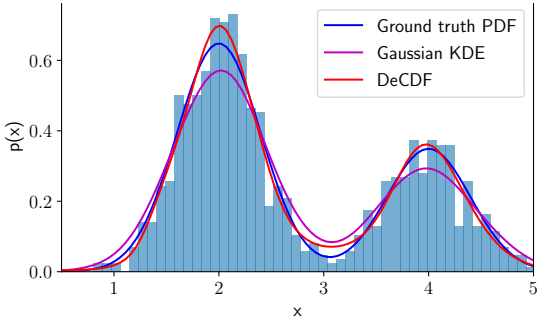
Classic idea

The solution of certain classes of **Ordinary Differential Equations** correspond to probability density functions.



Nonparametric CDF estimation





This "*Neural ODE*" construction allows us to model complex PDFs/CDFs with little functional constraints.

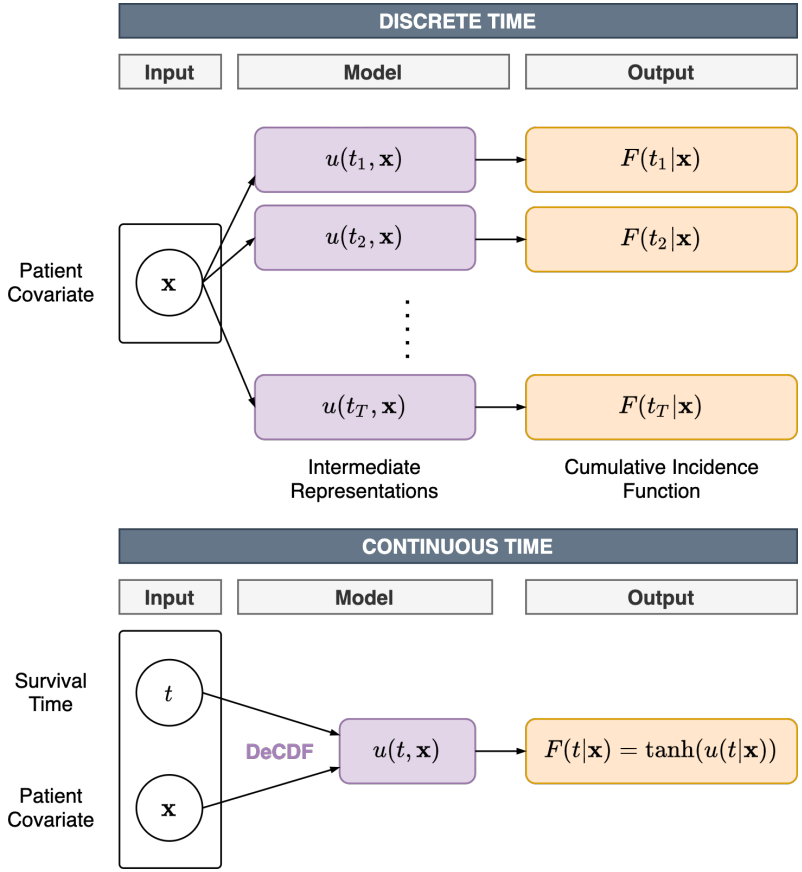
DeSurv Architecture

This construction allow us to:

1. Develop a **continuous time model** which avoids the need to specify a grid of time points for estimation.

2. Reduces **model complexity** (number of parameters) and improves stability.

3. Unifies a classical statistical model with modern ML machinery.



To appear in AISTATS 2022

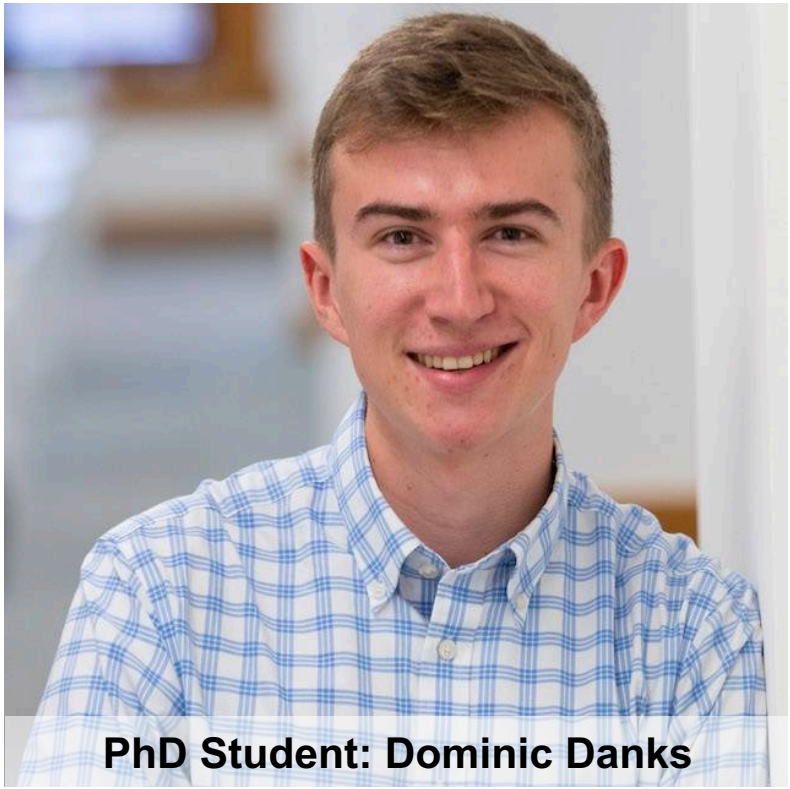
https://github.com/DeSurvival/DeSurv

Who does the work?

As an academic, I rarely do the work myself.

The majority of my research is performed by my PhD students.

The development of PhD students is something I have a particular interest in …



PhD Student: Dominic Danks

Health Data Research UK

The National Institute for Health Data Research.

I direct the PhD Programme in Health Data Science.

Recruit around 8-10 PhD students a year.

Partnership across 7 UK Universities and numerous industry partners.

Overview

In this talk, I will discuss:

- the relevance of PhDs in today's Data Science,
- tips on how to identify the right opportunities and
- how advanced studies in data science can lead to more choice and opportunities in your later career.

How many of you have considered/are doing a PhD in Data Science?

(where “Data Science” refers to anything in which the analysis of data forms the predominant part of the work/project)

What do you gain from a PhD in Data Science?

What do you gain from doing a PhD in Data Science?

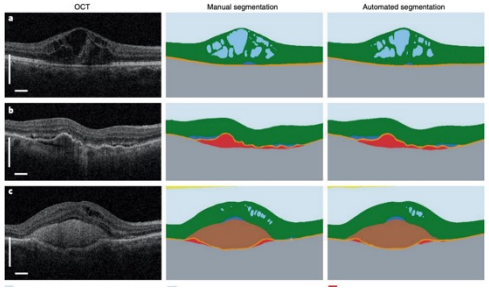
Traditional Academic View

1. A concentrated period of time for advanced study and learning.
2. Foundational knowledge in theory, methodologies and observational processes.
3. Technical skills (including programming) particularly emerging, non-industry standard ones.
4. Broader knowledge to connect to different areas.

Modern Popular View

1. Papers in top-tier data science conferences.
2. Learn to do “Deep Learning”.
3. Access internships at tech companies.
4. Opportunities to start your own business.

The rise of artificial intelligence



ARTICLES
https://doi.org/10.1038/s41591-018-0107-6
nature medicine

Clinically applicable deep learning for diagnosis and referral in retinal disease

Jeffrey De Fauw[1], Joseph R. Ledsam[1], Bernardino Romera-Paredes[1], Stanislav Nikolov[1], Nenad Tomasev[1], Sam Blackwell[1], Harry Askham[1], Xavier Glorot[1], Brendan O'Donoghue[1], Daniel Visentin[1], George van den Driessche[1], Balaji Lakshminarayanan[1], Clemens Meyer[1], Faith Mackinder[1], Simon Bouton[1], Kareem Ayoub[1], Reena Chopra[2], Dominic King[1], Alan Karthikesalingam[1], Cían O. Hughes[1,3], Rosalind Raine[3], Julian Hughes[2], Dawn A. Sim[2], Catherine Egan[2], Adnan Tufail[2], Hugh Montgomery[2], Demis Hassabis[1], Geraint Rees[3], Trevor Back[1], Peng T. Khaw[2], Mustafa Suleyman[1], Julien Cornebise[1,3,4], Pearse A. Keane[2,4]* and Olaf Ronneberger[1,4]*

The volume and complexity of diagnostic imaging is increasing at a pace faster than the availability of human expertise to interpret it. Artificial intelligence has shown great promise in classifying two-dimensional photographs of some common diseases and typically relies on databases of millions of annotated images. Until now, the challenge of reaching the performance of expert clinicians in a real-world clinical pathway with three-dimensional diagnostic scans has remained unsolved. Here, we apply a novel deep learning architecture to a clinically heterogeneous set of three-dimensional optical coherence tomography scans from patients referred to a major eye hospital. We demonstrate performance in making a referral recommendation that reaches or exceeds that of experts on a range of sight-threatening retinal diseases after training on only 14,884 scans. Moreover, we demonstrate that the tissue segmentations produced by our architecture act as a device-independent representation; referral accuracy is maintained when using tissue segmentations from a different type of device. Our work removes previous barriers to wider clinical use without prohibitive training data requirements across multiple pathologies in a real-world setting.

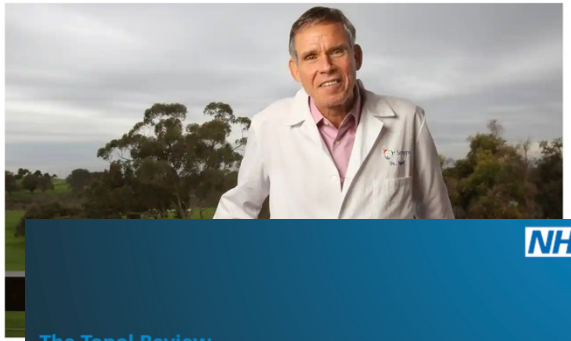
Google enters health

Interview

Cardiologist Eric Topol: 'AI can restore the care in healthcare'

Nicola Davis

The doctor, geneticist and author talks about his new book on the future of our relationship with medicine

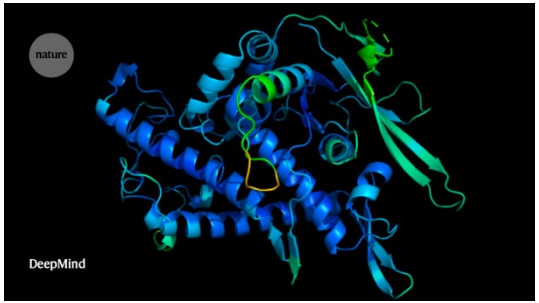


NHS

The Topol Review

Preparing the healthcare workforce to deliver the digital future

Mainstream advocacy



nature

Explore content · Journal information · Publish with us · Subscribe

nature › news › article

NEWS · 30 NOVEMBER 2020

'It will change everything': DeepMind's AI makes gigantic leap in solving protein structures

Google's deep-learning program for determining the 3D shapes of proteins stands to transform biology, say scientists.

Transformational research

Distorted expectations

How competitive is admission to top PhD programs in Computer Vision? What do they look for? Do you already need a conference publication like CVPR or NeurIPS to be admitted, e.g., MPI or EPFL or MILA?

[D] If the number of machine learning PhD graduate is increasing rapidly, wouldn't it get exponentially harder to be hired at machine learning related jobs without PhD?

[D] Why has machine learning become such a toxic field, know-it-all field?

Discussion

How I Got a Job at DeepMind as a Research Engineer (without a Machine Learning Degree!)

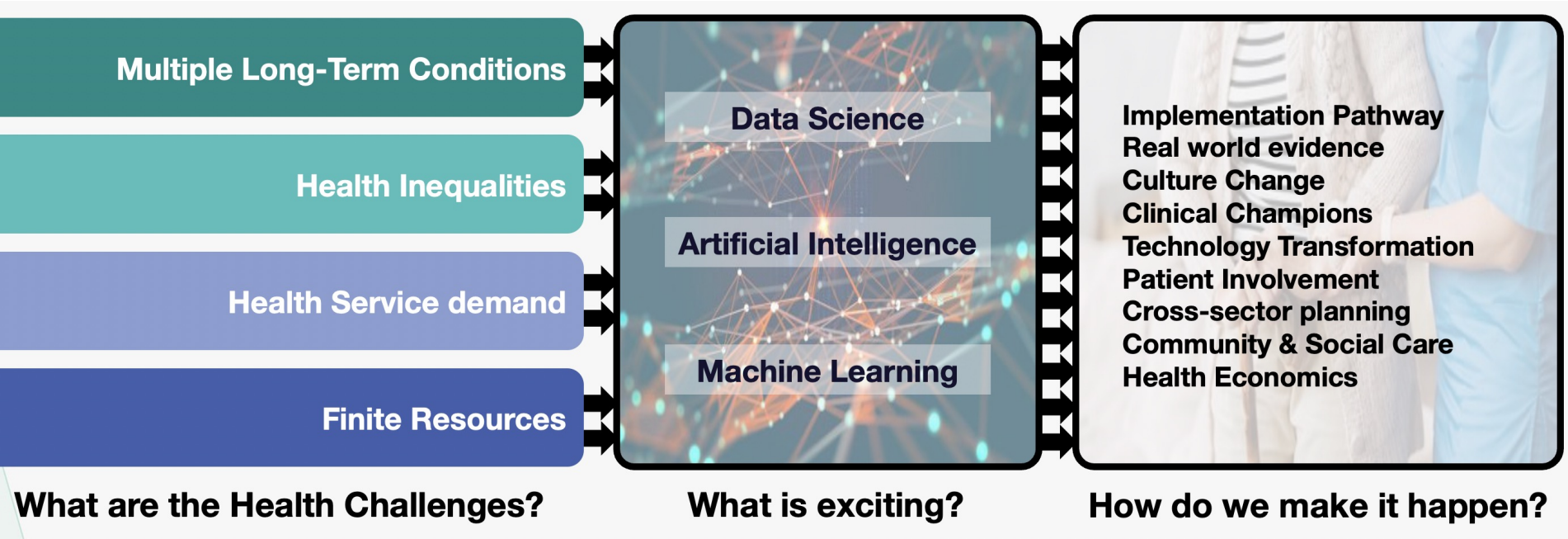
kgrounds, but none come close to
vibe from the machine learning
g the whole field.

What do you gain from doing a PhD in Data Science?

Industry perspective

1. Training in areas we have no internal expertise in.
2. More matured employees (oral communications, writing, etc).
3. Knowledge of working with certain types of (restricted) data types.
4. Innovation (R&D).

A PhD can help address gaps in expectations



There can be a mismatch between the perceived and actual needs of industry.

What might be the right PhD(/Postdoc) opportunity for you?

Components of a PhD

Topic

Supervisor

Environment

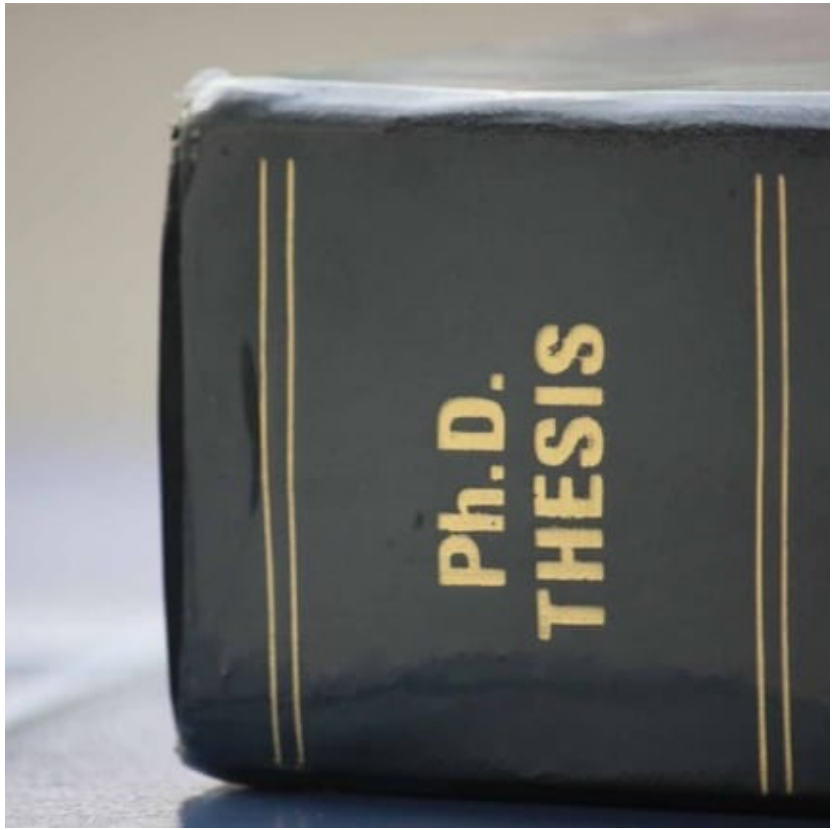
The Topic

The high-level goal of a PhD probably matters less than the methods and skills involved.

Will you learn something new?

Will the research will give you knowledge and skills longevity?

Is there flexibility and scope for expansion or retargeting of your research?



The Supervisor

Is the PhD supervisor actually a data scientist?

Do they have a track record in the areas you are interested in?

Have they worked with data before?

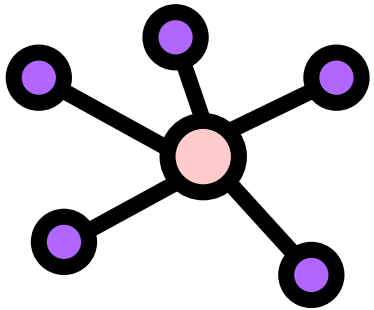
Do they collaborate with others?

What type of supervisor are they?

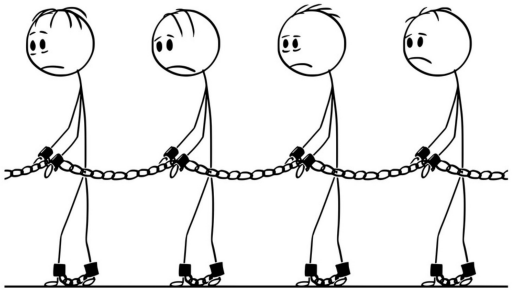
Master-Apprentice

Team

Invisible Man



Hub and Spoke



Slavery

The Environment

Is it about the reputation of the university or the people?

Does the department have established long-term relationships with industry?

What have PhD graduates gone on to do next?

Does the physical infrastructure exist to support you?

Is there a group of people all working at a high standard?

How can a PhD advance your career options?

State of the art skills today, are old skills tomorrow

I'm a data scientist/software developer and I keep longing for a simpler life. I'm getting tired of the constant need to keep up to date, just to stay in the game. Christ if an electrician went home and did the same amount upskilling that devs do to stay in the game, they'd be in some serious demand.

I'm sick to death of business types, who don't even try to meet you halfway, making impossible demands, and then being disappointed with the end result. I'm constantly having to manage expectations.

I'd love to become a electrician, or a train driver. Go in, do a hard days graft, and go home. Instead of my current career path where I'm having to constantly re-prioritize, put out fires, report to multiple leads with different agendas, scope and build things that have never been done, ect. The stress is endless. Nothing is ever good enough or fast enough. It feels like an endless [redacted] treadmill, and it's tiring. Maybe I'm misguided but in other fields one becomes a master of their craft over time. In CS/data science, I feel like you are forever a junior because your experience decays over time.

Anybody else feel the same way?

A PhD that involves a strong foundational learning component can give you more career stability.

State of the art skills today, are old skills tomorrow

Foundation knowledge

I can:

- derive a iterative-reweighted least squares algorithm for logistic regression (**linear algebra**, **calculus**).
- derive a likelihood for a given problem (**probability**).
- tell you when stochastic gradient descent will converge (**numerics**).
- work out the complexity of QuickSort? (**algorithms**)

Transient Skills

I can construct Natural Language Processing models using:

- LSTMs
- Recurrent neural networks
- Transformers / Attention

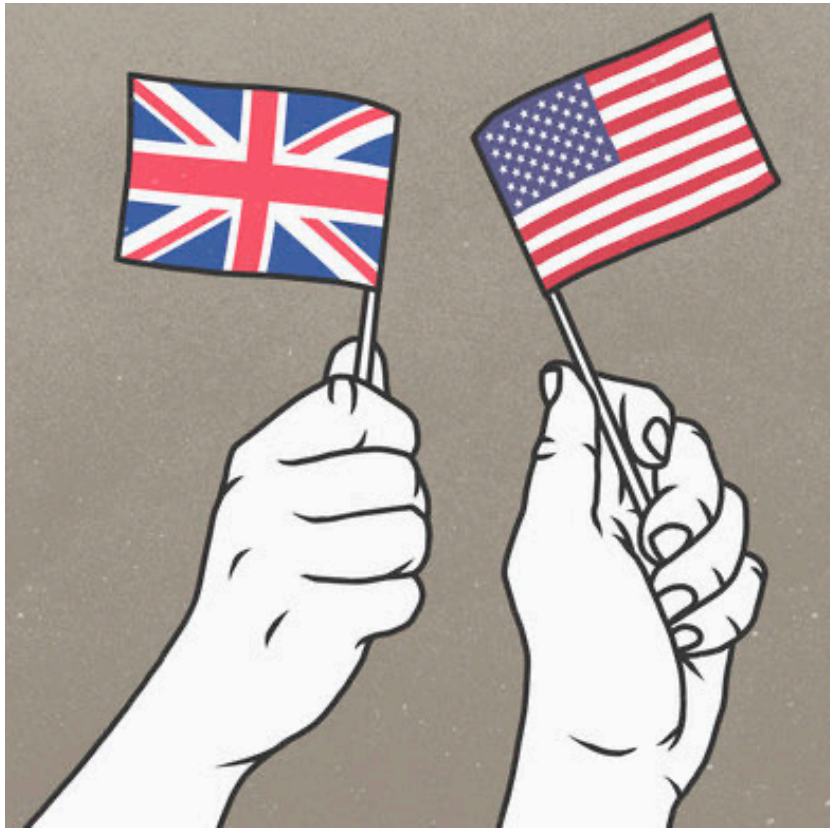
Global competition

Postgraduate education in the UK tends to be shorter and less comprehensive than in other major economic powers.

Example:

- UK MSc – 1 year, UK PhD – 3 years.
- US MSc – 2 years, US PhD – 6 years.

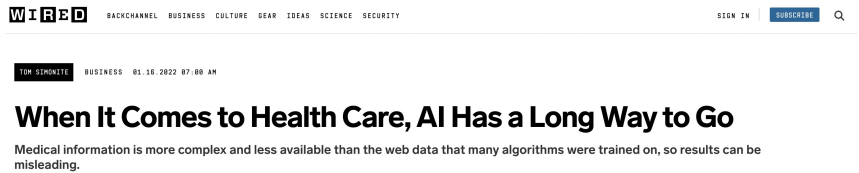
Outside of the UK, people are more likely to consider PhDs later in life.



Increasing regulation and emerging standards

Standards for data science regulation and practice are emerging based on academic principles.

> *"We can't take paradigms for developing AI tools that have worked in the consumer space and just port them over to the clinical space" - Visar Berisha*


WIRED BACKCHANNEL BUSINESS CULTURE GEAR IDEAS SCIENCE SECURITY SIGN IN SUBSCRIBE

TOM SIMONITE BUSINESS 01.16.2022 07:00 AM

When It Comes to Health Care, AI Has a Long Way to Go

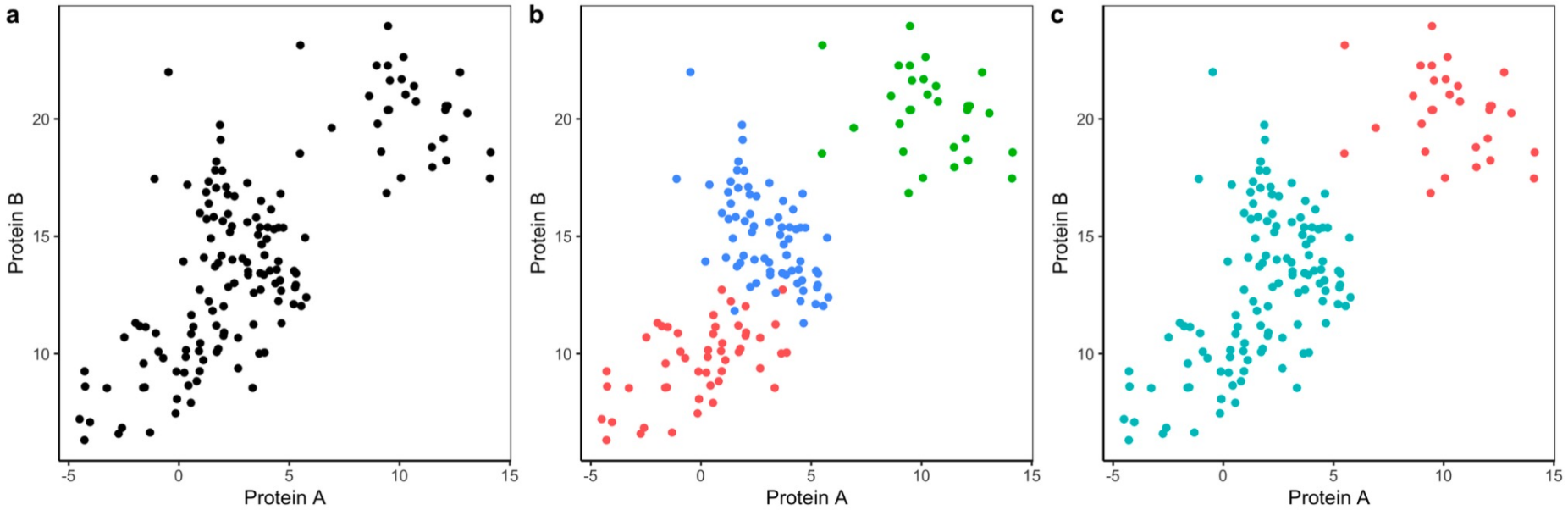
Medical information is more complex and less available than the web data that many algorithms were trained on, so results can be misleading.


TOM SIMONITE BUSINESS 08.08.2018 09:00 AM

When Bots Teach Themselves to Cheat

Even with logical parameters, AI programs can develop shortcuts and workarounds that humans didn't think to deem off-limits.

Gives you time to get to know your stuff



Two analysts come up with two different clustering structures (middle, right) for the same data set (far left). **Why might they have come up with different clusters?**

Typical questions

We can use k-means …

Can you explain how k-means works?

What are the assumptions that are made in k-means clustering?

How do you choose *k*?

Use a Gaussian mixture model …

Can you explain what a latent variable model is?

What approaches can you use to train a GMM?

What is the geometric interpretation of the covariance matrices in terms of the shapes of the clusters?

To End

In this talk, I said I would discuss:

- the relevance of PhDs in today's Data Science ***"time to mature"***
- tips on how to identify the right opportunities ***"not all PhDs are equal"***
- how advanced studies in data science can lead to more choice and opportunities in your later career ***"strong foundations can help you ride out trends"***

Any questions?

Thanks!

@cwcyau