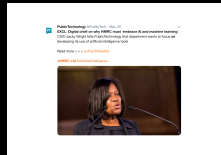Slava Jankin

Data Science Lab

Hertie School

AI for Common Good

France has just banned the use of judicial behavior data "to evaluate, analyze, compare or predict their professional practice, real or simulated". Violations punishable by jail time.

France Bans Judge Analytics, 5 Years In Prison For Rule Breakers
In a startling intervention that seeks to limit the emerging litigation analytics and prediction sector, the French Government has banned the publication of statistical...
artificiallawyer.com
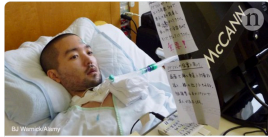
Brain signals translated into speech using artificial intelligence
Technology could one day be used to provide a voice for people who can't talk.
nature.com

HIGH RISK 10   LOW RISK 03

IEEE SPECTRUM   Topics ▾ Reports ▾ Blogs ▾ Multimedia ▾

24 Jan 2018 | 17:00 GMT

Michigan's MiDAS Unemployment System: Algorithm Alchemy Created Lead, Not Gold

A case study into how to automate false accusations of fraud for more than 34,000 unemployed people

By Robert N. Charette

Illustration: iStockphoto/IEEE Spectrum

The disastrous roll-out of the UK's digital welfare system is harming those m...
The roll-out of the UK government's new digitalised welfare system has fallen short of expectations, with disastrous consequences for the most in need of their benefits.
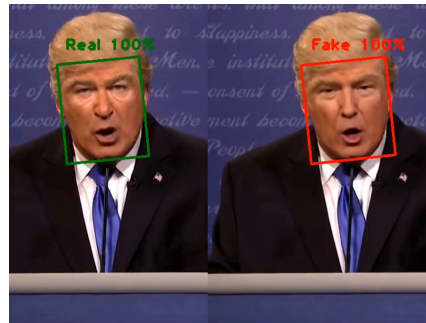euronews.com

Eric Monson and Elizabeth Honig

From Brueghel to Warhol: AI enters the attribution fray
Can artificial intelligence crack long-standing puzzles in art history? David Adam finds out.
nature.com

The Promise?

Real 100% Fake 100%

Deepfakes – this week's political scare

The Guardian

New AI fake text generator may be too dangerous to release, say cre...
The Elon Musk-backed nonprofit company OpenAI declines to release research publicly for fear of misuse
theguardian.com

MIT Technology Review
@techreview
Follow

It took two researchers 13 hours and $7.80 to train an AI model to fake UN speeches.

You can train an AI to fake UN speeches in just 13 hours
Deep-learning techniques have made it easier and easier for anyone to forge convincing misinformation.
technologyreview.com

3:24 PM - 7 Jun 2019

Cheapfakes we've seen before

What is AI?

"Actions that are indistinguishable from a human's." Alan Turing

"Systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals" European Commission, *AI for Europe* (2018)

The (Third) Coming of AI

- **1940-1950: Early days**
  - 1943: McCulloch & Pitts: Boolean circuit model of brain
  - 1950: Turing's "Computing Machinery and Intelligence"

- **1950—70: Excitement: Look, Ma, no hands!**
  - 1950s: Early AI programs, including Samuel's checkers program, Newell & Simon's Logic Theorist, Gelernter's Geometry Engine
  - 1956: Dartmouth meeting: "Artificial Intelligence" adopted
  - 1965: Robinson's complete algorithm for logical reasoning

- **1970—90: Knowledge-based approaches**
  - 1969—79: Early development of knowledge-based systems
  - 1980—88: Expert systems industry booms
  - 1988—93: Expert systems industry busts: "AI Winter"

- **1990—: Statistical approaches**
  - Resurgence of probability, focus on uncertainty
  - General increase in technical depth
  - Agents and learning systems… "AI Spring"?

- **2000—: Where are we now?**

AI Today

Figure: Yonatan Zunger, 12.10.17, @Medium
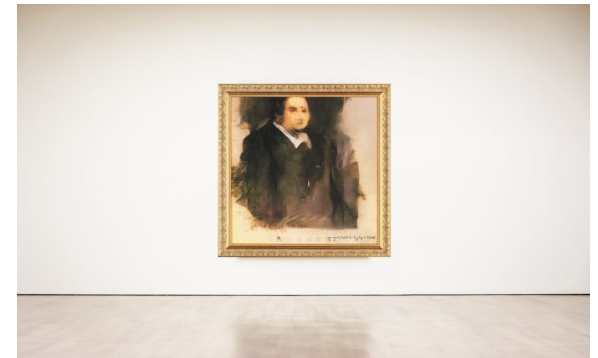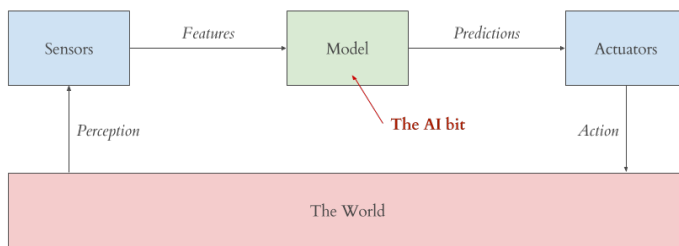
- Public discourse about "AI" largely refers to what is known as "Machine Learning" (ML).

- ML is a field mixing statistics, computer science, and many other disciplines to "design algorithms that process data, make predictions and help make decisions" (Michael J. Jordan).

- "Data Science" often used to refer to the link between ML and database and computing systems, and "reflecting the larger social and environmental scope of the resulting systems" (Michael J. Jordan).

- AI is not magic - a set of tools and technologies.

Subfields of AI

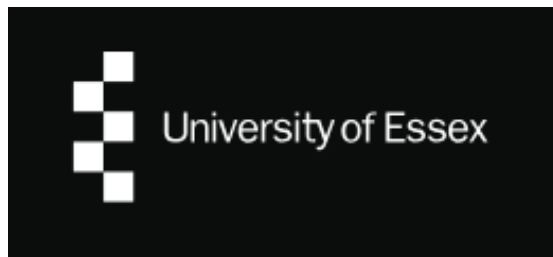- Perception: vision, speech understanding, etc.
- Machine Learning
- Robotics
- Natural Language Processing
- Reasoning and decision making (search, planning, decision theory, optimisation)

Essex Centre for Data Analytics (ECDA)

The Vision:
To make Essex national leaders using the power of data science and AI to tackle public policy challenges

The aims:
To make Essex a place that is an exemplar for the integration of data across public bodies.

To have the skill, capability and technology to undertake predictive analytics based on ethical, high standards.

To have a sustainable data infrastructure.

To have the best data science / analytical capabilities in the UK to benefit our people and communities.

A risk profile for school readiness

- Can we predict whether children will be school ready on starting school?

- What factors are most likely to cause this early indicator of "best start in life"?

- 511 households at risk in the pilot area;

- Including 280 previously unknown to public services;

- 74% prediction accuracy rate.

- Meet Bevan,

- He's two and a half, and a little shy!

- He has an amazing mum and grandad; Hayley and Ed, who set-up "Chicklets"

- This is the first community group in Vange, Basildon (Essex), to benefit from the school readiness work.

Improving outcomes for local people

Natural Language Processing in Public Policy

- Children's social care

- Given the text of the initial referral and assessment, and structured data relating to the case, could we predict whether the case would be re-referred and escalated if it were closed?

- ML and NLP

Figure 5: The inputs of the machine learning algorithm used to detect escalated closed cases.
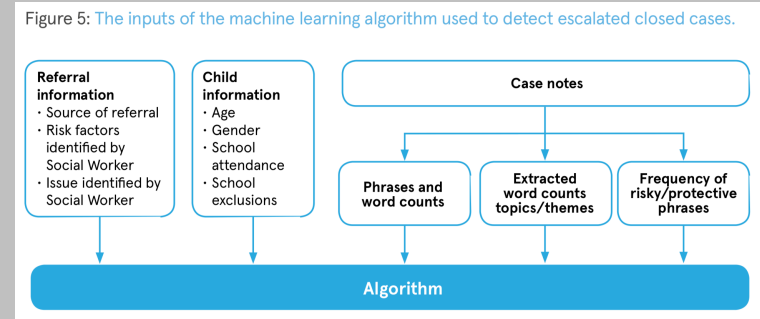
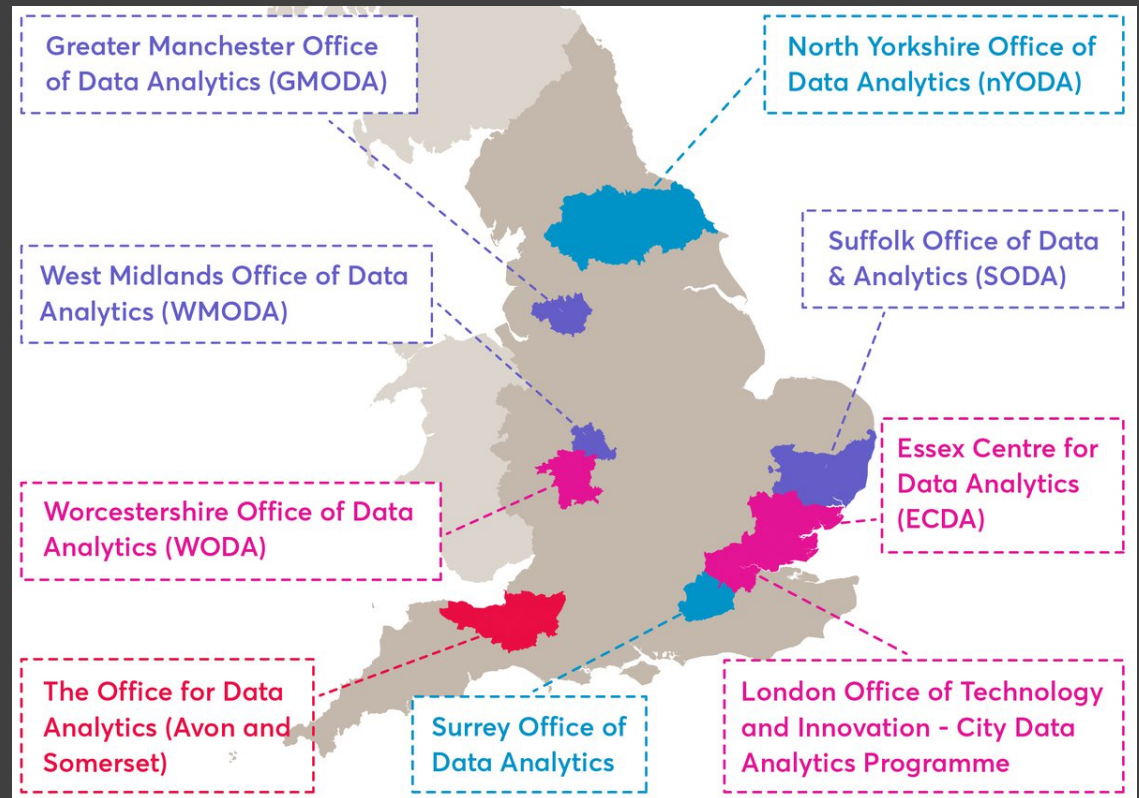# Predictive Modelling in Public Policy

- Hackney & Thurrock councils and Xantura

  - vulnerable families in need of additional support.

- Brent Council and IBM

  - children at risk of criminal exploitation.
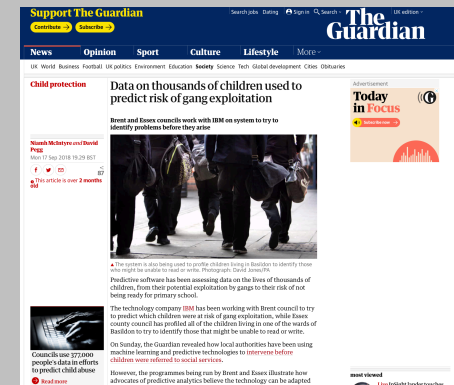
ECDA Essex Standard:
Openness, transparency, and
ethics underpin ECDA work

Nesta: State of Offices of Data Analytics in the UK

Greater Manchester Office of Data Analytics (GMODA)

North Yorkshire Office of Data Analytics (nYODA)

West Midlands Office of Data Analytics (WMODA)

Suffolk Office of Data & Analytics (SODA)

Worcestershire Office of Data Analytics (WODA)

Essex Centre for Data Analytics (ECDA)

The Office for Data Analytics (Avon and Somerset)

Surrey Office of Data Analytics

London Office of Technology and Innovation - City Data Analytics Programme

It's not always easy…

UK Government Industrial Strategy: Artificial Intelligence Sector Deal
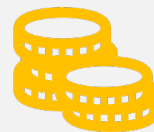
The world's most innovative economy ambition

Partnership between government, industry, and academia

Attract and retain AI talent;

Improve digital infrastructure;

Conducive business climate;

Prosperity of society through AI benefits.

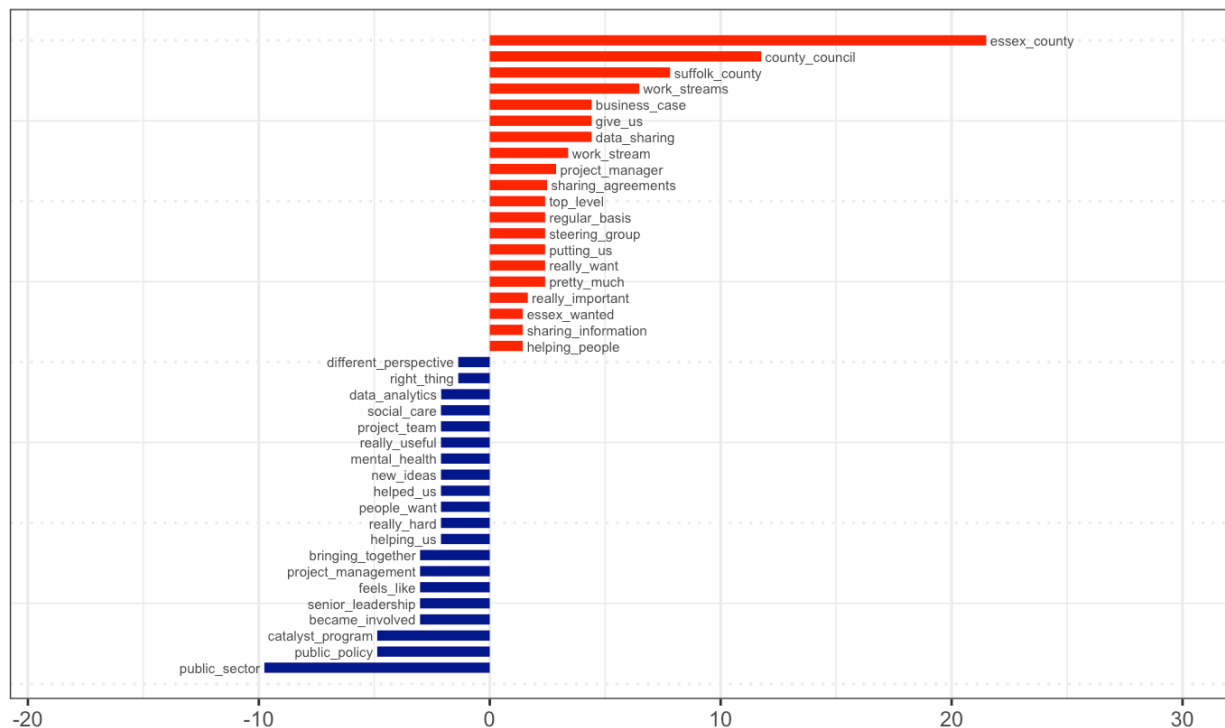£1bn + £1.7bn from the Industrial Strategy Challenge Fund

"Our ongoing partnership with the University of Essex puts us at the forefront of local authorities seeking innovative solutions to the challenges we face."

Councillor David Finch
LEADER OF ESSEX COUNTY COUNCIL AND CHAIR OF ESSEX PARTNERS

AI Collaboration for Common Good

Perspectives on Collaboration

- 24 interviews: SCC, ECC, and UoE stakeholders
- Core issues:
- Capacity and capability
- Collaborative mission vs organisational uniqueness
- Mechanics of data sharing
- Organisational priorities vs AI project priorities
- Routinising AI project success:
- Problem-centric redefinition
- Granularity and specificity
- Showing value

10 QUESTIONS TO ANSWER BEFORE USING AI IN PUBLIC SECTOR ALGORITHMIC DECISION MAKNG
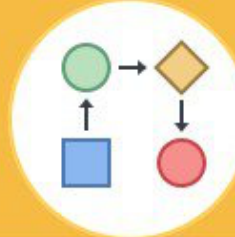
ASSUMPTIONS
What assumptions is the algorithm based on and what are their limitations and potential biases?

ETHICS
What assessment has been made of the ethics of using this algorithm?

OBJECTIVE
Why is the algorithm needed and what outcomes is it intended to enable?

DATA
What datasets is / was the algorithm trained on and what are their limitions and potential biases?

OVERSIGHT
What human judgement is needed before acting on the algorithm's output and who is responsible for ensuring its proper use?

USE
In what processes and circumstances is the algorithm appropriate to be used?

INPUTS
What new data does the algorithm use when making decisions?

EVALUATION
How, and by what criteria, will the effectiveness of the algorithm be assessed, and by whom?

IMPACTS
What impacts - good and bad - could the use of the algorithm have on people?

MITIGATION
What actions have been taken to migitate the negative impacts that could result from the algorithm's limitations and potential biases?
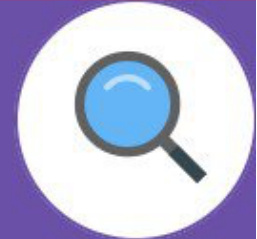
@EddieACopeland #IdeaOnAPage
@nesta_uk
Get this infographic at
http://bit.ly/IdeaOnAPage