# Algorithm Auditing

*Managing Legal, Ethical and Technological Risks of AI, ML and Associated Algorithms*

**Adriano Koshiyama, Emre Kazim, Philip Treleaven, et al.**

Department of Computer Science

**Corresponding Author:**
adriano.koshiyama.15@ucl.ac.uk

# Preface

# Our history and team

- **Automated Fraud Detection**
  - 25 years ago pioneered fraud detection, built first insider dealing detection system for LSEG
- **Algorithmic Trading**
  - 16 years pioneered algo trading systems with leading banks and funds
- **LawTech/RegTech and Digital Ethics**
  - Legal Services digital marketplace
  - ‘Computable’ (computer-executable) Legal Contract
- **Algorithm Auditing and Assurance**
  - Growing need by firms and customers

# Unique partnerships









# Publications

- **This presentation was based on**
  - Koshiyama, Adriano and Kazim, Emre and Treleaven, Philip et al. **Towards Algorithm Auditing**: A Survey on Managing Legal, Ethical and Technological Risks of AI, ML and Associated Algorithms (February 4, 2021). Available at SSRN: https://ssrn.com/abstract=3778998
  - Kazim, Emre and Koshiyama, Adriano, A High-Level Overview of AI Ethics (May 24, 2020). Available at SSRN: https://ssrn.com/abstract=3609292
  - **Upcoming:** Koshiyama, Adriano and Kazim, Emre and Treleaven, Philip et al. Foundations for Governing, Auditing and Assuring Algorithms: from Data Protection to AI Conduct (Sept, 2021).

# Presentation structure

- **Part I: Why Algorithm Auditing**
  - Long-term and near-term challenges
  - Algorithm auditing as a way to safeguard society
- **Part II: What is Algorithm Auditing**
  - Key components of algorithm auditing
  - Risk verticals and auditing levels
  - Mitigation strategies and assurance processes
- **Part III: On Auditing Algorithms**

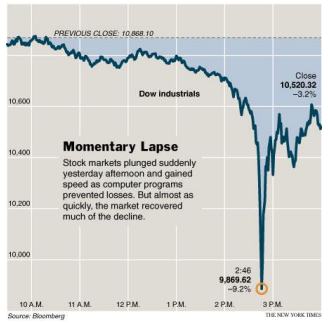
# Part I

*Why Algorithm Auditing*

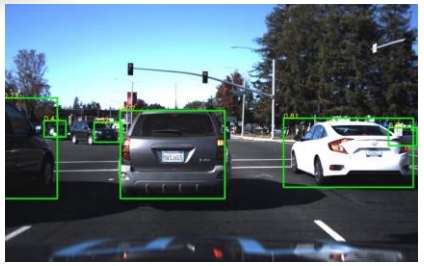
# Long-term Challenges: From ‘Big Data’ to ‘Big Algo’ age

- **Volume**: as resources and know-how proliferate, soon there will be ‘billions’ of algorithms
- **Velocity**: algorithms making real-time decision with minimal human intervention
- **Veracity**: reliability, legality, fairness, accuracy, etc. as critical characteristics
- **Variety**: from autonomous vehicles to medical treatment, employment, finance, and so on
- **Value**: new services, sources of revenue, cost-savings, and industries will be established

*Previous decade was about ‘**Data protection**’; this decade will focus on ‘**Algorithm Conduct**’*







# Near-term Challenges: fines, reputation and regulation

- **Companies are increasingly concerned about their algorithms**
  - Being illegal or unethical
  - Causing major financial loss
  - Reputational damage
- **Auditing algorithms is becoming mandatory**
  - Regulators are fining companies
  - Governments introducing legislation

THOMSON REUTERS

Amazon scraps secret AI recruiting tool that showed bias against women

FINANCIAL TIMES

Knight Capital glitch loss hits $461m

Bloomberg

Study of VW's Cheating on Diesels Examines Role of Bosch Code

BANK OF ENGLAND

Machine learning in UK financial services

Guidance on the AI auditing framework

Draft guidance for consultation

ico.

daten ethik kommission

EUROPEAN COMMISSION

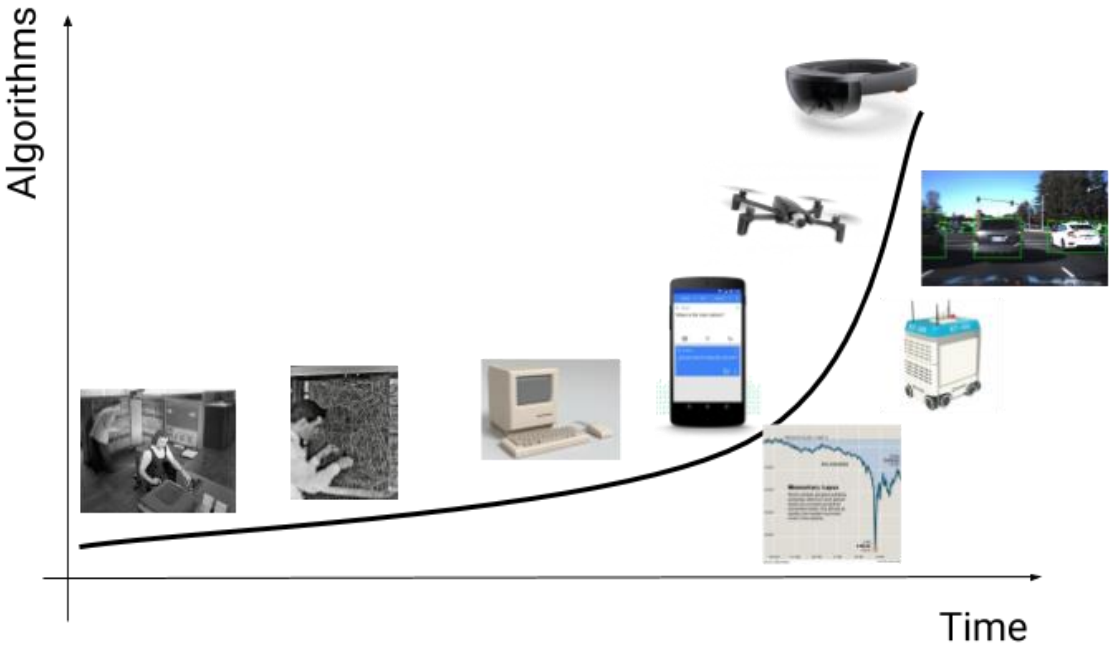
Brussels, 19.2.2020
COM(2020) 65 final

WHITE PAPER

On Artificial Intelligence - A European approach to excellence and trust

# Algorithm Auditing to Safeguard Society

- **How to face it: Algorithm Auditing**
  - the research and practice of assessing, mitigating, and assuring an algorithm's safety, legality, and ethics
  - As with Financial Audit, governments, business and society will require Algorithm Audit
- **This presentation outlines the principles and practices of Algorithm Auditing**

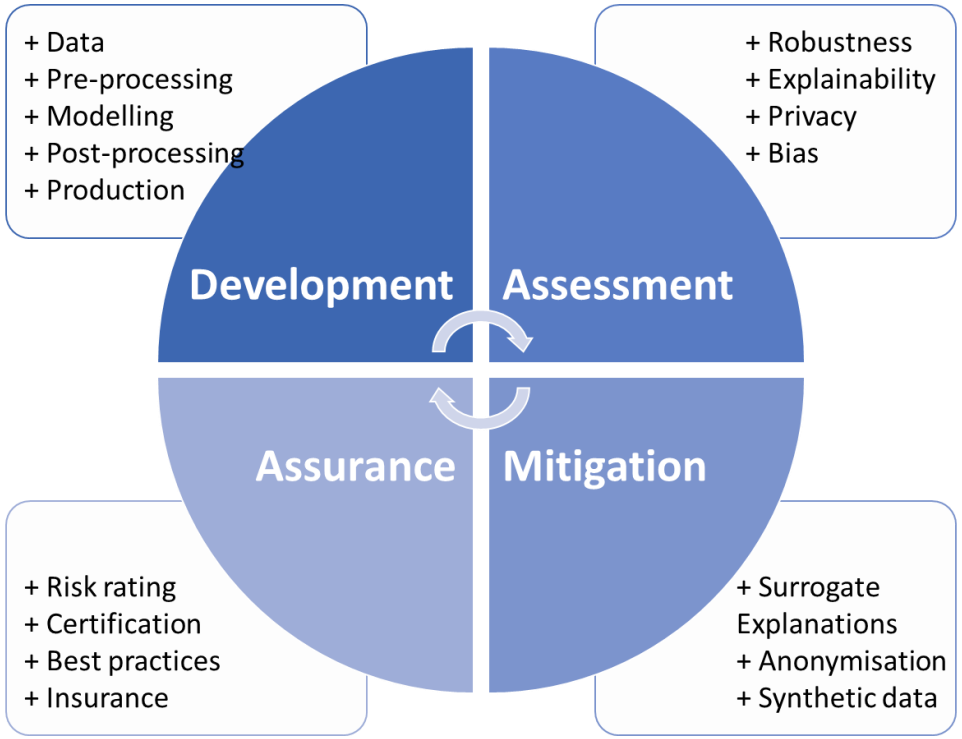


# Part II

*What is Algorithm Auditing*

# Elements of Algorithm Auditing

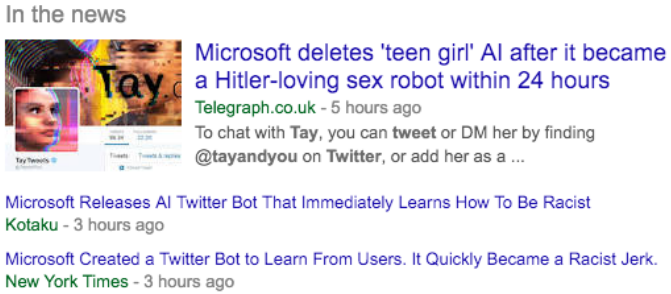
- **Development**: the process of developing an algorithmic system
- **Assessment**: the process of evaluating the algorithm behavior and capacities
- **Mitigation**: the process of servicing or improving an algorithm outcome
- **Assurance**: the process of declaring that a certain system conforms to pre-determined standards, practices or regulations



# Assessment: main risk verticals

- **Algorithm Privacy:** data minimization principles as well as adopt privacy-enhancing techniques to mitigate personal or critical data leakage.
- **Bias and Discrimination:** avoid unfair treatment of individuals given their protected characteristics.
- **Interpretability and Explainability:** provide decisions or suggestions that can be understood by their users and developers.
- **Performance and Robustness:** be safe and secure, not vulnerable to tampering or compromising of the data they are trained on.

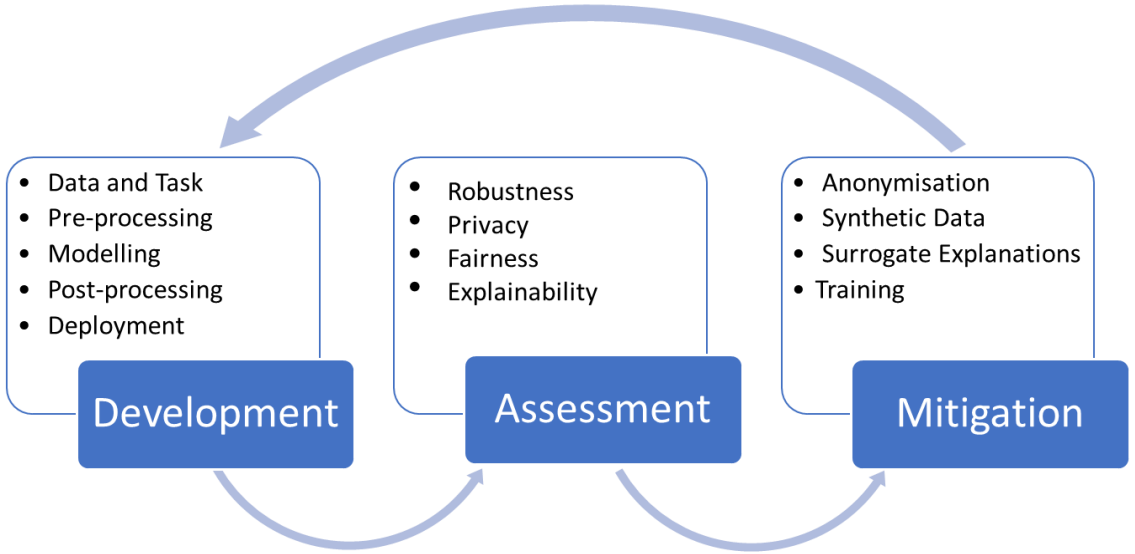
***To avoid these cases***





# Mitigation: intervention mechanism

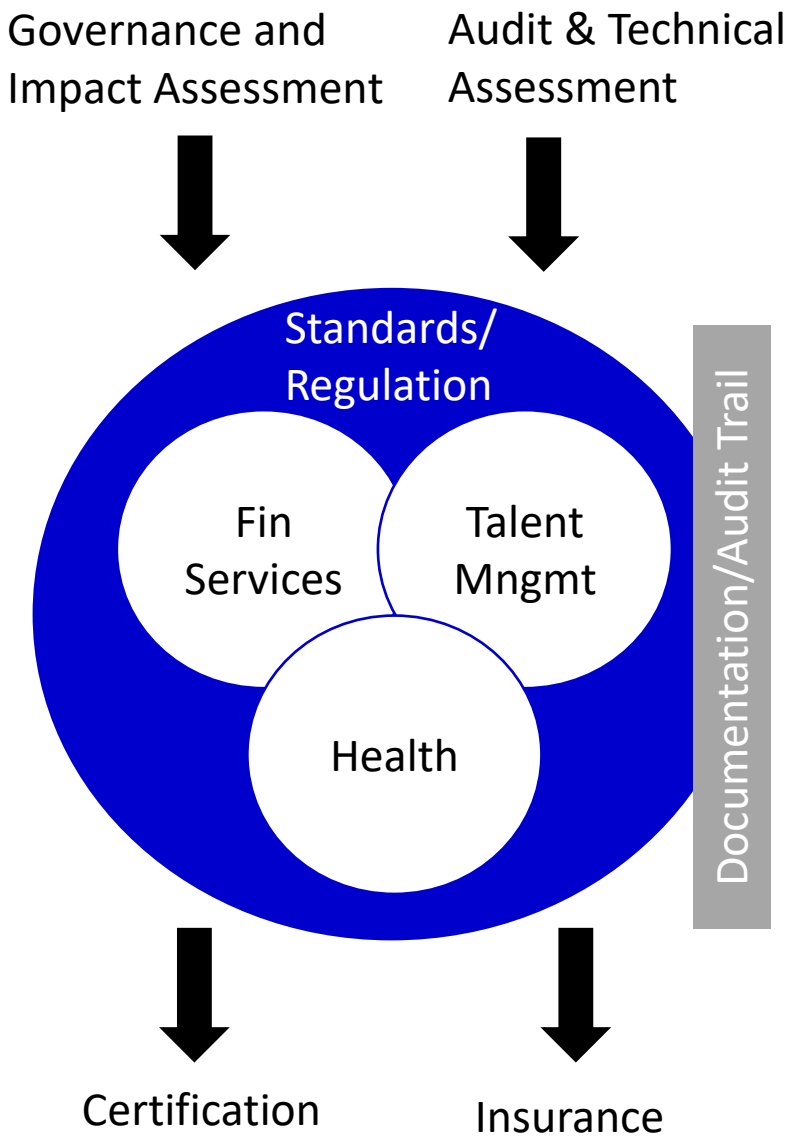
- **What are they**
  - specific procedures that can be used in conjunction in order to enhance an algorithm performance or solve issues like algorithm debiasing or establishing surrogate explanations




- **Main types**
  - **Human**: all procedures that involve how algorithm developers design, collaborate, reflect and develop algorithms, involving (re)training, impact assessment, etc.
  - **Algorithm**: all methodologies that can improve an algorithm current outcome.

# Assurance: final outcome

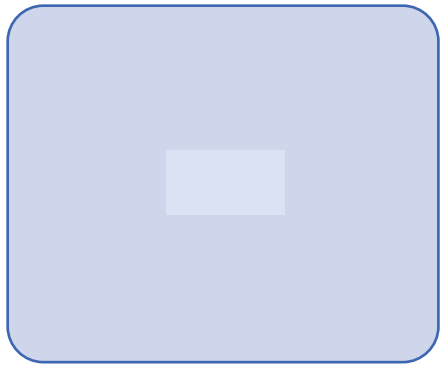
- **General and sector-specific assurance:** broad national regulation and standards, with sector specific ones
- **Governance:** technical assessments (robustness, privacy, etc.) and impact (risk, compliance, etc.) assessments
- **Unknown Risks:** risk schemes and procedures like ‘red teaming’ to mitigate unknown risks
- **Monitoring Interfaces:** like the use of ‘traffic-light’ user friendly monitoring interfaces
- **Certification:** such as of a system, of a team or unit, or AI engineers, etc
- **Insurance:** a subsequent service to emerge as a result of assurance maturing



# Part III

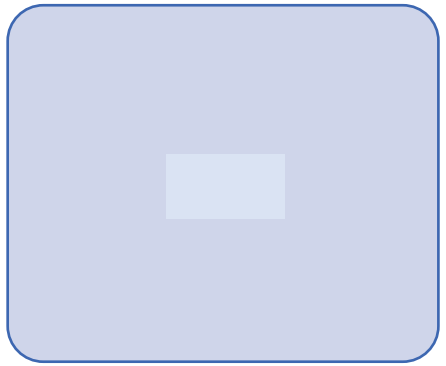
*On Auditing Algorithms*

# Some use cases in Financial Services



## Challenger Bank's Customer Onboarding

- AI Auditing and Improvement
- Key focus on performance and bias assessment
- Produce internal an external report
- Continuous support and lifelong engagement



## Challenger Bank's Risk Framework

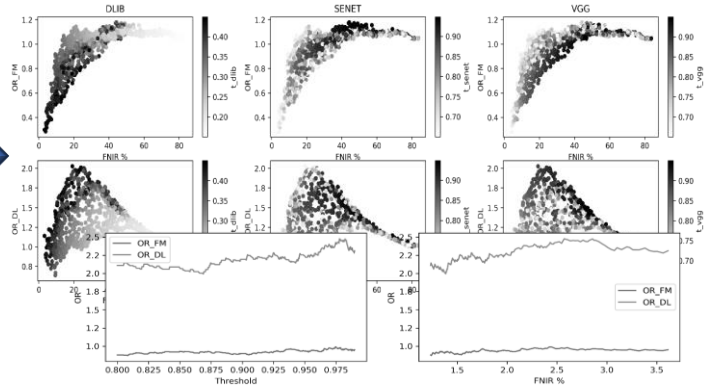
- MRM Risk Framework Adaption
- Scalable and flexible framework to improve innovation
- Be able to report to regulators and other stakeholders
- Continuous support and lifelong engagement

# Case study: Bank's Customer Onboarding

**Background**: Bank is a leading challenger bank. The interactions with its customers are all done digitally. The goal of the audit is to verify their customer onboarding process, and, in detail, the personal identity verification component.


**Review:** Holistic AI team performed a 3rd-party review of the face verification software components. We verified the systems performance and eventual gender and skin tone bias.

**Internal report**: Identified sources and components of underperformance.




**External**: Issued a report that is to be shared with BCR and other regulators.

# Case study: Bank's Risk Framework

**Background**: bank is a leading challenger bank. They are using AI to power several of their applications (onboarding, fraud analytics, etc.). They need a flexible and scalable framework to risk manage their AI applications across different jurisdictions and applications.

**Develop:** Holistic AI team gathered documents about their current framework, and scoped together which systems they want to manage, where, how they acquire and use them, etc.

**Internal framework**: Useful for technical and non-technical teams wanting to develop or buy AI systems

**External reporting**: Ready for regulators and stakeholders interested to know what they are doing about their use of AI

# Conclusions

# Conclusions

- **Promising area of research and practice**
  - Still in its inception phase
  - Many research institutions and large and small organizations manoeuvring at the moment
  - Will become a trade-barrier
- **This 20s we will see the emphasis on 'Algorithm Conduct' like 'Data Protection' was for the 10s**

THOMSON REUTERS

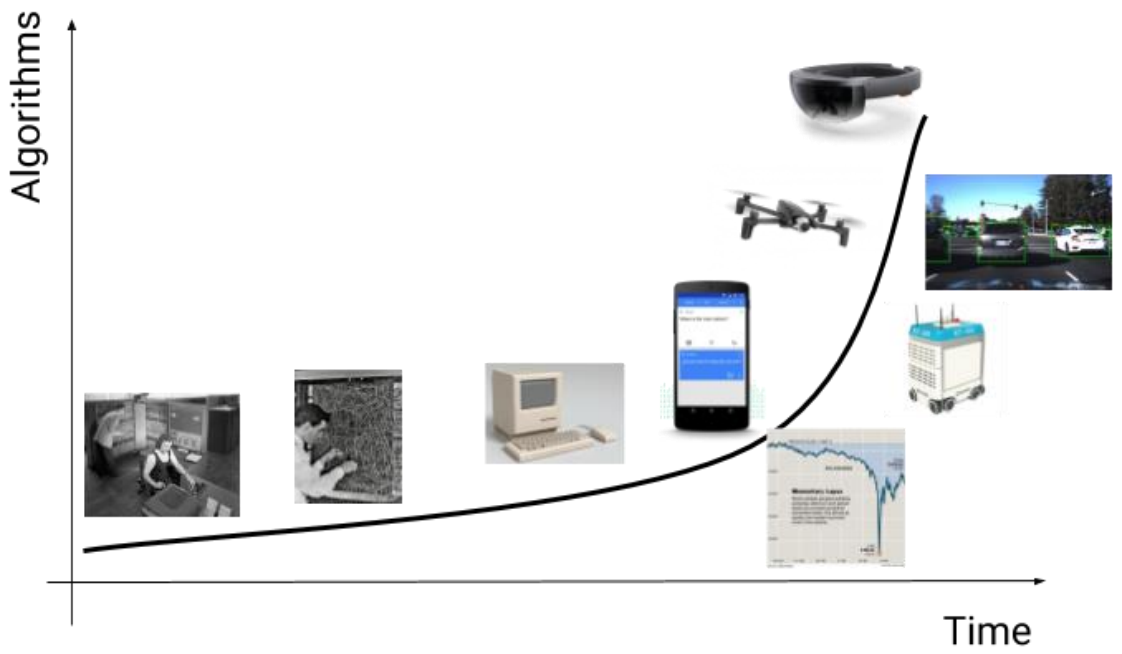
Amazon scraps secret AI recruiting tool that showed bias against women

FINANCIAL TIMES

Knight Capital glitch loss hits $461m

Bloomberg

Study of VW's Cheating on Diesels Examines Role of Bosch Code



# Algorithm Auditing

*Managing Legal, Ethical and Technological Risks of AI, ML and Associated Algorithms*

Koshiyama, Adriano and Kazim, Emre and Treleaven, Philip et al. Towards Algorithm Auditing: A Survey on Managing Legal, Ethical and Technological Risks of AI, ML and Associated Algorithms (February 4, 2021). Available at SSRN: https://ssrn.com/abstract=3778998