The Alan Turing Institute

The Good, the Bad, and the ~~Ugly~~ Necessary: AI in Finance

London, November 2024

Carsten Maple, The Alan Turing Institute

Carsten Maple

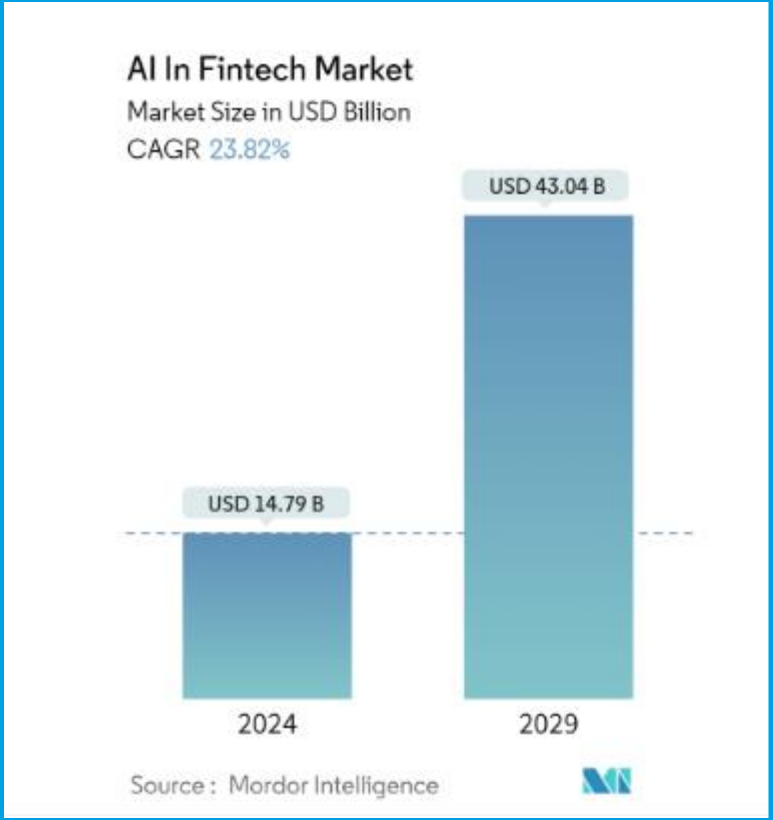
Principal Investigator of the NCSC-EPSRC Academic Centre of Excellence in Cyber Security Research, University of Warwick and Co-Investigator on FAIR Privacy & Security, The Alan Turing Institute

Director for Research and Innovation at the National Hub for Edge Artificial Intelligence

Background

Market Growth

The AI in fintech market is projected to grow from approximately **$14.79** billion in 2024 to **$43.04 billion** by 2029, at a compound annual growth rate (CAGR) of 23.82%.



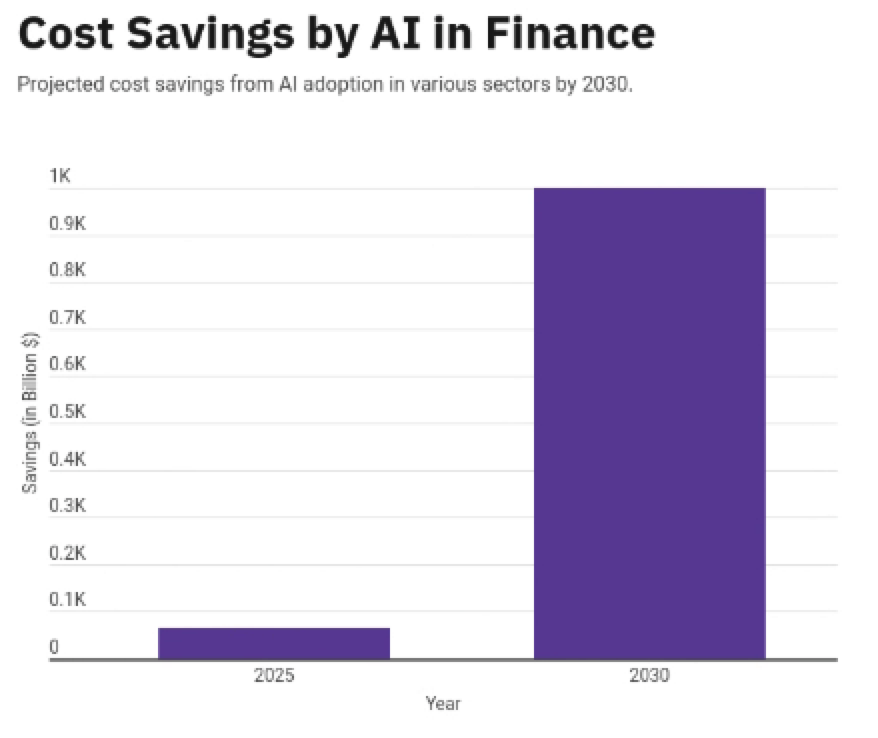
Source : Mordor Intelligence

Background

Cost Savings

By 2025, AI is expected to save banks between $200 billion and $340 billion, influencing $450 billion in revenue.

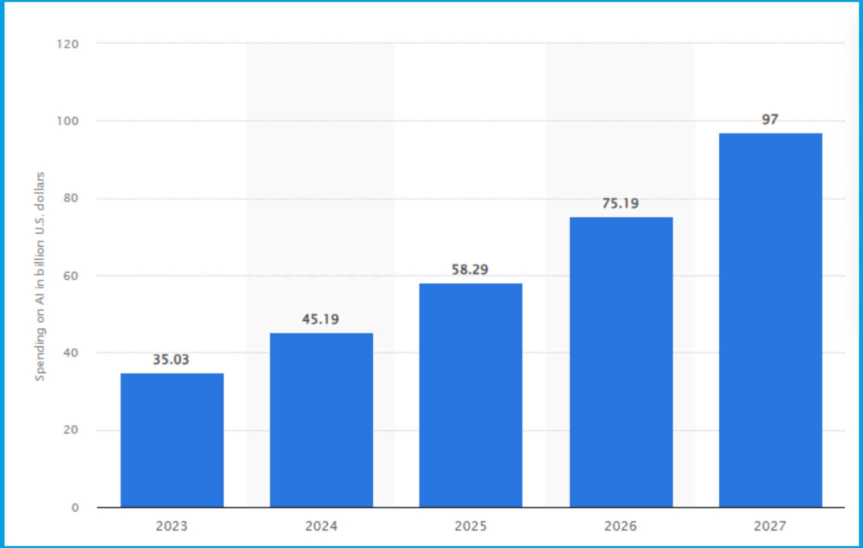
Source: AllAboutAI, 2024



Background

Investment in AI

In 2023, the financial services industry invested an estimated $35 billion in AI, with the banking sector accounting for approximately $21 billion of this investment.



Source: Statista, 2024

AI Technologies in Financial Service

- Machine Learning
- Expert Systems
- Natural Language Processing
- Neural Networks
- Robotics Process Automation

The Good

- **Fraud Detection**: Enhanced risk detection using PETs
- **Algorithmic Trading**: Leveraging predictive analytics for market insights
- **Customer Service**: AI-powered chatbots for personalised interaction
- **Credit Risk Assessment**: Using LLMs like FinBERT for financial sentiment analysis

The Bad

- **Algorithmic Bias**: Discriminatory outcomes in credit scoring
- **Data Misuse**: Over-reliance on opaque models leading to privacy violations
- **High-Stakes Errors**: Flash crashes caused by erroneous algorithmic trading

The challenge: The under realised potential of AI

- **AI is transforming the Finance Services Industry** across all core business domains: client experience, financial crime, risk management etc.
- **But significant delays and high-profile risks** exist in moving from experimentation to production due to lack of a coherent framework for deploying AI[1].
- **No clarity on how to maximise AI-driven innovation** with over 150 guidelines with principles for deploying AI safely, ethically and responsibly. Lack of consensus on how to operationalize these principles in practice.

[1] Surveys by: Financial Conduct authority/Bank of England and World Economic Forum

FAIR vision

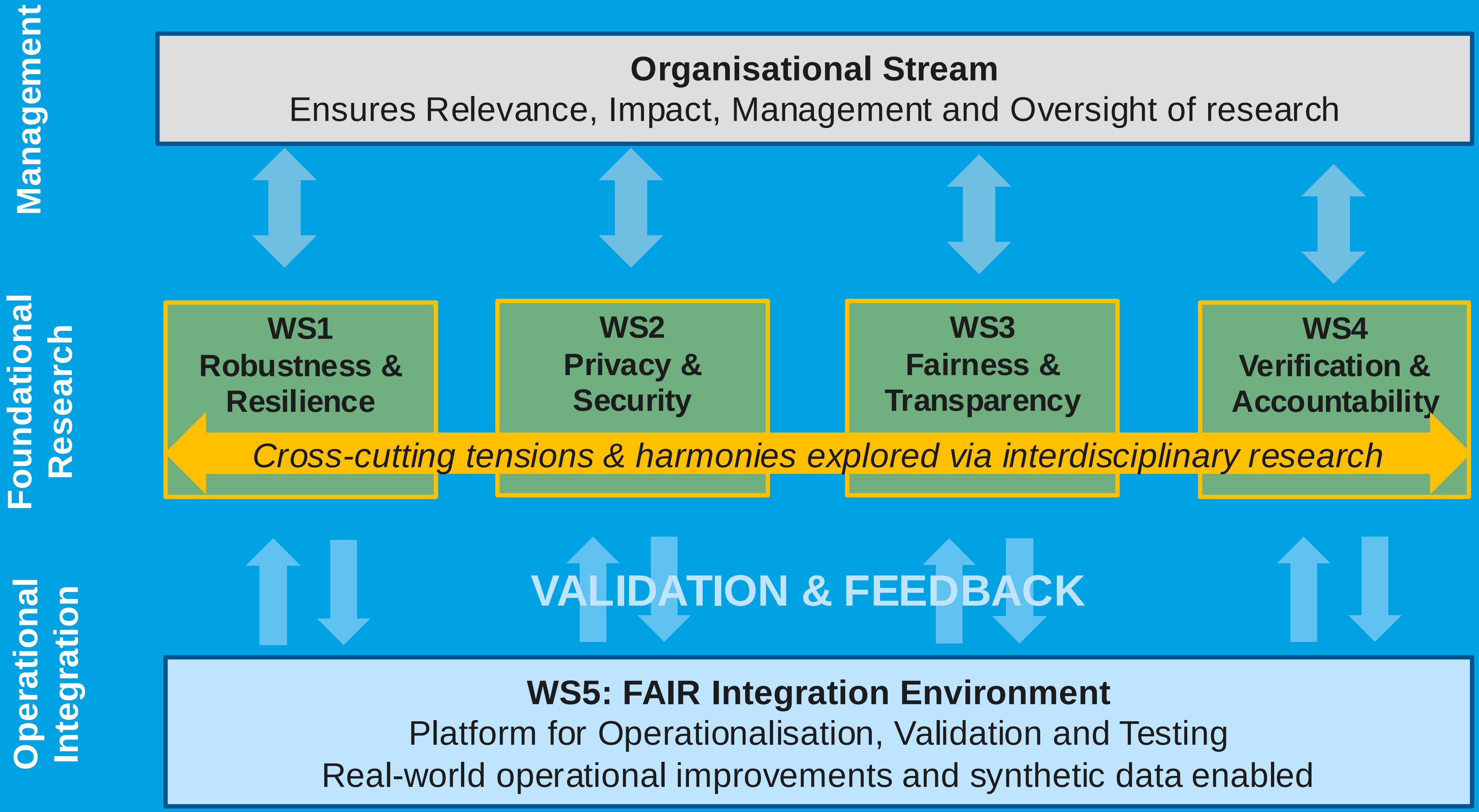
To unlock the transformational benefits of responsible adoption of AI across the financial services.

FAIR will:

- Develop scalable solutions for safe and trustworthy deployment of AI in Financial Services, underpinned by **foundational research**.
- Develop **digital sandbox** environments to enable validation, testing, and evaluation of emerging technologies.
- Identify **industry-wide standards** and processes to address trade-offs between regulatory and ethical dimensions facing industry and regulators.

Research challenges

- Enable **robust and resilient** data-driven decision-making.
- Develop **privacy and security methods** that provide safeguards against new risks.
- Deliver appropriate **fairness and transparency** algorithms addressing commercial, societal and regulatory goals.
- Provide methodologies for **verification and certification** of black-box models.
- Develop **digital sandboxes** to allow data sharing, testing and real time monitoring.



Management

Foundational Research

Operational Integration

Organisational Stream
Ensures Relevance, Impact, Management and Oversight of research

WS1
Robustness & Resilience

WS2
Privacy & Security

WS3
Fairness & Transparency

WS4
Verification & Accountability

Cross-cutting tensions & harmonies explored via interdisciplinary research

VALIDATION & FEEDBACK

WS5: FAIR Integration Environment
Platform for Operationalisation, Validation and Testing
Real-world operational improvements and synthetic data enabled

Partners and collaborators

Lead partners

UKRI Engineering and Physical Sciences Research Council

HSBC

Project partners

accenture

Academic collaborators

University of Oxford

UCL

Warwick
The University of Warwick

University of Cambridge

The University of Edinburgh

Affiliate partners

Smart Data Foundry

quantexa

Microsoft

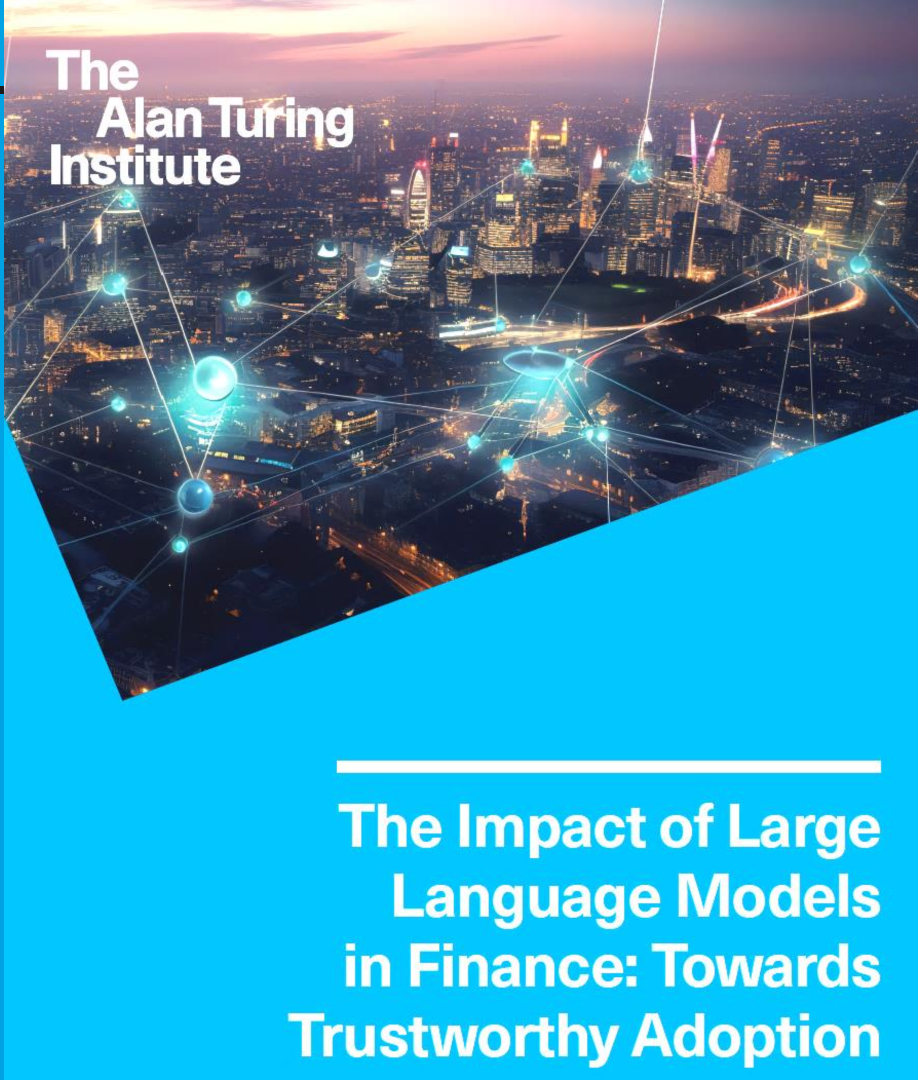
PRIVITAR

moz://a

Contributors

in Alphabetical Order:

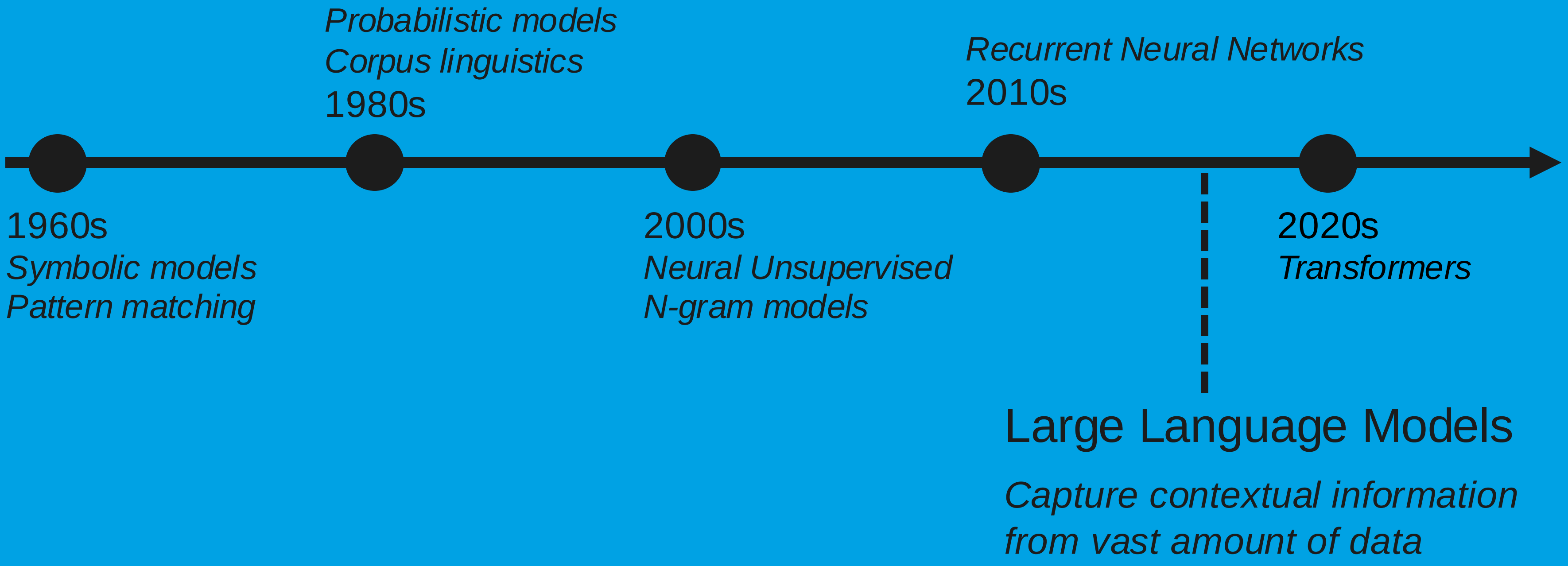
- *Andrew Elliott, The Alan Turing Institute*
- *Andrew Walters, Bank of England*
- *Anna Kharchenkova, Accenture*
- *Fern Watson, Financial Conduct Authority*
- *Gesine Reinert, The Alan Turing Institute*
- *Henrike Mueller, Financial Conduct Authority*
- *Isaac Bowers-Barnard, Accenture*
- *Jagdish Hariharan, University of Warwick*
- *Lukasz Szpruch, The Alan Turing Institute*
- *Marcus Turner, Allen & Overy*
- *Marie Briere, Amundi*
- *Matt Shelley, Accenture*
- *Oxana Samko, HSBC*
- *Paul Lickman, HSBC*
- *Pavle Avramovic, Financial Conduct Authority*
- *Praveen Selvaraj, The Alan Turing Institute*
- *Ray Eitel-Porter, Accenture*
- *Sapan Dogra, Accenture*
- *Todd Bose, Standard Chartered*
- *Vijay Jairaj, Standard Chartered*
- *Walter McCahon, UK Finance*



The
Alan Turing
Institute

The Impact of Large Language Models in Finance: Towards Trustworthy Adoption

A Brief History



1960s
Symbolic models
Pattern matching

Probabilistic models
Corpus linguistics
1980s

2000s
Neural Unsupervised
N-gram models

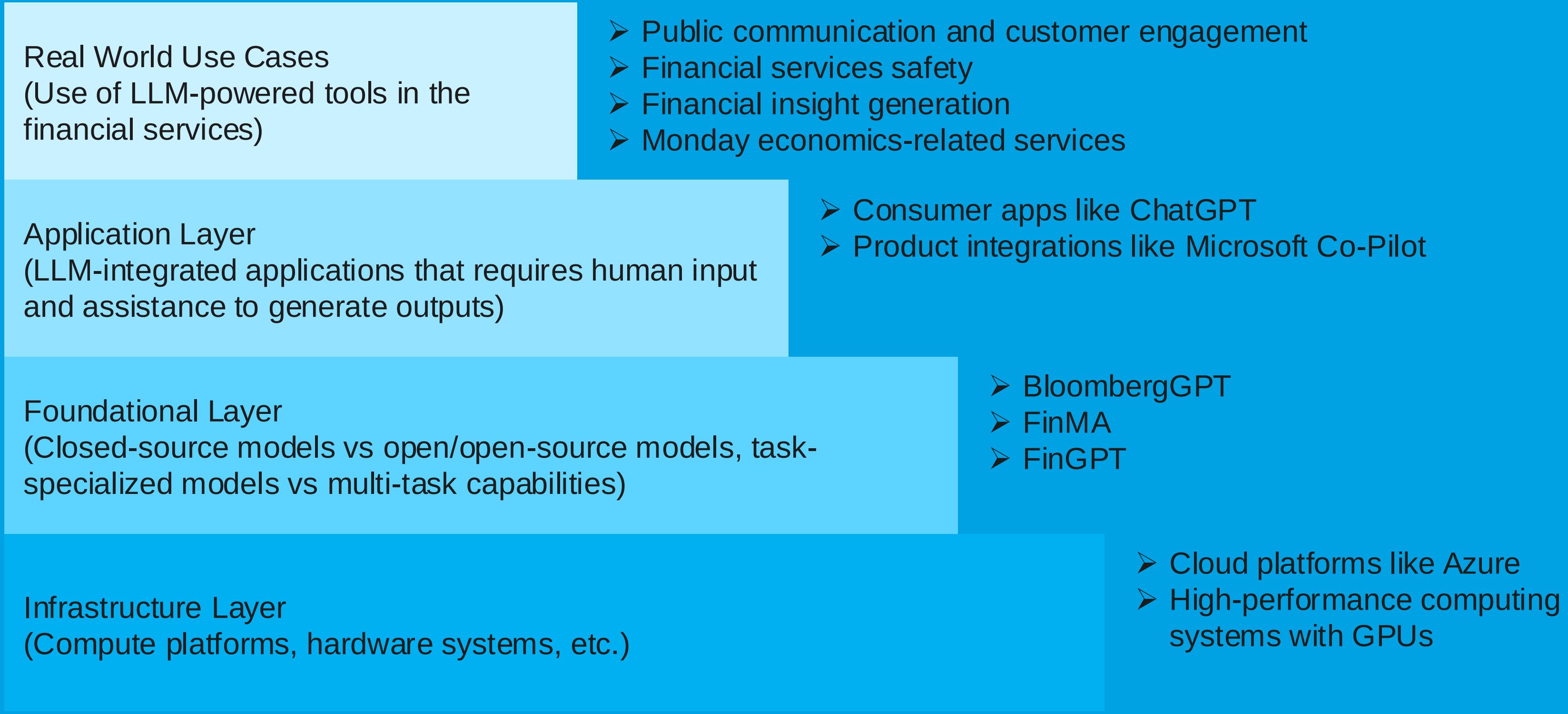
Recurrent Neural Networks
2010s

2020s
Transformers

Large Language Models

Capture contextual information from vast amount of data

LLMs integration into financial services



Real World Use Cases
(Use of LLM-powered tools in the financial services)

- Public communication and customer engagement
- Financial services safety
- Financial insight generation
- Monday economics-related services

Application Layer
(LLM-integrated applications that requires human input and assistance to generate outputs)

- Consumer apps like ChatGPT
- Product integrations like Microsoft Co-Pilot

Foundational Layer
(Closed-source models vs open/open-source models, task-specialized models vs multi-task capabilities)

- BloombergGPT
- FinMA
- FinGPT

Infrastructure Layer
(Compute platforms, hardware systems, etc.)

- Cloud platforms like Azure
- High-performance computing systems with GPUs

Functional Opportunities

Public Communication and Customer Engagement

- Financial Communication
- Customer service

Financial Service Safety

- Detecting and preventing fraud
- Product development
- Risk assessment

Financial Insight Generation

- Market surveillance
- Market insights and reports
- Business finance data insights
- Personal investment insights
- Generation of aggregate reports

Financing and Investment Activities

- Loan financing
- Investment banking
- Treasury optimisation
- Private equity and venture capital strategy development
- Asset allocation

Broader Opportunities

- **Operational:** Streamlining decision-making processes, risk profiling, benefit quantification, and prioritisation, improving investment research, and back-office operations.
- **Human-machine interaction:** Reducing complexity and accelerating productivity in credit analysis, client due diligence, and transaction monitoring.
- **Financial advisory:** Inform asset allocation decisions or develop actionable insights by processing signals from a broad range of inputs.
- **Financial literacy:** Personalised support based on an individual's literacy levels.

Risks

Data-related risks

- Bias
- Privacy
- Data transparency and security
- Violation of intellectual property

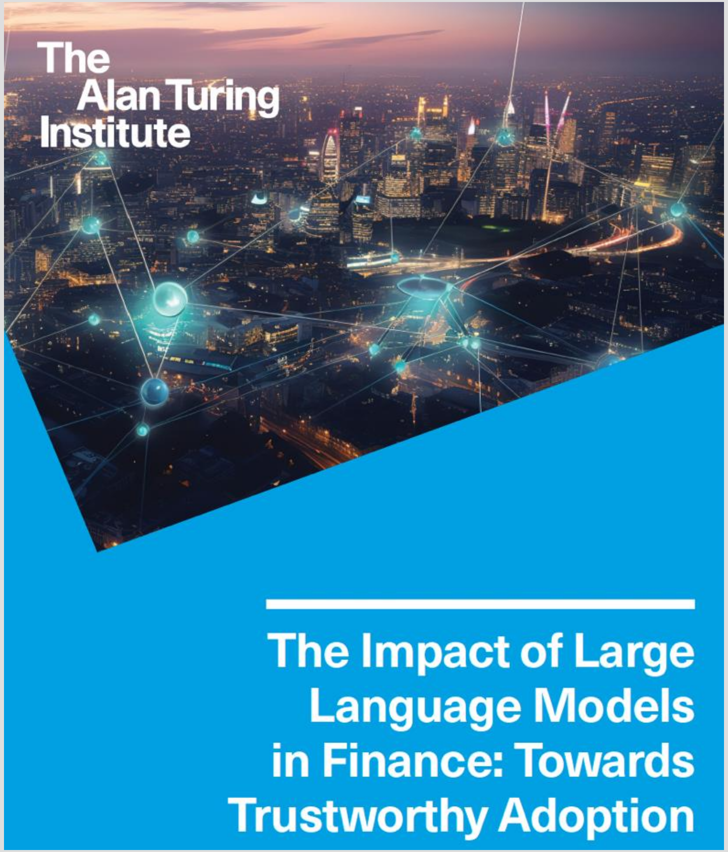
Model complexity-related risks

- Lack of explainability
- Reasoning errors
- Susceptibility to various attacks

Social behaviours and Human Values

- Alignment
- Information hallucination
- Toxic linguistic
- Environmental impacts
- Open vs close source impacts

Report (March '24)



The Alan Turing Institute

The Impact of Large Language Models in Finance: Towards Trustworthy Adoption

Authors:

Carsten Maple, Alpay Sabuncuoglu

Contributors in Alphabetical Order:

- *Andrew Elliott, The Alan Turing Institute*
- *Andrew Walters, Bank of England*
- *Anna Kharchenkova, Accenture*
- *Fern Watson, Financial Conduct Authority*
- *Gesine Reinert, The Alan Turing Institute*
- *Henrike Mueller, Financial Conduct Authority*
- *Isaac Bowers-Barnard, Accenture*
- *Jagdish Hariharan, University of Warwick*
- *Lukasz Szpruch, The Alan Turing Institute*
- *Marcus Turner, Allen & Overy*
- *Marie Briere, Amundi*
- *Matt Shelley, Accenture*
- *Oxana Samko, HSBC*
- *Paul Lickman, HSBC*
- *Pavle Avramovic, Financial Conduct Authority*
- *Praveen Selvaraj, The Alan Turing Institute*
- *Ray Eitel-Porter, Accenture*
- *Sapan Dogra, Accenture*
- *Todd Bose, Standard Chartered*
- *Vijay Jairaj, Standard Chartered*
- *Walter McCahon, UK Finance*

43 participants
21 contributing authors

The Necessary – Towards Safe Adoption

- Robustness and Resilience
- Data asymmetry
- Security
- Privacy
- Fairness
- Explainability
- Accountability and Transparency
- Integrity
- Skills

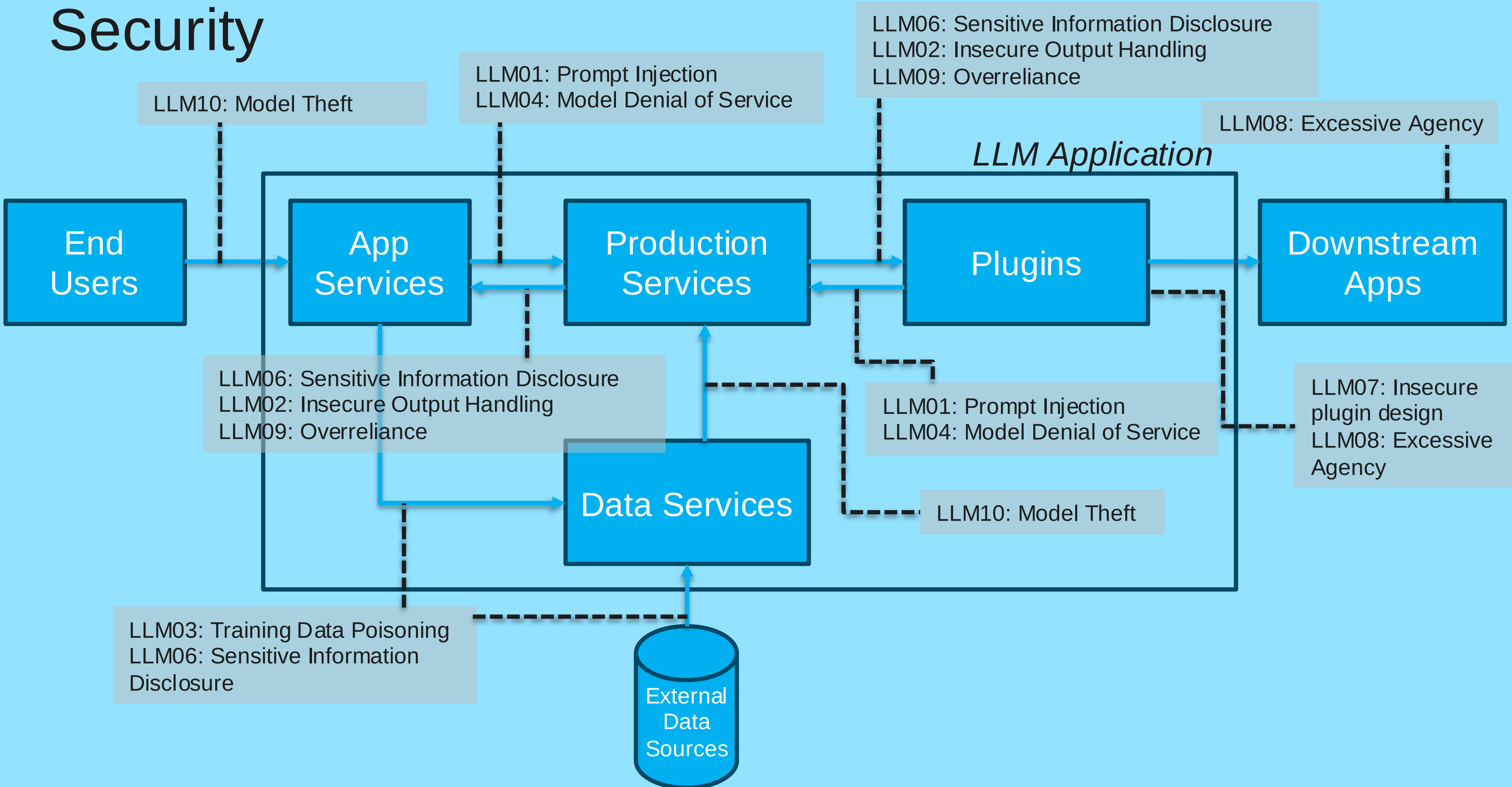
Robustness and Resilience

- Most services are bound to strict robustness checklists.
- Use case dependent, easier to evaluate for some use cases.
- Existing robustness checklists for ML systems can guide us through the LLM use cases.
- Human monitoring and red teaming activities with diverse teams is essential to achieve robustness.

Data Asymmetry

- It is a growing concern (concentration risk) in both tech and financial industries
- Current open banking practices can help practitioners navigate
- Sharing data might cause reputational and legal risks
- Defining incentives for big financial institutions is a challenge
- Can government-collected datasets support the practice of sharing?

Security

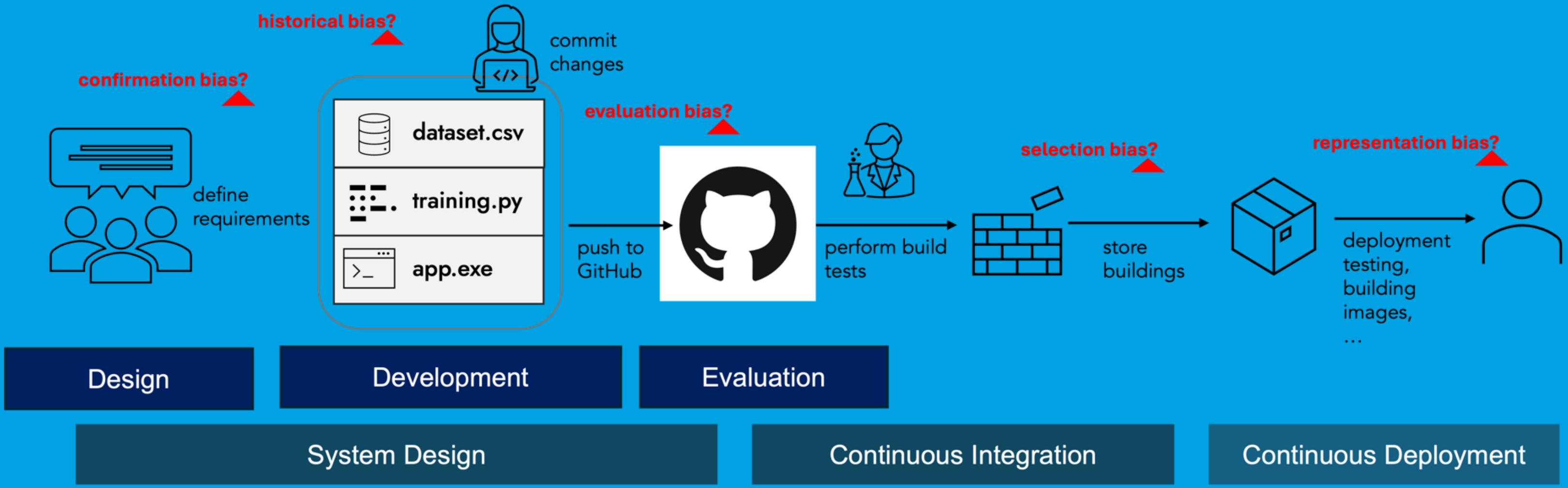

LLM10: Model Theft
LLM01: Prompt Injection
LLM04: Model Denial of Service
LLM06: Sensitive Information Disclosure
LLM02: Insecure Output Handling
LLM09: Overreliance
LLM08: Excessive Agency
LLM Application
End Users
App Services
Production Services
Plugins
Downstream Apps
LLM06: Sensitive Information Disclosure
LLM02: Insecure Output Handling
LLM09: Overreliance
LLM01: Prompt Injection
LLM04: Model Denial of Service
LLM07: Insecure plugin design
LLM08: Excessive Agency
Data Services
LLM10: Model Theft
LLM03: Training Data Poisoning
LLM06: Sensitive Information Disclosure
External Data Sources

Privacy

- Potential privacy impacts in a complex business logic is an active discussion.
- Evaluation frameworks of ethics councils need a re-evaluation.
- Assurance techniques are gaining importance.

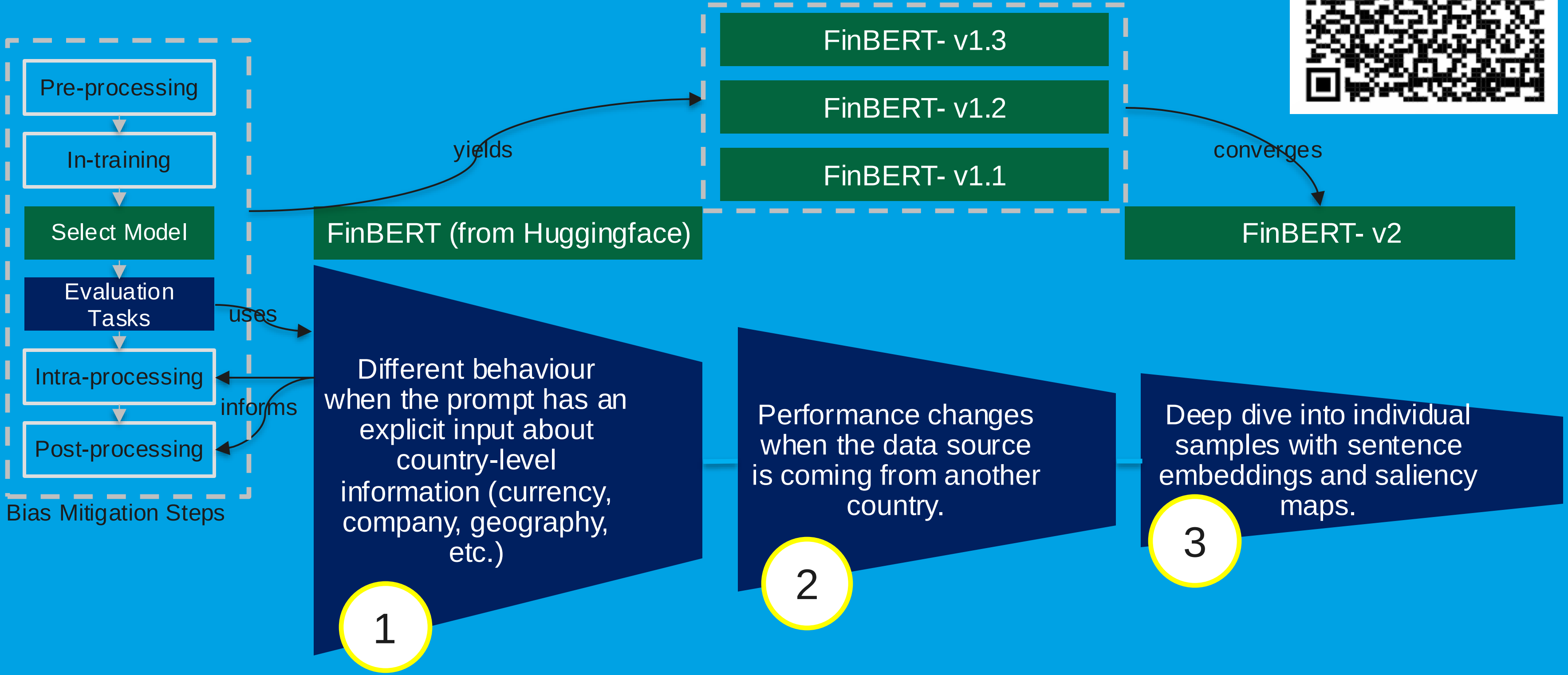
Fairness

- Should be approached from both social and technical perspectives
- A key challenge is projecting social concepts into mathematical formulations.
- Historical bias in the existing datasets is the most visible issue
- Current utilisation is under heavy human guidance.



Project Page

A sample financial sentiment tone classification model bias evaluation methodology from our Innovate UK Project:



1. Different behaviour when the prompt has an explicit input about country-level information (currency, company, geography, etc.)

2. Performance changes when the data source is coming from another country.

3. Deep dive into individual samples with sentence embeddings and saliency maps.

Explainability

- Having accurate but intuitively explainable models is more important than having complete explainability.
- The focus is auditability: Open data, corporate-level certification, etc.
- Employee training and attention to detail will gain importance

Accountability and Transparency

- Linked to customer trust, gained more importance with the recent incidents
- Clearly defining the processes and interactions is a challenge when the business logic becomes complex
- Re-evaluation of risk assessment frameworks is critical

Integrity

- It is a complex term linked with fitness, propriety, consistency, and adherence
- Measuring it is a challenge, but customer trust indicates a level of integrity
- Granularity in the operations can support evaluating overall integrity

Skills

- Fundamental literacy and critical thinking
- Internal training and recruitment have already started transforming
- Human-machine collaboration skills gained importance

Reflection and Recommendations

- Collaboration, sharing best practices, and defining levels of granularity are essential for the development of use-case dependant sector-wide analyses of LLM assessments
- Despite the security and privacy concerns, open model and data practices can bring diverse expertise from technical, legal, and ethical perspectives.

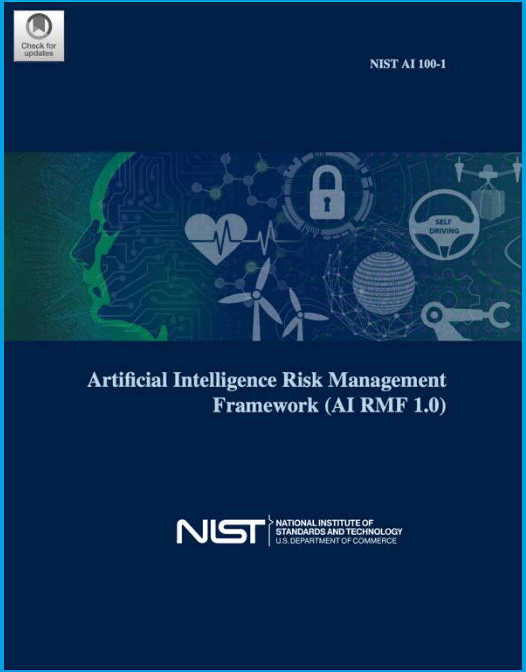
Privacy-Enhancing Technologies

- Traditional data protection mechanisms focus on threats from external adversaries.
- Privacy-Enhancing Technologies (PETs):
  - protect sensitive and confidential data in use against entities operating on the data.
  - enables multiple parties to safely collaborate on their data.
  - can provide the outcome on an analysis, without giving insight into the data.
- Examples of PETs:
  - Secure Multi-Party Computation (MPC)
  - Homomorphic Encryption
  - Federated Learning
  - Differential Privacy

Privacy-Enhancing Technologies

- Banks are very data-driven.
- More collaboration is needed, but is not always allowed.
- Bank data can be very personal and therefore very sensitive.
- Banks have an obligation to detect financial crime, but have a limited view.

Assurance, Risk and Governance Frameworks

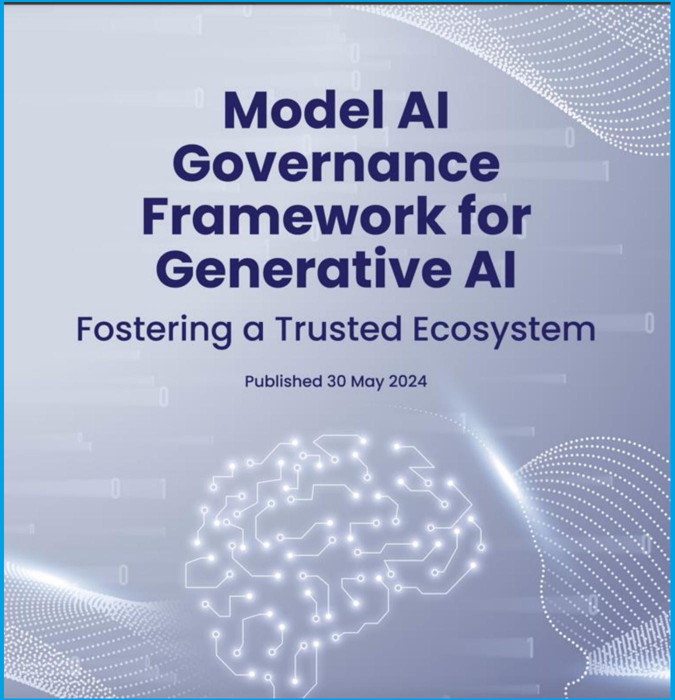


Check for updates

NIST AI 100-1

Artificial Intelligence Risk Management Framework (AI RMF 1.0)

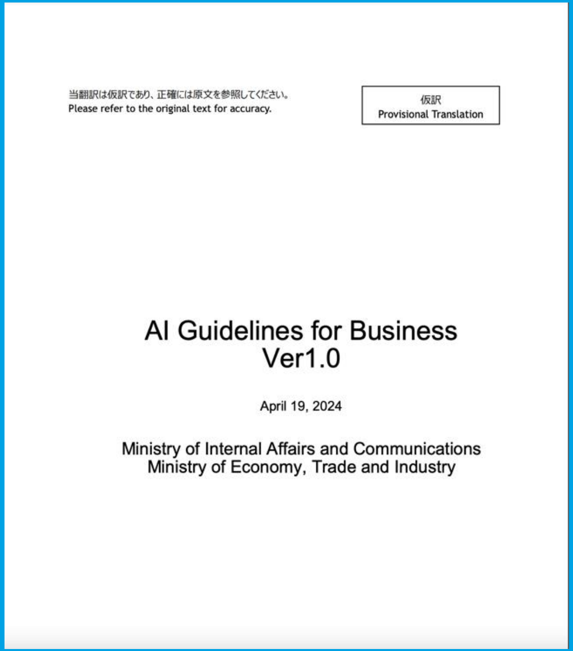
NIST NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
U.S. DEPARTMENT OF COMMERCE



Model AI Governance Framework for Generative AI

Fostering a Trusted Ecosystem

Published 30 May 2024



当翻訳は仮訳であり、正確には原文を参照してください。
Please refer to the original text for accuracy.

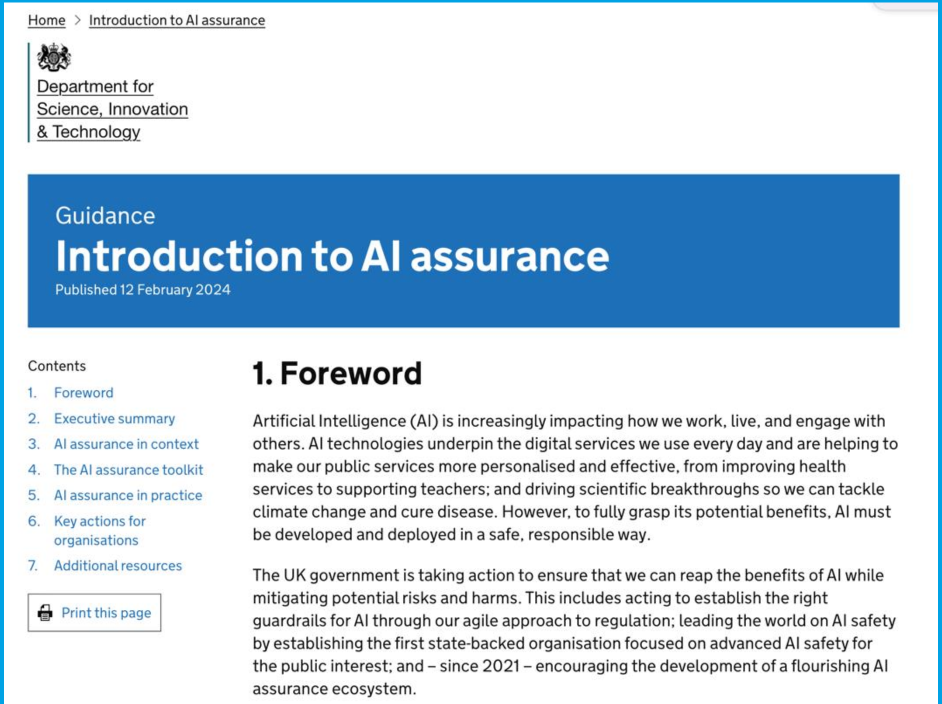
仮訳
Provisional Translation

AI Guidelines for Business
Ver1.0

April 19, 2024

Ministry of Internal Affairs and Communications
Ministry of Economy, Trade and Industry

AI Assurance UK



Home > Introduction to AI assurance

Department for Science, Innovation & Technology

Guidance

Introduction to AI assurance

Published 12 February 2024

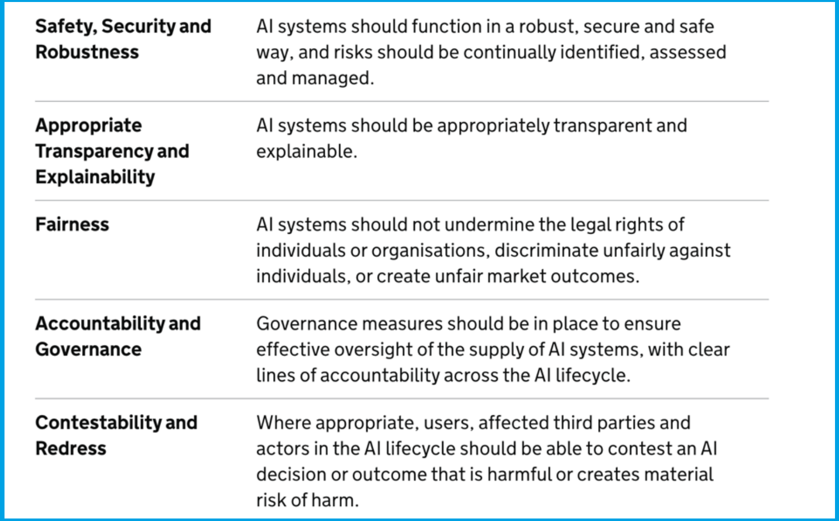
Contents

🖶 Print this page

1. Foreword

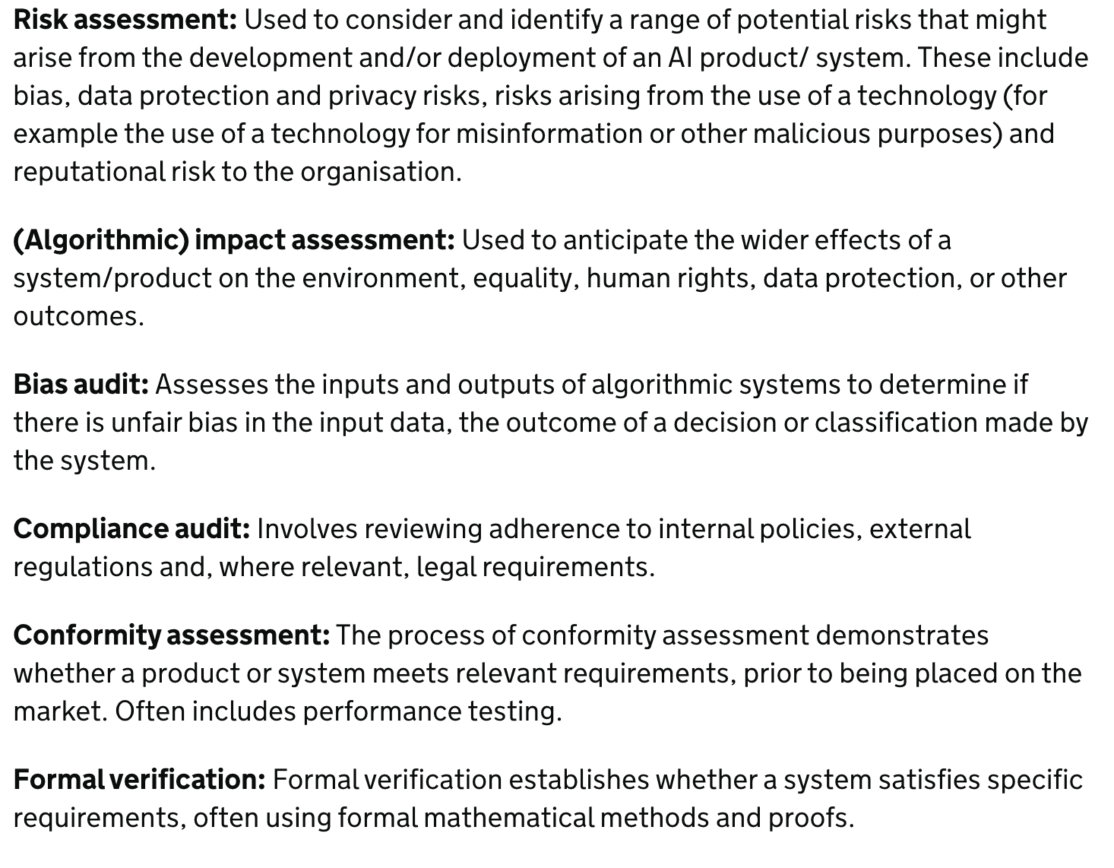
Artificial Intelligence (AI) is increasingly impacting how we work, live, and engage with others. AI technologies underpin the digital services we use every day and are helping to make our public services more personalised and effective, from improving health services to supporting teachers; and driving scientific breakthroughs so we can tackle climate change and cure disease. However, to fully grasp its potential benefits, AI must be developed and deployed in a safe, responsible way.

The UK government is taking action to ensure that we can reap the benefits of AI while mitigating potential risks and harms. This includes acting to establish the right guardrails for AI through our agile approach to regulation; leading the world on AI safety by establishing the first state-backed organisation focused on advanced AI safety for the public interest; and – since 2021 – encouraging the development of a flourishing AI assurance ecosystem.



Safety, Security and Robustness	AI systems should function in a robust, secure and safe way, and risks should be continually identified, assessed and managed.
Appropriate Transparency and Explainability	AI systems should be appropriately transparent and explainable.
Fairness	AI systems should not undermine the legal rights of individuals or organisations, discriminate unfairly against individuals, or create unfair market outcomes.
Accountability and Governance	Governance measures should be in place to ensure effective oversight of the supply of AI systems, with clear lines of accountability across the AI lifecycle.
Contestability and Redress	Where appropriate, users, affected third parties and actors in the AI lifecycle should be able to contest an AI decision or outcome that is harmful or creates material risk of harm.

AI assurance mechanisms



Risk assessment: Used to consider and identify a range of potential risks that might arise from the development and/or deployment of an AI product/ system. These include bias, data protection and privacy risks, risks arising from the use of a technology (for example the use of a technology for misinformation or other malicious purposes) and reputational risk to the organisation.

(Algorithmic) impact assessment: Used to anticipate the wider effects of a system/product on the environment, equality, human rights, data protection, or other outcomes.

Bias audit: Assesses the inputs and outputs of algorithmic systems to determine if there is unfair bias in the input data, the outcome of a decision or classification made by the system.

Compliance audit: Involves reviewing adherence to internal policies, external regulations and, where relevant, legal requirements.

Conformity assessment: The process of conformity assessment demonstrates whether a product or system meets relevant requirements, prior to being placed on the market. Often includes performance testing.

Formal verification: Formal verification establishes whether a system satisfies specific requirements, often using formal mathematical methods and proofs.

Thank You