New challenges for CISO: Artificial Intelligence, emerging technologies and regulations

COMMISSION
DE SURVEILLANCE
DU SECTEUR
FINANCIER

The regulator's viewpoint

10 May 2019

David HAGEN
CSSF, Head of IT supervision

Agenda

1. INTRODUCTION
2. AI OPPORTUNITIES
3. RISKS AND RECOMMENDATIONS
4. CONCLUSION & QUESTIONS

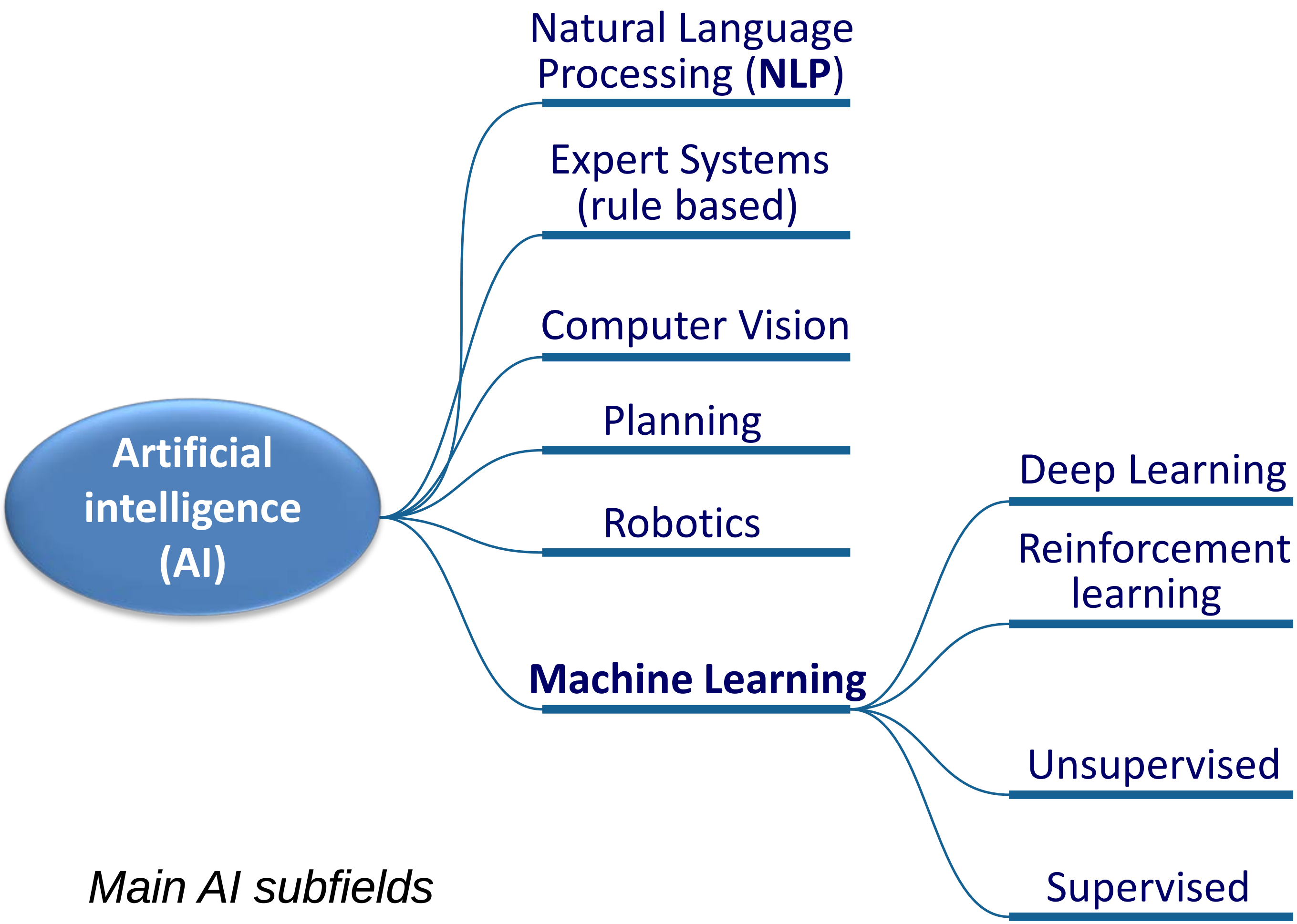
INTRODUCTION – WHAT IS AI?

- What is AI?
  - “The theory and development of computer systems able to perform tasks that traditionally have required human intelligence.”

    Financial Stability Board

  - Intelligent tasks:
    - Reasoning / Problem solving
    - Learning
    - Planning
    - Ability to understand language and speech
    - Ability to manipulate and move objects
    - etc...

INTRODUCTION – WHAT IS AI?



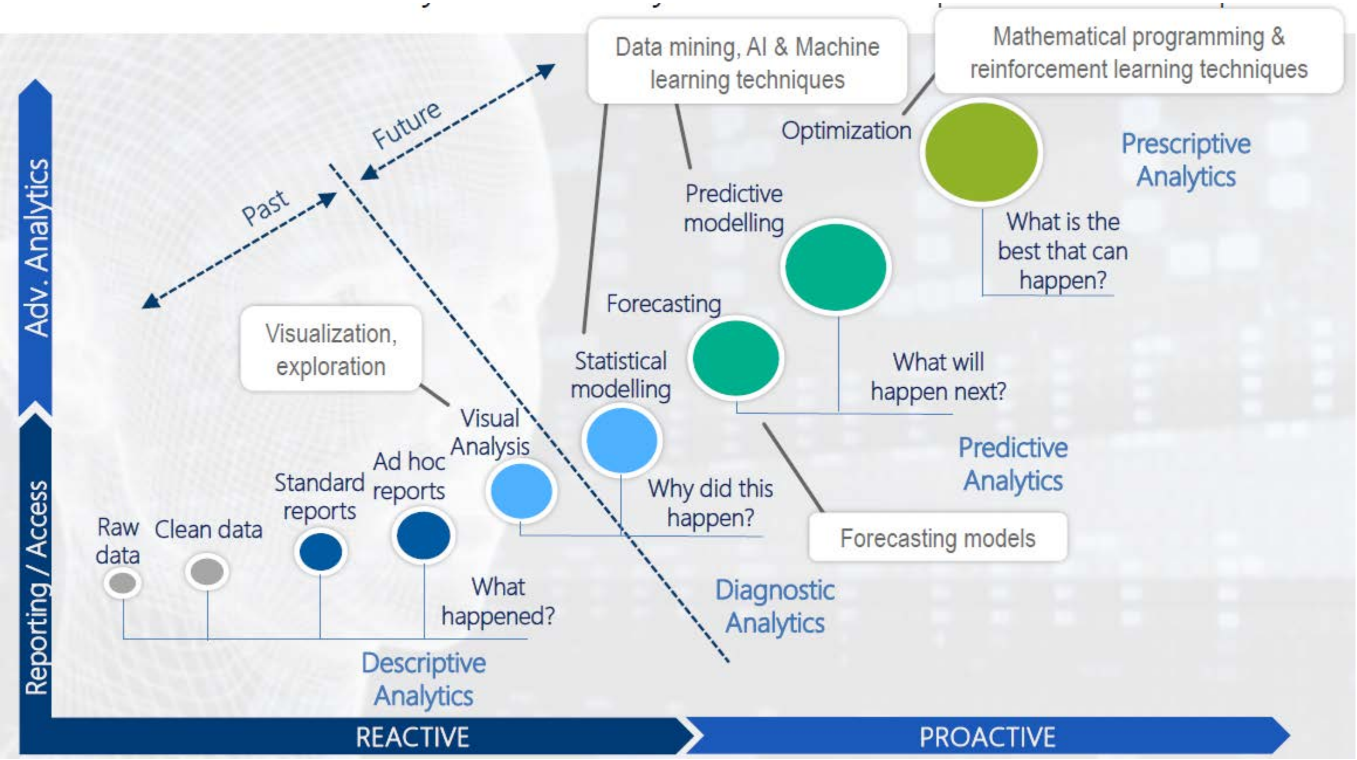
Main AI subfields

GOVERNANCE EXPECTATIONS ON AI FOR SUPERVISED ENTITIES

- The common principle underlying the supervised machine learning algorithms is:
  - Machine learning algorithms are described as learning a target function (f) that best maps input variables (X) to an output variable (Y): Y = f(X)
  - In other words, the goal is to learn the mapping Y = f(X) in order to be able to make predictions of Y for a new X. This is called **predictive modeling** or **predictive analytics**.

INTRODUCTION – WHAT IS AI?

- Data Analytics



Source: SAS (AI Luxembourg Summit 2018)

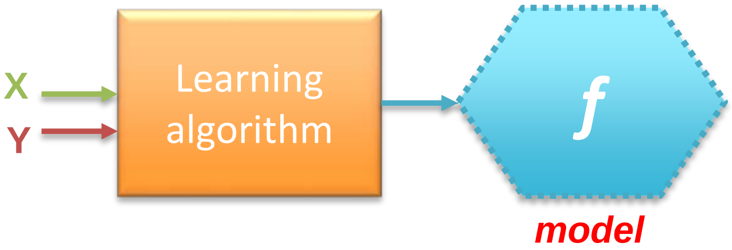
INTRODUCTION – WHAT IS AI?

Traditional programming

- The program (***f***) will , given some input x, calculate the output y (y= ***f***(x))

Machine Learning (ML)

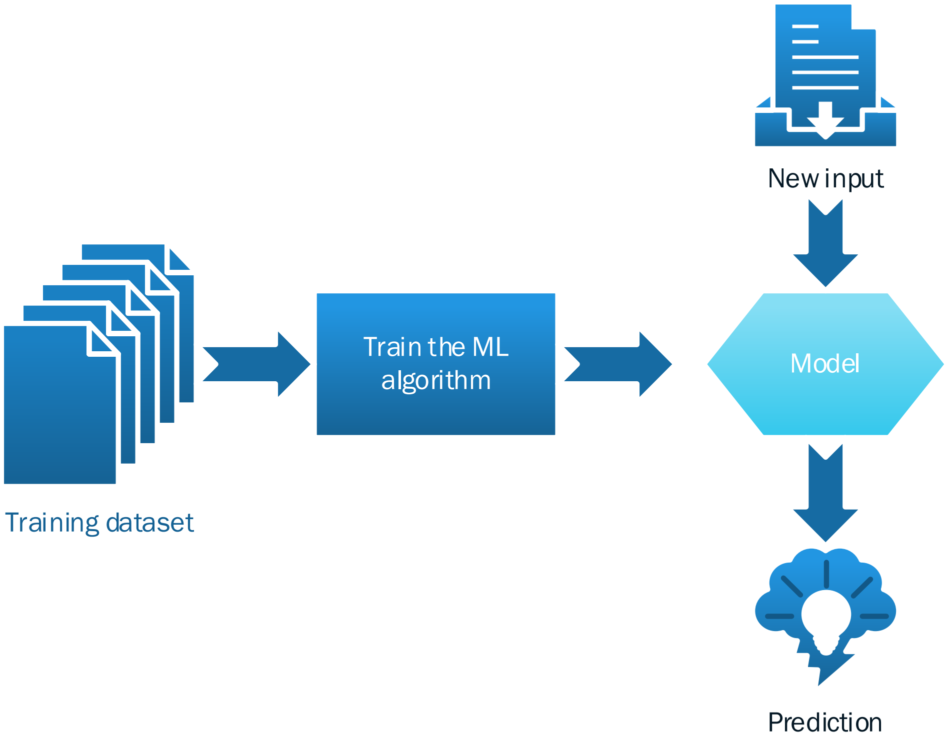
- The "program" (i.e. the function ***f*** mapping x into y) is "learnt" by the algorithm



- The mapping function ***f*** constitutes the **model**

INTRODUCTION – MACHINE LEARNING

- A **model** is a representation of what the algorithm has learnt from the training data and is used to make prediction on new input data



AI OPPORTUNITIES

AI OPPORTUNITIES – AI DEMOCRATIZATION

- AI is not new, but is now more accessible due to several factors:
  - More powerful and dense processors (GPU)
  - Lower storage cost
  - Cloud computing
  - Big data: large data sets available for “learning”
  - AI tools and platforms (e.g. DataRobot, DataIKU, Microsoft, Google, Amazon, etc..)

AI OPPORTUNITIES – USE CASES

- RPA (Robotic Process Automation) and IPA (*Intelligent* Process Automation)
- Chatbots
- Robo-advisors
- **Fraud detection**
- Terrorism Financing detection
- Credit scoring
- Other (NLP/text mining, algorithmic trading, facial recognition in KYC processes, IT security, etc...)

RISKS AND RECOMMENDATIONS

RISKS AND RECOMMENDATIONS

- Key Risk areas:
  - Data
  - Governance
  - Ethics
  - Technology
  - External providers

RISKS AND RECOMMENDATIONS - DATA

Risks

- Difficult to find the right data
- **Data quality** issues
- External data not appropriate/ not reliable

Recommendations

- **Data governance**:
  - clear roles & responsabilities for data ownership;
  - data dictionaries,
  - data quality management,
  - etc...
- Involve business data owners
- Due diligence of data source providers
- Verify adequacy of data for the target context

RISKS AND RECOMMENDATIONS - GOVERNANCE

Risks	Recommendations
• No human in the loop / uncontrolled automated actions • Lack of AI specific skills (e.g. data scientists; AI auditor,…) or over-reliance on few key staff • Fear of change/ lack of adoption by business users • Lack of understanding of AI results	• Never leave a machine to decide on critical tasks alone (**human oversight** / dual validation) • Involve Internal Audit, Risk and Compliance functions in AI projects since the beginning (+ training) • Use external AI experts and ensure knowledge transfer • Involve business users from the start (key success factor)

RISKS AND RECOMMENDATIONS - ETHICS

Risks	Recommendations
• **Bias** (within training/validation datasets, algorithms,…) • **Discrimination** (e.g. populations not fairly represented in the training data) • **Personal data collected/processed without consent** (e.g. behavioral data) • **Accountability** of AI actions	• AI code of conduct (incl. fairness) • Identify and **remove bias** (during data preparation) • Active inclusion: seek for diversity in training / validation data • Create specific datasets to **test against discrimination** • Challenge the need for personal data/ **data privacy by design** • **Accountability cannot be delegated to a machine**: ultimate responsibility relies with senior management

RISKS AND RECOMMENDATIONS - ETHICS

Risks	Recommendations
• Lack of **explainability**/ «black box» models • Lack of **auditability**	• **Document the data preparation** process (model blueprint) • Document the choice of the algorithm; choose more **interpretable algorithms** (e.g. decision trees) depending on the criticality of the system • Use **explainable AI** techniques (e.g. interpreter) when required • Implement detailed **audit logs** • Implement technical means to **simulate the input data** into the AI to perform investigations in case of need

RISKS AND RECOMMENDATIONS - TECHNOLOGY

Risks

- Change management:
  - Lack of involvement of business users
  - **data leakage** (output information in input data)
  - Lack of documentation/ traceability
- Poor results/ **model not accurate**
- **Predictive power of ML is limited to what can be learnt from past observations: cannot predict something never seen before!**

Recommendations

- **Document** the choices made at each step of the development process (e.g. feature selection, choice of algorithm,...)
- Prefer using integrated platforms
- Monitor model performance (via **accuracy** metrics and business KPIs, etc.) and **update the model** (re-training) when needed
- Perform **parallel runs** (old Vs new AI model)

RISKS AND RECOMMENDATIONS - TECHNOLOGY

Risks	Recommendations
• Insufficient error and incident management • Technical operational issues (e.g. interfaces with legacy systems) • Security vulnerabilities/ robustness to attacks	• Plan for **error and incident management** (e.g. RPA processes can generate frequent operational errors) • Apply **security by design** • Test **model robustness** • Perform independent security reviews according to the criticality of the system • **Technological watch**: monitor improvements in the attack and defense techniques (remember that **attackers are also using AI to improve their attacks!)**

RISKS AND RECOMMENDATIONS – EXTERNAL PROVIDERS

Risks	Recommendations
• **Dependency on few providers** • General outsourcing risks • **Systemic risks**: if the same model is used by many institutions, market movements and errors may be amplified	• Plan for the **maintenance** of the AI solution (e.g. have the right AI staff internally Vs SLA with external provider) • **Apply best practices** and regulatory recommendations on IT outsourcing (e.g. circular CSSF 12/552) • **Customize** the AI product • **Monitor systemic effects**

KEY SECURITY ISSUES

- Why should someone hack an AI?
  - To stop the service based on AI
  - To hijack the AI
    - For personal needs
      - To bypass the analysis (i.e. KYC/AML, biometrics)
      - To Influence the outcome in favor of the hacker (i.e. Asset management, credit scoring, political elections)
    - To harm the provider
      - To influence the outcome used by the provider (i.e. wrong investments)
      - To fuzz the results in an random way that it will lead to a loss of trust

KEY SECURITY ISSUES

- How could a hacker corrupt an AI?
  - By acting on the data
    - Initial data
    - Learning process
  - By acting on the model
    - Tuning parameters
    - Mathematical model substitution
    - Code modification
  - By acting on the explainability tools
    - Hiding the biais

RISKS AND RECOMMENDATIONS

Risks

- **Data quality** issues
- External data not appropriate/ not reliable
- No human in the loop / uncontrolled automated actions
- Lack of AI specific skills (e.g. data scientists; AI auditor,…) or over-reliance on few key staff
- **Bias** (within training/validation datasets, algorithms,…)

Threats

- Corruption of data
- Security threats on external data providers
- External data providers phishing (MITM, fake data provider by redirection)
- Undetected corruption (no human, no audit)
- Modified datasets, algorithms
- Traceability of AI actions

RISKS AND RECOMMENDATIONS

Risks

- **Discrimination**
- **Personal data collected**
- **Accountability** of AI actions
- Lack of **explainability**/ «black box» models
- Lack of **auditability**

Threats

- Corruption of data
- Security threats on external data providers
- External data providers phishing (MITM, fake data provider by redirection)
- Undetected corruption (no human, no audit)
- Modified datasets, algorithms
- Personal data leakage (also in AI output)
- Traceability of AI actions

RISKS AND RECOMMENDATIONS

Risks

- Change management:
  - Lack of involvement of business users
  - **data leakage** (output information in input data)
  - Lack of documentation/ traceability
- **Predictive power of ML is limited to what can be learnt from past observations: cannot predict something never seen before!**

Threats

- Corruption of data
- Security threats on external data providers
- External data providers phishing (MITM, fake data provider by redirection)
- Undetected corruption (no human, no audit)
- Modified datasets, algorithms

CONCLUSION

CONCLUSION

- Plan for **error and incident management**
- Apply **security by design**
- Test **model robustness**
- Perform **independent security reviews** according to the criticality of the system
- **Technological watch**: monitor improvements in the attack and defense techniques (remember that **attackers are also using AI to improve their attacks!**)
- **Use only the necessary data**

- Key controls: **data governance, human in the loop**
- Key challenges: **fairness, explainaibility, auditability**

CONCLUSION

Reference:

- CSSF whitepaper "Artificial Intelligence: opportunities, risks and recommendations for the financial sector" www.cssf.lu/fileadmin/files/Publications/Rapports_ponctuels/CSSF_White_Paper_Artificial_Intelligence_201218.pdf

QUESTIONS ?