Ethical and legal issues in the design and use of AI systems in health

Thomas Guyet

Ethical and legal issues in the design and use of AI systems in health

eHealth & ethics day, 13/04/2023

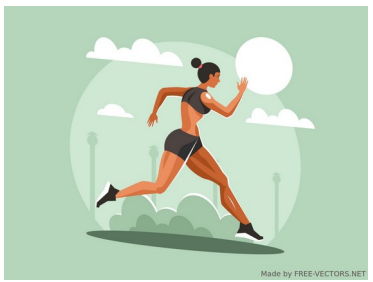
Pôle Léonard de Vinci in Paris-La Défense

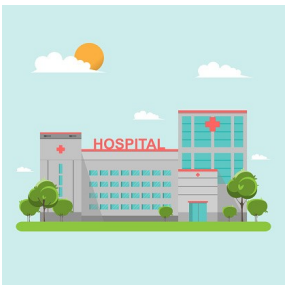
Thomas Guyet
AIstroSight, Inria/Theranexus/UCBL/HCL

An example ...

A healthcare software company is developing a patient triage tool to assess the level of risk of stroke based on a machine learning algorithm. The AI system was trained on data sold by a provider and included only stroke cases in patients >70 years old, mostly men.

A 35-year-old woman lost consciousness while jogging, was brought to hospital by a friend and the hospital prioritised other patients who arrived after him. The medical diagnosis then identifies a stroke for which earlier management would have limited the long-term consequences. The hospital used the AI system to prioritise patients in emergency.





An example ...

A healthcare software company is developing a patient triage tool to assess the level of risk of stroke based on a machine learning algorithm. The AI system was trained on data sold by a provider and included only stroke cases in patients >70 years old, mostly men.

A 35-year-old woman lost consciousness while jogging, was brought to hospital by a friend and the hospital prioritised other patients who arrived after him. The medical diagnosis then identifies a stroke for which earlier management would have limited the long-term consequences. The hospital used the AI system to prioritise patients in emergency.

- What are the ethical issues involved in this situation?
- What questions for the law?
- What are the specificities of using an AI system?

→ in the context of development of AI as an epidemiological studies!

AI Systems -- UNESCO definition

AI systems are information-processing technologies that integrate models and algorithms that produce a capacity to learn and to perform cognitive tasks leading to outcomes such as prediction and decision-making in material and virtual environments. AI systems are designed to operate with varying degrees of autonomy by means of knowledge modelling and representation and by exploiting data and calculating correlations. AI systems may include several methods, such as but not limited to:

- machine learning, including deep learning and reinforcement learning;
- machine reasoning, including planning, scheduling, knowledge representation and reasoning, search, and optimization.

Annex VII of the Proceedings of the General Conference of UNESCO, Resolution of the 41st session, Nov. 2021

AI questions on machine learning

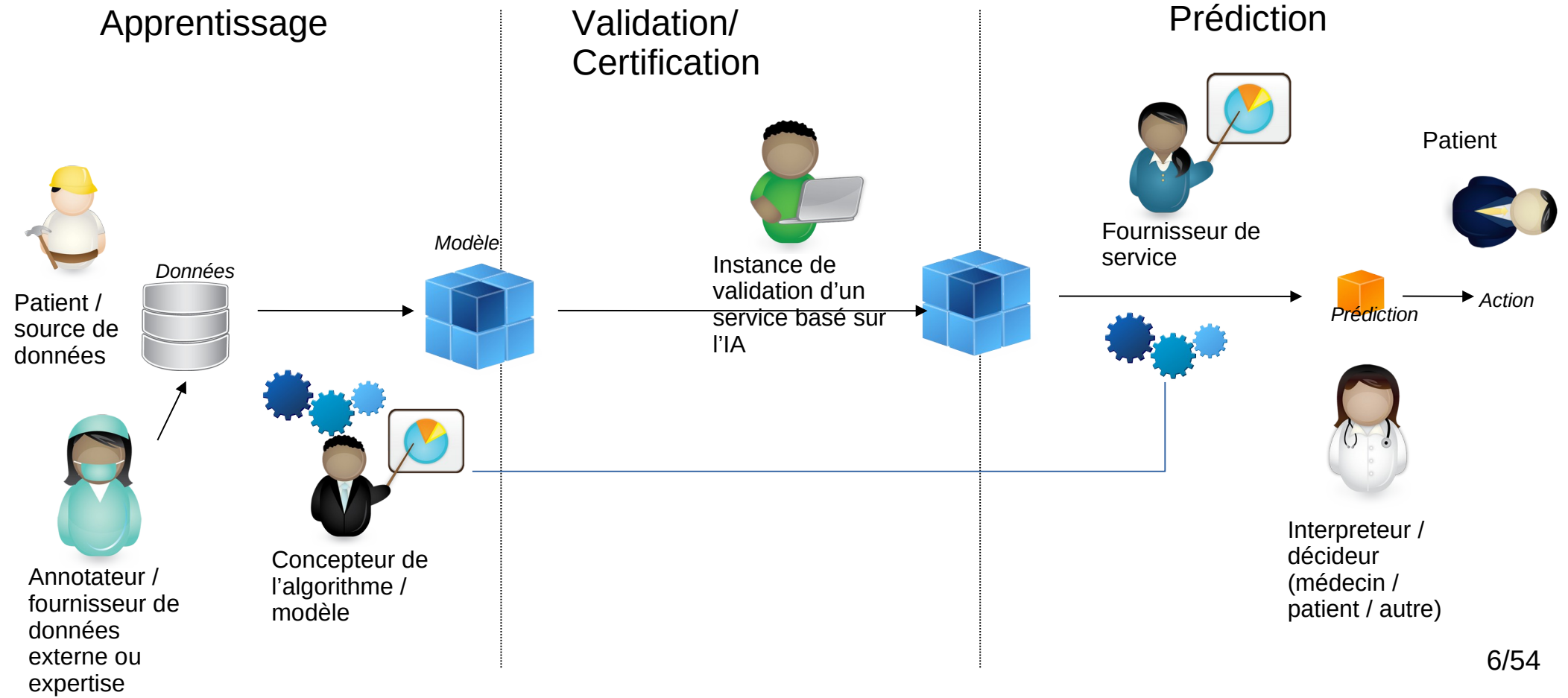
Diversity of Artificial Intelligence

- Automatic reasoning
- Collective intelligence
- Knowledge engineering
- Machine learning
- ...

AI systems with learning capabilities concentrate ethical and legal issues

- Learning AI system = Decision-making system
  - Representable as any computer system (Turing machine)
- The design of such a system changes
  - It is no longer a programmer who describes how to make the decision
  - The program is built indirectly *from examples*
  - Abandonment of full control over the decision-making process

Quelques acteurs (personne morale ou physique) autour des systèmes d'IA en Santé (cycle de vie)



AI in health

AI intervention in health care systems ...

- prescription (e.g. medication)
- diagnosis (radiological image analysis more efficient than humans)
- surgery assistance (Da Vinci robot)
- the organisation of care (prediction of hospitalisation times, organisation of cares)
- the organisation of the healthcare system (epidemiology based on massive health data)

Inserm, https://www.inserm.fr/dossier/intelligence-artificielle-et-sante/

The promise of AI in health

Improved diagnostic/prognostic accuracy

- The use of AI technologies could improve decision making
- e.g. in radiology/dermatology today

Improving efficiency

- Improvement of the care system
- Focusing caregivers' time on useful tasks
- Speed/timeliness of diagnosis and intervention

Care personalisation

- Tailoring care to patients
- Prevention through monitoring

More objective decisions

- Rationalisation of the decision

At the extreme, Sam Altman (Twitter, 4th Feb 2022)

- Techno-optimism is the only good solution to our current problems (...) We can build AGI. We can colonize space. We can get fusion to work and solar to mass scale. **We can cure all human disease**. We can build new realities.

Risks and fears of AI in health

There are many situations where one can imagine "damaging facts".

- for patients
  - misdiagnosis, diagnostic delays, intervention error
  - limitation of possible remedies
- for caregivers
  - loss of jobs
  - reduced freedom of prescription and practice
  - loss of technical skills
- for citizens
  - Privacy
  - inequity of treatment
  - respect for informed choice
  - Misinformation
  - privatisation of care

Work carried out within the AIRacles Chair

Objective: to exploit the AP-HP Health Data Warehouse (EDS) with three axes

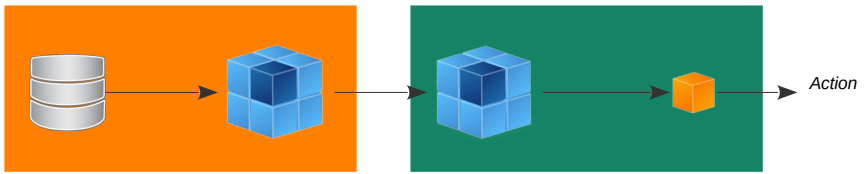
- Axis 1 - Integration of heterogeneous data at a given time and/or from longitudinal follow-up, including data of a clinical, biological or textual nature (hospitalization reports, imaging, anatomopathology, etc.)
- Axis 2 - Identification of frailty phenotypes and care trajectories involving the conduct of unsupervised analyses (patient representation and identification of typical profiles and trajectories)
- Axis 3 - Construction and integration of predictive tools useful for clinical practice involving the conduct of supervised analyses based in particular on machine learning approaches

Study example: OPTISOINS, understanding and improving the care pathways of patients hospitalised for lung resection.

Work carried out within the AIRacles Chair

Specificity: use of data from the APHP data warehouse

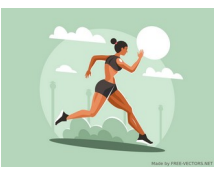
- Secondary data (idem SNDS): *Data collected in the context of care for medical or administrative reasons*
- Massive data
  - APHP = 9 hospitals
  - Largest hospital structure in Europe



- Heterogeneous but structured data
- Available data: no delay in data collection
- Data representative of the information available in the care setting

Exploitation of these data by machine learning methods

- Identification of care trajectories (phenotyping)
- Prognostic algorithms
- Deep learning: non-selective exploitation of data
- Framework with similar objective as in our example



Some ethical and legal questions?

Development of AI systems in health

- Strong expectations to improve our care systems and care
- Potential risks: personal and social

How to address the many questions that arise about the opportunities and limitations of developing AI systems in health

Need for an inter-disciplinary approach between

- Philosophy/epistemology: thinking about the evolution of technology and science induced by AI
- Law/Ethics: defining a collective and responsible framework for the development of AI systems
- Computer science: identification of real risks linked to the use of AI (today and tomorrow)
- Health: a source of practical cases and questions

Outline

Epistemological perspective

- What (re)evolution does AI bring?
- Is there a real need to treat AI differently from any technical object?
- How can we distinguish between myths, fantasies and reality?

AI development: which ethical principles?

- Ethical principles of AI (in health)
- A critical look at principalism and epidemiology

AI development: what regulatory framework (in Europe)?

- Motivation of the European regulatory framework
- Use of the data
- Responsibility framework

Emergence and transformation of AI

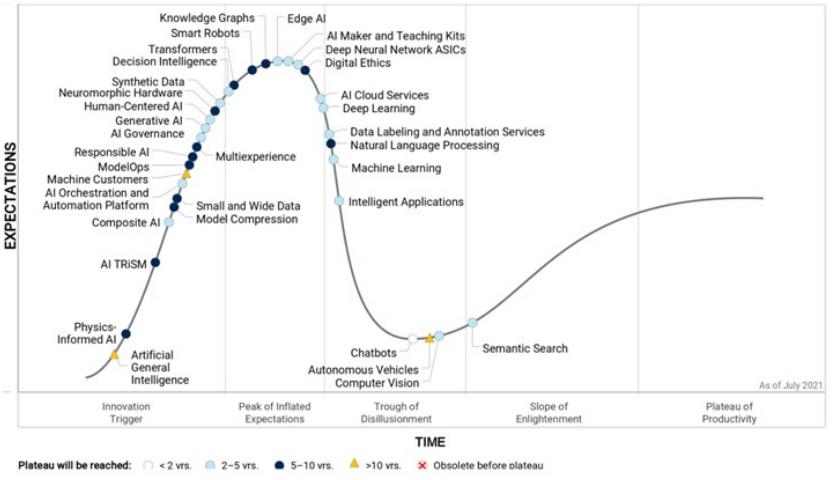
A little history

- Pascal ... Babbage ...
- 50's : Early work on *cybernetics*
  - Neural networks McCulloch & Pitts, Hebb)
  - Macy conferences
- 70-80's: expert systems
  - Knowledge representation and modelling of reasoning
  - Diagnostic support (MYCIN)
- 80-90's: Return of connectionism and the subsymbolic movement
  - Multi-layer perceptron (1985)
  - Distributed AI (Genetic Algorithms, Artificial Life, ...)
- 90-00's: revival of machine learning
  - Preferences for Big Data approaches
  - *1997 Deep blue vs Kasparov*
- 00-10's: emergence of Big Data
  - Automates the identification of features useful to a model
  - Spectacular advances in image classification
- 10-20's: Era of deep learning (AI = ML)
  - Increasing the size of the architectures
  - Wide dissemination of technologies
- 20's: Generative models
  - *ChatGPT / MidJourney*
  - Content (text, image, sound) generation

Emergence and transformation of AI

A little history

- Pascal ... Babbage ...
- 50's : Early work on *cybe*
  - Neural networks McCulloch
  - Macy conferences
- 70-80's: expert systems
  - Knowledge representation reasoning
  - Diagnostic support (MYCIN
- 80-90's: Return of conne subsymbolic movement
  - Multi-layer perceptron (198
  - Distributed AI (Genetic Algo Life, ...)
- 90-00's: revival of machine learning
  Data approaches
  Kasparov
  ce of Big Data
  tification of features useful
  es in image classification
  ep learning (AI = ML)
  of the architectures
  of technologies



multilevel hype cycles

Gartner

models

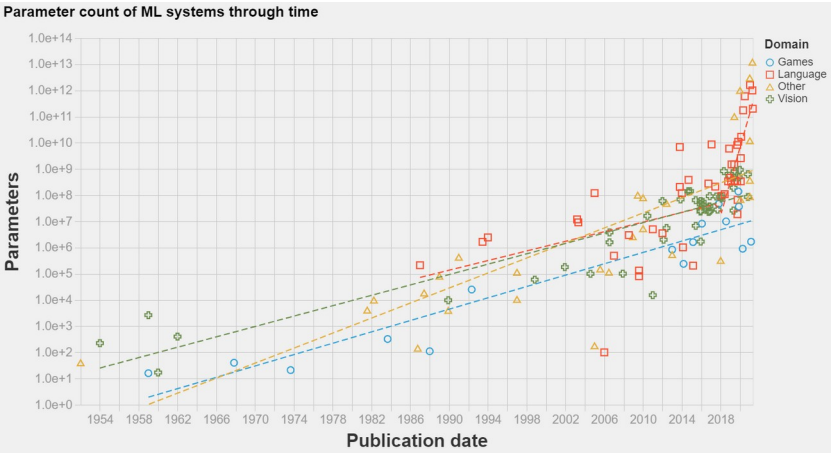
irney

ge, sound) generation

Emergence and transformation of AI

Rising of new machine learning thanks to the convervence of

- Calculation capabilities
- Available data
- Research and investment efforts



Recent accelerations

- Diffusion of AI technologies
- Acceleration of spectacular results
  - Go game
  - Autonomous driving
  - Machine translation
  - Content generation

Tends to suggest

- Increasingly abstract automation of cognitive tasks
- An unstoppable and limitless progression
- towards Artificial General Intelligence -- AGI

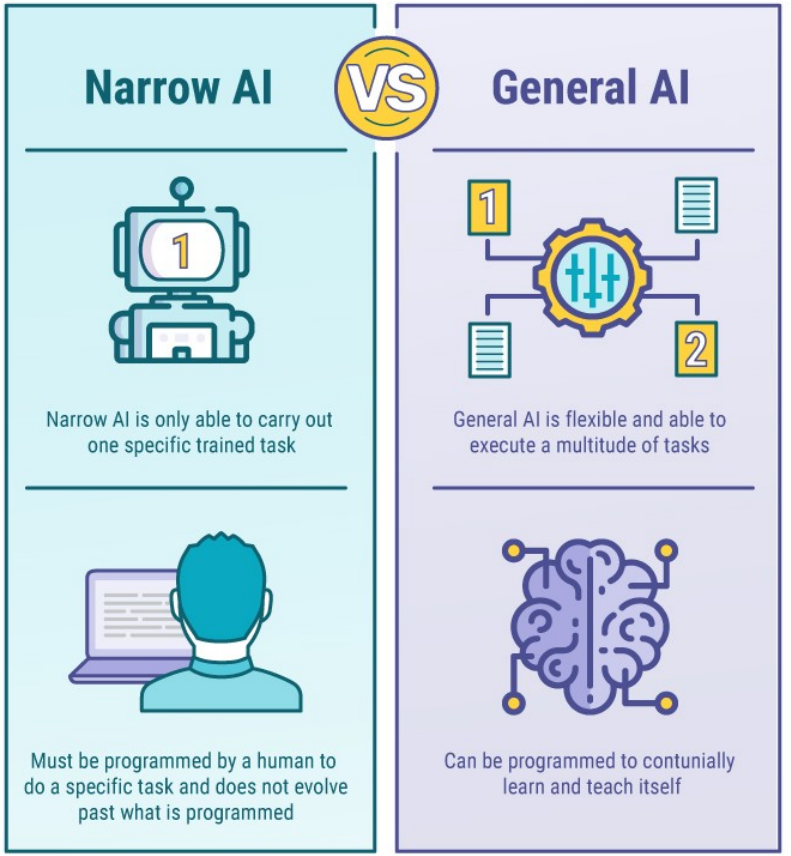
Myth of singularity

Singularity

- Transition to a "strong" AI that would
  - Autonomous
  - Able to deal with many problems
  - Even with a 'conscience'.
- Several times predicted ...
- OpenAI: Planning for AGI and beyond
  - Our mission is to ensure that artificial general intelligence-AI systems that are generally smarter than humans-benefits all of humanity.Sam Altman: Future of AI will "create new knowledge

"AI does not exist" L. Julia

- In the sense of AGI (Artificial General Inteligence)
- Rethinking human/machine comparisons in terms of energy

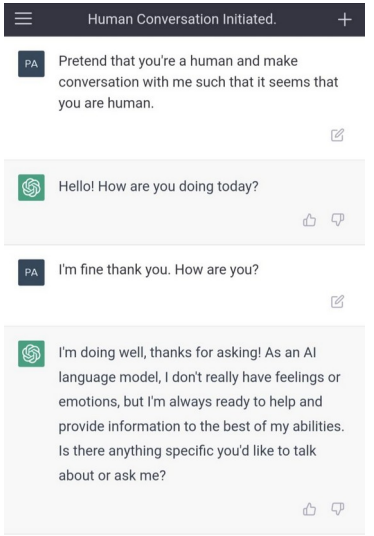


https://openai.com/blog/planning-for-agi-and-beyond

Role of the term 'Artificial Intelligence' in the myth

The term helps to convey an imaginary

- Today, no AI system, no matter how complex our computational capabilities, can be called intelligent
- Wide visibility in media
- Dump AI ?



Experiment by Pascal Hitzler with chatGPT

A marketing buzzword

- Term carrying the promise of innovation
- AudikaFull "all the sounds that are important to you are played back thanks to the power of artificial intelligence

Should a more neutral term be preferred for these technologies?

The contribution of epistemology to thinking about AI

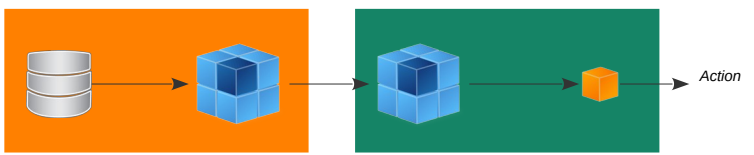
Is AI a scientific 'revolution'?

- Specific concepts in epistemology (T. Kuhn)
- "paradigm shift" = change of world view / relationship to the world
- The new paradigm has to solve more puzzles than the old one
  - What was the old paradigm

The contribution of epistemology to thinking about AI

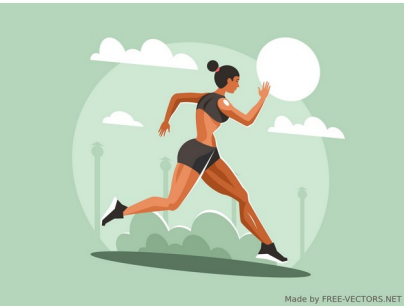
Is AI a science? A technique?

- AI system as a tool
  - Performs a function
  - Its design is put at a distance (unknown)
  - *Do I need to understand the technique to use it?*
- AI as a technique for science (*AI is seen in its full cycle?)*
  - In health, more than anywhere else: the discovery of new techniques is governed by the scientific approach
  - AI: empowerment tool for epidemiologists (analyse more data, more accurately)
  - Essentially this tool in many other sciences
  - *Application of ethical/legal principles of research?*



An AI system for patient triage

(...) The hospital used the AI system to prioritise patients in emergency.. (...)



- What characterises an AI system?
- Three ‘systems’ of stroke stratification
  - NIHSS score (Brott et al, 1989) - reference score
    - Evaluation of 14 criteria (Level of consciousness, Simple order, Facial paralysis, ...)
    - Severity classification
      - » 1-4 points : Minor stroke
      - » 5-15 points : Moderate stroke
      - » 15-20 points : Severe stroke
      - » > 20 points : Severe stroke
  - Quick-NIHSS (Geiser, 2019)
    - 5 binary variables + global score from 0 to 5
    - Constructed by ANOVA on a mono-centric database, 3995 patients
  - Machine learning model (Yu et al., 2019)
    - Learning a C4.5 model from 227 patients

[1] Brott T, Adams HP; Olinger CP, Marler JR, Barsan WG, Biller J, Spilker J, Holleran R, Eberne R, Hertzberg V, Rorick M, Moomaw CJ, Walker M. Measurements of acute cerebral infarction: a clinical examination scale. Stroke 1989;20:864-70
[2] Geiser A. Elaboration of the Quick-NIHSS: A simple and rapid score for prehospital triage of suspected stroke, PhD Thesis, 2019
[3] Yu J. et al, Semantic Analysis of NIH Stroke Scale using Machine Learning Techniques, International Conference on Platform Technology and Service (PlatCon), 2019

The contribution of epistemology to AI design

What would be new?

1) AI: data-driven approach (Elodie Giroux)

- Pure" discovery science from data (knowledge building without theory)
  - Data neutrality
  - Broadening the concept of data
  - The truth would be in the mass of data (either hidden or debiased)

- Relativisation of causality in favour of correlation
  - Sufficiency of the empirical result
  - No need to research causes

- Overcoming traditional logics
  - Idea of better exploitation of digital data by digital
  - Questioning the explanation of decision logic

The contribution of epistemology to AI design

What would be new?

1) AI: data-driven approach (Elodie Giroux)

Pure" discovery science from data (knowledge building without theory)

- Data neutrality
- Broadening the concept of data
- The truth would be in the mass of data (either hidden or debiased)
- **There is no such thing as "pure data" (S. Leonelli): assumptions implicit in the data (+PIML)**

Relativisation of causality in favour of correlation

- Sufficiency of the empirical result
- No need to research causes
- **Explicability of models, identification of causes required in health research**

Going beyond traditional logics

- Idea of better exploitation of digital data by digital
- Questioning the explanation of decision logic

The contribution of epistemology to AI design

What would be new?

2) AI: towards autonomous systems

- Continuous learning: giving up some control by the designer
  - Raises many new questions

- Issue of the emergence of functional systems without control
  - Treated by Cybernetics (1950s)
    - Cybernetics: science of control (Norbert Wiener)
    - Second order cybernetics: theory of self-organising systems (von Foerster)
  - Autopoietic systems (H. Maturana)

Connected watches that measure physiological constants and make personalised recommendations

→ online learning of user habits

Take home message

Does AI introduce revolutions in technology?

- Little **change in the nature of the** technology
  - Developments to be expected as a result of technological change
  - The usual battle between technophobes and technolaters
  - Epistemological developments that need attention
    - Data-driven epistemology
    - Cybernetics epistemology
- But **changes in magnitude**
  - On the number of people affected by a single technology
  - On the speed of their dissemination
  - On ease of access to use

These changes, particularly in their application in health, require the promotion of ethics and regulation

- Different issues for the design and use of AI systems

Outline

Epistemological perspective

- What (re)evolution does AI bring?
- Is there a real need to treat AI differently from any technical object?

AI development: which ethical principles?

- Ethical principles of AI (in health)
- A critical look at principalism and epidemiology

AI development: what regulatory framework (in Europe)?

- Motivation of the European regulatory framework
- Use of the data
- Responsibility framework

Framework of the ED-AIM project
https://www.mfo.ac.uk/article/ed-aim-ethical-design-artificial-intelligence-models-patient-management-and-treatment

CNRS PRIME80 with Maison Française d'Oxford
Philosophy thesis by **Éric Pardoux** (in progress)

Ethical considerations in the design of an AI system in health

There are many specific ethical issues related to the introduction of AI systems in healthcare

- Acceptability of errors
- Change in the doctor-patient relationship
- Adbiquation in front of the machine
- Informed consent in the face of opaque algorithms
- ...

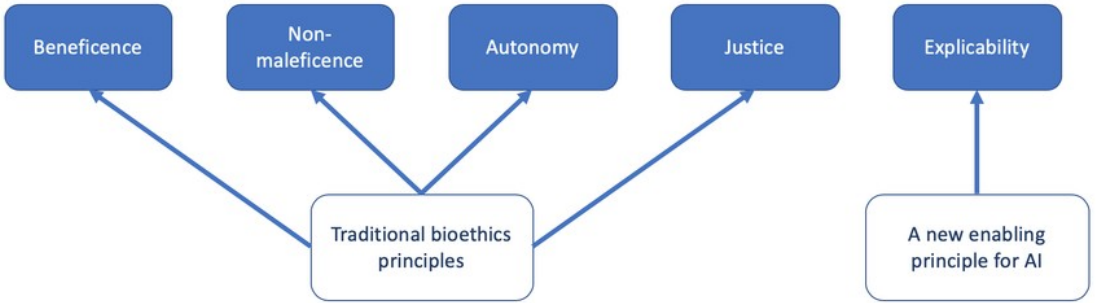
There are other issues related to the design of an AI (learning) system

- Benefit/risk balance between public health interest and open data
- Loss of competence of doctors
- Self-interest vs. general interest

How to guide or drive an ethical reflection?

Principled artificial intelligence

What's already exists?



- Principles for ethical foundation of AI
  - 4 principles of biomedical ethics by Beauchamp and Childress (2001)
    - Autonomy
    - Beneficence
    - Non-maleficence
    - Justice
  - + Explainability (Fioridi and Cowls, 2019, A Unified Framework of Five Principles for AI in Society)
- Seven key requirements for Trustworthy AI (Ethics Guidelines for Trustworthy AI, 2019)
  - human agency and oversight,
  - technical robustness and safety,
  - privacy and data governance,
  - transparency,
  - diversity, non-discrimination and fairness,
  - environmental and societal well-being and
  - accountability
- In fact ... there are many documents that set out ethical principles for AI design, from :
  - scientific societies
  - private companies
  - NGOs

Principled artificial intelligence

A Map of Ethical and Rights-Based Approaches to Principles for AI, 2020

Jessica Fjeld, Nele Achten, Hannah Hilligoss, Adam Nagy, Madhulika Srikumar

Privacy
Accountability
Safety and Security
Transparency and Explainability
Fairness and Non-discrimination
Human Control of Technology
Professional Responsibility
Promotion of Human Values
Human rights

Implementing ethical principles
The case of digital epidemiology

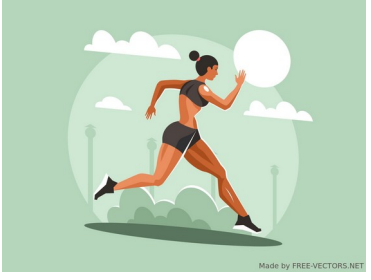
Process of designing AI from research health data

- Ethical principles applied in research (+ regulatory framework)
- Establishment of scientific and ethical evaluation committees
  - CEREES for the Health Data Hub
  - CSE for APHP
- A priori evaluation criteria
  - The purpose and methodology of the research
  - The need to use personal health data
  - Ethical relevance
  - The scientific quality of the project
  - If applicable, on the public interest character of the project
- Ex-post validation of research results through scientific publication
- Towards the integration of user representatives in committees

In this context of medical research, reasonable application of standard and adapted ethical principles

How can similar practices be encouraged for products resulting from industrial developments?

A healthcare software company is developing a patient triage tool to assess the level of risk of stroke based on a machine learning algorithm. The AI system was trained on data sold by a provider and included only stroke cases in patients >70 years old, mostly men.



Limits

For the question of the effective application of the principles in the framework of companies: move from principles to implementation!

- Goes through regulation (see part three)
- Information (fair, understandable and undistorted)
- Device/system certification

Limits of principalism

- Limits of ethics that are reduced to compliance with guidelines
- Who defines the principles to be respected?
  - e.g. AI design principle in China: in line with the Party's ambitions

Ethical design of AI systems vs. Design of ethical AI systems

- e.g. Bloom: his "ethical" conception does not guarantee his behaviour

On going work

Pro-ethical design of an artificial intelligence systems (ED-AIM, E. Pardoux)

- Reflecting on and proposing a framework for the ethical design of AI systems in health
- How to promote and encourage the implementation of good practices
  - Throughout the AI life cycle
  - Throughout the AI design process (from training to validation)

- Example of ongoing work on the design of epidemiological studies
  - Towards the formalisation of epidemiological studies leading to the design of AI systems (models) from data
  - According to the FAIR principles
    - FAIR: Findable, Accecccible, Interoperable and Reusable
    - Transposed to the notion of study
  - Desired inputs
    - Objectivise the studies to facilitate the identification of potential problematic steps
    - Accompanying (and tracing) the design of an AI system

Outline

Epistemological perspective

- What (re)evolution does AI bring?
- Is there a real need to treat AI differently from any technical object?

AI development: which ethical principles?

- Ethical principles of AI (in health)
- A critical look at principalism and epidemiology

AI development: what regulatory framework (in Europe)?

- Motivation of the European regulatory framework
- Use of the data
- Responsibility framework

Work carried out in the framework of the CNRS/DRIAS project
http://drias.irisa.fr/

P. Saurel, C. Paillard, H. Muscat, D. Bourcier, N. Le Meur, E. Oger, T. Alalrd

General context

Development of AI systems in health

- Strong expectations to improve our care systems and care
- Potential risks: personal and social

What guides the (European) regulator: excellence and trust strategy

- Providing protection for people
- Promote technological innovation
- Ensuring free competition

Excellence and trust in artificial intelligence

Trustworthy AI can bring many benefits, such as better healthcare, safer and cleaner transport, more efficient manufacturing, and cheaper and more sustainable energy.

Artificial intelligence (AI) can help find solutions to many of society's problems. This can only be achieved if the technology is of high quality, and developed and used in ways that earns peoples' trust. Therefore, an EU strategic framework based on EU values will give citizens the confidence to accept AI-based solutions, while encouraging businesses to develop and deploy them.

General context

Development of AI systems in health

- Strong expectations to improve our care systems and care
- Potential risks: personal and social

A central question: *That of responsibility in a context of medical use and involving an AI?*

Need for an inter-disciplinary approach between

- Law: the concept of liability is governed in part by law
- Computer science: identification of risks related to the use of AI (today and tomorrow)
- Health: a source of practical cases and questions

General context

Development of AI systems in health

- Strong expectations to improve our care systems and care
- Potential risks: personal and social

A central question: *That of responsibility in a context of medical use and involving an AI?*

Need for an inter-disciplinary approach between

- Law: AI raises very topical questions (law under construction, eg by the EU)
- Computer science: liability drives current AI research + legal design
- Health: how AI methods should look in future medical practice

The framework of accountability in law

A **prejudice** is analysed as a consequence of a damage that the law will apprehend. It is a subjective notion and is likely to give rise to a right to compensation. Accountability must then be imputed.

A **Damage** is an impairment of a situation or a person (e.g. privacy issue, personal injury). Damage is an objective concept which is subject to a finding. Damage is the potential source of the loss.

E.g.: personal injury (paralysis)

The « **fait générateur »**: For a person to be liable, the event that gives rise to liability must be the direct cause of the damage. It may be *at fault* or *not* (no-fault liability).

E.g. misuse of an AI technology (faulty), overloading of emergencies requiring prioritisation (therapeutic hazard, non-faulty, Kouchner Law 2002)

For our example ...

A healthcare software company is developing a patient triage tool to assess the level of risk of stroke based on a machine learning algorithm. The AI system was trained on data sold by a provider and included only stroke cases in patients >70 years old, mostly men.

A 35-year-old woman lost consciousness while jogging, was brought to hospital by a friend and the hospital prioritised other patients who arrived after him. The medical diagnosis then identifies a stroke for which earlier management would have limited the long-term consequences. The hospital used the AI system to prioritise patients in emergency room.

- Is there a fault?
- Who should take accountability? The AI tool? The regulating doctor?
- Is the AI tool accountable? Is it the AI as a person? Its designer(s)? ... and how to identify the responsibilities of each?
- How can the patient make a claim against the AI tool if he/she does not know that it was used?

Our research questions

Questions

- Does the current legal framework allow these questions to be answered?
- What are the new legal issues raised by AI?
- What are the current avenues of reflection?
- What AI research challenges do these questions raise?

Evolution and innovation in law

What exists in law?

- product liability
  - This covers certain situations related to a device defect with an AI
  - Nevertheless, the notion of defect defined in the context of products may need to be revised (does an AI learned on ill-fitting examples have a defect?).
- civil liability
  - Allows to deal with the case of malice (e.g. modification of an AI)
- the bioethics laws (+ health code) revised in 2021
- the GDPR
  - Addresses harms specifically related to the use of personal data (privacy) and, in particular, in the context of AI learning
  - Contradictions between consent requirement, dataset requirement, requirement to use state-of-the-art methods

Evolution and innovation in law: specificity of AI

Multiple actors

- An AI system involves a multitude of actors → usually no single manager

An AI is a system that will evolve

- A challenge in relation to the function of medical devices based on certification (at a time t)

An AI is a complex system

- Difficulty in proving fault, identifying the cause and tracing responsibility

Giving an AI a legal personality?

=> The problems are not in the identification of new losses or new events, but rather in the modalities defining the distribution of accountabilities in the case of damage and in the qualification as a harmful event (*direct link*).

Questions addressed in the DRIAS project

1) liability for the use of personal data

- framework of the new bioethics laws (2021)
- regulatory difficulties
- anonymisation and pseudonymisation of data

2) liability and automated decision support

Regulatory framework for the use of AI in healthcare

Modification of the Health code (Art. L. 4001-3)

- I. Le professionnel de santé qui décide d'utiliser, pour un acte de prévention, de diagnostic ou de soin, un dispositif médical comportant un traitement de données algorithmique dont l'apprentissage a été réalisé à partir de données massives s'assure que la personne concernée en a été informée et qu'elle est, le cas échéant, avertie de l'interprétation qui en résulte.
  - Dans notre exemple : utilisation d'une IA dans le cadre du triage (acte diagnostic ?) ?
  - Pas de mention d'opposition, juste une information : « La personne concernée a le droit de ne pas faire l'objet d'une décision fondée exclusivement sur un traitement automatisé, y compris le profilage, produisant des effets juridiques la concernant ou l'affectant de manière significative de façon similaire » (GDPR, Art 22.1)
    - L'interprétation actuelle est qu'une intervention humaine « mineure » n'est pas suffisante.
  - Capacité d'explication par le médecin ?
- III. Les concepteurs d'un traitement algorithmique mentionné au I s'assurent de l'explicabilité de son fonctionnement pour les utilisateurs.
  - Explicabilité du fonctionnement vs explication d'une décision !

Questions for *AI research*

The challenge of ensuring the development of efficient/accurate AI methods while preserving personal rights

- tension between privacy and the need for doctors to be state of the art

User data protection strategies of the HAS (P.-A. Jachiet)

- Limited dissemination / statistical aggregation / remote access / pseudonymisation
- Need to adapt the strategy to the use of data

The regulatory framework is under construction: between legislation and technique (CNIL/LNE/AFFNOR)

Three specific examples :

- Existing contradictory injunction in these studies due to the RGPD
- What anonymisation/pseudonymisation methods can provide practical solutions to these health data issues
- Do models learned from personal data contain personal information?

Questions addressed in the DRIAS project

1) liability for the use of personal data

2) liability and automated decision support

- Context / Example: the European framework
- Product defects / certification: what regime for AIs
- The issue of interpretability of AIs for accountability

What the European Commission proposes (1/2)

A variety of recent proposals/legal texts

- DRIAS worked on the EP resolution of 20 Oct 2020 on "a civil liability regime for artificial intelligence"
  - https://www.europarl.europa.eu/doceo/document/TA-9-2020-0276_FR.html
- Recent new documents
  - **Proposed AI act November 2022**
    - https://www.artificial-intelligence-act.com/
  - Adaptation of existing law, 28/9/2022
    - Revision of product liability (with specificities fors AIS)
    - adaptation of the rules on extra-contractual civil liability to the field of artificial intelligence
- Developments still to be expected in 2023

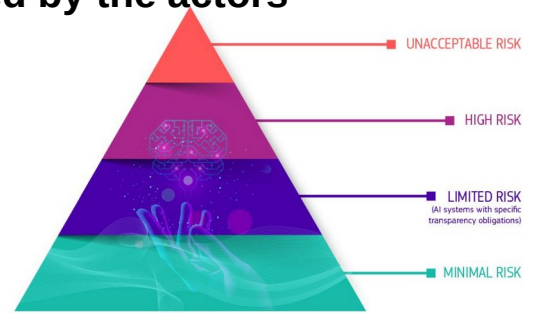
Definition of an AI system in November 2022

"'Artificial intelligence system' (AI system) means a system that is designed to operate with elements of autonomy and that, based on machine and/or human-provided data and inputs, infers how to achieve a given set of objectives using machine learning and/or logic- and knowledge based approaches, and produces system-generated outputs such as content (generative AI systems), predictions, recommendations or decisions, influencing the environments with which the AI system interacts"

What the European Commission proposes (2/2)

Risk-based liability

- **Accountability: according to the proportionality of the risks assumed by the actors**
- **Raises the issue of traceability and interpretability of decisions**



Separation of AI systems into two classes

- High risk systems
  - **AIs identified and listed by a commission**
  - **Strict liability regime**
  - AI systems intended to be used as safety components of products with third party ex-ante conformity assessment; or Other autonomous AI systems raising fundamental rights issues (list in Annex III: 8 AI systems)
  - 2020 text "Health is an area where a high risk ranking can be expected".
- Other systems
  - **Fault-based liability regime : burden of proof and of causality** to the claimant
  - Less user protective than the proposal in 2020

A private insurance system

Raises questions about the principles of assessing contributions (risk assessment)

The issue of interpretability of AIs for accountability

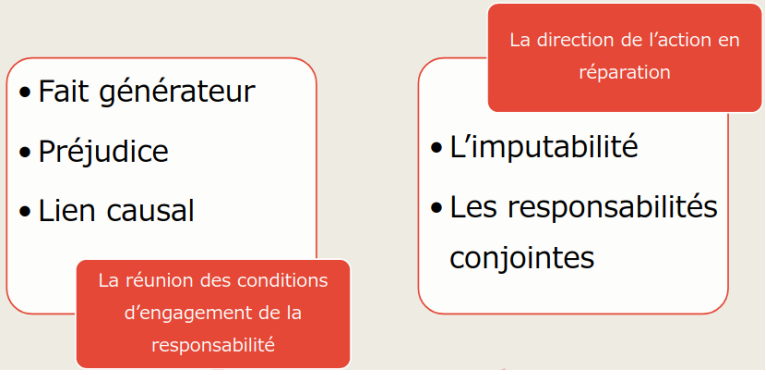
Explicability/interpretability of AI tools

Many AI researches programs aim at bringing explicability to

- of the models or
- decisions made on the basis of these models

The distinction between classes of methods is also reflected in the legal issues raised by AI systems

The issue of interpretability of AIs for accountability



H. Muscat and C. Paillard

Explicability of models in case of injury : A crucial issue for lawyers in the redress of damages related to the use of AI

- Question of the **causal link**: need for elements that relativise or remove the causality between the event and the damage
- Question of **imputability**: How to apportion responsibility in case of injury?

=> the explicability of the models must allow the reasons for a decision to be explained post-hoc

The issue of interpretability of AIs for accountability

Explicability of decisions to users

A strong expectation from health service users,

- Principle 7 France Asso Santé: "Tout système algorithmique utilisé dans le cadre d'une aide à la décision médicale est explicable à un usager sous une forme intelligible".

RGPD obligations on automated decisions: tension on the role of explicability between dis-accountability and informed information

Conditions of (moral) accountability: control and epistemics [Habli, 2020].

- Does the use of AI (and the inexplicability of some decisions) mitigate the accountability of decision-makers?
- Does this shift the accountability to the prescriber of the tool?
- What training obligations can be imposed on AI users? What are the consequences for the attribution of their liabilities?

Take home message

Artificial intelligence act and small adaptation of liability

- Supplement the RGPD (for data risk management)
- Legal framework in Europe
  - Mainly for high-risk AIS
  - AIS for Health are not considered as high risk systems
  - Does not constraint so much the AIS designers
- To be further investigated …

Close connections between computer science and

- Personal data risk management
- Explicability and burden of the proof

Conclusion

Epistemological perspective

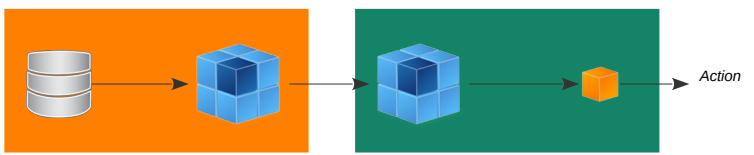
- What (re)evolution does AI bring?
- Is there a real need to treat AI differently from any technical object?
- How can we distinguish between myths, fantasies and reality?

AI development: which ethical principles?

- Ethical principles of AI (in health)
- A critical look at principalism and epidemiology
- A framwork for pro-ethical design of artificial intelligence systems

AI development: what regulatory framework (in Europe)?

- Motivation of the European regulatory framework
- Use of the data
- Responsibility framework



An example ...

A healthcare software company is developing a patient triage tool to assess the level of risk of stroke based on a machine learning algorithm. The AI system was trained on data sold by a provider and included only stroke cases in patients >70 years old, mostly men.

A 35-year-old woman lost consciousness while jogging, was brought to hospital by a friend and the hospital prioritised other patients who arrived after him. The medical diagnosis then identifies a stroke for which earlier management would have limited the long-term consequences. The hospital used the AI system to prioritise patients in emergency.

- What are the ethical issues involved in this situation?
- What questions for the law?
- What are the specificities of using an AI system?

Thank you

Principled artificial intelligence

Privacy:

Privacy
Control over Use of Data
Consent
Privacy by Design
Recommendation for Data Protection Laws
Ability to Restrict Processing
Right to Rectification
Right to Erasure

Accountability:

Accountability
Recommendation for New Regulations
Impact Assessment
Evaluation and Auditing Requirement
Verifiability and Replicability
Liability and Legal Responsibility
Ability to Appeal
Environmental Responsibility
Creation of a Monitoring Body
Remedy for Automated Decision

Safety and Security:

Security
Safety and Reliability
Predictability
Security by Design

Transparency and Explainability:

Explainability
Transparency
Open Source Data and Algorithms
Notification when Interacting with an AI
Notification when AI Makes a Decision about an Individual
Regular Reporting Requirement
Right to Information
Open Procurement (for Government)

Fairness and Non-discrimination:

Non-discrimination and the Prevention of Bias
Fairness
Inclusiveness in Design
Inclusiveness in Impact
Representative and High Quality Data
Equality

Human Control of Technology:

Human Control of Technology
Human Review of Automated Decision
Ability to Opt out of Automated Decision

Professional Responsibility:

Multistakeholder Collaboration
Responsible Design
Consideration of Long Term Effects
Accuracy
Scientific Integrity

Promotion of Human Values:

Leveraged to Benefit Society
Human Values and Human Flourishing
Access to Technology

Example 1: Regulatory difficulties (REPERES)

Epidemiological study conducted on several hospitals (multi-centric) requiring access to health data

Regulatory framework for access and processing authorisations

- CNIL authorisations by each university hospital of its warehouse
- CNIL authorisation for the project including the multicentre study between the university hospitals + matching with the historical SNDS
- Two distinct legal regimes
  - for the creation of health warehouses
  - for the re-use of data from the warehouse for research, studies or evaluations based on data held in the warehouse
- Extract from the CNIL deliberation: *La Commission relève par ailleurs que les patients sont informés de l'éventualité de l'utilisation des données à des fins de recherche dans le domaine de la santé. Elle rappelle que cette information générale ne peut se substituer à l'information individuelle prévue par les dispositions de l'article 58 de la loi Informatique et Libertés et qui devra être réalisée pour chaque traitement de données réalisé à partir des données de l'entrepôt, en application des dispositions du chapitre IX de la loi informatique et libertés*

source: CNIL Deliberation 2018-295 of 19 July 2018 https://www.legifrance.gouv.fr/cnil/id/CNILTEXT000037509951/

An example of a regulatory difficulty (REPERES)

Practical implications for this multicentre study

- To have an individual means of informing the patient about the study
- Allocation of a specific number from the NIR in each university hospital
  - Risk of patients known to several institutions => deduplication to be carried out
  - Could be avoided, but CNIL recommends an "RGPD compatible" data model for the warehouse making it more difficult to access personally identifiable information.

Contradictory injunction in these studies due to the RGPD

- The data must be pseudonymised (impossible to trace back to individuals)
- But the controller must be able to escalate to the patient concerned to exercise his or her right of withdrawal

« si les finalités pour lesquelles des données à caractère personnel sont traitées n'imposent pas ou n'imposent plus au responsable du traitement d'identifier une personne concernée, celui-ci n'est pas tenu de conserver, d'obtenir ou de traiter des informations supplémentaires pour identifier la personne concernée à la seule fin de respecter le présent règlement. Lorsque, dans les cas visés au paragraphe 1 du présent article, le responsable du traitement est à même de démontrer qu'il n'est pas en mesure d'identifier la personne concernée, il en informe la personne concernée, si possible. En pareils cas, les articles 15 à 20 ne sont pas applicables, sauf lorsque la personne concernée fournit, aux fins d'exercer les droits que lui confèrent ces articles, des informations complémentaires qui permettent de l'identifier. » (RGPD, Art 11)

Example 2: Access to data with privacy guarantees

What anonymisation/pseudonymisation methods can provide practical solutions to these health data issues (S. Gambs)

First issue of privacy qualification: a very active IT issue

- k-anonymity
- Differential Privacy (DP) → being adopted by industry

More and more open source implementations are available to the machine learning world:

- Access to data via perturbed query results (Laplace, Gaussian mechanisms, etc.)
- Provision of summary data

The guarantee of non-re-identification reduced to limited situations (easily overcome in practice):

- Which model for certification?
- What is the liability in case of re-identification?

Example 3: Access to the learned model (on non-anonymised data) with privacy guarantees

Do models learned from personal data contain personal information? (CNIL, F. Vallet)

Startup XXX: request for advice from a company marketing a tool implementing automatic learning methods for the coding of medical acts (PMSI coding)

- OK to deploy to an X centre if the learning and model remain on the centre's premises.
- Can the startup transfer the model learned in centre X to centre Y and adapt it there?
- If possible, what measures and practices should be recommended to minimise the risks?

Development of learning methods (on non-anonymised data) with guarantees : DP-SGD, DP-GAN, ... DP-*

Adaptation de la responsabilité

Type de mécanisme

- Responsabilité objective pour défaut

Source de responsabilité

- Art. 6 Caractère défectueux du produit

Charge de la preuve

- Art. 9.1 et 9.2 Charge à la victime de produire tous les éléments au soutien de la plausibilité de l'action et de prouver le défaut du produit et le dommage subi (lien de causalité présumé)

Type de mécanisme

- Responsabilité pour faute

Source de responsabilité

- Art. 4.1.a Manquement à un devoir de vigilance

Charge de la preuve

- Art. 4.1, 4.2 et 4.3 Charge à la victime de de prouver la faute et du lien de causalité avec le dommage (lien de causalité présumé sous conditions, notamment pour les SIA à haut risque, et réfragable)

https://www.actu-juridique.fr/ntic-medias-presse/les-nuances-des-nouveaux-regimes-de-responsabilite-des-ia-en-europe/

Defective Products / certification: what risk management for AIs?

Look at the possible damaging facts

- Injury caused by a surgical robot
- Disease caused by AI-assisted diagnostic error
- Disclosure of sensitive data
- Failure to inform the patient
  - Information on the use of AI
- Respecting the wish not to use an AI
- Equity of AI based on machine learning
- Unequal access to technology
- Decision bias on a model learned from unbalanced data

Complex burden of the proof (EP Directive adapting liability of defective products)

- Burden of the proof belongs to the claimant in general
- Defectiveness and causal link presumption in case of « technical or scientific complexity »

Very different time frames for decisions :

- no distinction for AI
- How important is it to qualify responsibilities?

Product defects / certification: what regime for AIs?

French policy guided by the Innovation Council

- Grand Challenge "Securing, making reliable and certifying systems that use Artificial Intelligence
- White Paper (Oct 2022): https://www.confiance.ai/

AI certification (G. Avrin, ex-LNE)

- AI compliance assessment is a major issue in the practical implementation of a liability regime
- A company developing a service seeks to assess the legal risks of its products
- How to 'evaluate' AI: many questions still open

Certification issues for evolving systems (by learning/customisation)

- FDA experience with evolving DMs
  - Classification of 4 families of modifications + methodology for evaluating Software as a Medical Device

Insurance approach

Proposal in 2020

"*Uncertainty about risks should not lead to prohibitively high insurance premiums, which would act as a brake on research and innovation*".

A private insurance system

- "the Commission should work closely with the insurance industry to consider how to use data and innovative models
  - It is a question of facilitating access to data by insurers: what are the risks with regard to personal data? How to ensure the separation between personal data (usable by the insurer) and the need to quantify risks?
- Should (wish) to be complemented by a special compensation fund (for specific needs)
  - What is the relationship with ONIAM (Office National d'Indemnisation des Accidents Médicaux) in the case of health damage?

In the Artificial Intelligence Act, 2022

- No compulsory insurance mentioned in AI legislation
- The appropriateness of this issue is subject to a future report (Directive on liability in the field of AI)