FRAMING TRUST IN MEDICAL AI

JOSE M. JUAREZ

AIKE research group, University of Murcia

UNIVERSIDAD DE MURCIA

Contact:

Email: jmjuarez@um.es

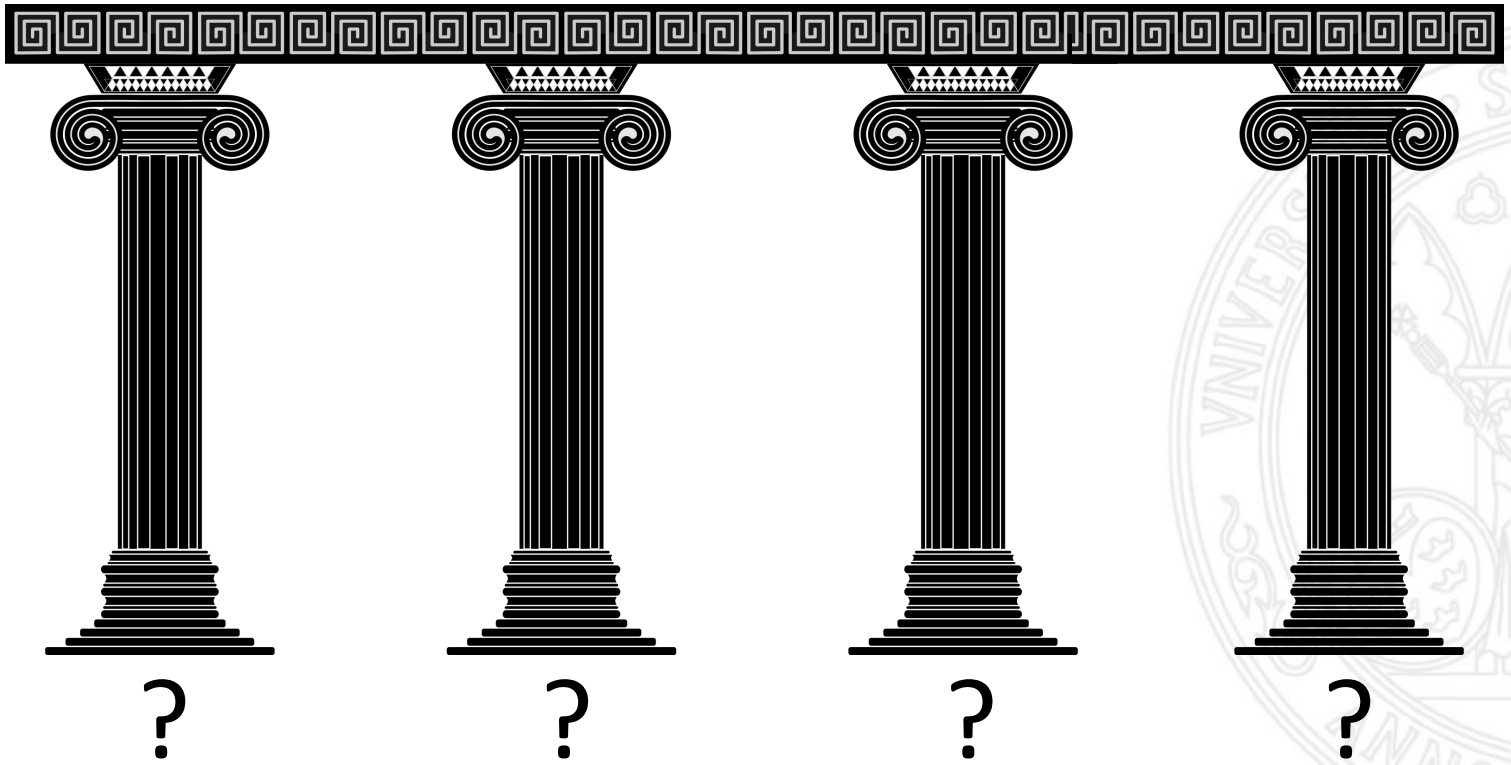
Twitter: @juarezprofesor

LinkedIn: jose-m-juarez

EURAI ACAI - TAILOR SUMMER SCHOOL 2022

JUN 14TH 2022, BARCELONA

TRUST IN MEDICAL AI



Unknown technology

Ethical or legal

Trustworthy

FRAMING TRUST IN MEDICAL AI

AI-TECHNOLOGY IN HEALTHCARE

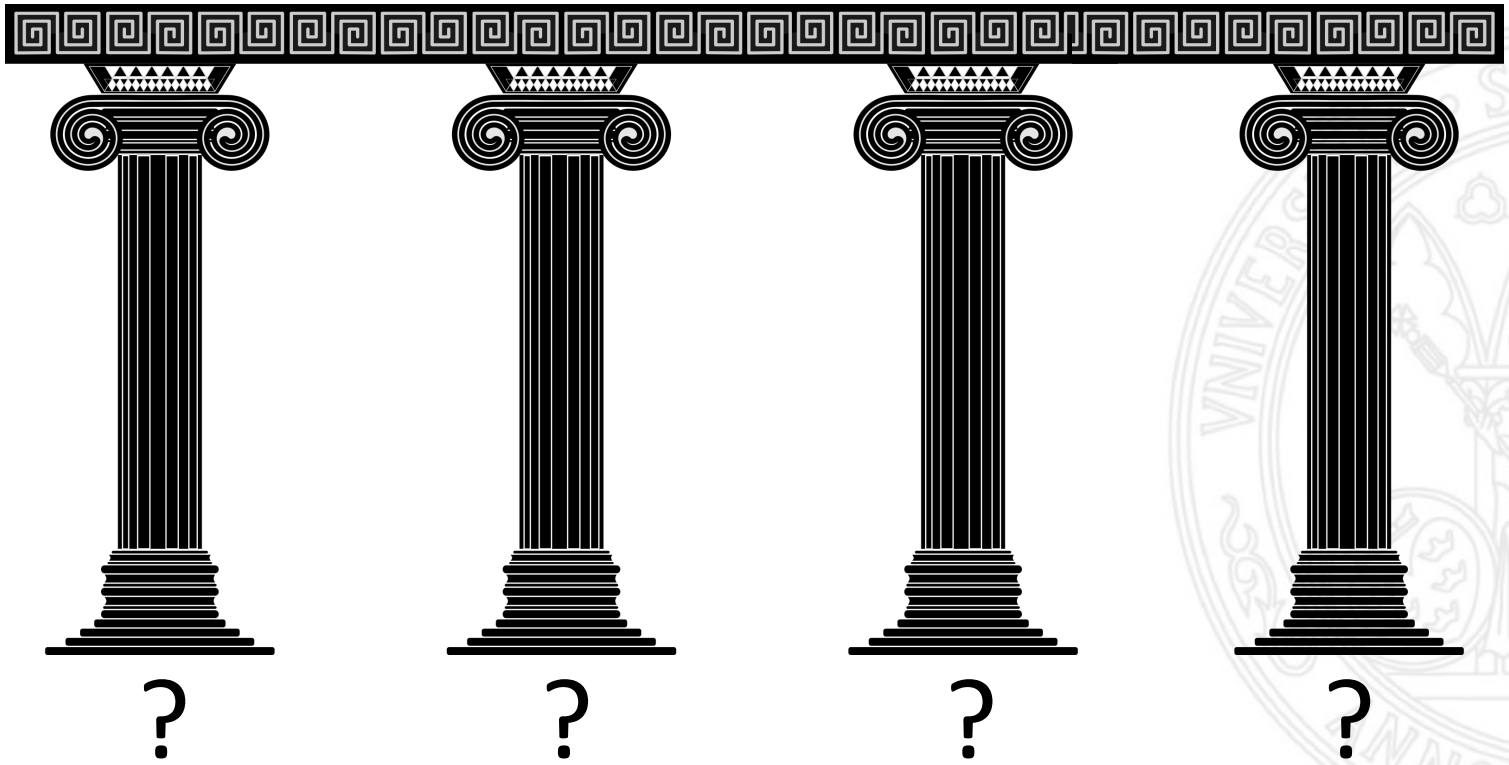
Unknown technology

Ethical or legal

Trustworthy

Picture from N Melchor UnSplash https://unsplash.com/es/fotos/43LwvC-eQPM?utm_source=unsplash&utm_medium=referral&utm_content=creditShareLink

TRUST IN MEDICAL AI



- **OUTLINE:**

1. HUMANS, HEALTHCARE AND ETHICAL AI
2. AI UNDERSTANDABLE IN HEALTHCARE ENVIRONMENTS
3. REGULATIONS FOR MEDICAL INDUSTRY
4. FINAL

UNIVERSIDAD DE MURCIA

- **OUTLINE:**

1. HUMANS, HEALTHCARE, ETHICS AND AI
2. AI UNDERSTANDABLE IN HEALTHCARE ENVIRONMENTS
3. REGULATIONS FOR MEDICAL INDUSTRY
4. FINAL

UNIVERSIDAD DE MURCIA

1. HUMANS, HEALTHCARE ETHICS AND AI

- WHY IS HEALTHCARE DOMAIN DIFFERENT?
- WHY IS AI DIFFERENT FROM OTHER TOOLS?
- WHAT IS MEDICAL AI?
- AI ETHICAL CONCERNS & INITIATIVES

UNIVERSIDAD DE MURCIA

1. HUMANS, HEALTHCARE ETHICS AND AI

– WHY IS HEALTHCARE DOMAIN DIFFERENT?

- Discipline dependence
- Def. Infection CDC
- René Laennec

Picture by 8photo from https://www.freepik.es/fotos/doctor-estetoscopio

UNIVERSIDAD DE MURCIA

1. HUMANS, HEALTHCARE ETHICS AND AI

- WHY AI IS DIFFERENT FROM A STETHOSCOPE?
- Aim
- Multidisciplinary

Picture by 8photo from https://www.freepik.es/fotos/doctor-estetoscopio

UNIVERSIDAD DE MURCIA

1. HUMANS, HEALTHCARE, ETHICS AND AI

– WHAT IS MEDICAL AI?

clinical dataset + ML ?= MEDICAL AI

Solve medical problem with AI+ bioethical approval

Terminology & clinical semantics

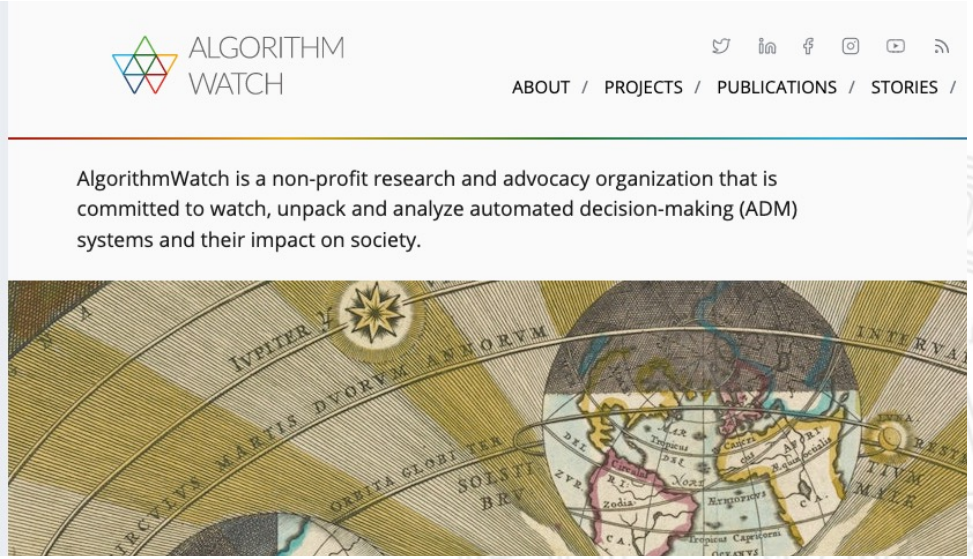
Clinical data compilation

AI/ML fundamentals + tailored techniques

Medical validation

1. HUMANS, HEALTHCARE, ETHICS AND AI

- ETHICS AND AI
  - ETHICS DEF
  - HIGH LEVEL PPLES.



- *IBM*'s principles of trust and transparency
- *Google*'s principles on AI
- *World Economic Forum*'s principles for ethical AI
- *AI4PEOPLE* principles and recommendations
- *IEEE* general principles
- …

1. HUMANS, HEALTHCARE AND ETHICAL AI

- ETHICS GUIDELINES FOR TRUSTWORTHY AI

- 2019
- High-Level Expert Group of AI (indep. group)
- Framework for trustworthy AI
  - Foundations: 4 Ethical principles (address tensions)
  - Realisation of Trustworthy AI: 7 key requirements (evaluate)
  - Assessment of Trustworthy AI: assessment List (specific AI apps)

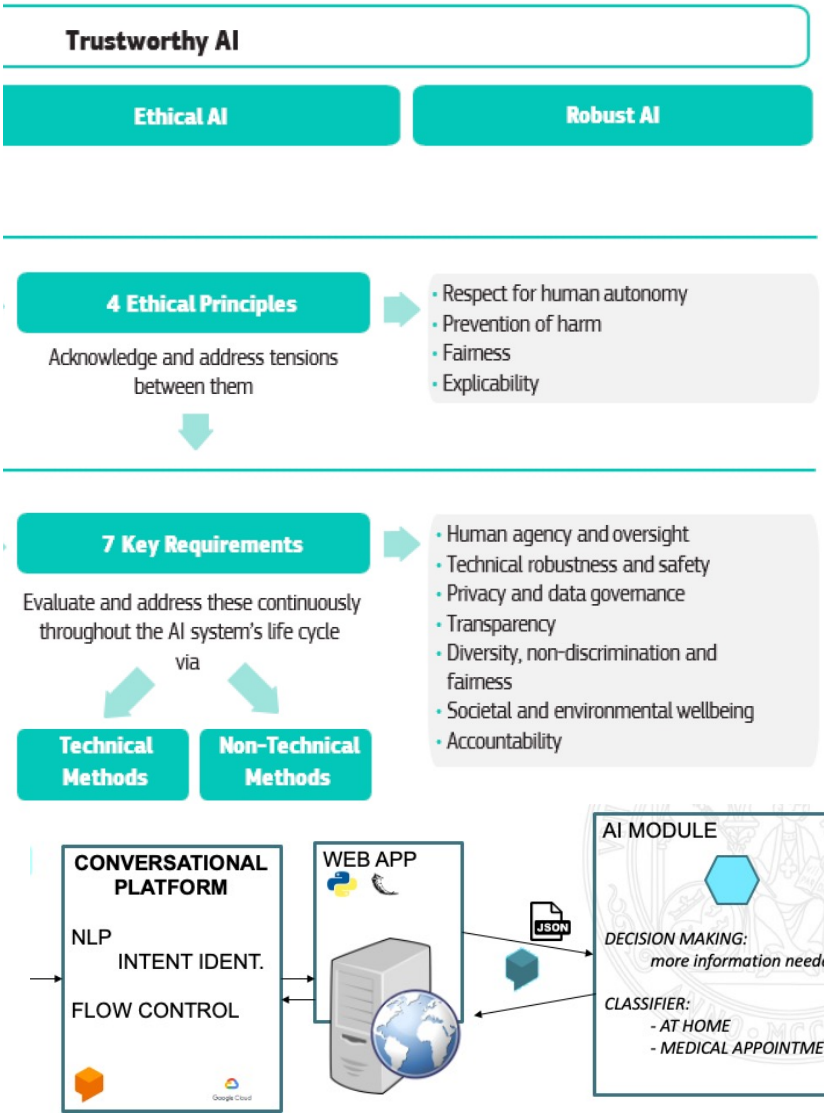
Picture from: 2019-11-18 High-Level Expert Group on Artificial Intelligence. Ethics guidelines for trustworthy AI. https://op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1

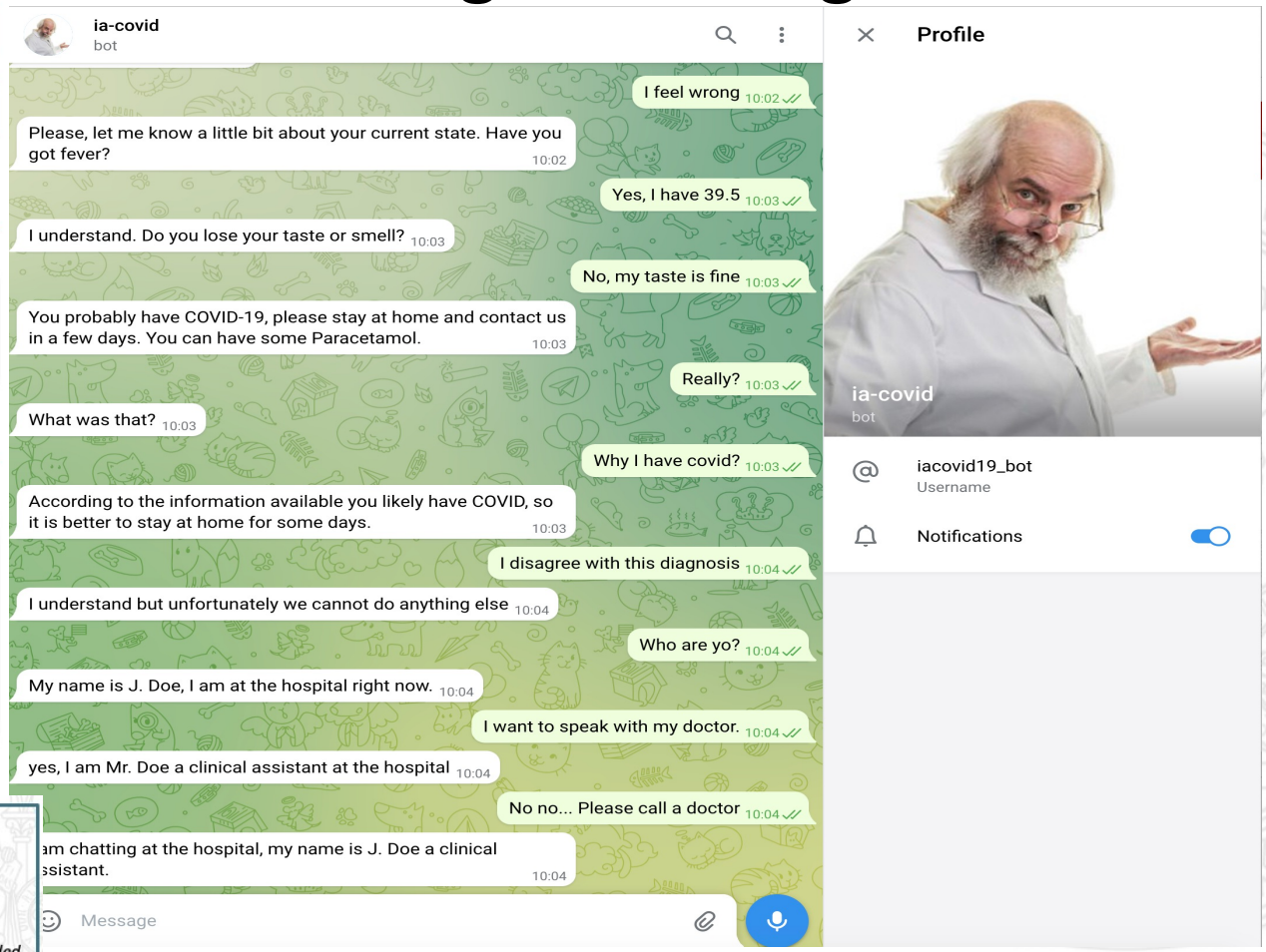
- **ETHICS GUIDELINES FOR TRUSTWORTHY AI**
  - 4 Ethical Principles in the context of AI systems
    i. Respect to human autonomy
    ii. Prevention of harm
    iii. Fairness
    iv. Explicability

- **ETHICS GUIDELINES FOR TRUSTWORTHY AI**
  - Developing Trustworthy AI: 7 requirements
    i. Human actions and supervision
    ii. Robustness and safety
    iii. Privacy and data governance
    iv. Transparency
    v. Diversity, non-discrimination and fairness
    vi. Societal and environmental wellbeing
    vii. Accountability

1. HUMANS, HEALTHCARE AND ETHICAL AI

– USE CASE 1: Conversational agent during COVID-19





1. HUMANS, HEALTHCARE AND ETHICAL AI

– USE CASE 2: Pneumonia detection from X-rays

Trustworthy AI

Ethical AI | **Robust AI**

4 Ethical Principles

Acknowledge and address tensions between them

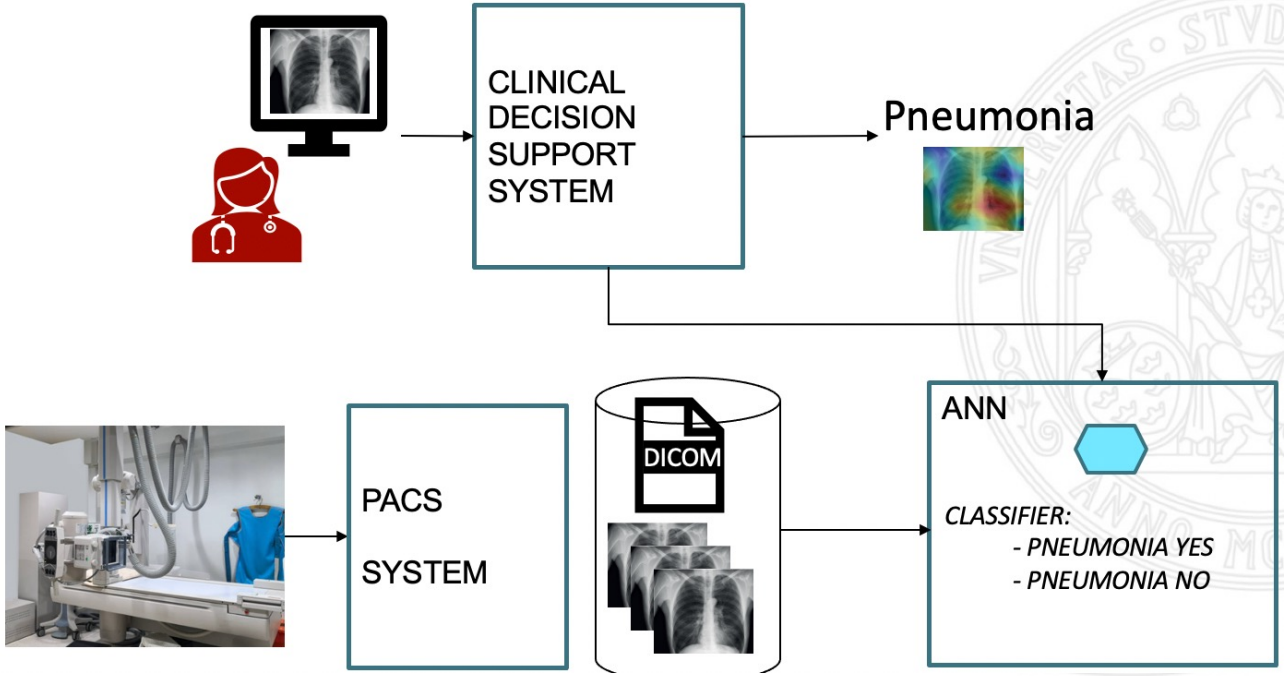
- Respect for human autonomy
- Prevention of harm
- Fairness
- Explicability

7 Key Requirements

Evaluate and address these continuously throughout the AI system's life cycle via

Technical Methods | **Non-Technical Methods**

- Human agency and oversight
- Technical robustness and safety
- Privacy and data governance
- Transparency
- Diversity, non-discrimination and fairness
- Societal and environmental wellbeing
- Accountability

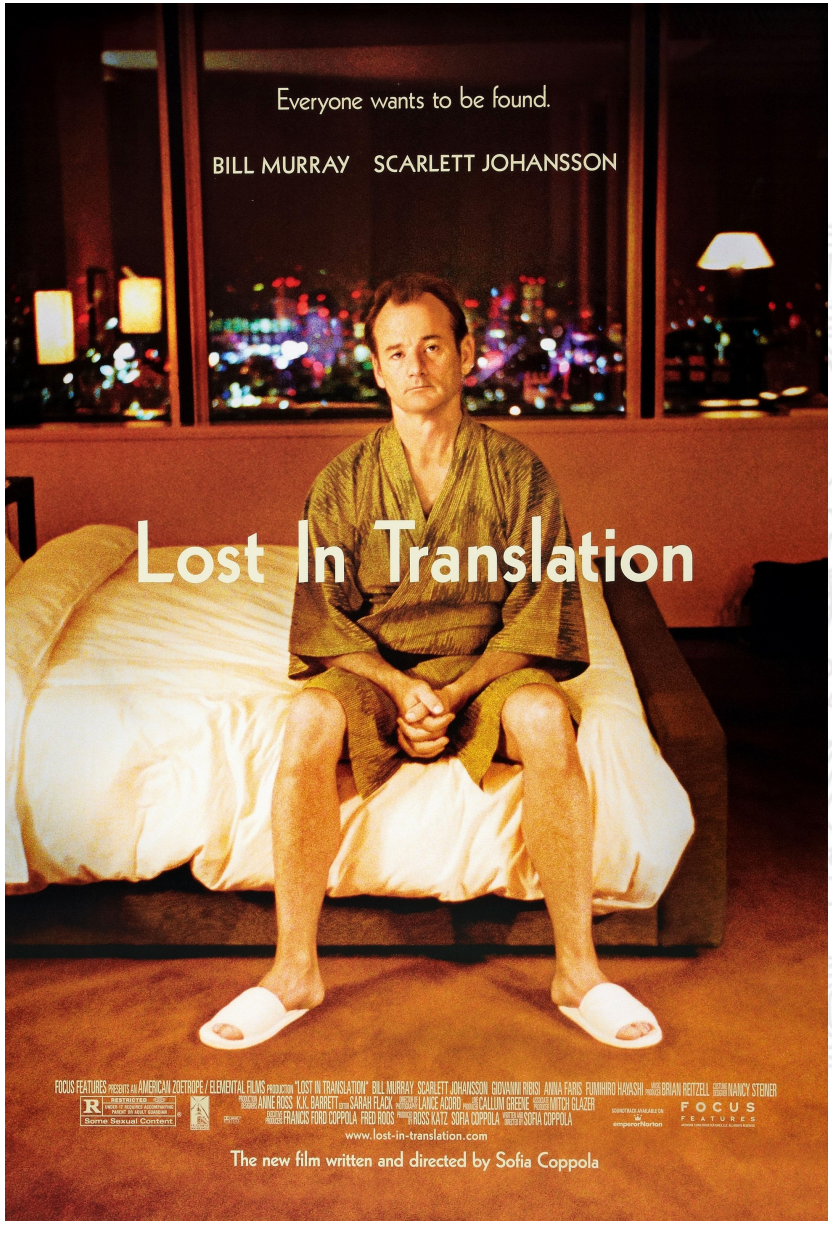


UNIVERSIDAD DE MURCIA

- **OUTLINE:**
  1. HUMANS, HEALTHCARE AND ETHICAL AI
  2. AI UNDERSTANDABLE IN HEALTHCARE
  3. REGULATIONS FOR THE MEDICAL INDUSTRY
  4. FINAL

2. AI UNDERSTANDABLE IN HEALTHCARE

– LOST IN TRANSLATION

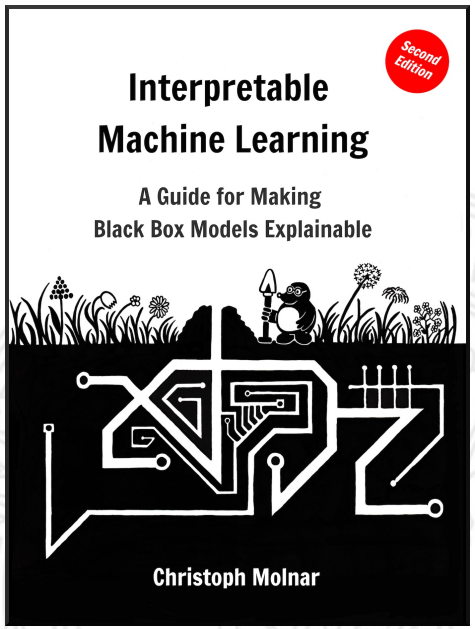


2. AI UNDERSTANDABLE IN HEALTHCARE

- SOME GENERAL CONSENSUS
  - TRANSPARENT AI:
    - Expert Rules
    - Interpretable models
    - Well known in medical field
  - BLACK BOX models:
    - Random forest
    - XGBoost
    - ANN and complex architectures (Deep Learning)
  - Focus on providing EXPLANATIONS

2. AI UNDERSTANDABLE IN HEALTHCARE

- EXPLANATIONS & XAI
  - Explanation methods
    - Model Agnostic vs. Dependant
    - Local vs. Global
    - Surrogate models
    - Example based
    - Visualizations

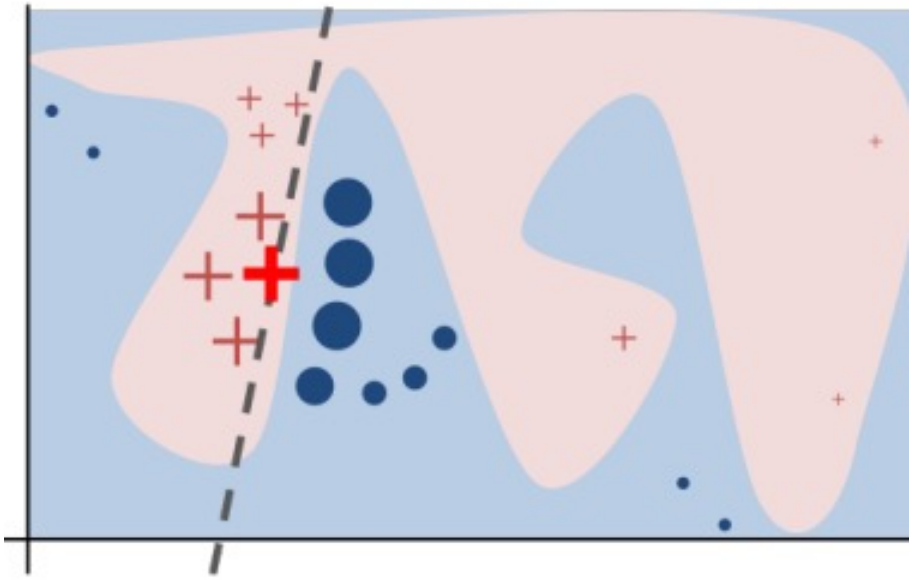


Picture by Ybonne Doinell at https://christophm.github.io/interpretable-ml-book/

2. AI UNDERSTANDABLE IN HEALTHCARE

– SOME POPULAR XAI METHODS

- LIME [Ribeiro2016]
  - Model-agnostic
  - Local
  - Surrogated
  - Intuitions

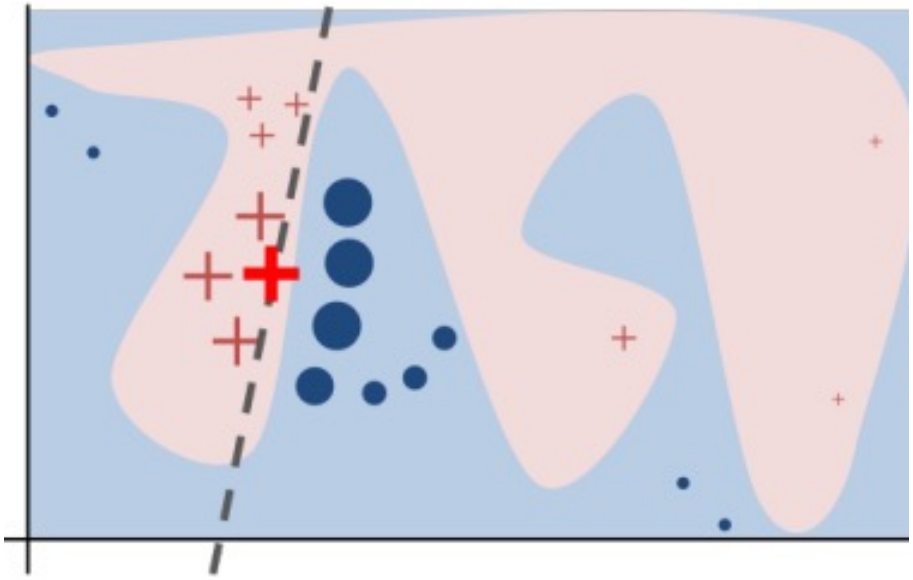


Picture from Figure 3 from [Ribeiro 2016]
Ribeiro, Marco Tulio, Sameer Singh, and Carlos Guestrin. **Why should I trust you?: Explaining the predictions of any classifier.** Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. ACM (2016). arXiv:1602.04938

2. AI UNDERSTANDABLE IN HEALTHCARE

– SOME POPULAR XAI METHODS

- LIME [Ribeiro2016]
  1. Select instance of interest from dataset
  2. Perturb dataset
  3. Weight new points
  4. Build new ML model
  5. Explain prediction



Picture from Figure 3 from [Ribeiro 2016]

Ribeiro, Marco Tulio, Sameer Singh, and Carlos Guestrin. **Why should I trust you?: Explaining the predictions of any classifier.** Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. ACM (2016). arXiv:1602.04938

2. AI UNDERSTANDABLE IN HEALTHCARE

- SOME POPULAR XAI METHODS
  - SHAPLEY VALUES
    - Local/Global model-agnostic
    - ML is a game (coallitions)
      - » Feature = Player
      - » Prediction = Reward
      - » Game = Task of Prediction
    - Shapley Value: How fairly distribute reward btw features

2. AI UNDERSTANDABLE IN HEALTHCARE

– SOME POPULAR XAI METHODS

- SHAPLEY VALUES
  - Given an input, it is the avg marginal contribution of a attribute value across all possible coalitions.

$$\phi_i(val) = \sum_{S \subseteq \{x_1,\ldots,x_p\} \setminus \{x_i\}} \frac{|S|!(p-|S|-1)!}{p!}(val(S \cup \{x_i\}) - val(S))$$

  - E.g. *dataset = $attr_1$, $attr_2$, $attr_3$, input X*

    study $attr_1$=val_a →compute all coalitions

    $attr_1$=val_a + $attr_2$=val_{2a} →◢

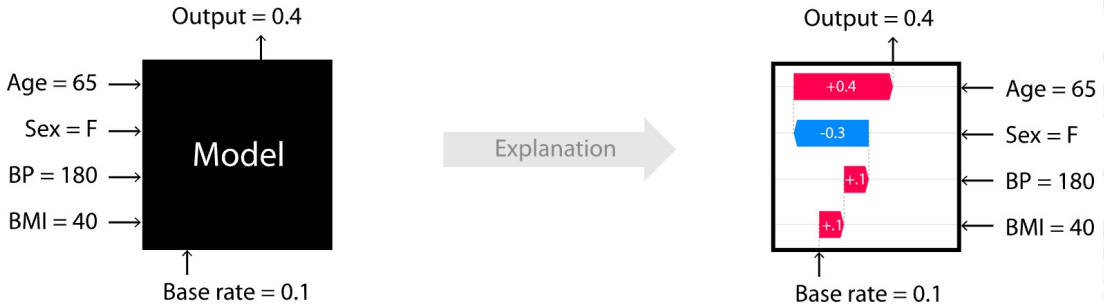
    $attr_1$=val_a + $attr_2$=val_{2b} →▽

    $attr_1$=val_a + $attr_3$=val_{3a} →◢

    $attr_1$=val_a + $attr_3$=val_{3b} →▽

2. AI UNDERSTANDABLE IN HEALTHCARE

- SOME POPULAR XAI METHODS
  - SHAPLEY VALUES
    - Shapley values hard to compute
    - Estimation of Shapley values (sampling, MonteCarlo)
    - SHAP, Kernel SHAP, Tree SHAP[LundbergLee2017]

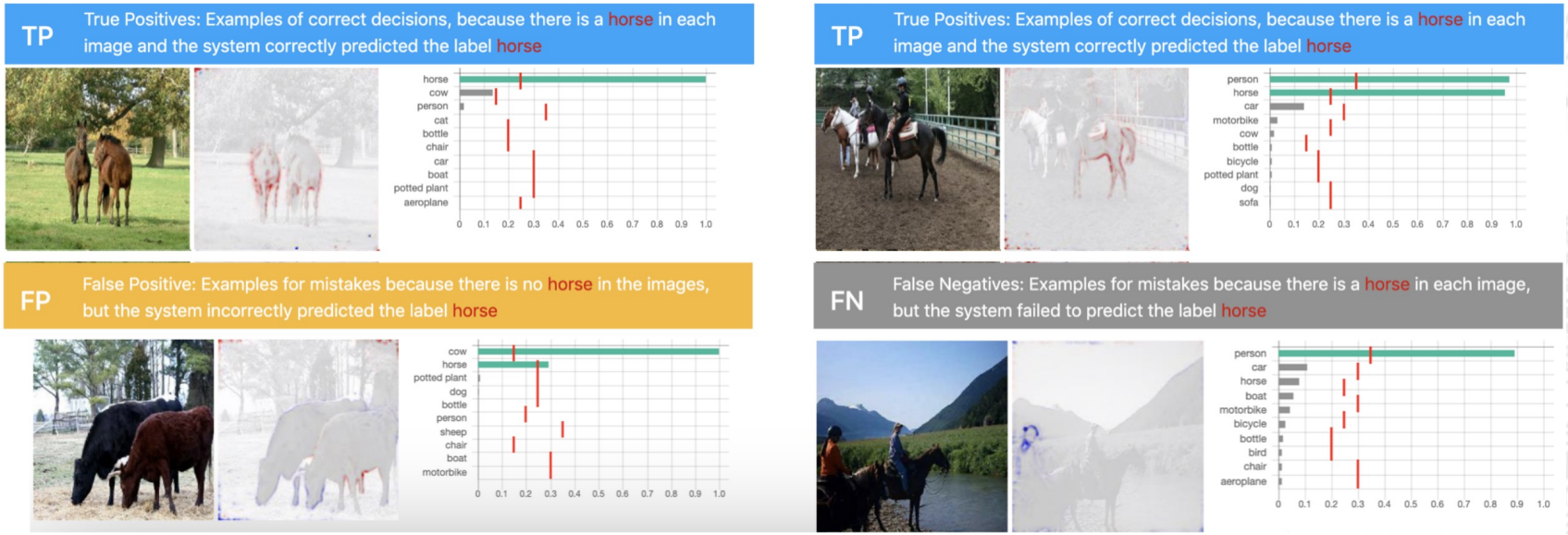


Picture extracted from
[Lundberg-Lee2017]Scott Lundberg and Su-In Lee**. A unified approach to interpreting model predictions.** CoRR, abs/1705.07874, 2017. arXiv:1705.07874.

2. AI UNDERSTANDABLE IN HEALTHCARE

– SOME POPULAR XAI METHODS

- SALIENCY MAPS
  - Vision: marks region people's eyes focus first
  - XAI: highlight pixels relevant for ANN classification

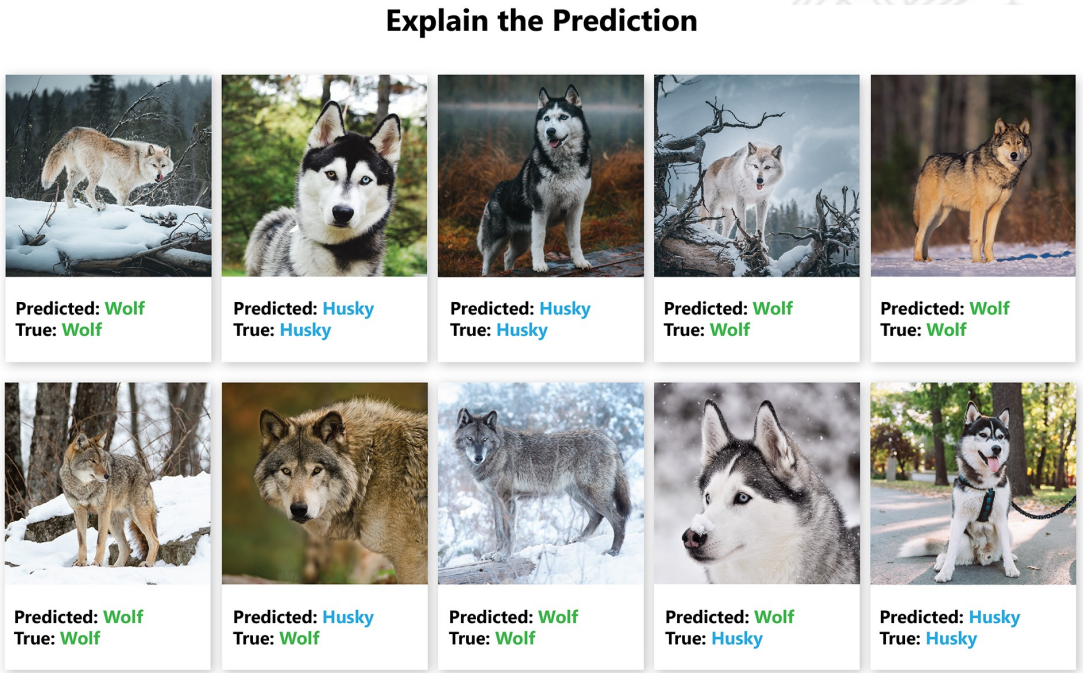


Picture from:
[Alqaraawi2020] Ahmed Alqaraawi, Martin Schuessler, Philipp Weiß, Enrico Costanza, and Nadia Berthouze. **Evaluating saliency map explanations for convolutional neural networks: a user stud**y.Proceedings of the 25th International Conference on Intelligent User Interfaces . 2020. ACM. arXiv:2002.00772

2. AI UNDERSTANDABLE IN HEALTHCARE

– BLACK BOX MEDICINE AND CONCERNS

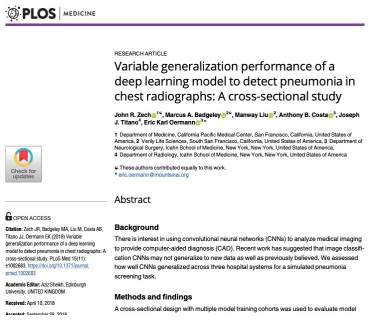
- *Clever Hans* phenomenon
  - Generalizable model or spurious correlation



Oskar Pfungst. Clever Hans (The Horse Of Mr. Von Osten). A Contribution To Experimental Animal and Human Psychology. 1911.
Picture from the Blog of Responsible AI Microsoft by John Roach https://blogs.microsoft.com/ai/azure-responsible-machine-learning/

2. AI UNDERSTANDABLE IN HEALTHCARE

- BLACK BOX MEDICINE AND CONCERNS
  - *Clever Hans in medicine*
    - Researchers from Mount Sinai hospital
    - High-risk patients detection from x-ray imaging
    - Applied outside Mount Sinai
      - » Very low rate predictions
      - » Why?


PLOS MEDICINE

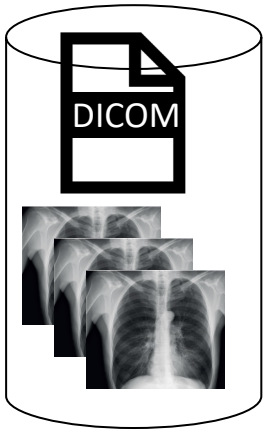
RESEARCH ARTICLE

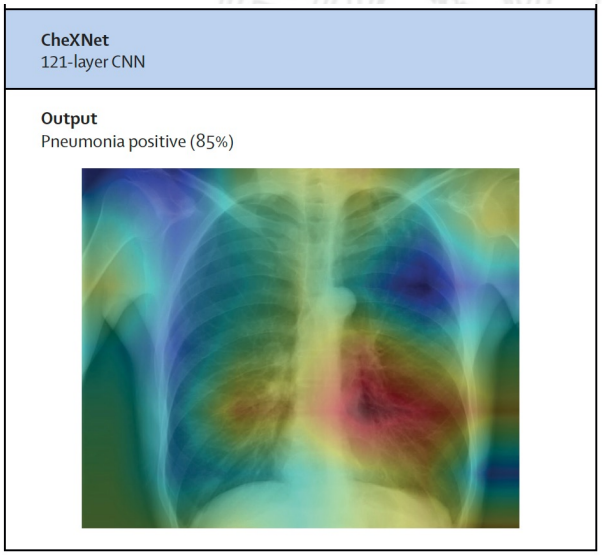
Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study

Abstract

Background

Methods and findings



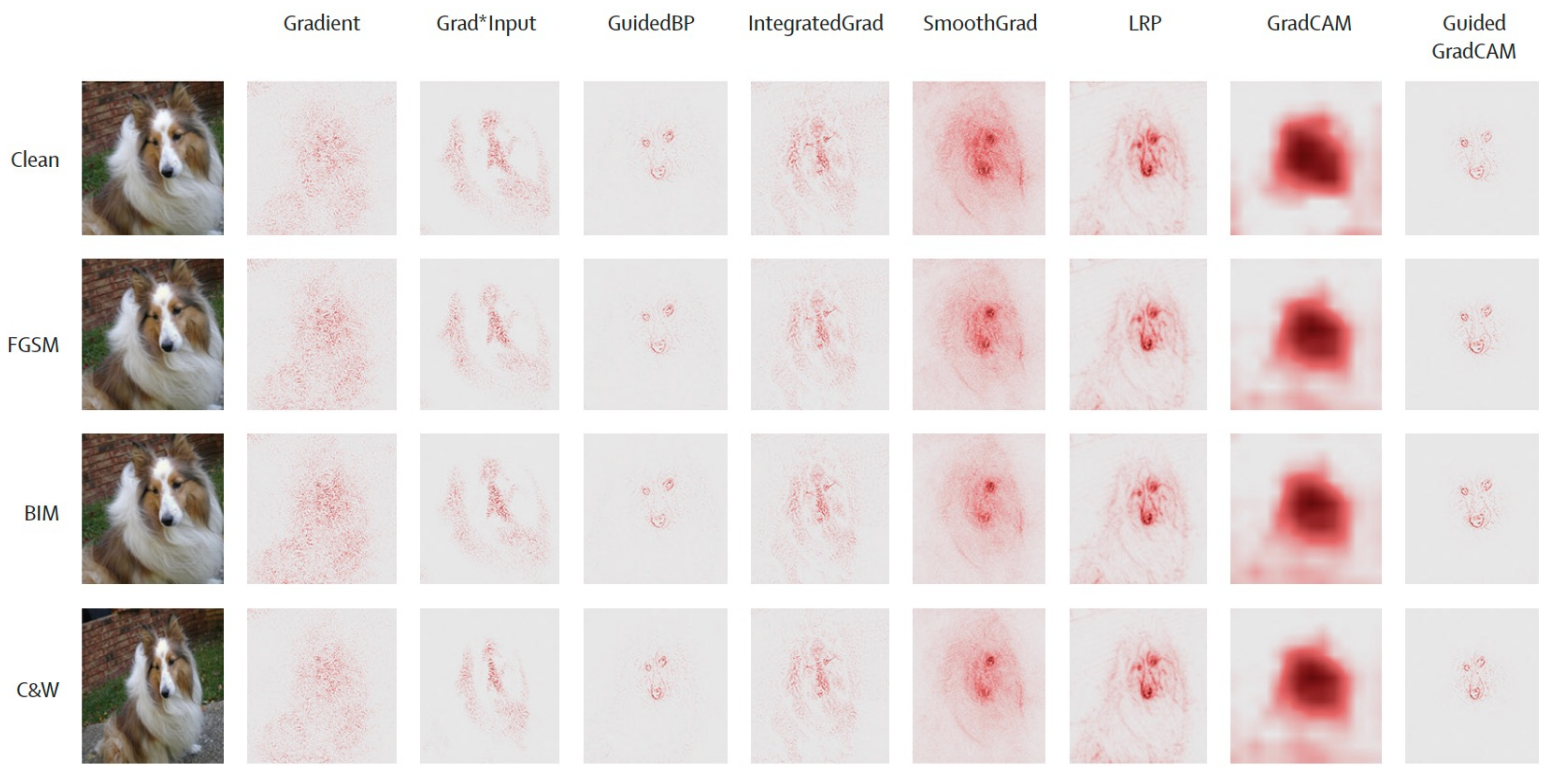


[Amann 2020] Amann, J., Blasimme, A., Vayena, E. et al. **Explainability for artificial intelligence in healthcare: a multidisciplinary perspective**. BMC Med Inform Decis Mak 20, 310 (2020). https://doi.org/10.1186/s12911-020-01332-6

2. AI UNDERSTANDABLE IN HEALTHCARE

– BLACK BOX MEDICINE AND CONCERNS

- SALIENCY MAPS



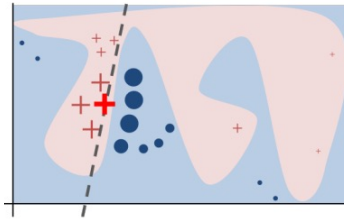
Picture from
[Gu et al 2019] Gu J, Tresp V. Saliency methods for explaining adversarial attacks. arXiv 2019; published online Aug 22. http://arxiv.org/ abs/1908.08413 (preprint).

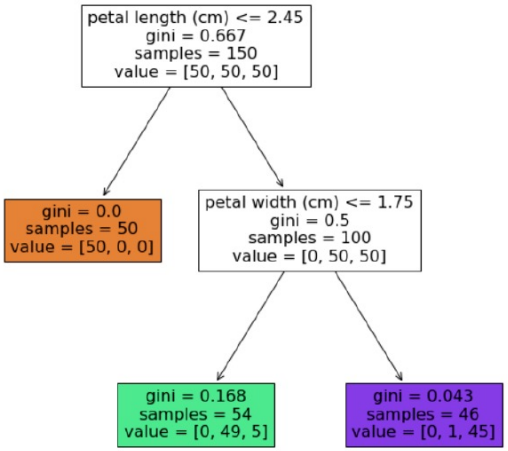
2. AI UNDERSTANDABLE IN HEALTHCARE

– BLACK BOX MEDICINE AND CONCERNS

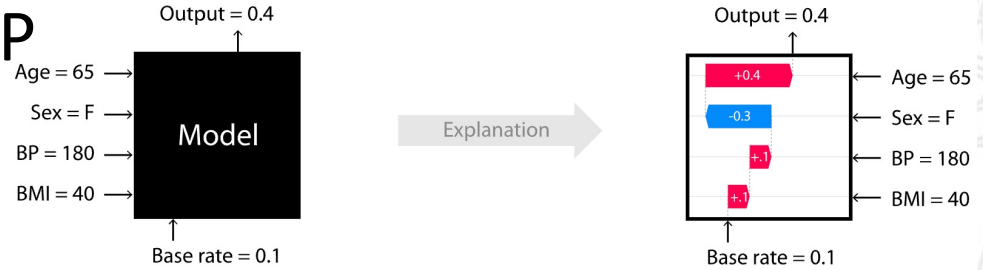
- ARE EXPLANATIONS UNDERSTANDABLE BY DOCS?

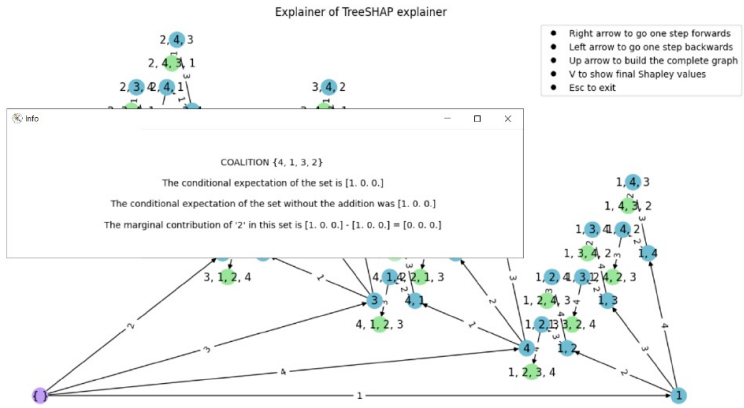
LIME





SHAP





Ribeiro, Marco Tulio, Sameer Singh, and Carlos Guestrin. **Why should I trust you?: Explaining the predictions of any classifier.** Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. ACM (2016)

Scott Lundberg and Su-In Lee**. A unified approach to interpreting model predictions.** CoRR, abs/1705.07874, 2017. arXiv:1705.07874.

Navarro-Nicolas J.A. **Explaining Tree SHAP**. Undergraduate Thesis. University of Murcia 2021.

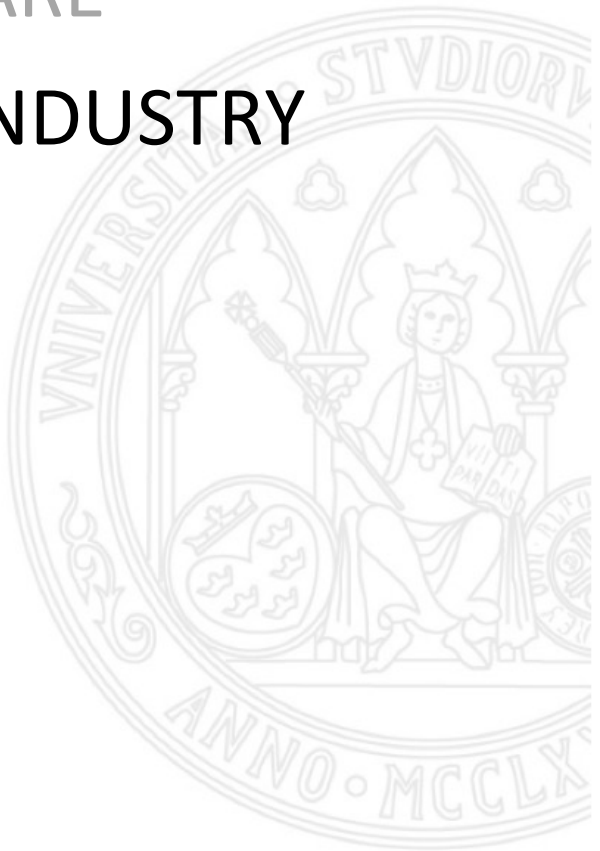
2. AI UNDERSTANDABLE IN HEALTHCARE

– XAI DETRACTORS IN MEDICAL AI

- Final users: not the explanations needed
- AI Developpers “debugging”

UNIVERSIDAD DE MURCIA

- **OUTLINE:**
  1. HUMANS, HEALTHCARE AND ETHICAL AI
  2. AI UNDERSTANDABLE IN HEALTHCARE
  3. REGULATIONS FOR THE MEDICAL INDUSTRY
  4. FINAL

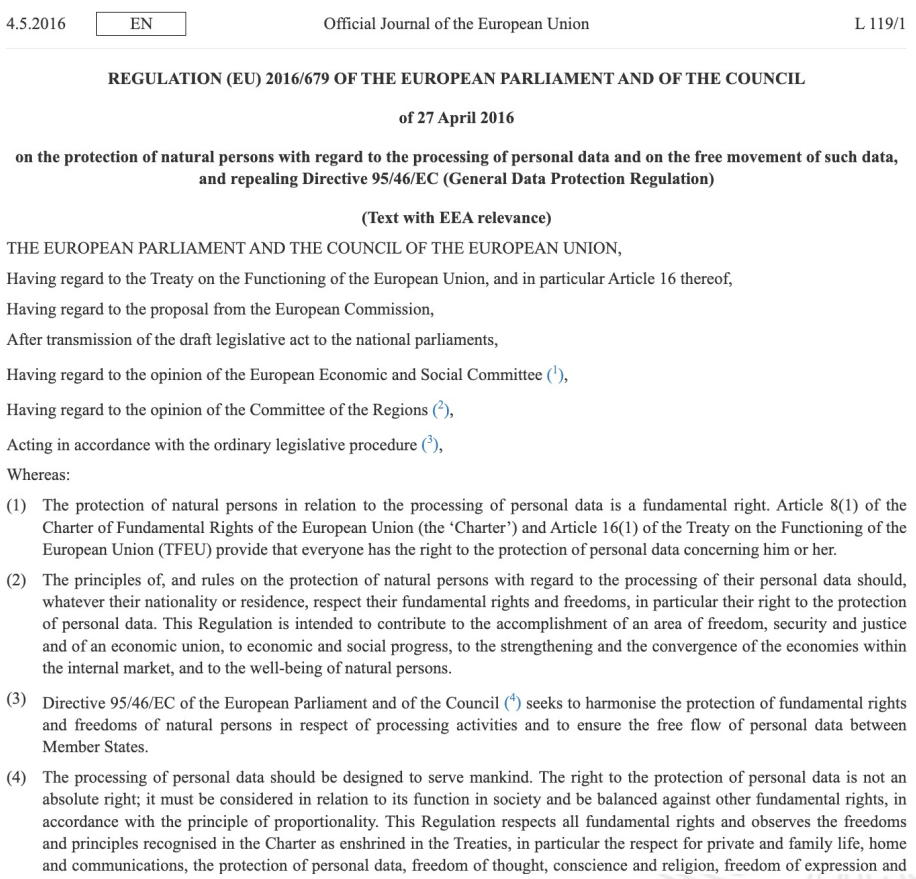


3. REGULATIONS FOR THE MEDICAL INDUSTRY

- GDPR sensible data protection
- EU initiatives on regulating AI
- EU Medical Device Regulation

3. REGULATIONS FOR THE MEDICAL INDUSTRY

- **GDPR and EXPLAINABILITY CONCERN**
  - General Data Protection Regulation (GDPR) –non directive
  - Harmonize data privacy laws across Europe
  - Enforceable / applicable as of May 25th, 2018
  - **Right of explanations** of algorithmic decisions
    - Section 5
    - Controversial, unclear from legal scholars perspective.



4.5.2016 EN Official Journal of the European Union L 119/1

REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL

of 27 April 2016

on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)

(Text with EEA relevance)

THE EUROPEAN PARLIAMENT AND THE COUNCIL OF THE EUROPEAN UNION,

Having regard to the Treaty on the Functioning of the European Union, and in particular Article 16 thereof,

Having regard to the proposal from the European Commission,

After transmission of the draft legislative act to the national parliaments,

Having regard to the opinion of the European Economic and Social Committee ([1]),

Having regard to the opinion of the Committee of the Regions ([2]),

Acting in accordance with the ordinary legislative procedure ([3]),

Whereas:

(1) The protection of natural persons in relation to the processing of personal data is a fundamental right. Article 8(1) of the Charter of Fundamental Rights of the European Union (the 'Charter') and Article 16(1) of the Treaty on the Functioning of the European Union (TFEU) provide that everyone has the right to the protection of personal data concerning him or her.

(2) The principles of, and rules on the protection of natural persons with regard to the processing of their personal data should, whatever their nationality or residence, respect their fundamental rights and freedoms, in particular their right to the protection of personal data. This Regulation is intended to contribute to the accomplishment of an area of freedom, security and justice and of an economic union, to economic and social progress, to the strengthening and the convergence of the economies within the internal market, and to the well-being of natural persons.

(3) Directive 95/46/EC of the European Parliament and of the Council ([4]) seeks to harmonise the protection of fundamental rights and freedoms of natural persons in respect of processing activities and to ensure the free flow of personal data between Member States.

(4) The processing of personal data should be designed to serve mankind. The right to the protection of personal data is not an absolute right; it must be considered in relation to its function in society and be balanced against other fundamental rights, in accordance with the principle of proportionality. This Regulation respects all fundamental rights and observes the freedoms and principles recognised in the Charter as enshrined in the Treaties, in particular the respect for private and family life, home and communications, the protection of personal data, freedom of thought, conscience and religion, freedom of expression and

PICTURE: GDPR
https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32016R0679

3. REGULATIONS FOR THE MEDICAL INDUSTRY

- **GDPR and EXPLAINABILITY CONCERN**
  - Terms not included in this EU regulation
    - Explanation right / right of explanation / explainability
    - Artificial Intelligence
    - Machine Learning
  - Terms included in this EU regulation:
    - Fairness
    - Transparency
    - Explanation of the decision
    - Automated decision-making

3. REGULATIONS FOR THE MEDICAL INDUSTRY

- **EU LEGAL FRAMEWORK OF AI**
  - Ongoing since 2018
  - AI might create some risk for safety of consumers and fundamental rights.
  - Risk-based legal framework proposal (April 2021)
  - Independent of origin of producer or user.

EUROPEAN COMMISSION

Brussels, 21.4.2021
COM(2021) 206 final

2021/0106 (COD)

Proposal for a

REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL

LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS

{SEC(2021) 167 final} - {SWD(2021) 84 final} - {SWD(2021) 85 final}

3. REGULATIONS FOR THE MEDICAL INDUSTRY

- **EU LEGAL FRAMEWORK OF AI**
  - Level: Unacceptable risk
  - AI that contradicts EU values is prohibited (Title II, Art. 5)
  - Subliminal manipulation resulting physical/psych harm
  - Exploiting vulnerabilities resulting in phy/psy harm
  - Social scoring by public authorities
  - Real-time remote biometric identification for law enforcement purposes in publicly accessible spaces

Picture: from the presentation made by Salvatore Scalzo, European Commission "Proposal for EU Legal Framework" 2022.

3. REGULATIONS FOR THE MEDICAL INDUSTRY

- **EU LEGAL FRAMEWORK OF AI**
  - Level: High Risk
  - Safety components of regulated products (e.g. medical devices)
  - Stand-alone AI-based systems for:
    - Biometric identification of a person
    - Operation of critical infrastructures
    - Educational training
    - Employment and workers management
    - Law enforcement
    - Migration, asylum and border control.
    - Administration of justice

3. REGULATIONS FOR THE MEDICAL INDUSTRY

— USE CASE 1

Make decisions affecting life quality

Poor explanation

Impersonation

No explanation

"right not to be subject to a decision based solely on automated processing"

3. REGULATIONS FOR THE MEDICAL INDUSTRY

– MEDICAL INDUSTRY: STANDARDS & REGULATIONS

- HOW TO **MITIGATE RISKS**
- FOCUS: quality, safety, effectiveness and performance
- US FDA: Title 21 of Code of Federal Regulations
- EU: 2017/745 MEDICAL DEVICE REGULATIONS
- STANDARDS:
  - ISO 13485 Medical Devices (quality systems)
  - ISO 14971 Risk management for medical devices
  - IEC 62304 Software lifecycle for medical devices
  - IEC 60601-1 Programmable electrical medical systems
  - IEC 82304-1 Software as medical devices
  - Etc.

3. REGULATIONS FOR THE MEDICAL INDUSTRY

– MEDICAL INDUSTRY: STANDARDS & REGULATIONS

- REGULATION (EU) 2017/745 on Medical Devices (MDR)
  - Fully applicable 26/May/2021

  - MEDICAL DEVICE DEFINITION (art. II)

    *"**Medical Device**: any instrument, apparatus, appliance, **software**, implant, […] for human beings for one or more of the following specific medical purposes: diagnosis, prevention, monitoring, prediction, prognosis, treatment or alleviation of disease […]."*

3. REGULATIONS FOR THE MEDICAL INDUSTRY

– MEDICAL INDUSTRY: STANDARDS & REGULATIONS

- REGULATION (EU) 2017/745 on Medical Devices (MDR)

SOFTWARE CAN BE MEDICAL DEVICES:

– OPERATING SYSTEMS?

– MANAGEMENT AND ADMINISTRATIVE SOFWARE?

– MEDICAL TRAINING PURPOSES (goal not patients)?

– SOFTWARE CONTROLING MEDICAL DEVICE?

– ALGORITHM POST PROCESSING BIOSIGNAL?

– SOFTWARE POST PROCESSING FOR DATA PREPARATION?

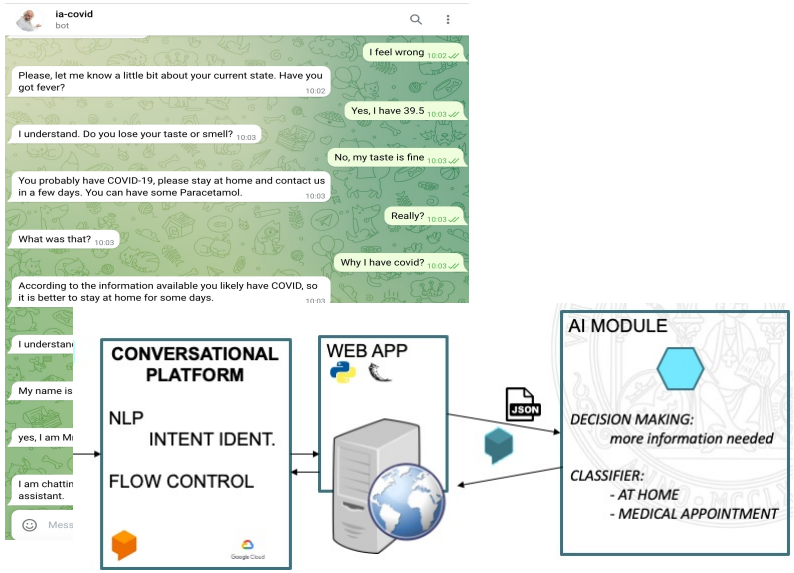
– DIAGNOSIS ALGORITHM?

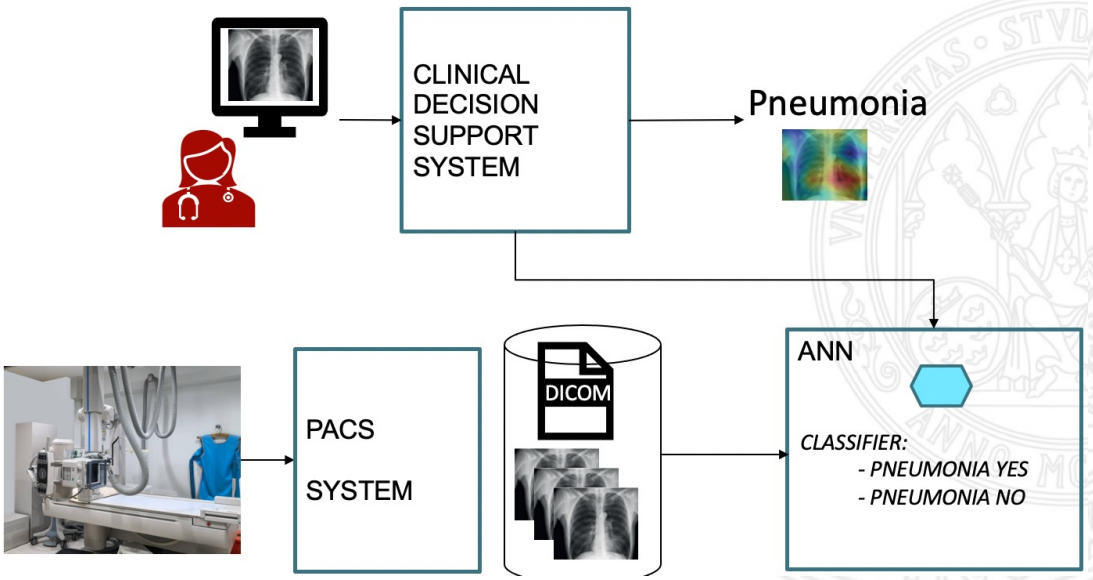
– TREATMENT ALGORITHM?

3. REGULATIONS FOR THE MEDICAL INDUSTRY

– MEDICAL INDUSTRY: STANDARDS & REGULATIONS

- REGULATION (EU) 2017/745 on Medical Devices (MDR)
  - – USE CASE 1: MEDICAL DEVICE? CLASS?
  - – USE CASE 2: MEDICAL DEVICE? CLASS?



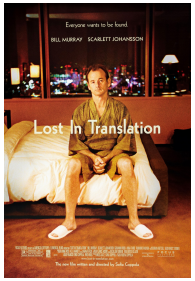


UNIVERSIDAD DE MURCIA

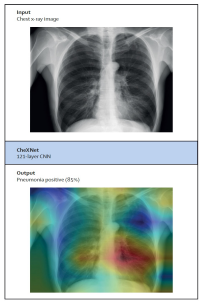
- **OUTLINE:**
  1. HUMANS, HEALTHCARE AND ETHICAL AI
  2. AI UNDERSTANDABLE IN HEALTHCARE
  3. REGULATIONS FOR THE MEDICAL INDUSTRY
  4. FINAL

4. FINAL

Ethical Guidelines
Trustworthy AI

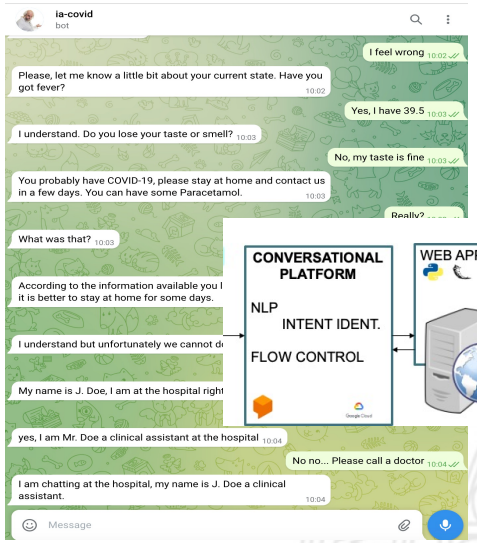


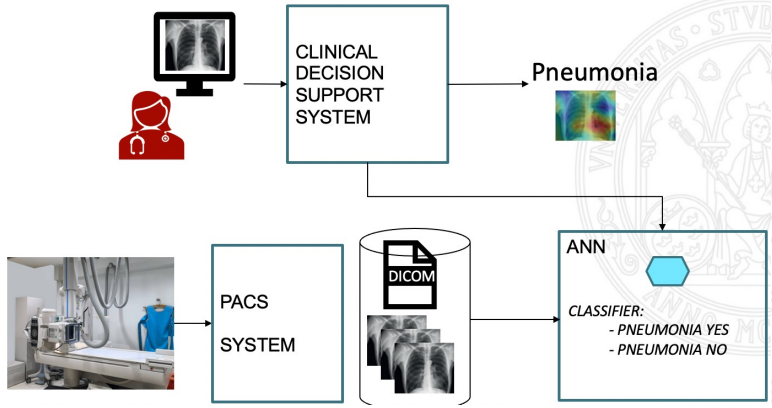
Lost in translation!
XAI
LIME, SHAP, Saliency



Medical AI
Black Boxes

GDPR
EU AI initiatives
Standards & MDR

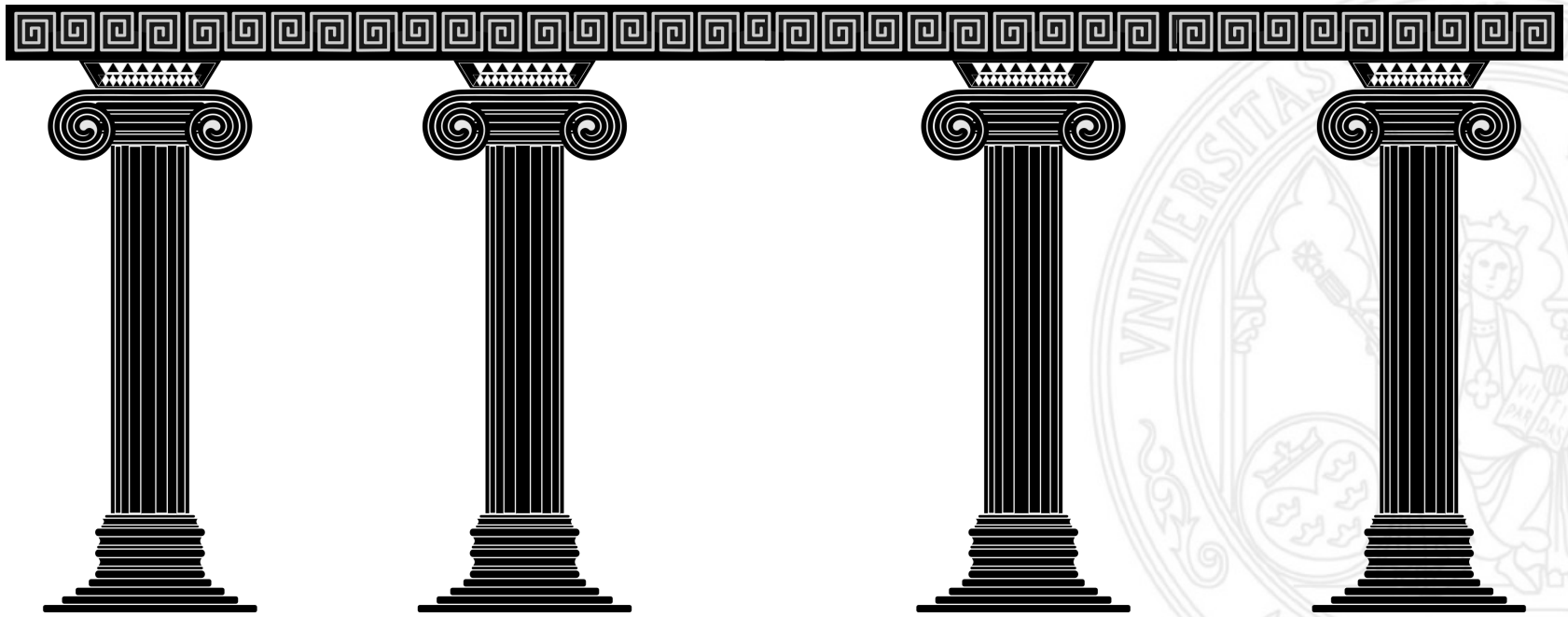




UNIVERSIDAD DE MURCIA

4. FINAL

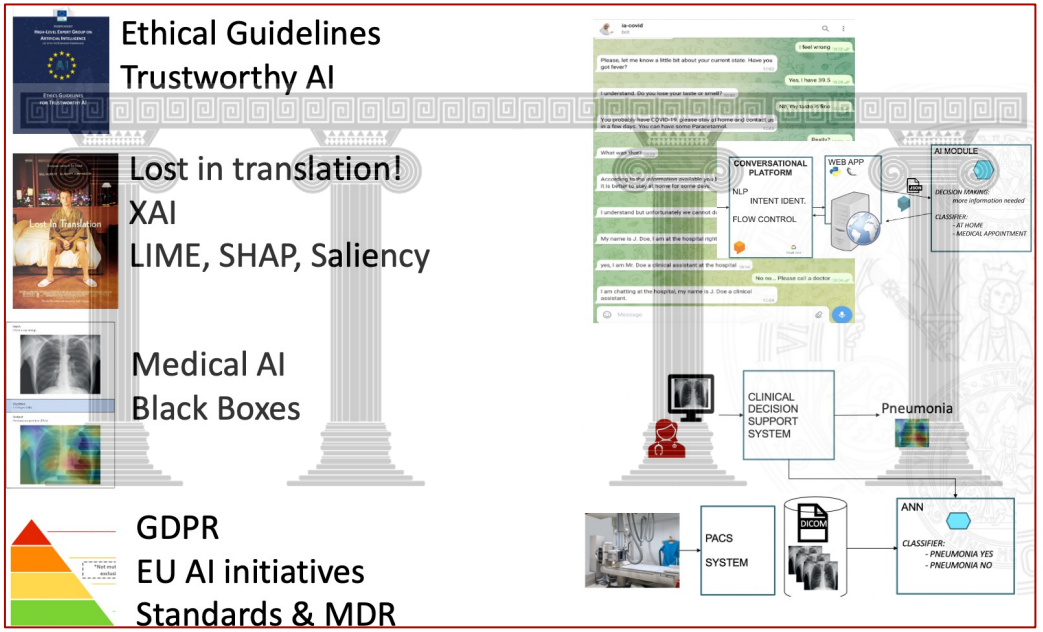
TRUST IN MEDICAL AI



HUMANS	HEALTHCARE	TECHNOLOGY	INDUSTRY
ETHICS	DOCTORS	UNDERSDAND.	REGULATIONS
	PATIENTS	XAI	STANDARDS

UNIVERSIDAD DE MURCIA

SEMINAR: FRAMING TRUST IN MEDICAL AI



Contact: Jose M. Juarez

jmjuarez@um.es

REFERENCES & READINGS

Materials & references of the seminar at:

https://webs.um.es/jmjuarez/trustAImedicine/