FITec
Inovações Tecnológicas

IFIP WG 10.4 event

Praia do Forte - BA

Feb 2025

Critical and non-critical AI applications

the need for interpretability issues

Giovanni M. de Holanda

FITec
Inovações Tecnológicas

FITec
Technological Innovations

Artificial Intelligence solutions for R&D and innovation projects

Research, development and applications of artificial intelligence approaches, methods and techniques to overcome technological challenges and meet current business demands from various sectors of society.

Member of

BRAZILIAN INSTITUTE OF DATA SCIENCE

BIDS

a living *Artificial Intelligence* laboratory connecting researchers, corporations and startups in an ecosystem of innovation

FITec
Inovações Tecnológicas

Artificial Intelligence

Some applications

AI applications

Critical applications, mission critical
(e.g., as in Draft Law of the Brazilian Federal Senate)

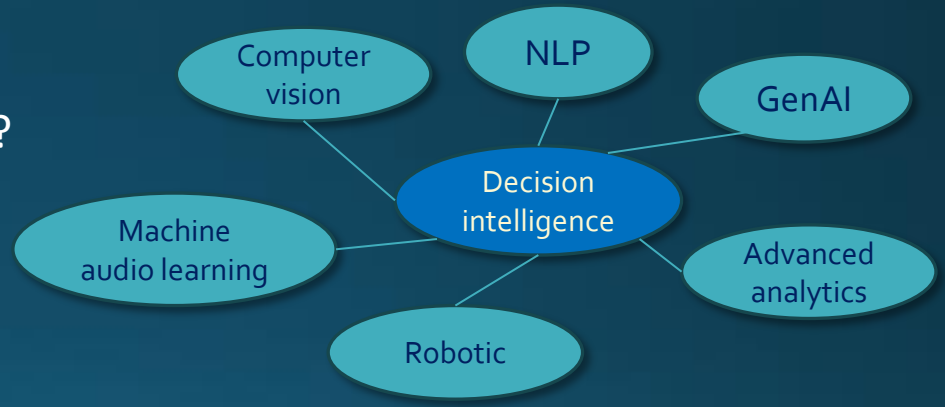
- Critical services (high risk) with high algorithmic impact
- Autonomous vehicles
- Diagnosis and medical procedures
- Management and operation of critical infrastructures (transit, water and energy supply networks)
- Support for the administration of justice and law
- Assessment of access criteria to essential services
- Support logistics for emergency services, such as firefighters and emergency services
- Biometric identification

Non-critical applications: what else?

Important but-non-critical applications

There are a lot of applications that fall into we can call as “important but-non-critical applications”

- Computer vision for identifying and classifying objects
- Predictive analysis for power performance
- Fault detection, but to some extent it is non-critical?
- LLM for chatbots, but GenAI (?)
- Immersive experiences
- …



Important but-non-critical applications
to what extent?

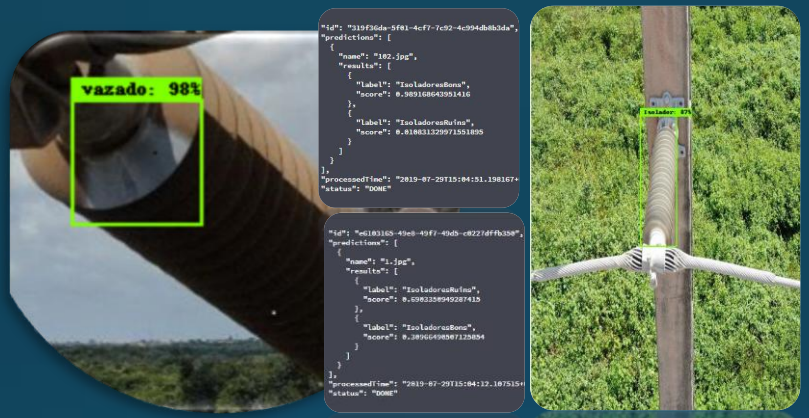
Some FITec AI projects

FITec
Inovações Tecnológicas

IoT energy

fleet management

chatbot



ML Identification

ML Inspection

production control

resources monitoring

Insulators' identification

R&D project

FITec
Inovações Tecnológicas

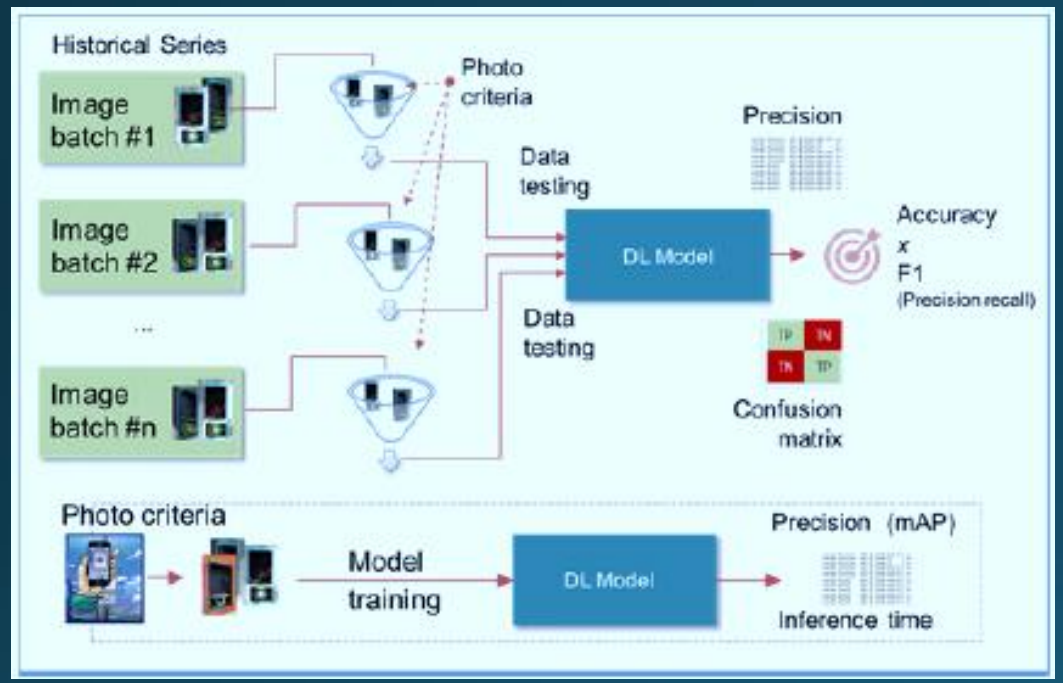

```
"id": "319f36da-5f01-4cf7-7c92-4c994db8b3da",
"predictions": [
  {
    "name": "102.jpg",
    "results": [
      {
        "label": "IsoladoresBons",
        "score": 0.9891686439[illegible]1416
      },
      {
        "label": "IsoladoresRuins",
        "score": 0.010831329971551895
      }
    ]
```


```
"id": "e6103165-49e8-49f7-49d5-c0227dffb350",
"predictions": [
  {
    "name": "1.jpg",
    "results": [
      {
        "label": "IsoladoresRuins",
        "score": 0.6903350949287415
      },
      {
        "label": "IsoladoresBons",
        "score": 0.3096649050712585
      }
    ]
  }
],
```

Detection and classification of power meter

FITec
Inovações Tecnológicas

R&D project



Source: (Finardi et al., 2021)

Digital Transformation

R&D projects

Engineer Asset Management
Predictive analysis and RCM

Power Performance Assessment
Prediction, identification and RUL

Soundscape Monitoring
Machine audio learning

Digital Transformation

R&D projects

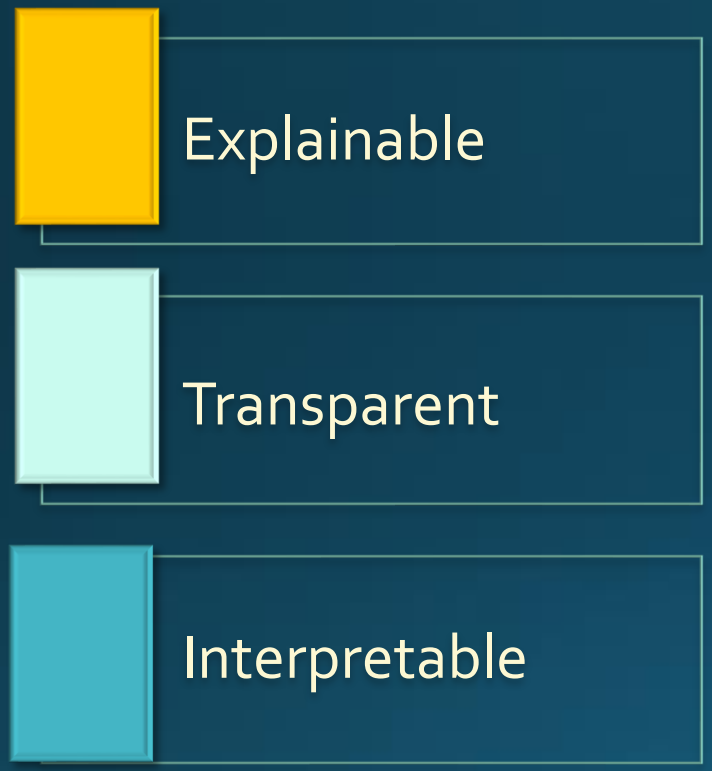
Project New Energy Connection (Equatorial Energia)

LLM
WhatsApp
Usability and UX
New interaction process

Project ACGM

includes a Virtual Showroom: immersive environment for displaying and interacting with products such as laptops and monitors, mediated by a virtual assistant (GenAI) that responds to technical details of the products.

Critical or non-critical: the interpretability need



Cf. (Holanda & Pfeiffer, 2023)

XAI – eXplainable Artificial Intelligence

The decision of the model can be comprehended *post-hoc* by experts using tools and considerations

Transparency

refers to the characteristic of a model being, in itself, understandable by a human

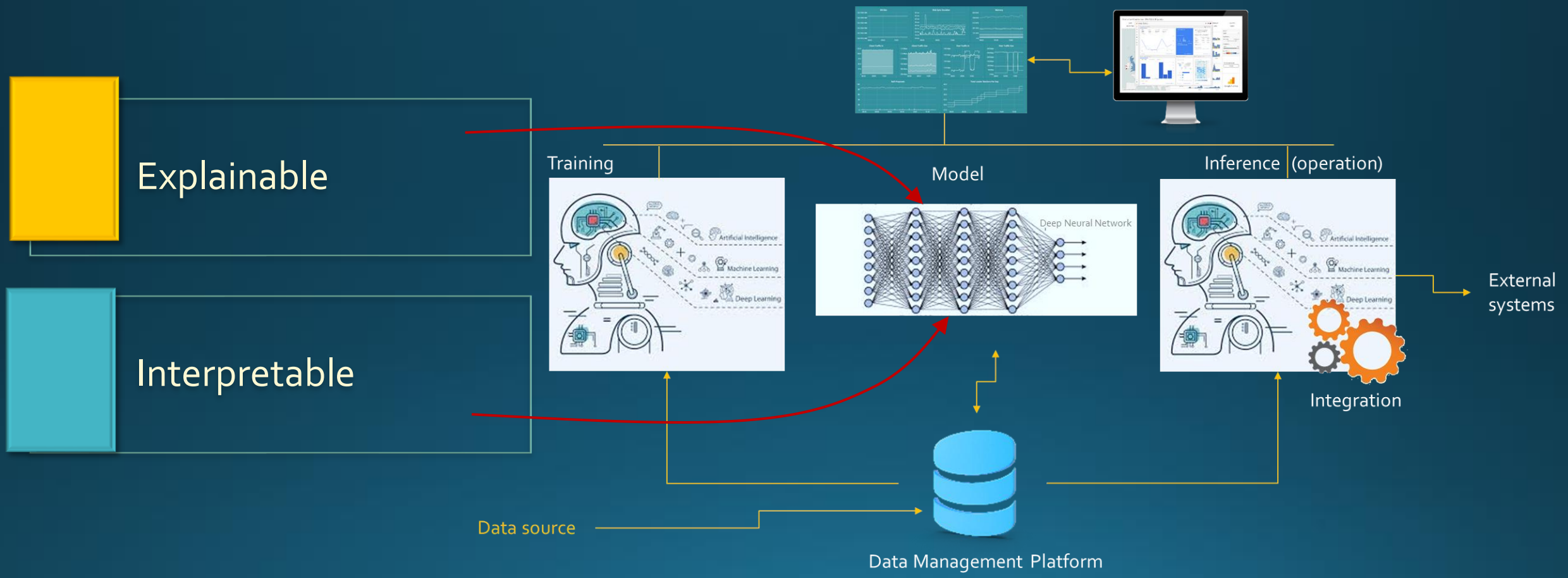
Arrieta et al. (2020)

IML – Interpretable Machine Learning

The decision of the model can be easily comprehended by experts according to the *ante-hoc* model design and their domain knowledge

Rudin et al. (2022)

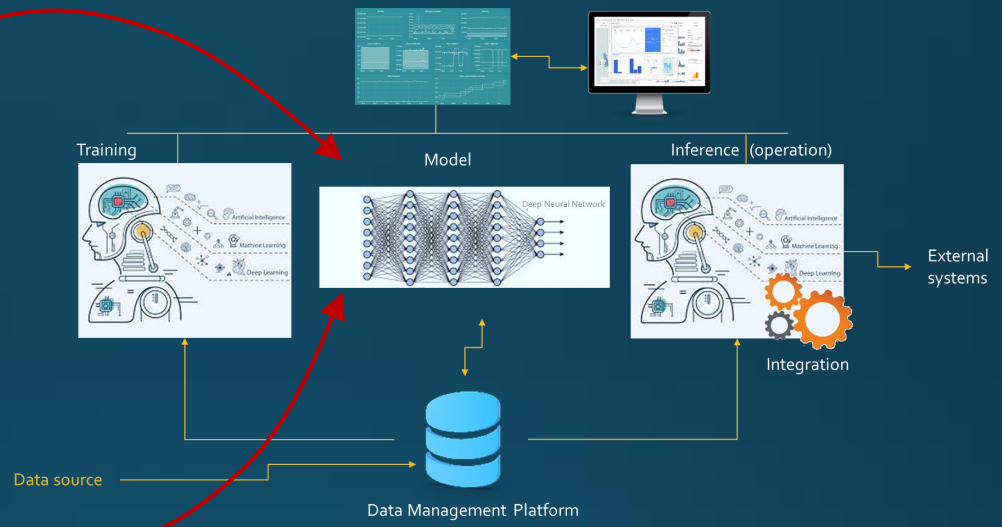
Critical or non-critical: the interpretability need



Critical or non-critical: the interpretability need

Explainable (XAI)

Interpretable (IML)



- There is a long discussion on IML *vs.* XAI: which is more trustworthy, or which may be less accurate
- There is no consensus in the literature or in practice regarding this

Critical or non-critical: the interpretability need

- Black boxes are generally unnecessary, given that their accuracy is generally not better than a well-designed interpretable model (*sic*)
- Explanations for black boxes are often problematic and misleading, potentially creating misplaced trust in black box models
- Explainability techniques give authority to black box models rather than suggesting the possibility of models that are understandable in the first place
- An interpretable model is constrained, obeying a set of domain-specific constraints that make judgment processes understandable

(Rudin et al., 2022); (Rudin and Radin, 2019)

References

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ... & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information fusion*, 58, 82-115.

Rudin, C., Chen, C., Chen, Z., Huang, H., Semenova, L., & Zhong, C. (2022). Interpretable machine learning: Fundamental principles and 10 grand challenges. *Statistics Surveys*, 16, 1-85.

Rudin, C. and Radin, J. (2019). Why are we using black box models in AI when we don't need to? A lesson from an explainable AI competition. *Harvard Data Science Review*, 1(2).

Holanda, G. M., Pfeiffer, C. C. (2023). *Sentimento da Inteligência Artificial - novas tecnologias, antigos conceitos*. Pontes editores.

Finardi, F.A.R., G.M. Holanda, G.M., Adorni, C.Y.K.O., Nader, M.V.P, and K.R.C Pinheiro (2021) Evaluating the application of a computer vision model in the customer service chatbot of an electric utility. *Proc. of the ICECCME*. **IEEE** *Xplore®*

Thanks!

Giovanni M. de Holanda
Lead Researcher and Data Scientist
gholanda@fitec.org.br

FITec
Inovações Tecnológicas