ODSC EU 2024

LONDON I SEPTEMBER 5-6

MERVE ALANYALI, PHD

Head Of Data Science Research
Allianz Personal

Beyond Interpretability: An Interdisciplinary Approach to Communicate Machine Learning Outcomes

TALK

How to explain ML outcomes?

XAI often referred to as **a solution to social, ethical and regulatory considerations** of using ML models

XAI

↙ ↘

Technical explanations | **Socio-technical explanations**

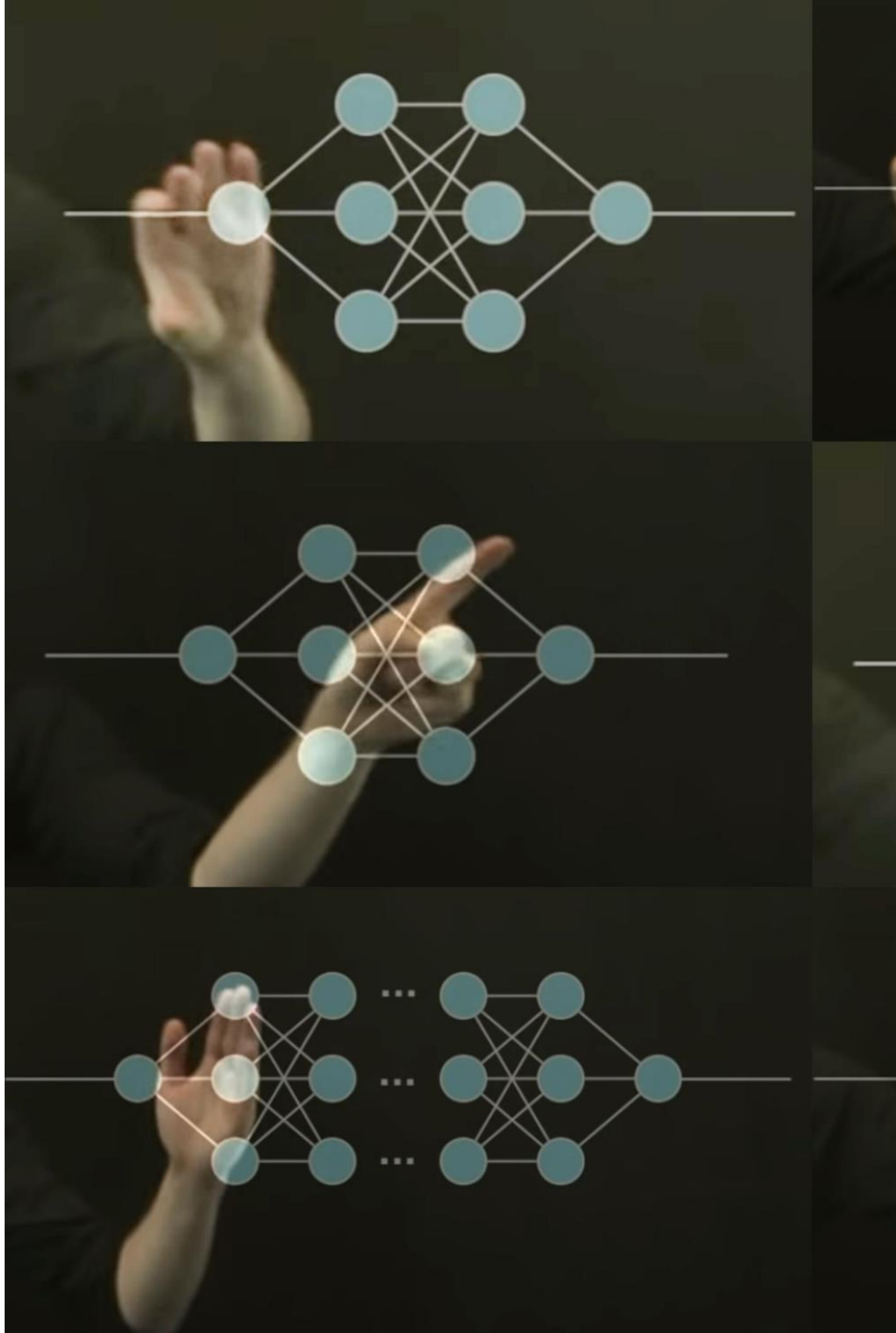
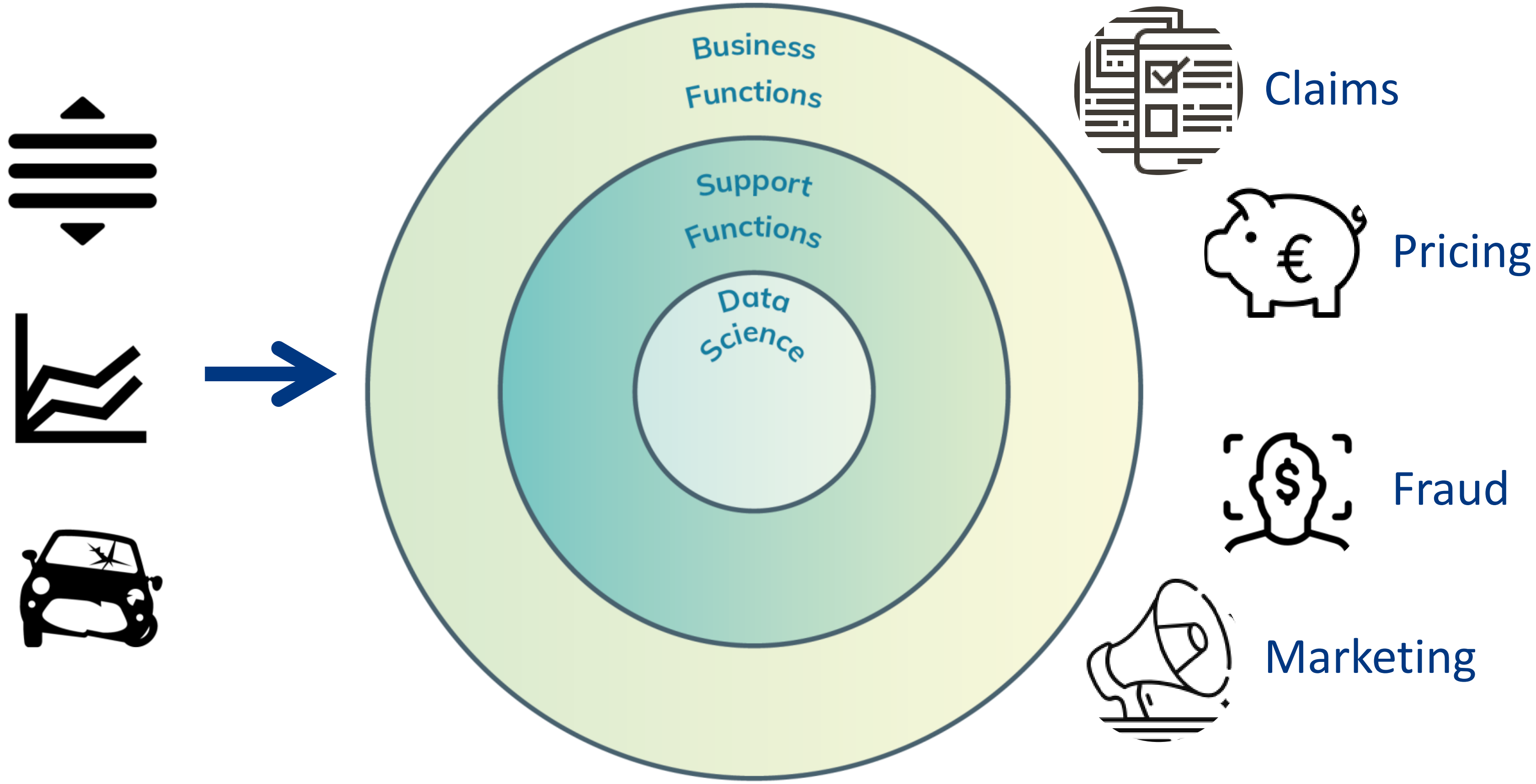


Image credit: Alexa Steinbrück / Better Images of AI / Explainable AI / CC-BY 4.0

Data Science at Allianz Personal

Putting machine learning into practice



7+ years

since
team inception

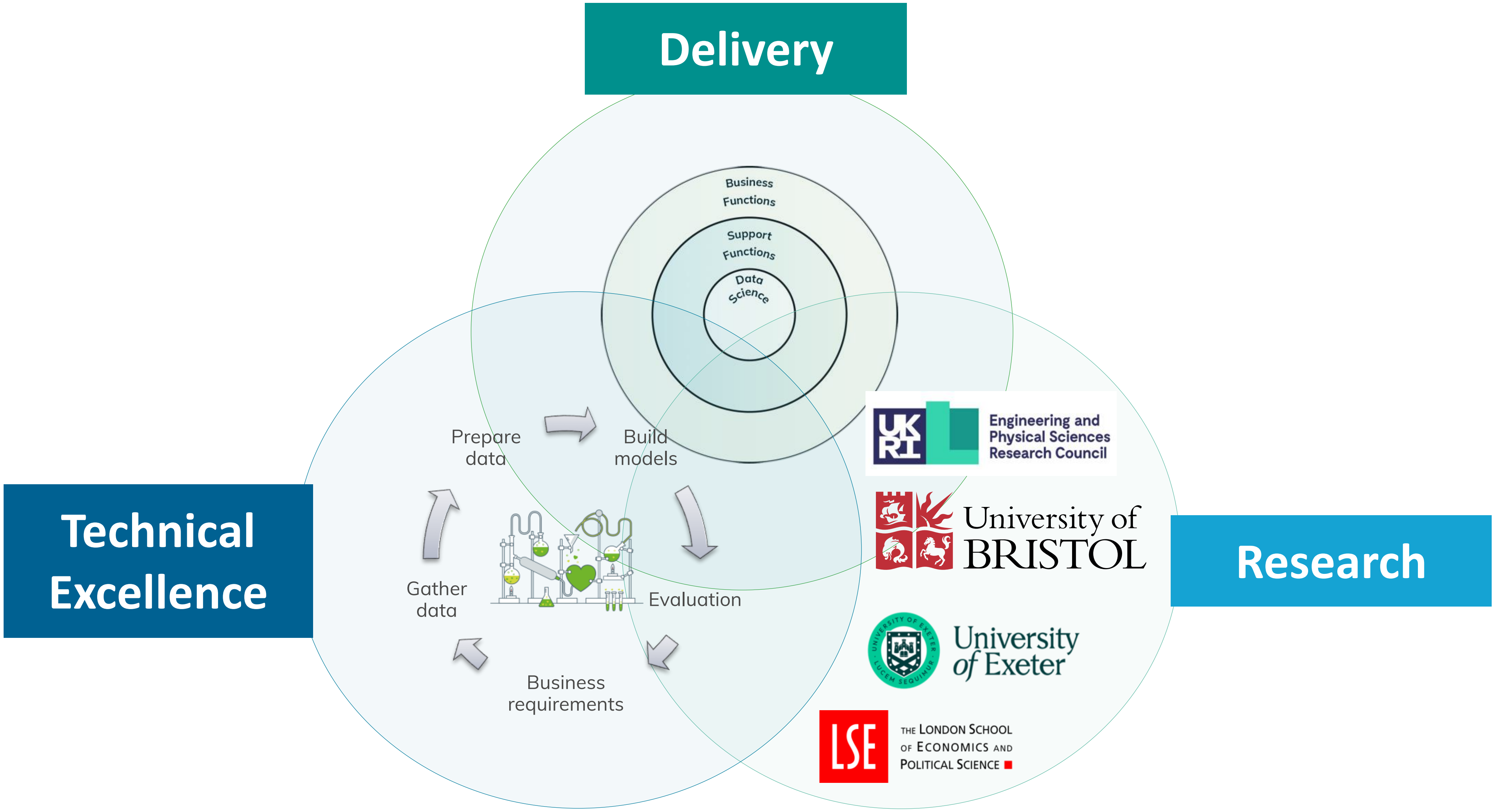
50+

decisions influenced
by mid 2023

55+

team members

Where do we focus our effort?



Interpretability vs Explainability

Interpretability

"… to understand exactly why and how the model is generating predictions,… observe the inner mechanics of the model."

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

Ex: Keeping everything else same, one unit increase in x_1 will make β_1 change on y.

Explainability

"… take an ML model and explain the behaviour in human terms."

Ex: Why does this model predict that I would find this film boring? (see the next slide for the model details☺)

Source: Kin, J., Zhang, b., Mahboobi, H. and Roy, S., 2021. Model Explainability with AWS Artificial Intelligence and Machine Learning Solutions. *AWS Whitepaper*.

Interpretability vs Explainability

Allianz

Interpretability

"… to understand exactly why and how the model is generating predictions,… observe the inner mechanics of the model."

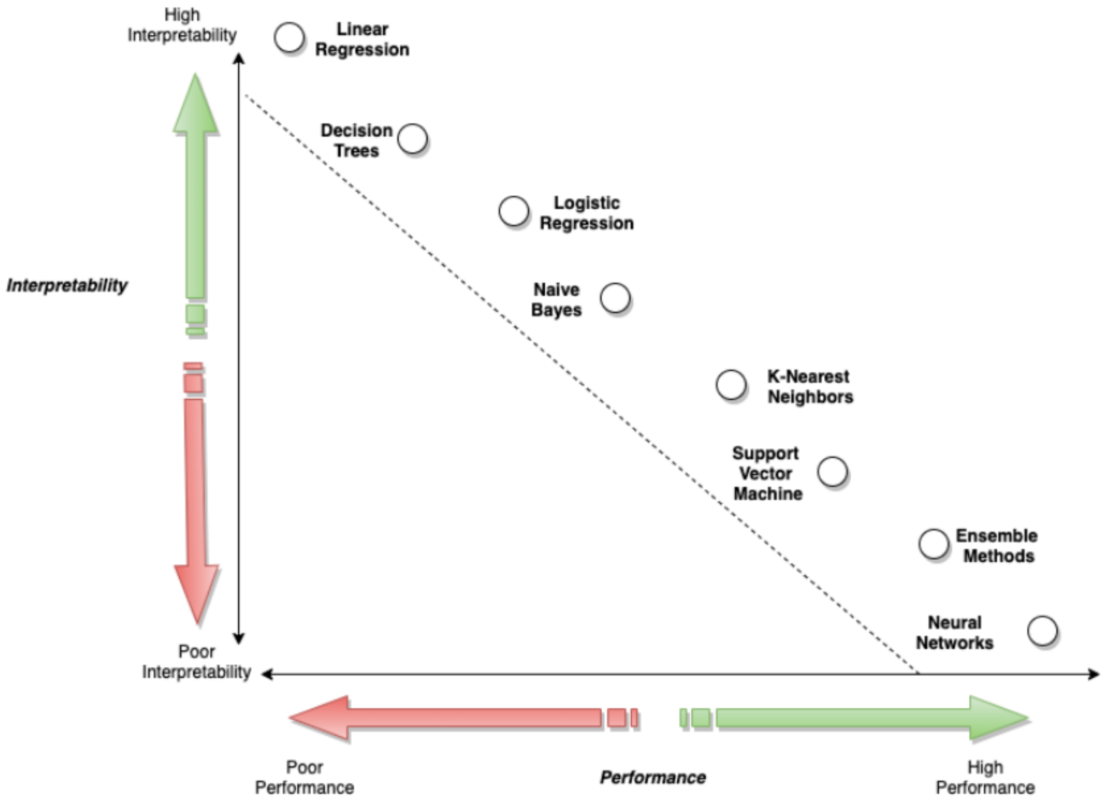
$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

Ex: Keeping everything else same, one unit increase in x_1 will make β_1 change on y.

Explainability

"… take an ML model and explain the behaviour in human terms."

Ex: Why does this model predict that I would find this film boring? (see the next slide for the model details☺)



Source & image credit: Kin, J., Zhang, b., Mahboobi, H. and Roy, S., 2021. Model Explainability with AWS Artificial Intelligence and Machine Learning Solutions. *AWS Whitepaper*.

Glass-box vs Black-box models

Glass-box (white-box) models

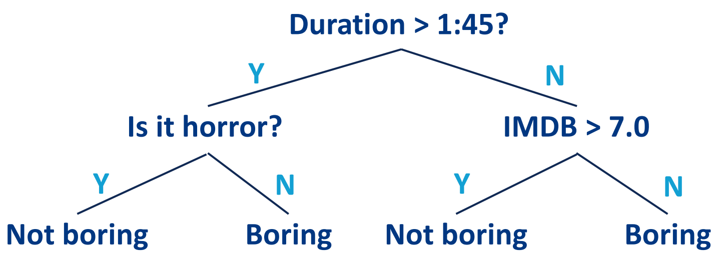
Models that are built directly for interpretability.

Examples include:

- Linear models (linear and logistic regression)
- Decision trees
- Explainable boosting machines (paper, code)
- Automatic piecewise linear regression (paper, code)

Great for interpretability however usually fall behind in performance comparison.

Is this film boring? (for me)



Black-box models

Models that we cannot directly extract how the model components and inner mechanics of the model impact its outcome.

Examples include:

- Ensemble methods (ex: Random forest, xgboost…)
- Deep learning models

Quite often performs better than glass-box models, however they are not easy to interpret. We need to use additional tools to provide “post-hoc” explanations.

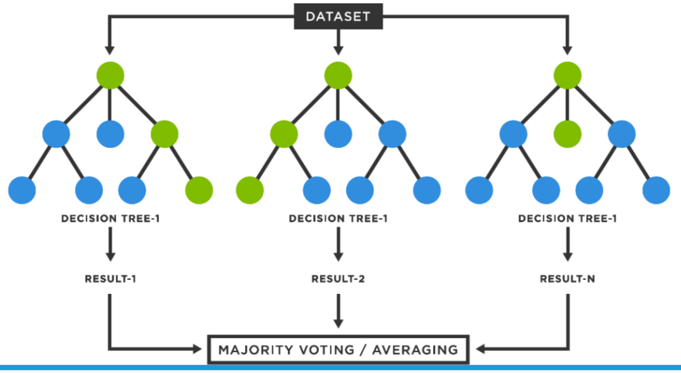


Image credit (right): https://medium.com/@denizgunay/random-forest-af5bde5d7e1e

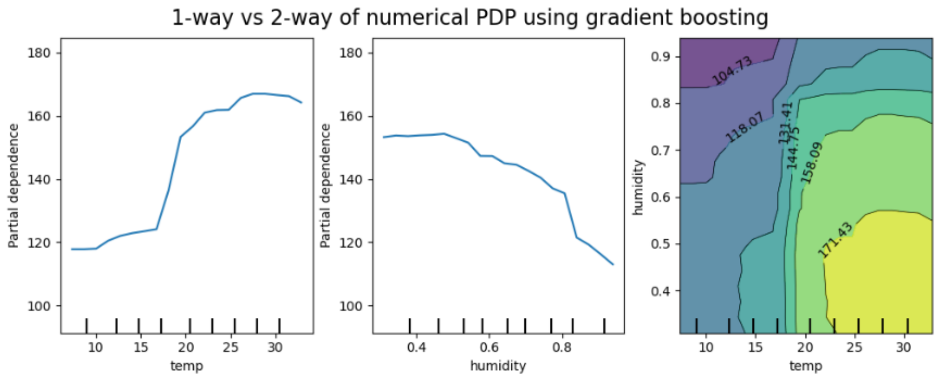
Model-agnostic methods

Global model agnostic methods

Great at creating a summary of an ML model. Useful to provide high-level explanations when communicating with stakeholders

Methods include:

- Partial dependence plots (PDPs)
- Feature importance and interaction
- Global surrogate methods



Local model agnostic methods

Provide explanations for individual data points. Useful for live-models to provide explanations

Methods include:

- Individual conditional expectation (ICE)
- Local interpretable model-agnostic explanations (LIME)
- Shapley additive explanations (SHAP)

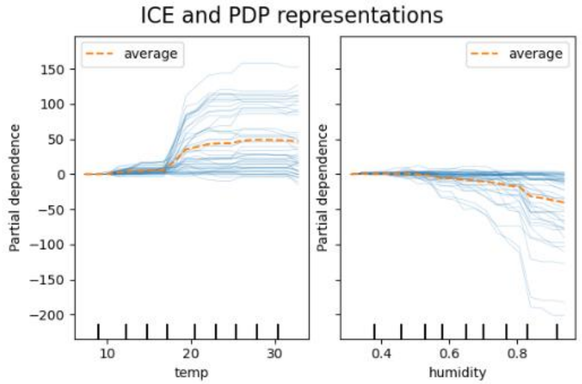
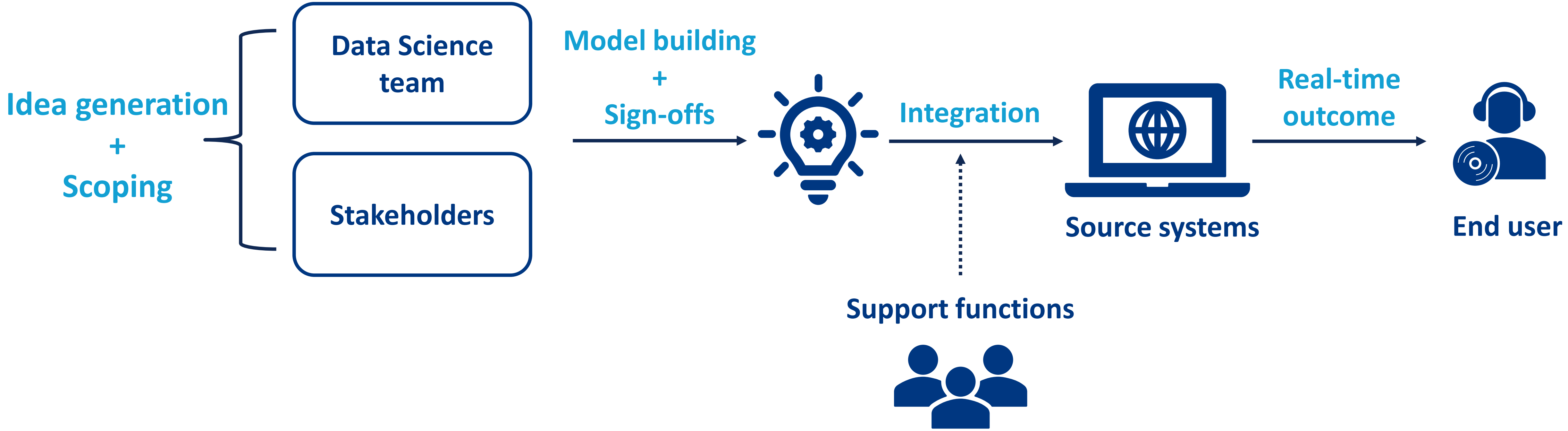


Image credit: https://scikit-learn.org/stable/modules/partial_dependence.html

End-to-end process



Bristol Digital Futures Institute collaboration

Allianz

University of BRISTOL

/BDFI

Bristol Digital Futures Institute (BDFI), is one of the University of Bristol's five research institutes.

The institute is working to fundamentally **transform digital innovation**, and create more inclusive, sustainable and prosperous futures for all.

- Understand the implications of innovative technologies powered by ML
- Create new knowledge and understanding of sociotechnical innovation
- Shape new ways of working across disciplines and sectors through inclusive conversations

- **Explainable AI decision making; Dr Marisela Gutierrez, Prof Susan Halford**
- Machine learning "design" and "use"; Kate Byron, Prof Susan Halford

* Information used with the permission of Dr Marisela Gutierrez

What makes AI explainable?

Allianz

Reimagining "Explainable AI": de-centring ML models from explanations

Explaining **ML models** → Explaining **"ML practices"**

Research methodology:

- Organisational ethnography at AZP
- Co-produced with BDFI community partners

/BDFI + Black South West Network KWMC KNOWLE WEST MEDIA CENTRE

What makes AI explainable?

Allianz

Reimagining "Explainable AI": de-centring ML models from explanations

Research met

- Organisatio
- Co-produce

Organisational ethnography – being a "fly on the wall"

- **Shadowing** different teams and employees and **observing** various meetings
- **Interviews** to gather qualitative data
- **Review of internal documents**, communications, and public information

/BDFI + Black South West Network KWMC KNOWLE WEST MEDIA CENTRE

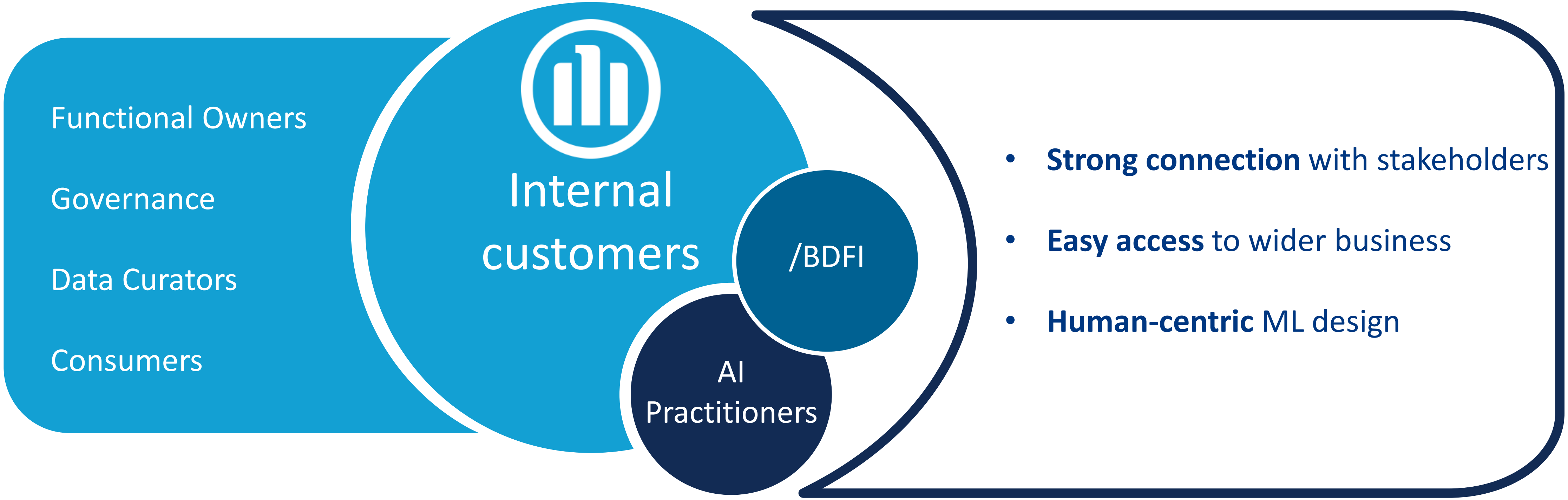
The missing link

Allianz

- **Strong connection** with stakeholders
- **Easy access** to wider business
- **Human-centric** ML design

The missing link

Allianz



Our academic partners **bridge the gap** between our AI Practitioners and the AI Public

Co-designed = better designed

Allianz

Impact from qual work

Internal workshops and launched an open-source explainbility package.

Closer collaboration

End-users are co-designers, and open communication creates efficient feedback loops.

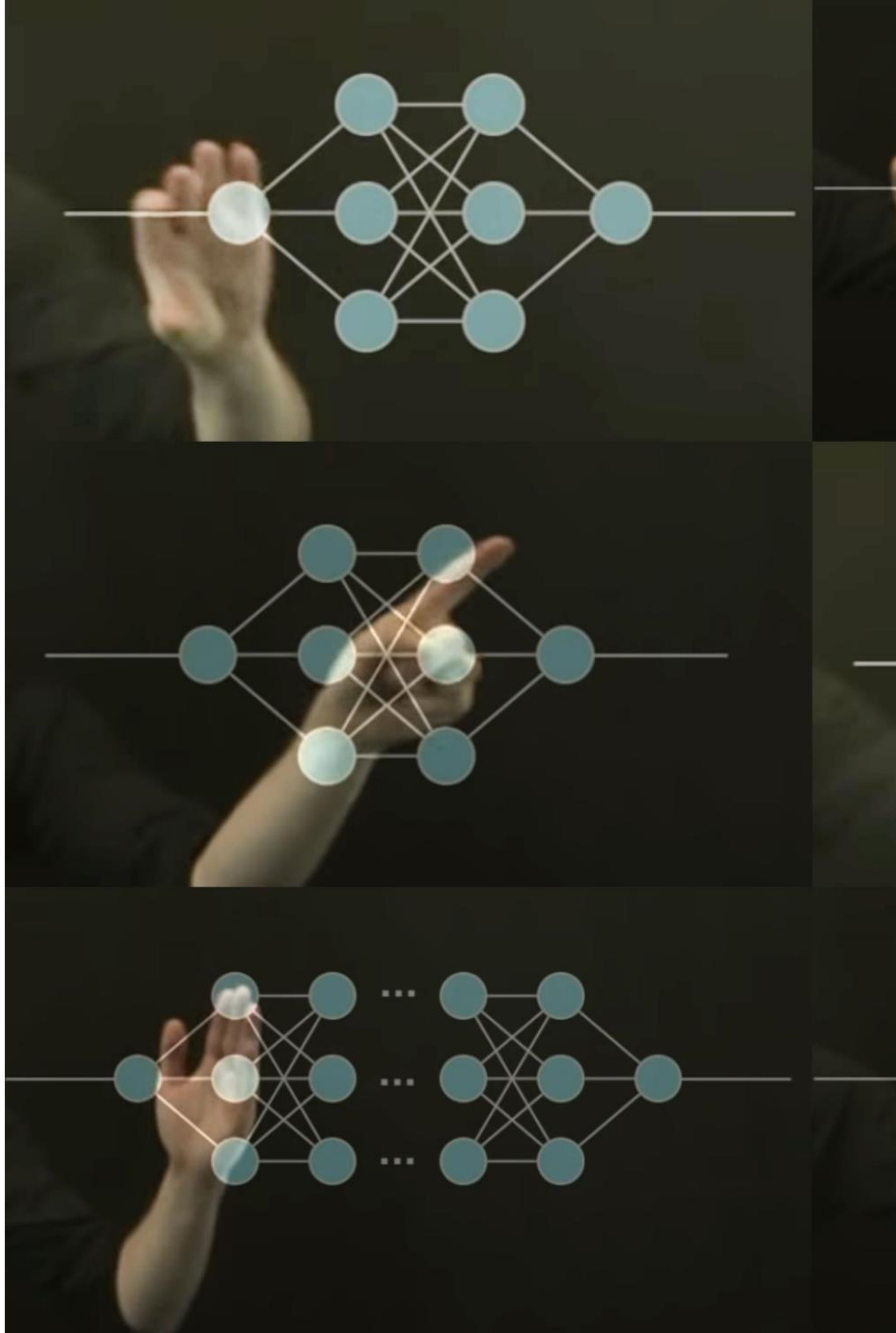
Raised awareness

Enhanced and meaningful knowledge sharing between DS team and internal customers.

Why explain ML models?

Reasons beyond regulation include (but not limited to):

- Sanity check
- Identifying bias
- Social acceptance



Resources and further reading

Allianz

- AZP model explainer Python package (GitHub repo)
- BDFI Seminar Series Dr Marisela Gutierrez Lopez and LV= General Insurance (video)
- Interpretable Machine Learning (book)
- InterpretML package (documentation)
- Miller, T., 2019. Explanation in artificial intelligence: Insights from the social sciences. *Artificial intelligence*, 267, pp.1-38 (paper)

Special thanks to Marisela and Kieran!

Dr Marisela Gutierrez

Senior Research Associate
University of Bristol

Kieran Billingham

Senior Data Scientist
Allianz Personal

Our talk: BDFI Seminar Series Dr Marisela Gutierrez Lopez and LV=General Insurance

Image credit: Nick Riddle/University of Bristol/Flickr

Thanks for listening!

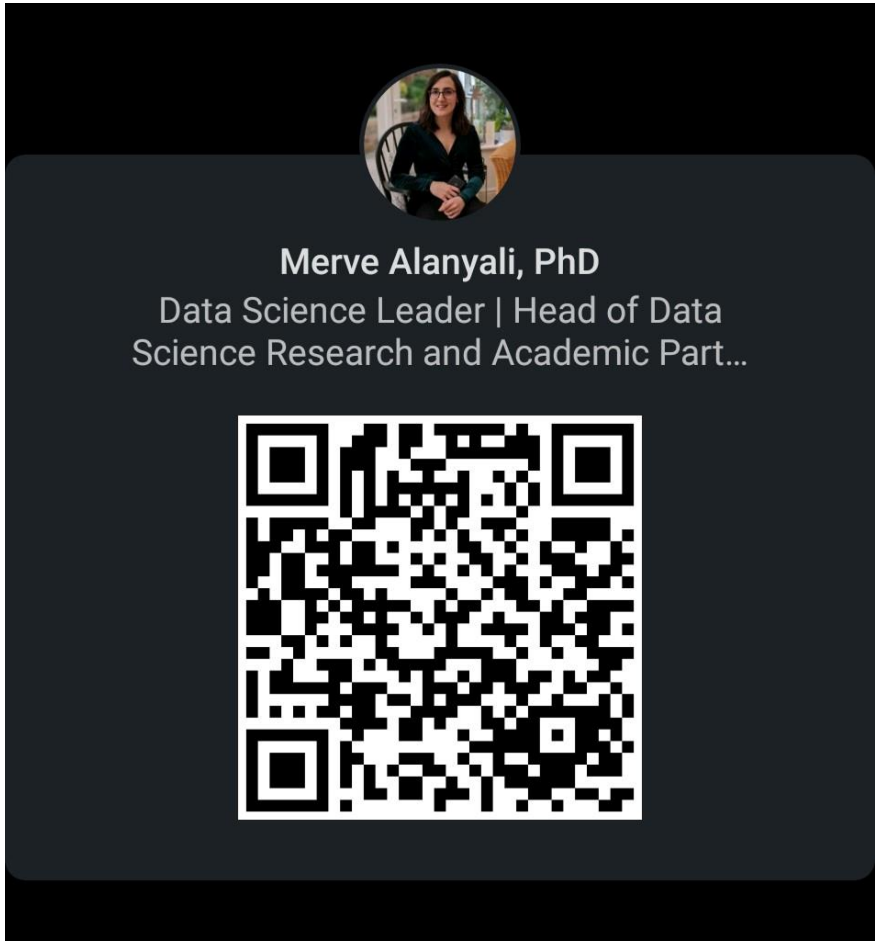


Image credit: Nick Riddle/University of Bristol/Flickr