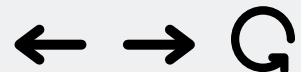Misk DSI 2022-01 Lightning Talk

Explainable Artificial Intelligence (xAI)

Presented by:
Sara Al-Qahtani

Agenda
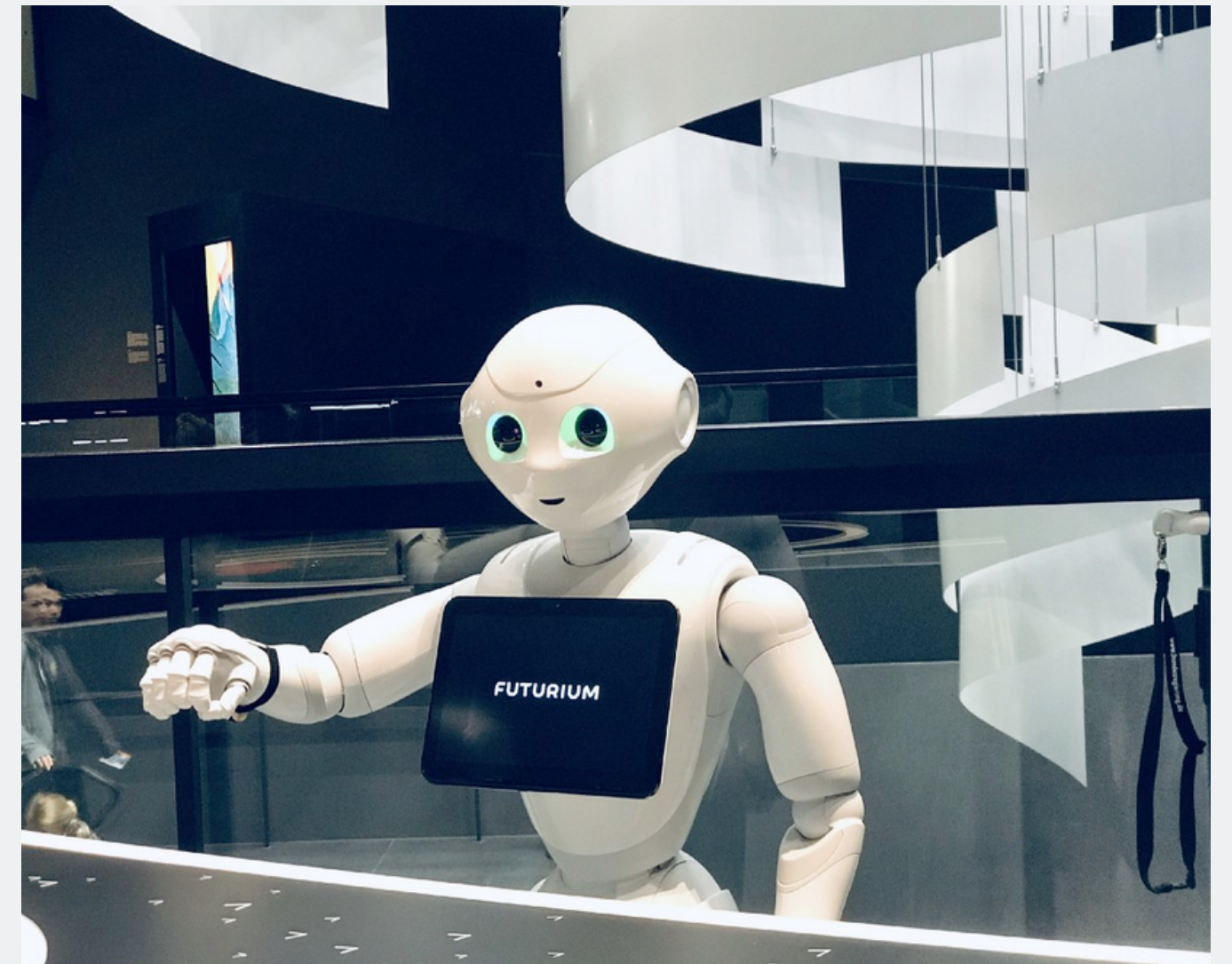
1) What's the difference between AI and xAI?
2) Why does explainability matter?
3) DEMO
4) Accuracy vs. Explainability
5) Solutions
6) References

What's the difference?

AI Versus xAI

AI produces **blackbox** models (i.e., unexplainable outcome)

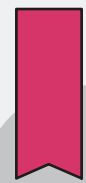xAI produces **glassbox** models (i.e., interpretable outcome)

Why does explainability matter?

Why does explainability matter?

The First POV

Applicant/customer

The Second POV

Regulator/supervisor

The Third POV

Developer/Data scientist

Explainability can help developers ensure that the system is working as expected, it might be necessary to meet regulatory standards, or it might be important in allowing those affected by a decision to challenge or change that outcome.
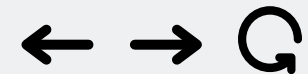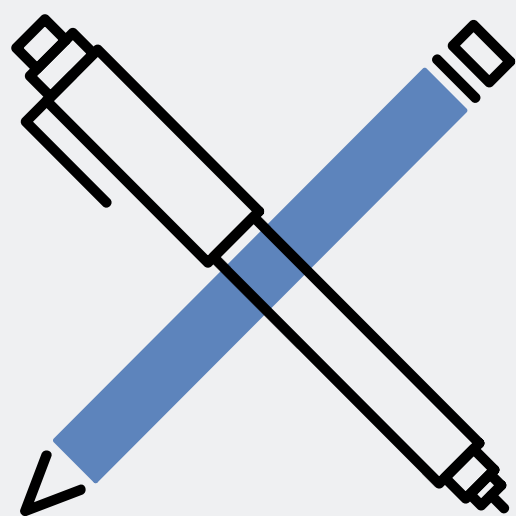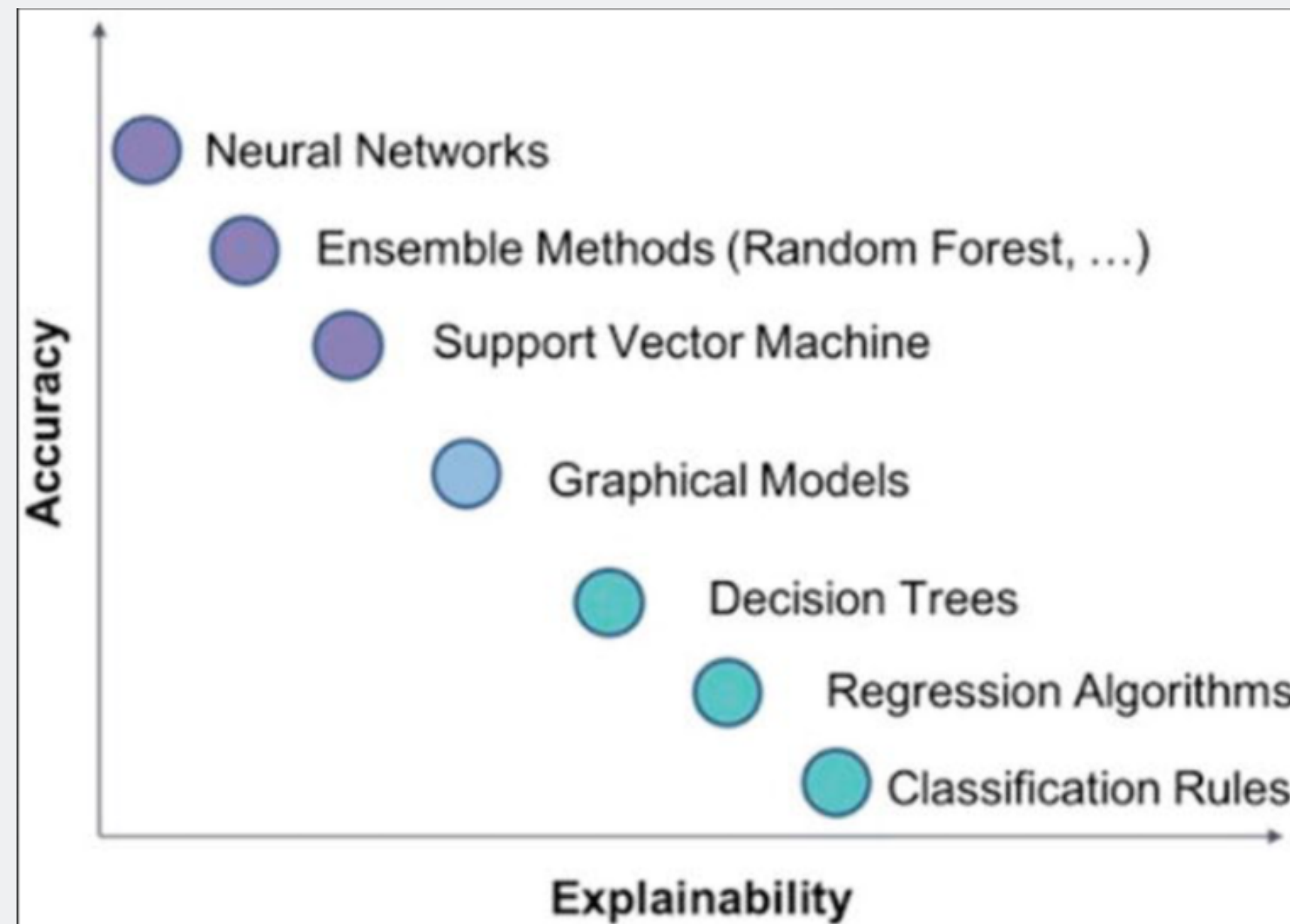
Demonstration for LIME

DEMO

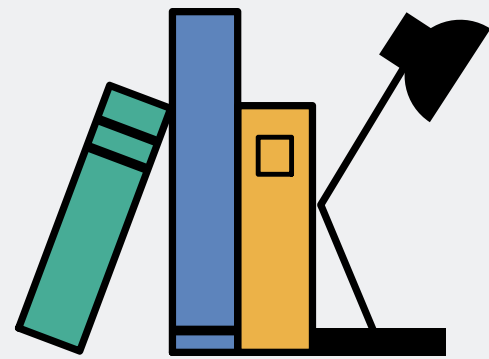# A better decision is not the same as higher accuracy scores or lower RMSE. It's using the right inputs to derive a good answer.

Python packages for explainability

Solutions

-LIME
-Shapley Values
-IBM AI Explainability 360
-Shapash
-ExplainerDashboard
-Dalex
-Explainable Boosting Machines (EBM)
-ELI5
-iModels

-Captum
-InterpretML
-Alibi Explain
-Aequitas
-DeepVis
-Tensorflow:lucid
-Tensorflow:cleverhans
-Tensorflow:WhatIf

Sources

References

01 Gunning, David; Stefik, Mark; Choi, Jaesik; Miller, Timothy; Stumpf, Simone; Yang, Guang-Zhong (2019). XAI—Explainable artificial intelligence. Science Robotics, 4(37), eaay7120–. doi:10.1126/scirobotics.aay7120

02 https://en.m.wikipedia.org/wiki/Explainable_artificial_intelligence

03 https://www.ibm.com/blogs/research/2019/08/ai-explainability-360/

04 Hind, Michael (2019). Explaining explainable AI. XRDS: Crossroads, The ACM Magazine for Students, 25(3), 16–19. doi:10.1145/3313096

05 https://www.intechopen.com/chapters/72398

06 https://www.researchgate.net/publication/338184751_Explainable_Artificial_Intelligence_XAI_Concepts_Taxonomies_Opportunities_and_Challenges_toward_Responsible_AI

07 https://towardsdatascience.com/what-is-explainable-ai-xai-afc56938d513

08 https://towardsdatascience.com/explainable-ai-xai-a-guide-to-7-packages-in-python-to-explain-your-models-932967f0634b