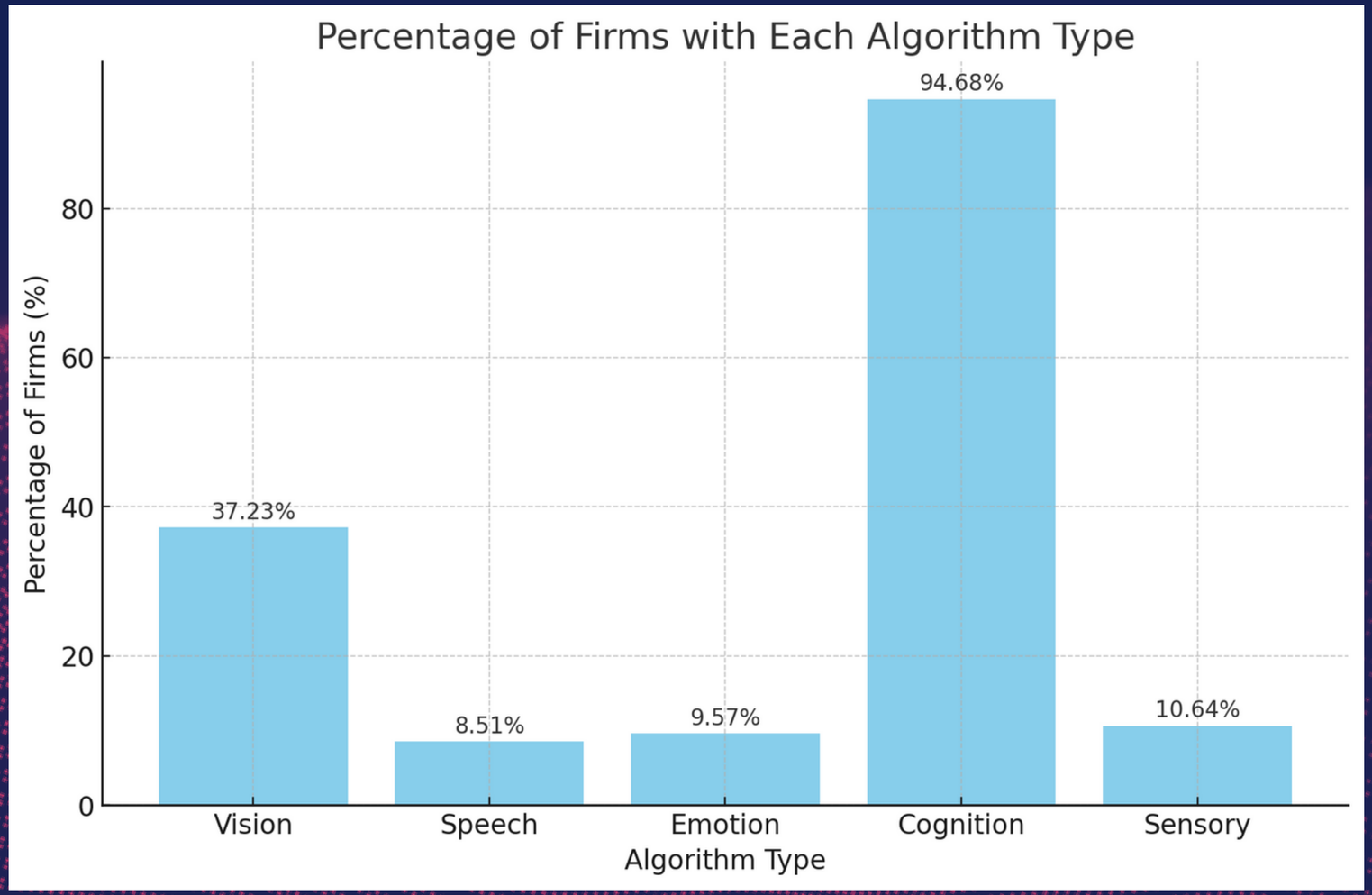
TEXAS McCombs
The University of Texas at Austin
McCombs School of Business

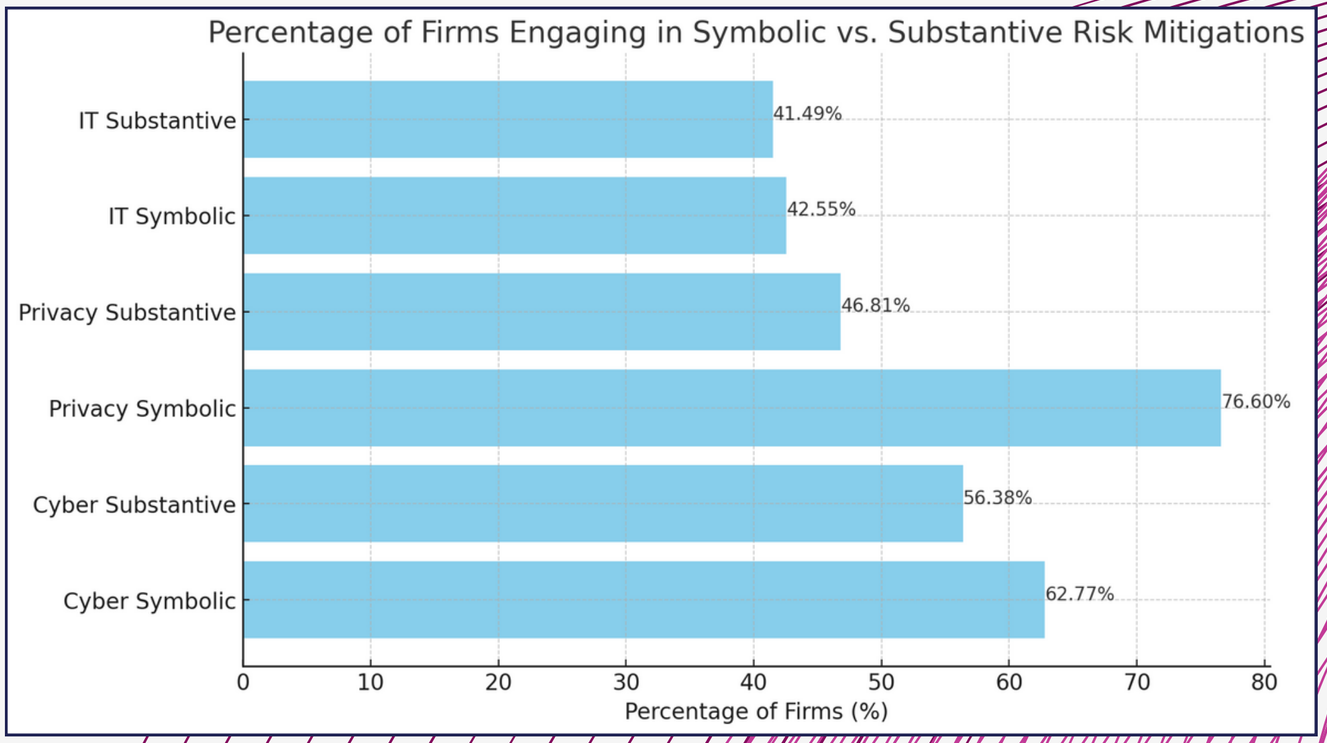
AI RISK MANAGEMENT

Consulting Group 3: Jack Hu, Jenna Kim, Ethan Sonnenreich

ALGORITHM TYPES

Of the 5 algorithm types, most of the companies employed cognitive algorithms. The **emotion-based** algorithms proved to be the **least problematic**.





SYMBOLIC VERSUS SUBSTANTIVE RISK MITIGATIONS

Symbolic mitigations: risk disclosures

Substantive mitigations: oversight and design considerations

76.60% employed symbolic privacy risk mitigations

41.49% employed substantive IT risk mitigations

DESCRIPTIVE ANALYSES

- **70.21%** of firms **disclose AI's cyber risks on average**.
- **84.04%** of firms **disclose AI's privacy risks on average**.
- **55.85%** of firms **disclose AI's IT risks on average**.
- **48.94%** of firms **disclose all three types** of risks on average: cyber, privacy, and IT risks.
- **22.34%** of firms do AI risk mitigations **symbolically on average**.
- **33.33%** of firms do AI risk mitigations **substantively on average**.
- **17.02%** of firms experienced **financial loss**
- **12.76%** of firms **harmed people**
- **6.38%** of firms faced a **lawsuit**
- **4.26%** of firms **discontinued their algorithm**
- **48.94%** of firms **did not** have **Ground Truth**.
- **29.79%** of firms employed **supervised learning**
- **37.23%** of firms employed **hybrid learning**
- **74.47%** of firms discovered algorithm issues from a **3rd-party source**, **21.28%** from the **User Org**, and **4.26%** from the **Dev Org**

CAUSE AND EFFECT ANALYSES

01

Correlation Analyses

02

Logistic Regression

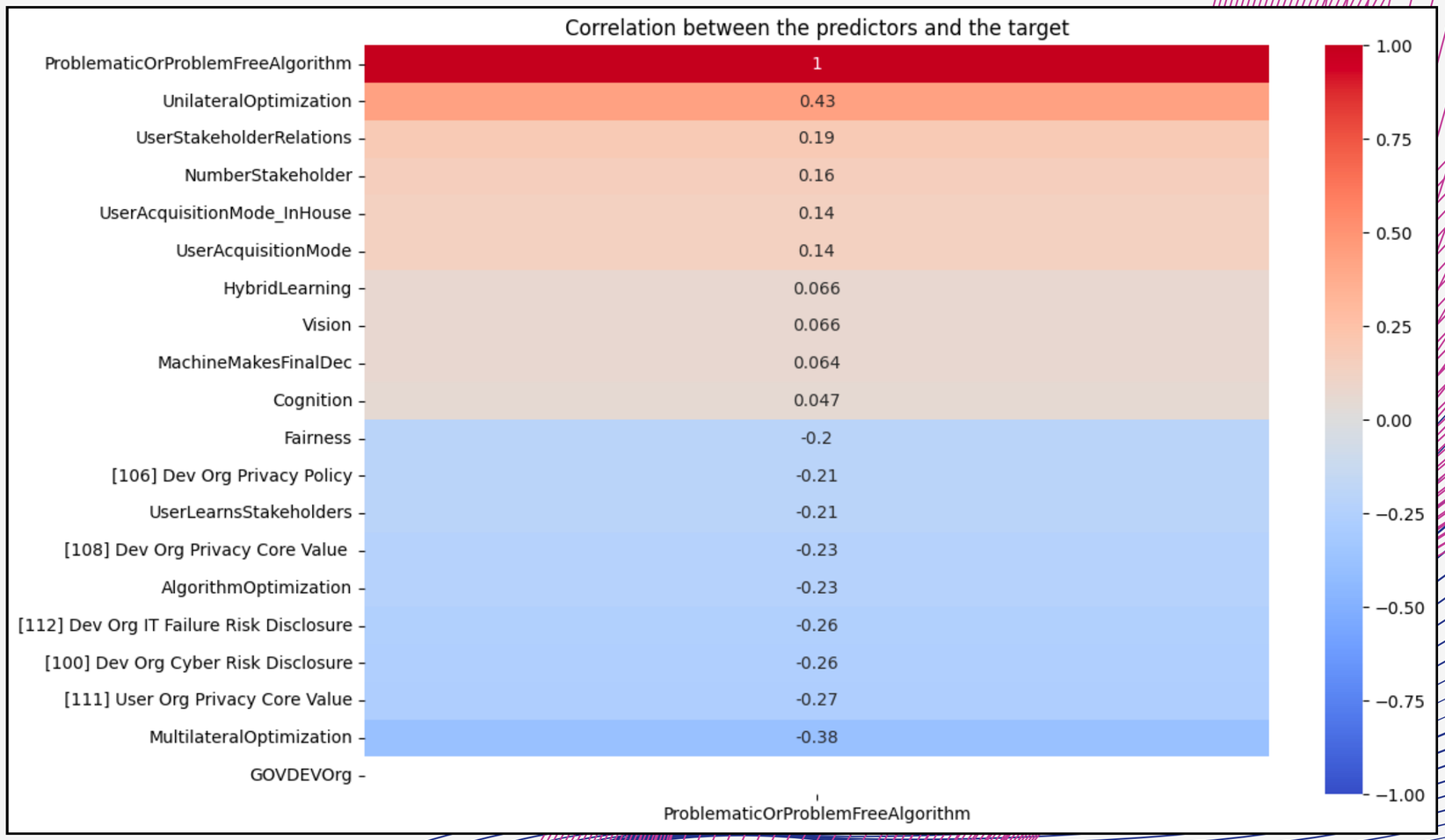
03

Random Forest Model

CORRELATION TO TARGET

These are the 10 most positive and negative correlations to the “ProblematicOrProblem FreeAlgorithm” variable.

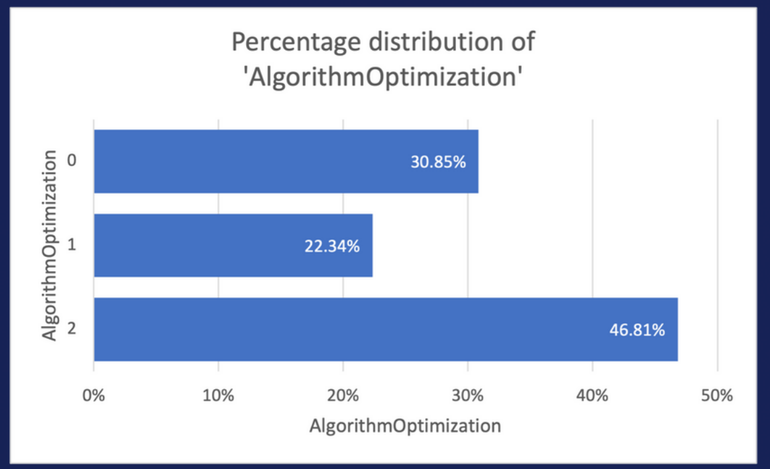
A negative correlation contributes more to having a problem-free algorithm, and vice versa.



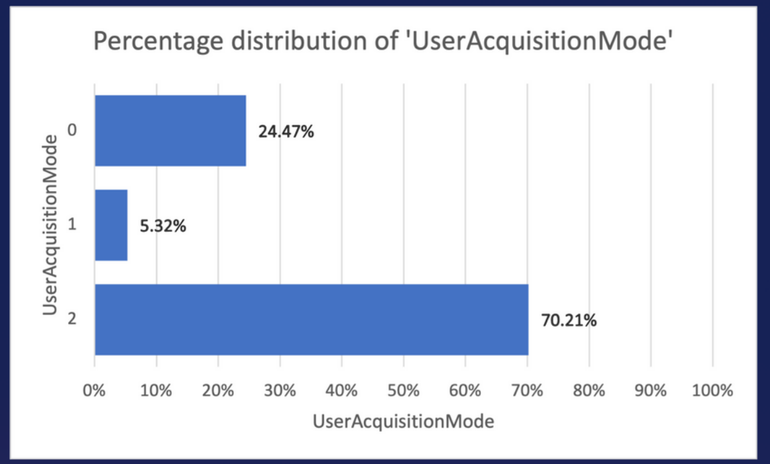
INFLUENTIAL FACTORS

Algorithms incorporating supervised machine learning, multilateral optimization, and inter-organizational collaboration are the least likely to present issues.

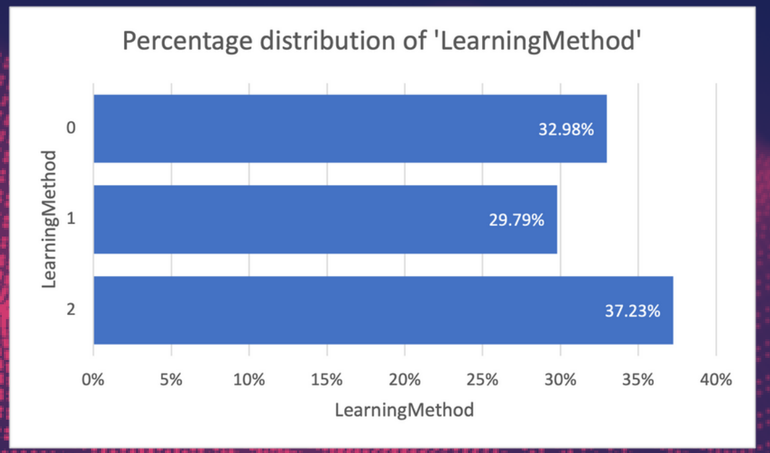
Among the firms observed:
29.79% implemented a supervised learning approach
46.81% engaged in multilateral optimization
5.32% employed collaborative acquisition.



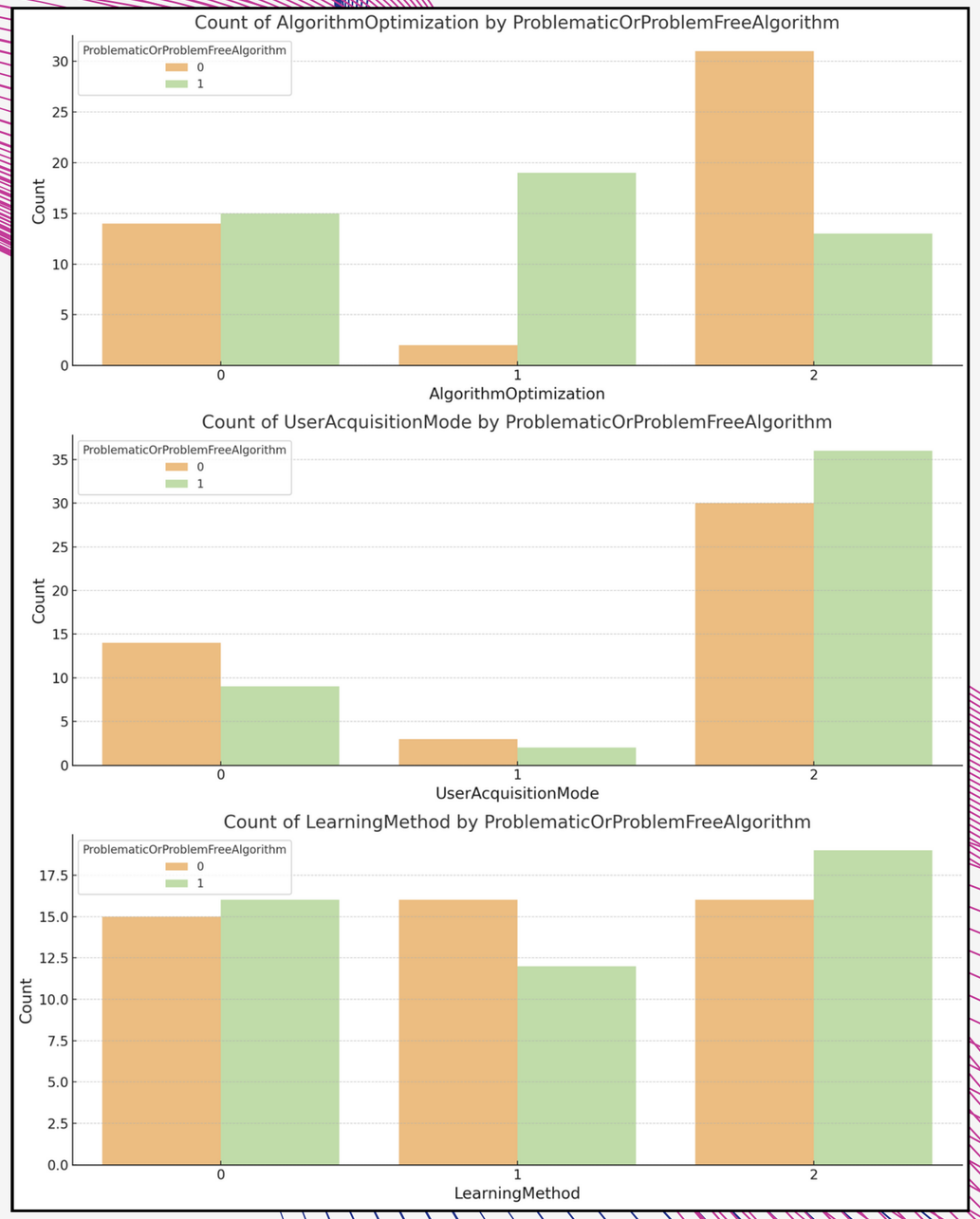
0 - No optimization
1 - Unilateral
2 - Multilateral



0 - Off the shelf
1 - Collaborative
2 - In-House



0 - Unsupervised
1 - Supervised
2 - Hybrid



INFLUENTIAL FACTORS TO TARGET VARIABLE

Orange: Problem-free
Green: Problematic

Algorithm Optimization:

- A **majority of the problematic algorithms used unilateral optimization**, while a majority of the problem-free algorithms used multilateral optimization

User Acquisition Mode:

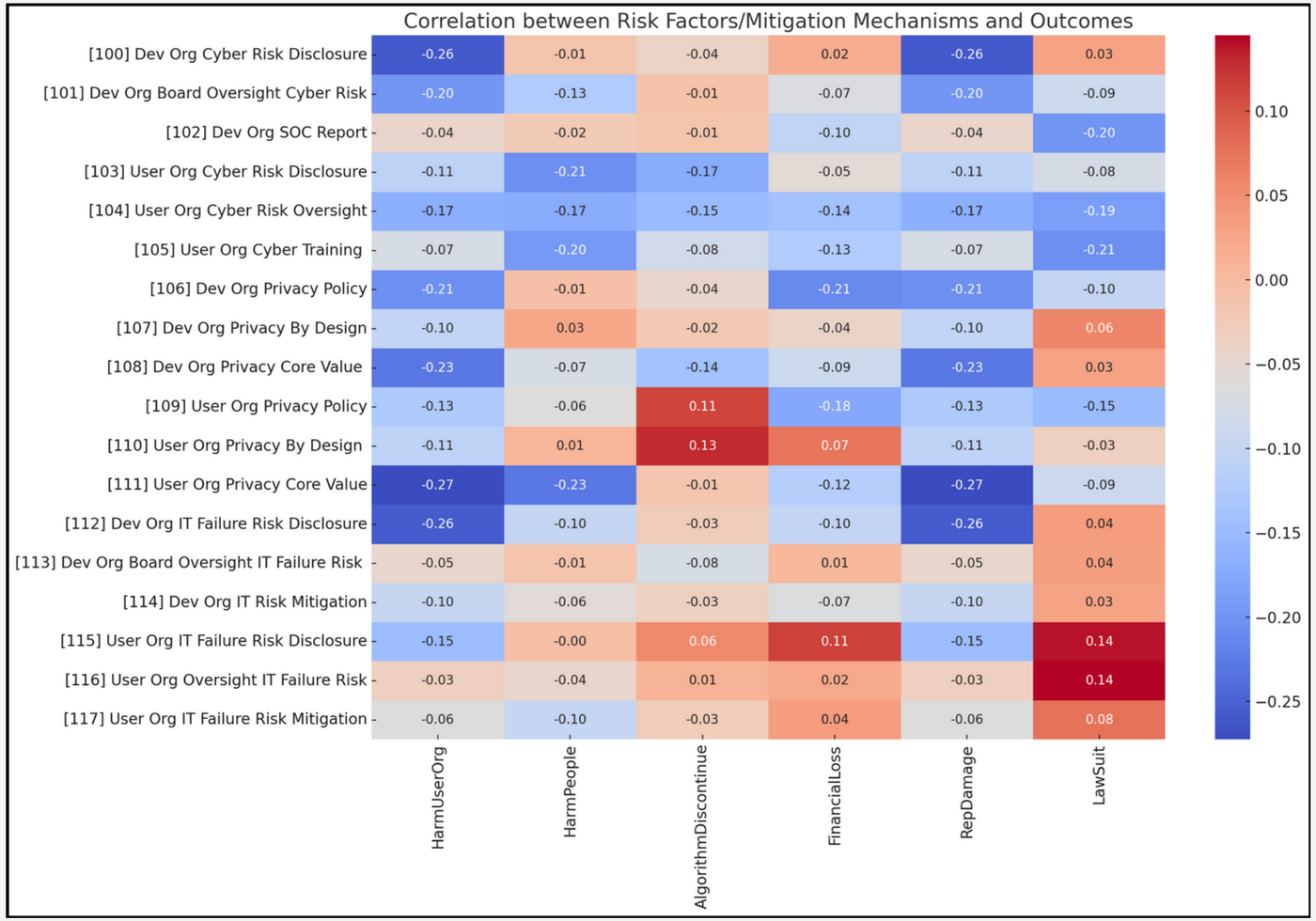
- A **majority** of both problem-free and problematic algorithms relied on **in-house acquisition**

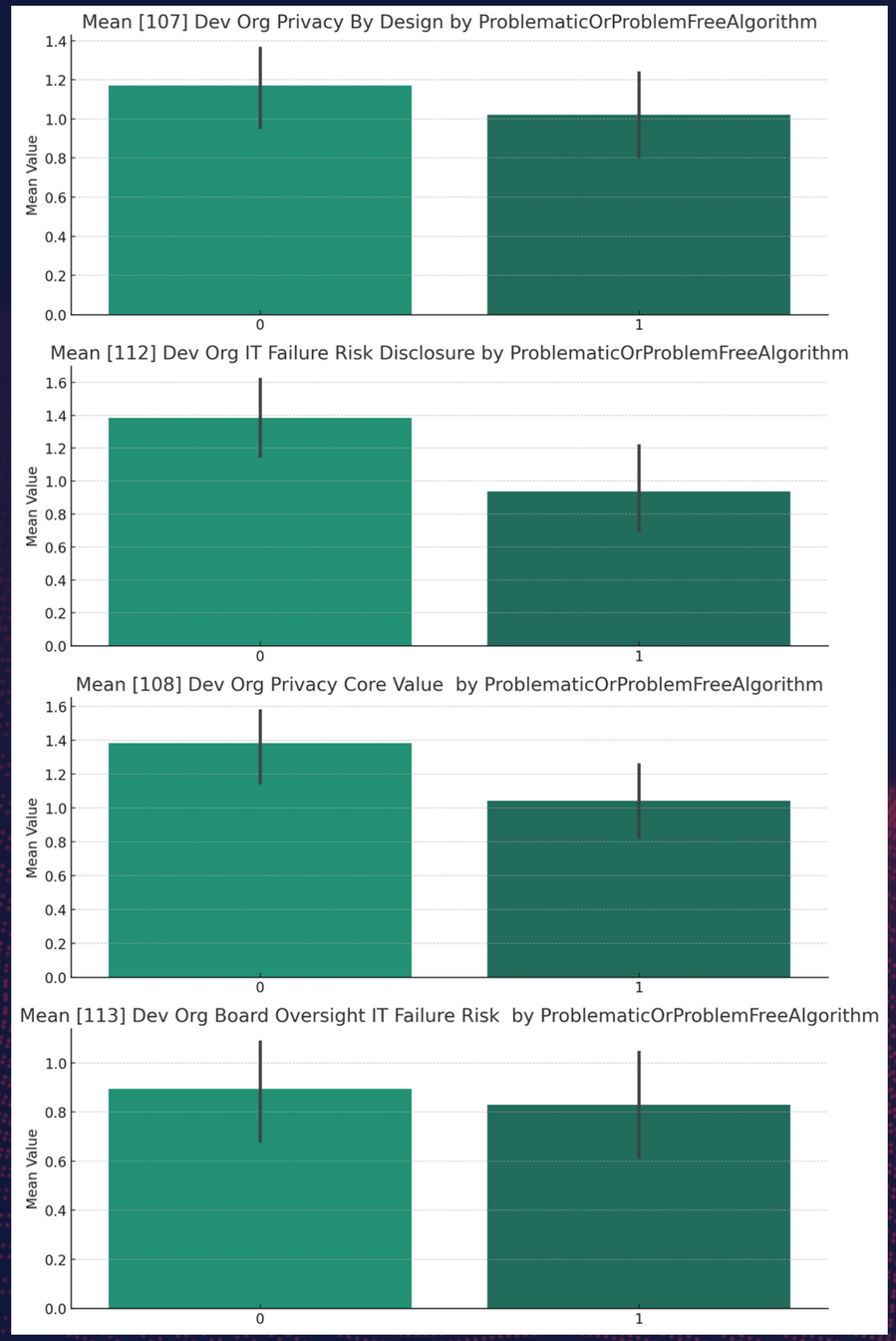
Learning Method:

- The learning methods of the **problem-free** algorithms are somewhat evenly spread, but **supervised learning** appears to have the most.
- Most of the **problematic** algorithms used **hybrid learning**.

CORRELATION TO OUTCOMES HEATMAP

Dark blue colors imply potential effectiveness in risk mitigation. Dark red colors suggest a potential increase in risk or impact associated with the factor.





MEAN VALUES

These variables had the lowest (yet not statistically significant) p-values from our logistic regression analysis.

Across all four variables, the problem-free algorithms exhibited higher means, indicating that incorporating substantial risk mitigation mechanisms reduces the likelihood of encountering issues.

MACHINE LEARNING MODEL METHODOLOGY

01

Cleaning and Splitting Data

Data leakage problem, use 80-20 split, and determine target variable

02

Fit Model on Training Set

Fine tuning parameters and check in-sample performance using cross-validation

03

Use Model on Testing Set

Check out-of-sample performance using AUC-ROC (area under ROC curve)

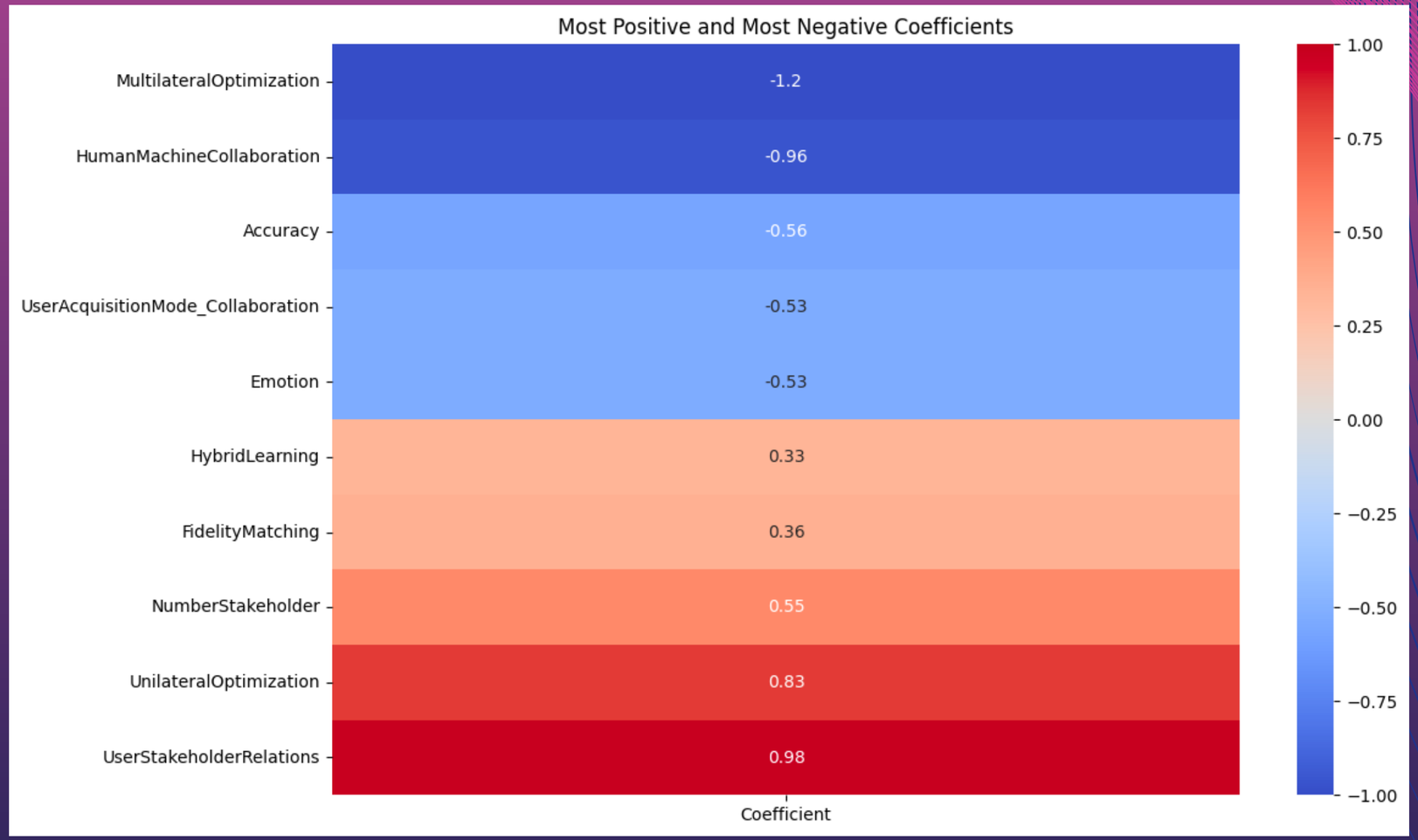
04

Evaluate the Models

Determine which ML model performs the best

LOGISTIC REGRESSION

	Coefficient
MultilateralOptimization	-1.159472
HumanMachineCollaboration	-0.962988
Accuracy	-0.562545
UserAcquisitionMode_Collaboration	-0.529096
Emotion	-0.526137
Fairness	-0.452300
[111] User Org Privacy Core Value	-0.432240
[112] Dev Org IT Failure Risk Disclosure	-0.383535
AlgorithmRepurposed	-0.318356
[100] Dev Org Cyber Risk Disclosure	-0.275138
MachineMakesFinalDec	-0.201714
Sensory	-0.167695
GOVUserOrg	-0.163223
MANUFUserOrg	-0.149773
Speech	-0.142254
SERVICEDEVOrg	-0.098080
OnPlatformOrOffPlatform	-0.090538
LearningMethod	-0.076610
[116] User Org Oversight IT Failure Risk	-0.054758
GOVDEVOrg	0.000000
[106] Dev Org Privacy Policy	0.046963
MANUFDEVOrg	0.098004
SERVICEUserOrg	0.312920
HybridLearning	0.333878
FidelityMatching	0.356372
NumberStakeholder	0.549299
UnilateralOptimization	0.832454
UserStakeholderRelations	0.984086

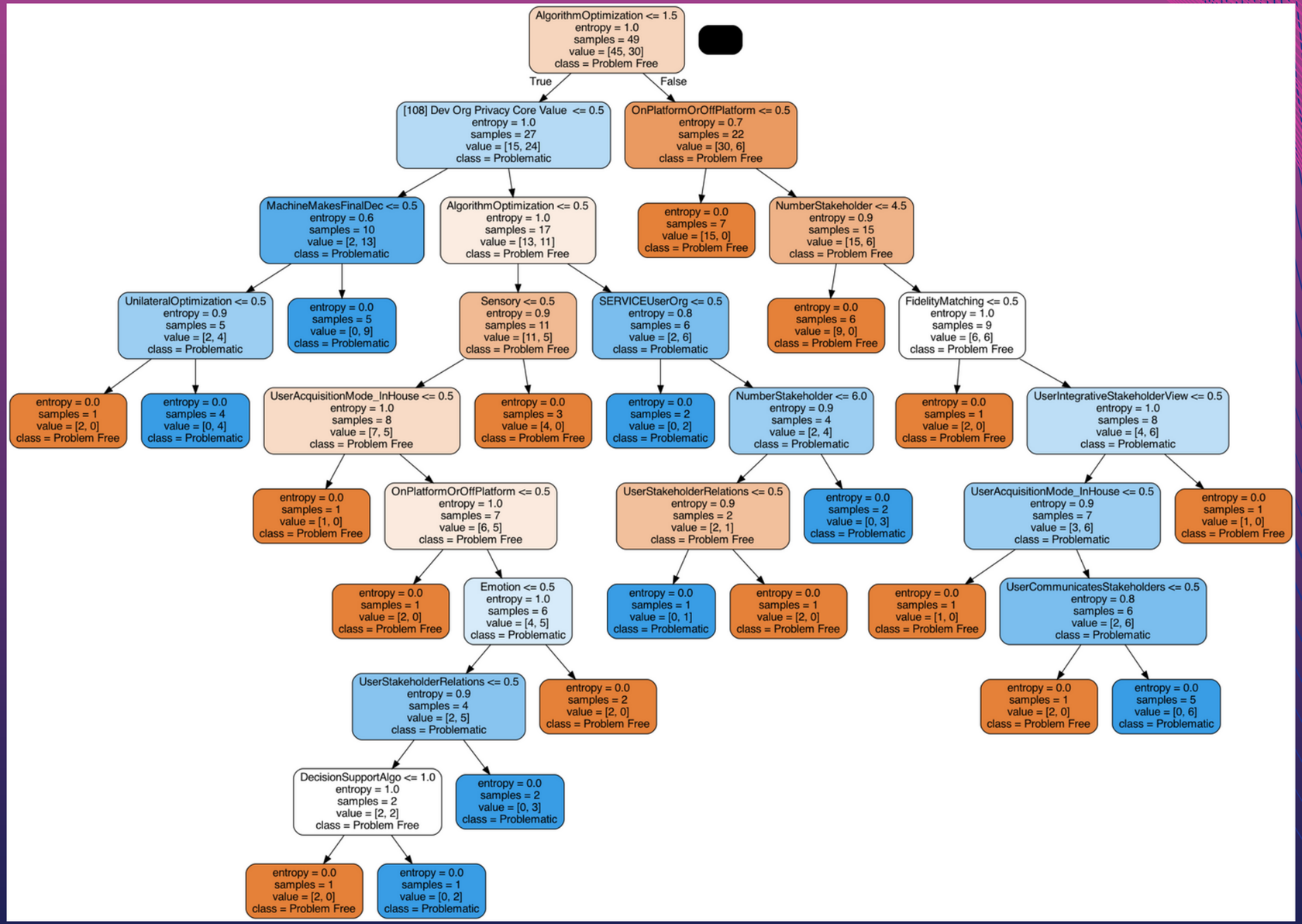


The model resulted in 85.5% out-of-sample accuracy.

RANDOM FOREST CLASSIFICATION

The higher up the tree, the more influential the variable is in making classification predictions.

The model resulted in 89% out-of-sample accuracy.



RECOMMENDATIONS

Multilateral Optimization

Multilateral optimization is the leading factor for having a problem-free algorithm. Aim to identify a range of solutions to reflect a balance between objectives.

Stakeholders

Organizations should **assess the scope of their stakeholders and adjust their strategy accordingly.** Emphasize stakeholder **values**, as **robust communication channels** and **learning** from stakeholders are less prone to produce problematic algorithms.

Follow Risk Mitigation

Disclose **cybersecurity, privacy, and IT risks**. Ensure proper **oversight** by following board frameworks such as NACD. Emphasize safeguarding **user privacy**. Implement **risk mitigation frameworks** such as COSO or NIST.

Type of Algorithm

Algorithms possessing **visual and cognitive** abilities, while **unsupervised** and developed **in-house**, are more sensitive to impact. Understand the type of algorithm in use and the risk levels.