WER2025

Specifying Intelligence: RE for the Next Generation of AI Systems

Prof. Marcos Kalinowski, PUC-Rio

WER 2025, Rio de Janeiro, Brazil

Marcos Kalinowski

kalinowski@inf.puc-rio.br
linkedin.com/in/kalinowski
@prof_kalinowski
www.inf.puc-rio.br/~kalinowski

Professor of Software Engineering at PUC-Rio

Research Interests: AI Engineering, Empirical Software Engineering, and Human Aspects in Software Engineering.

Coordinator of the ExACTa PUC-Rio Lab

R&D in Artificial Intelligence, AI Engineering, Software Engineering, and Human Computer Interaction.

Member of ACM, IEEE, ISERN and the Brazilian Computer Society

Context

Among the top-ranked universities and **leader in industry integration in Latin America**, according to Times Higher Eduction

First Brazilian graduate program in Computing (1967), assessed with the **highest research quality score** by the Brazilian Federal Government

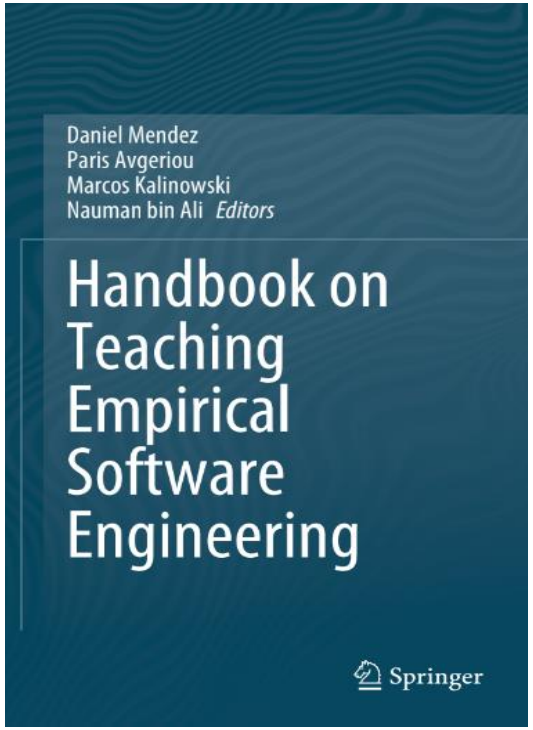
R&D in Artificial Intelligence, **AI Engineering**, Software Engineering, and Human-Computer Interaction

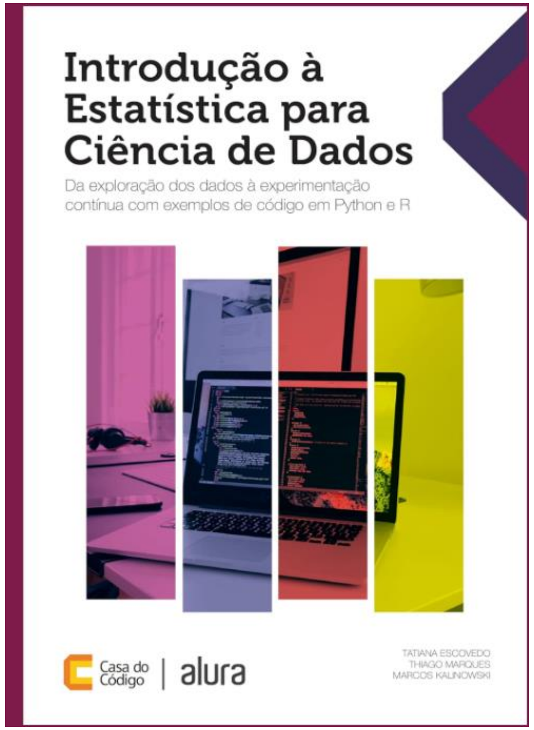
ExACTa PUC-Rio

R&D Partners (selection)

...

Recent Books





200+ scientific papers, with free author versions available at http://www.inf.puc-rio.br/~kalinowski

Agenda

1. AI Engineering: Challenges in RE
2. AI Engineering: Advancements in RE
3. RE in Multi-Paradigm AI Engineering: The Road Ahead

AI Engineering: Challenges in RE

Naming the Pain in ML-Enabled Systems Engineering

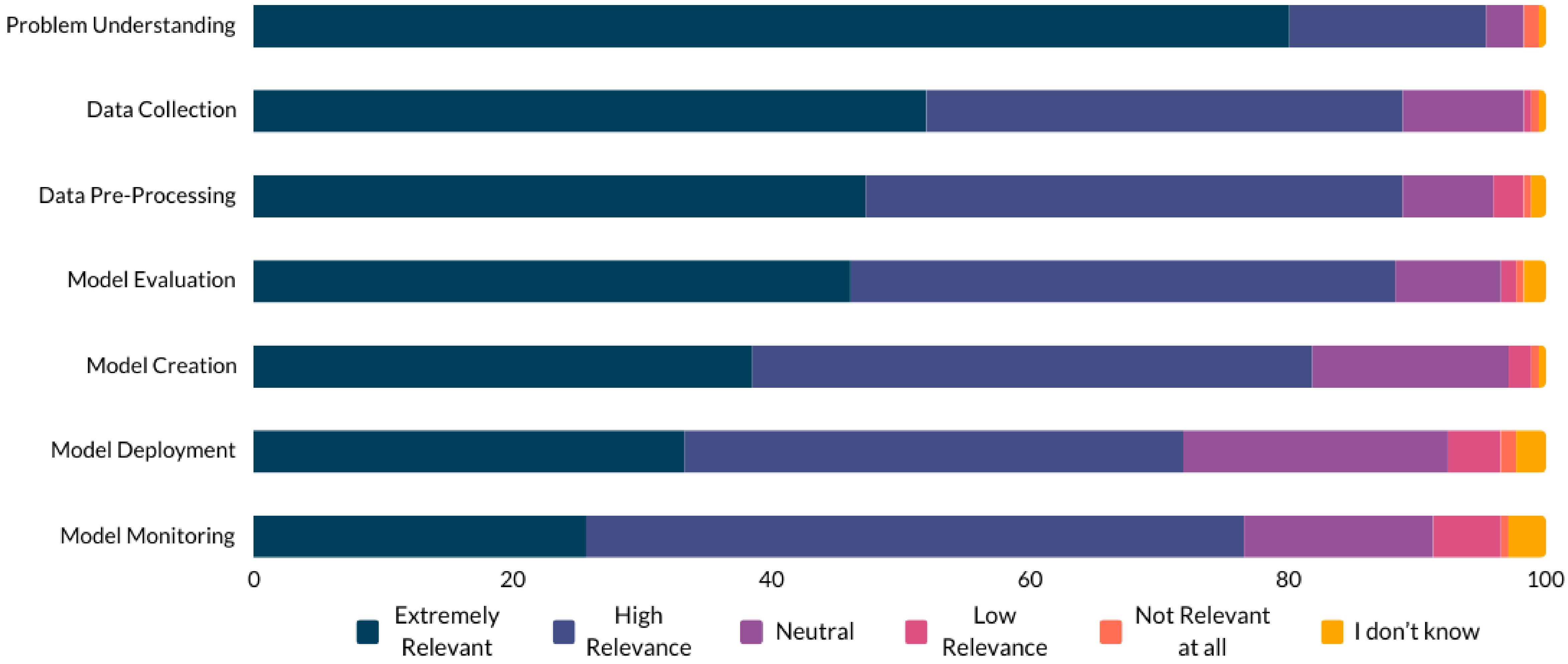
International Collaboration → Survey Naming the Pain in ML-Enabled Systems Development

Target Population: Professionals involved in building ML-enabled systems

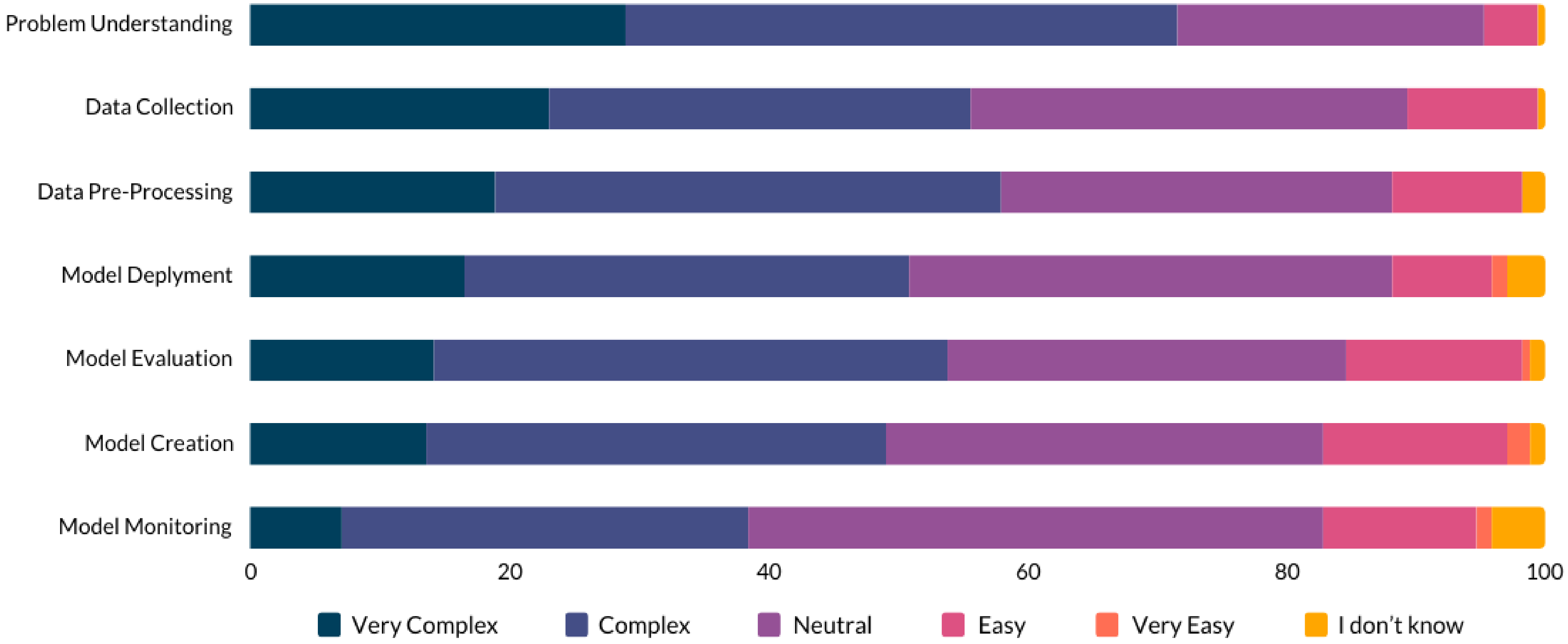
Total of Answers: 276 professionals (188 complete answers)

Kalinowski, M., Mendez, D., Giray, G. *et al.*, **Naming the Pain in Machine Learning-Enabled Systems Engineering**. Information and Software Technology, vol. 187, 2025.

Perceived Relevance of ML Lifecycle Stages



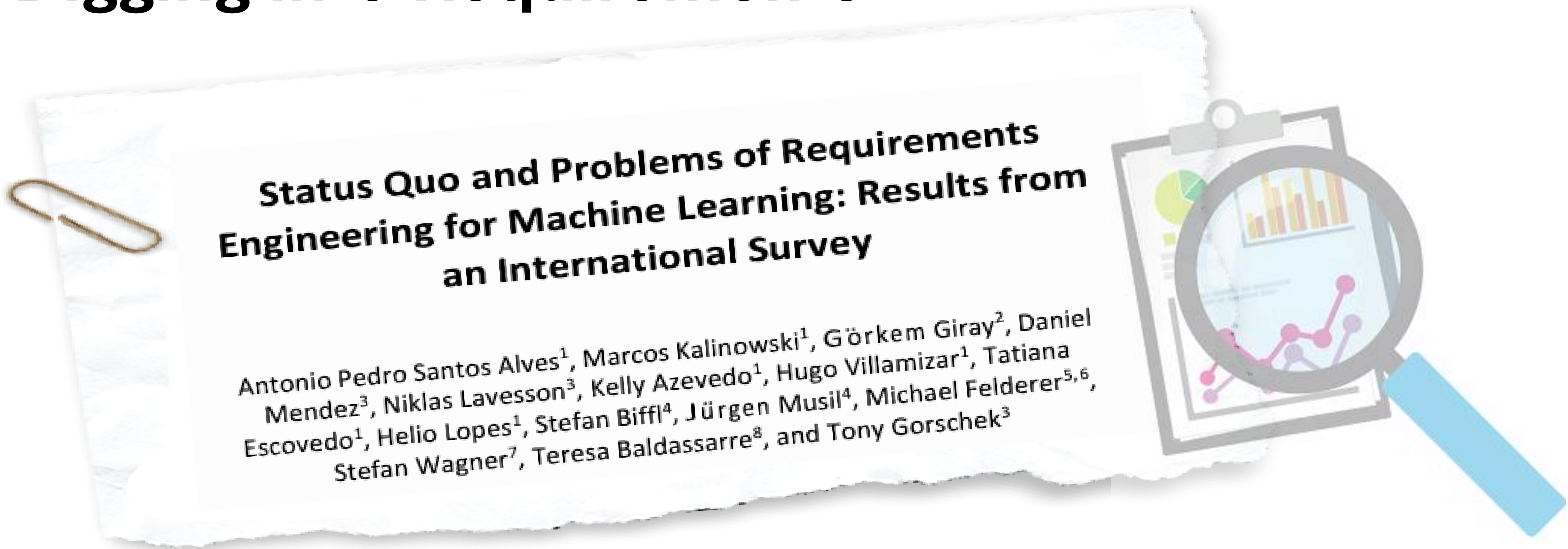
Perceived Complexity of ML Lifecycle Stages



The hardest single part of building a software system is deciding precisely what to build.

— Fred Brooks —

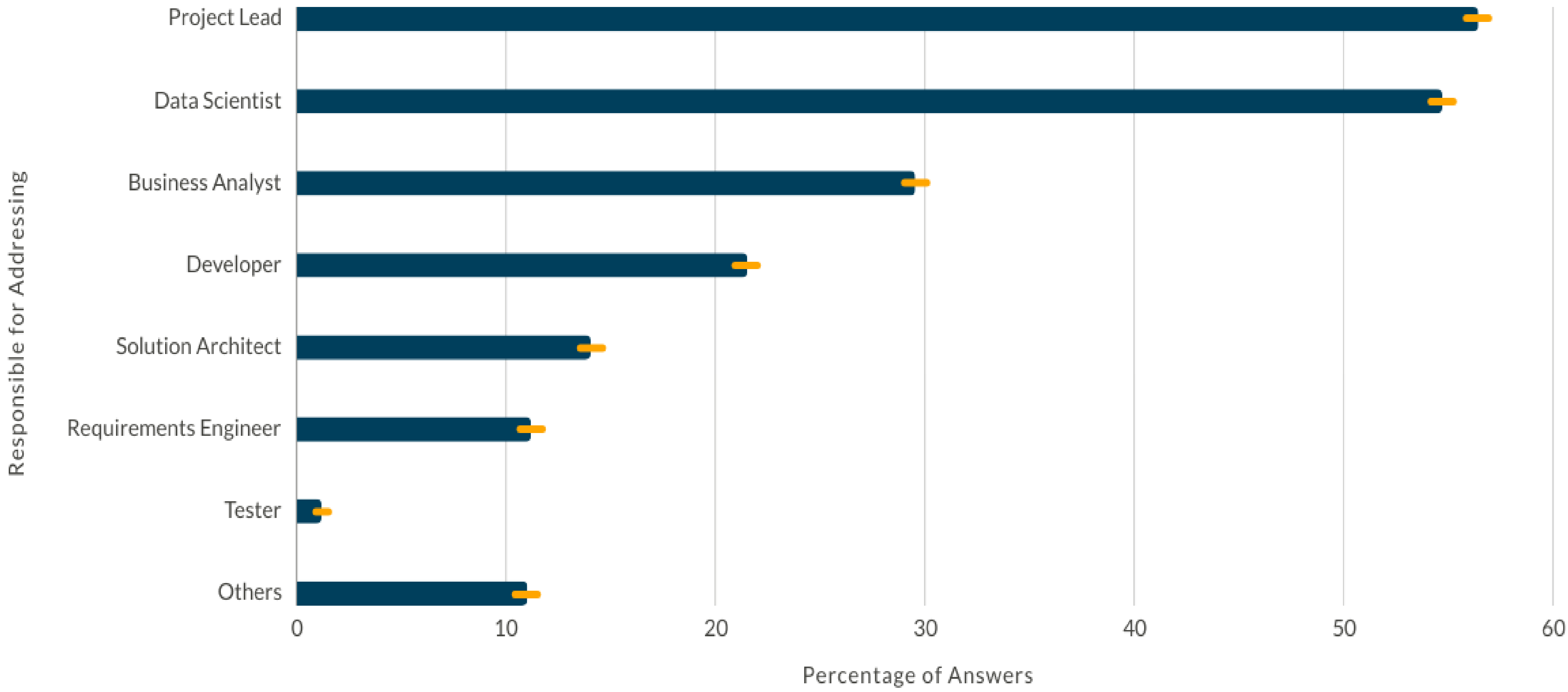
Digging into Requirements


Status Quo and Problems of Requirements Engineering for Machine Learning: Results from an International Survey

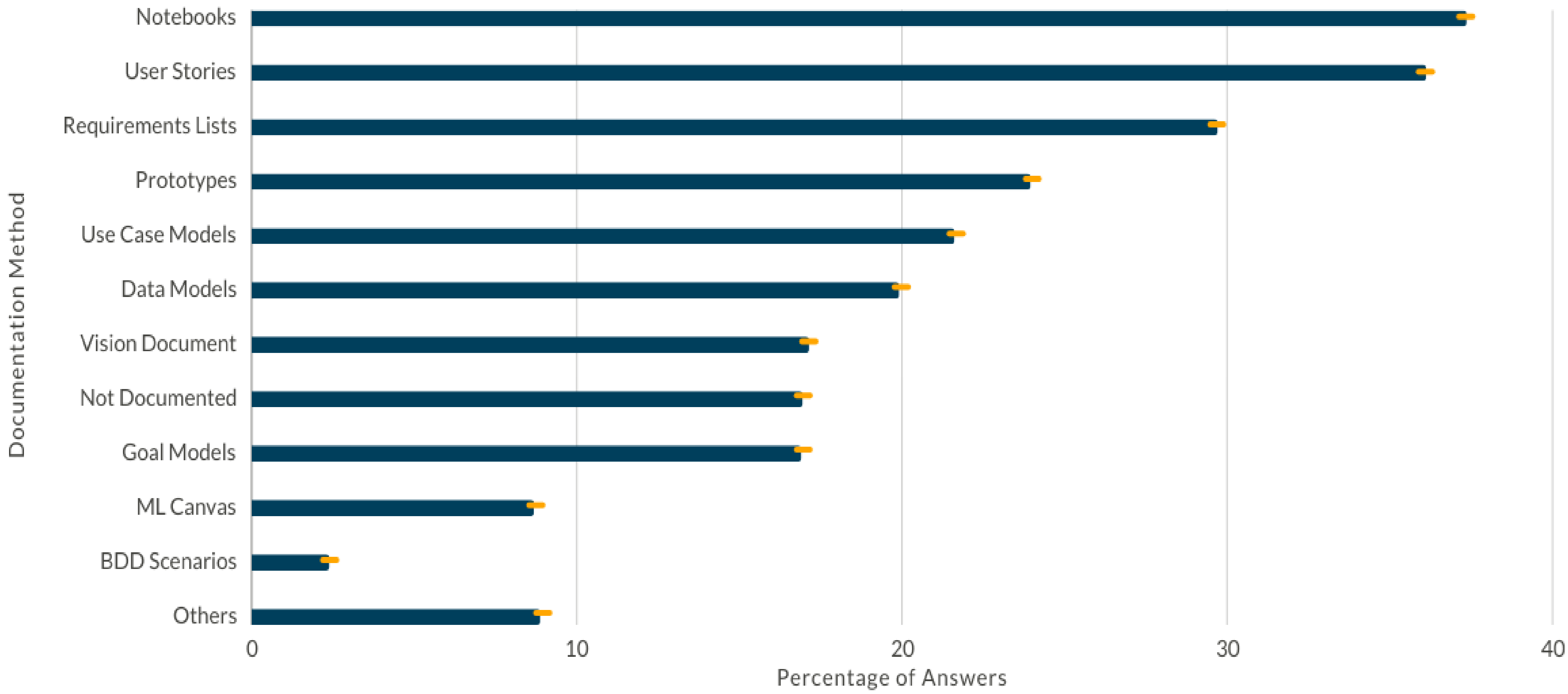
Antonio Pedro Santos Alves[1], Marcos Kalinowski[1], Görkem Giray[2], Daniel Mendez[3], Niklas Lavesson[3], Kelly Azevedo[1], Hugo Villamizar[1], Tatiana Escovedo[1], Helio Lopes[1], Stefan Biffl[4], Jürgen Musil[4], Michael Felderer[5,6], Stefan Wagner[7], Teresa Baldassarre[8], and Tony Gorschek[3]

Alves, A.P.S., Kalinowski, M., Giray, G., *et al.* **Status Quo and Problems of Requirements Engineering for Machine Learning: Results from an International Survey**. International Conference on Product-Focused Software Process Improvement (PROFES), pp. 159-174, 2023.

Who is Responsible for Requirements?

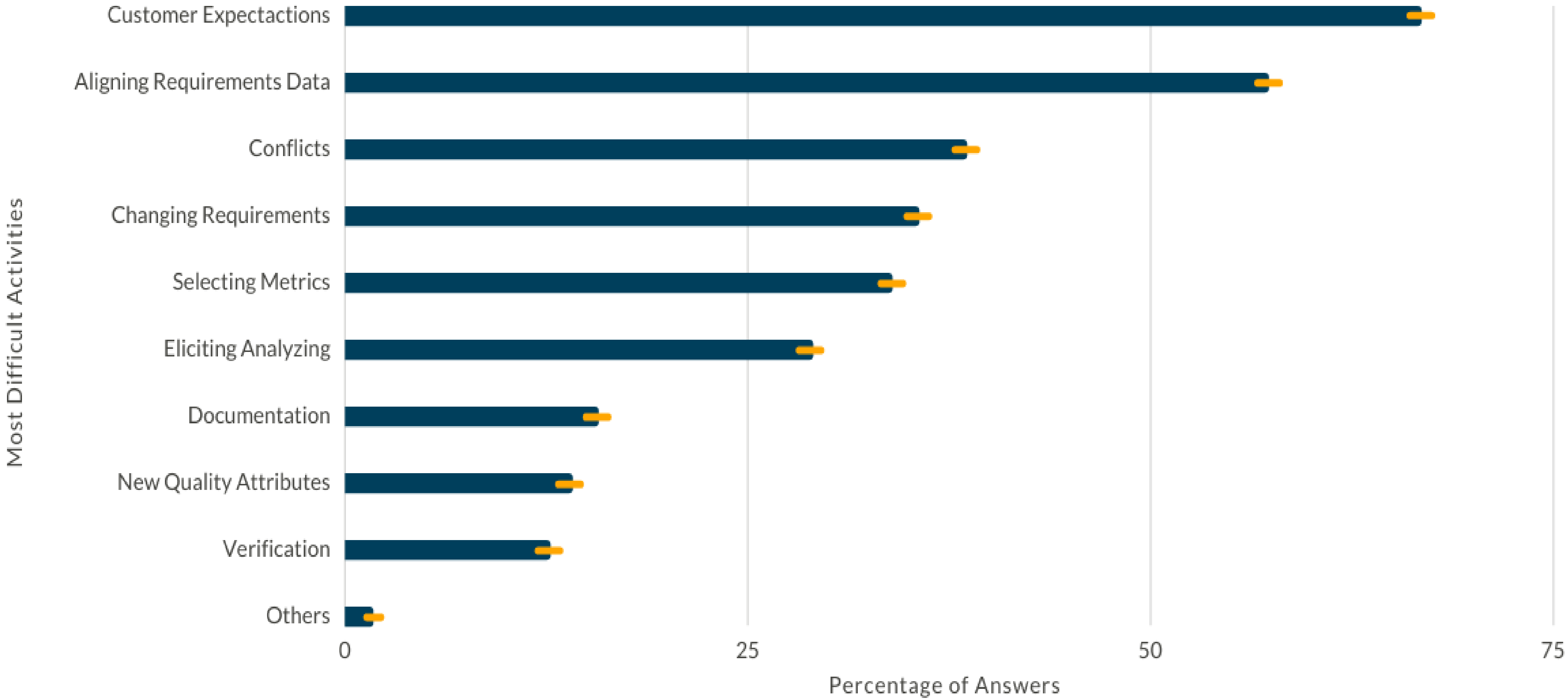


How are Requirements Documented?

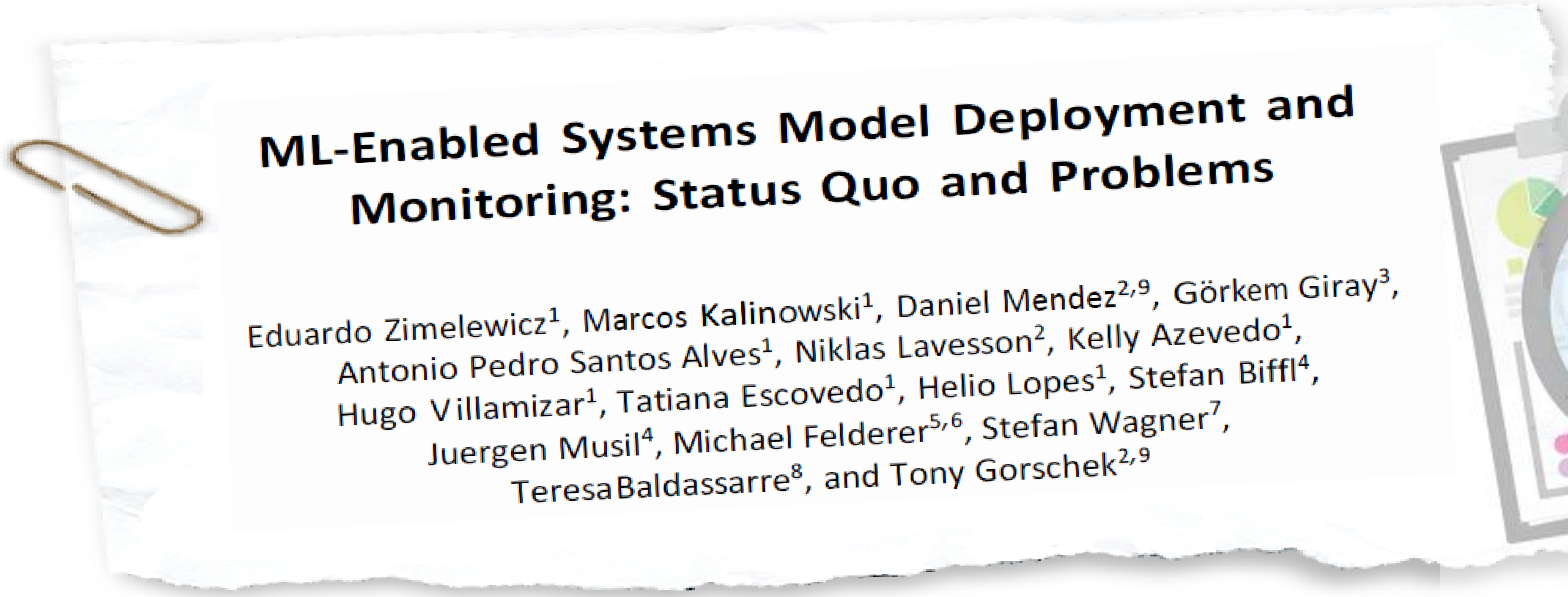



MIND THE GAP

Main Requirements Related Challenges



Digging into Model Deployment and Monitoring



ML-Enabled Systems Model Deployment and Monitoring: Status Quo and Problems

Eduardo Zimelewicz[1], Marcos Kalinowski[1], Daniel Mendez[2,9], Görkem Giray[3], Antonio Pedro Santos Alves[1], Niklas Lavesson[2], Kelly Azevedo[1], Hugo Villamizar[1], Tatiana Escovedo[1], Helio Lopes[1], Stefan Biffl[4], Juergen Musil[4], Michael Felderer[5,6], Stefan Wagner[7], Teresa Baldassarre[8], and Tony Gorschek[2,9]

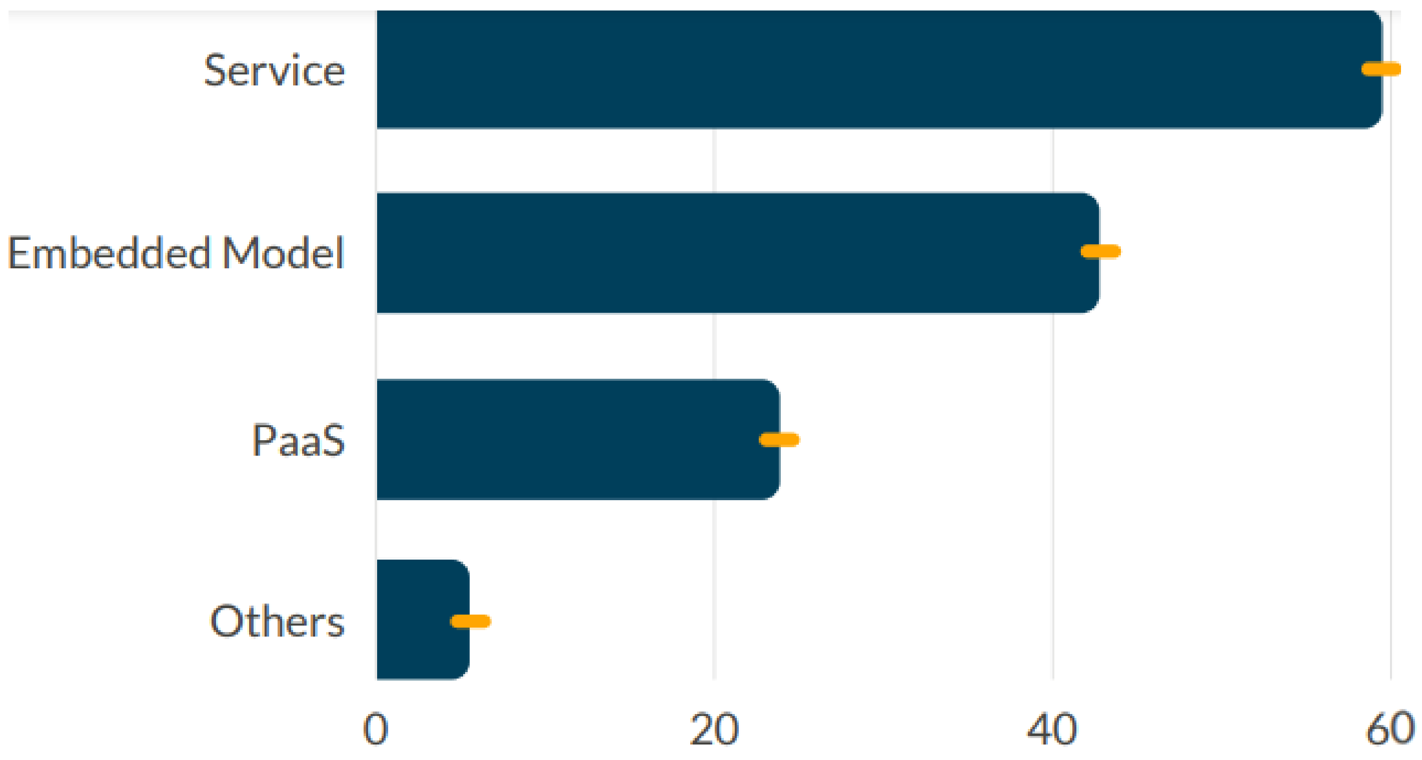
Zimelewicz, E., Kalinowski, M., Mendez, D., *et al.* **ML-Enabled Systems Model Deployment and Monitoring: Status Quo and Problems**. International Conference on Software Quality (SWQD), pp. 112-131, 2024.

How are Models being Deployed?

<u>Only 45%</u> of the projects made it to production!

Mostly deployed as a separate service.

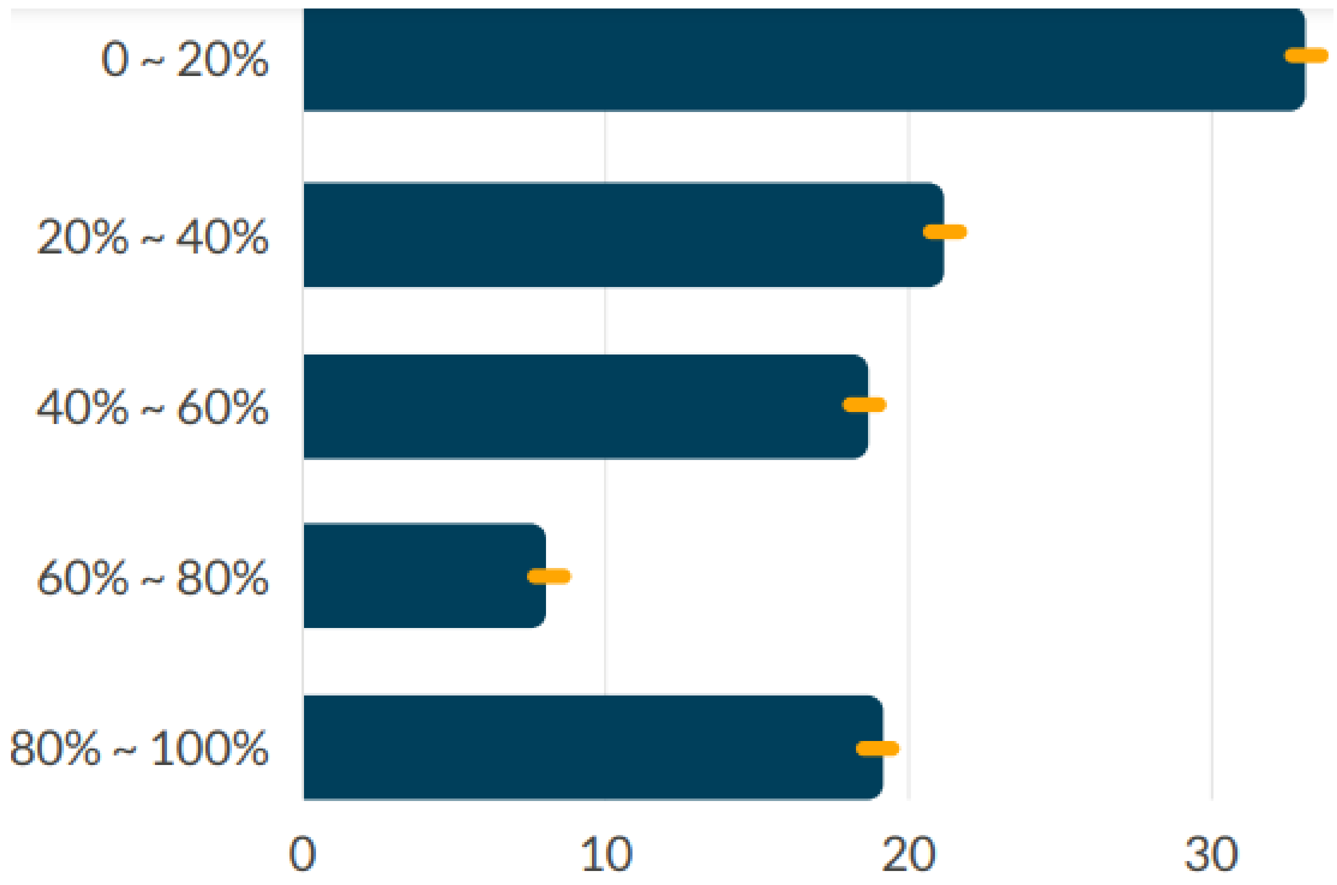
MLOps principles are rarely applied.



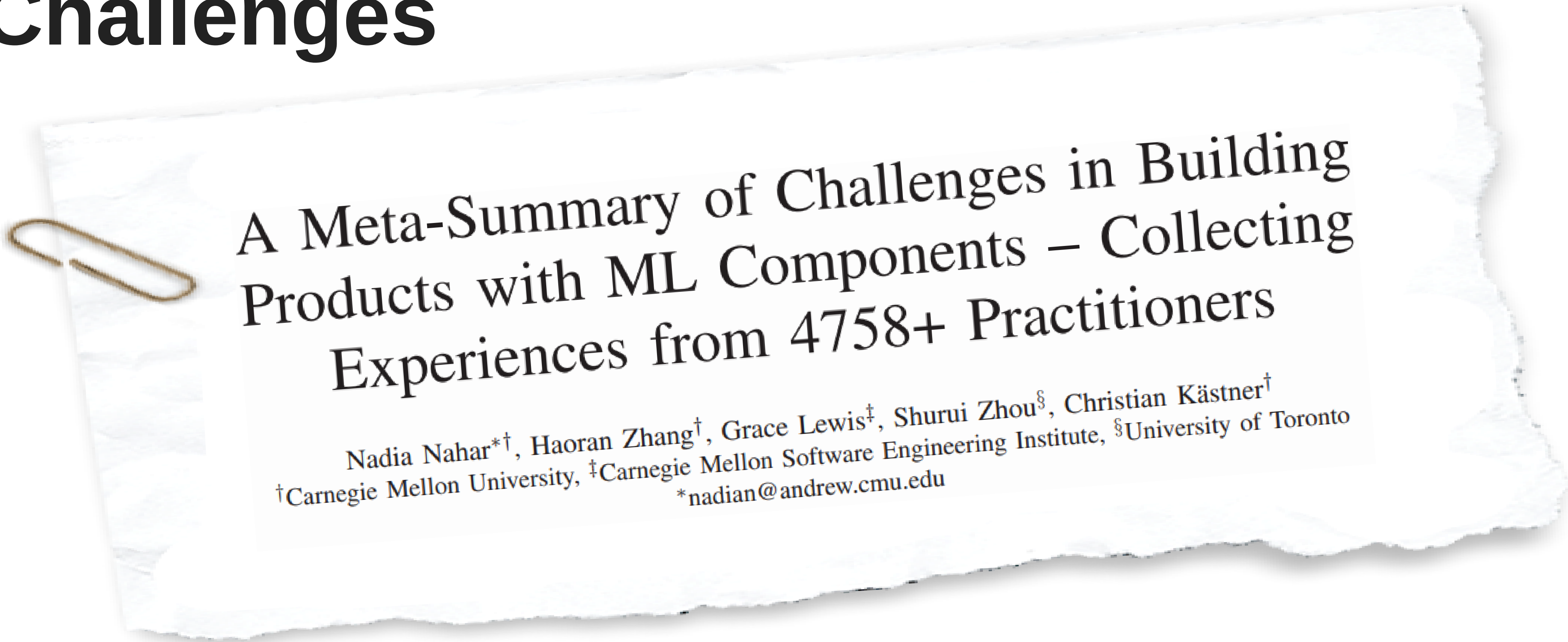
Are Models being Monitored in Production?

Most Deployed Models are not being Monitored!

Monitoring mostly only involves inputs and outputs!



Challenges



A Meta-Summary of Challenges in Building Products with ML Components – Collecting Experiences from 4758+ Practitioners

Nadia Nahar[*†], Haoran Zhang[†], Grace Lewis[‡], Shurui Zhou[§], Christian Kästner[†]
[†]Carnegie Mellon University, [‡]Carnegie Mellon Software Engineering Institute, [§]University of Toronto
[*]nadian@andrew.cmu.edu

Nahar, N., Zhang, H., Lewis, G., Zhou, S., and Kästner, C., **A meta-summary of challenges in building products with ml components – collecting experiences from 4758+ practitioners**. International Conference on AI Engineering–Software Engineering for AI (CAIN), pp. 171-183, 2023.

The good thing about science is that it's true whether or not you believe in it.

— Neil deGrasse Tyson —

We <u>need</u> to adapt and disseminate software engineering practices to enable better AI Engineering!

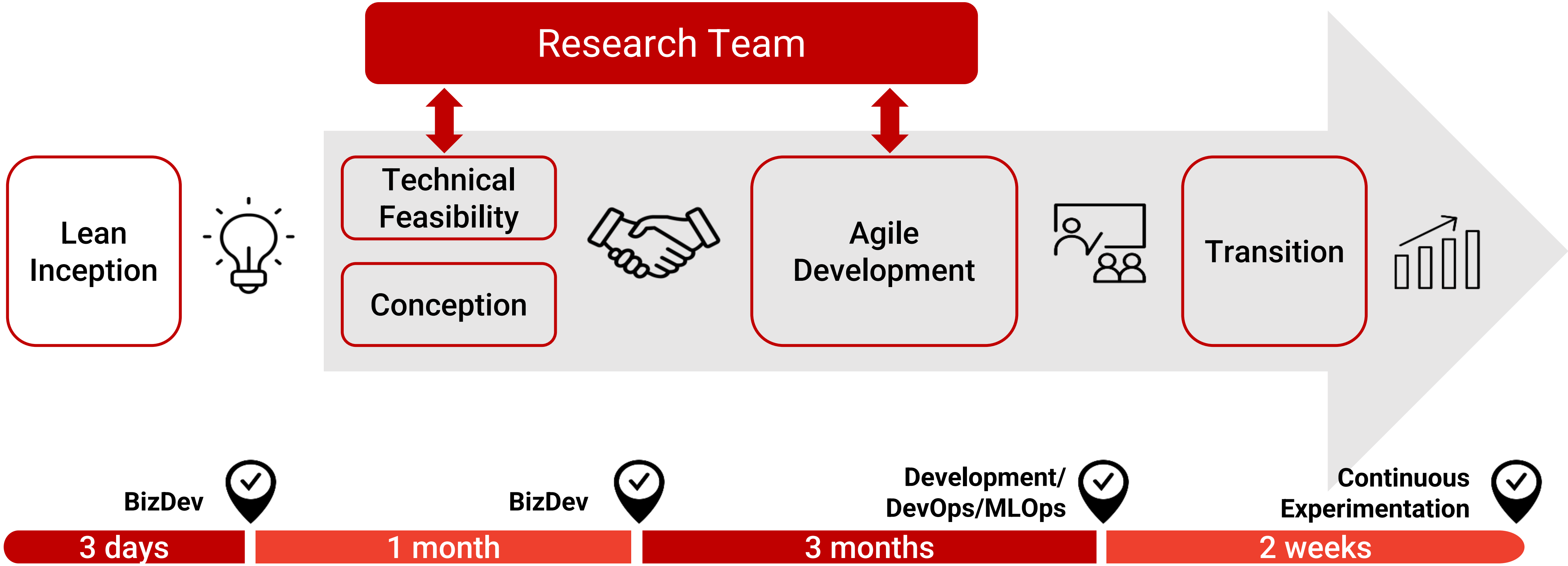
CAIN

“[…] the main challenge is not to develop the best models or algorithms but rather to provide support for the entire system life cycle […] **there is a clear need to advance the field of Software Engineering for AI.**”

AI Engineering: Advancements (in RE)

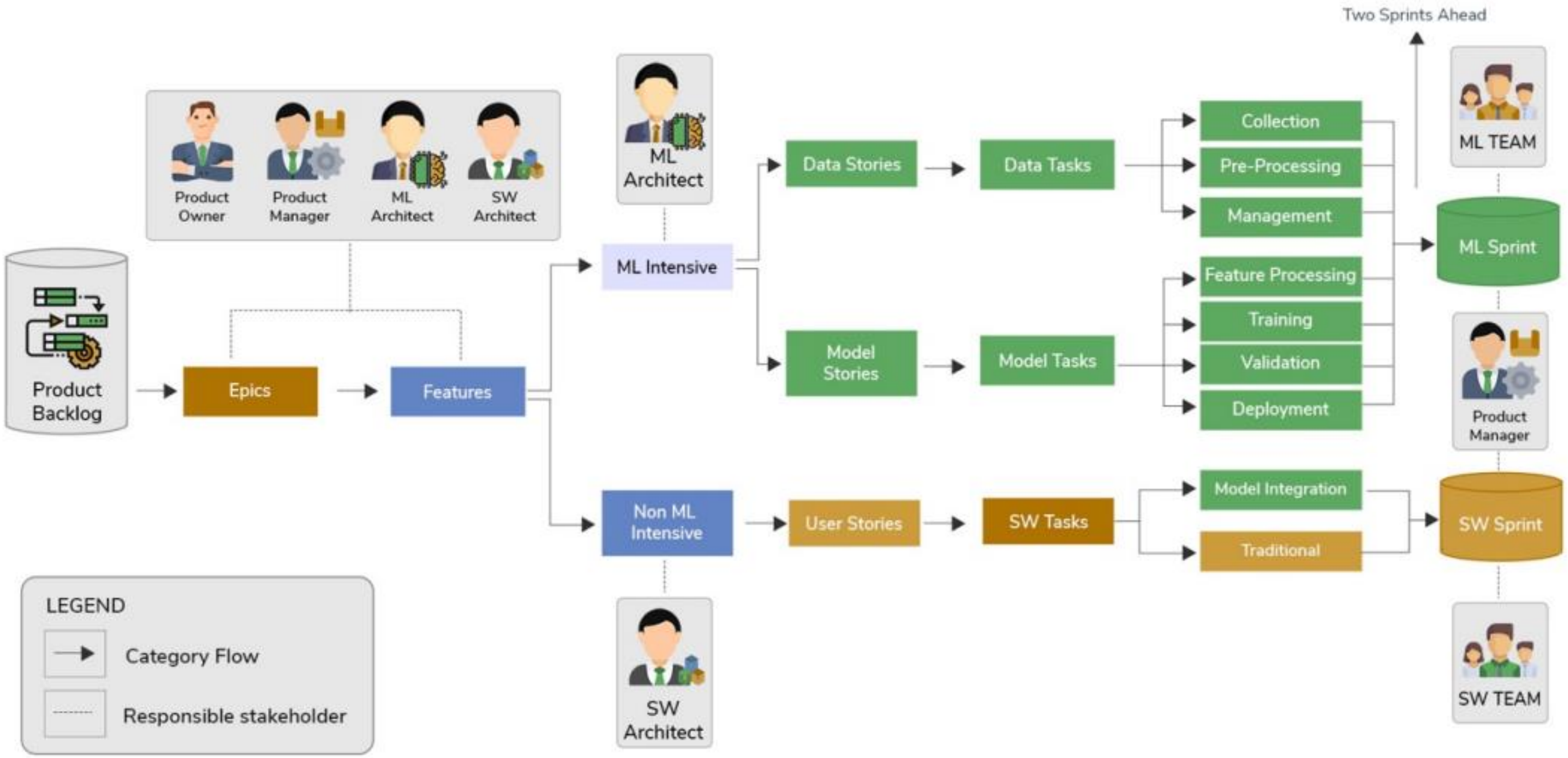
Approaches

Lean R&D



Kalinowski, M., Romao, L., Rodrigues, A., Barbosa, C., Villamizar, H., Barbosa, S.D. and Lopes, H. **Experiences Applying Lean R&D in Industry-Academia Collaboration Projects**. In *International Conference on Software Quality* (pp. 109-123). Springer, 2025.

Agile4MLS



Vaidhyanathan, K., Chandran, A., Muccini, H. and Roy, R. **Agile4MLS - Leveraging Agile Practices for Developing Machine Learning-Enabled Systems: An Industrial Experience**. IEEE Software, 39(6), 2022.

Agile Management of ML-Enabled Systems

There are at least 27 papers reporting approaches/practices/recommendations/challenges

Iteration Flexibility / Decoupled Ceremonies
- Capability-based iterations
- Flexible sprints for handling long experimentation cycles
- Asynchronous ML progress
- Sequential ML tasks with Agile's iterative flow

ML-Specific Artifacts
- Model/Data Stories
- Ethical User Stories

Demo API & MVM
- Deliver lightweight, testable model versions
- Maintain continuous value delivery

Business Alignment
- Collaborative planning to reduce wasteful ML experiments

Ethical Backlog Management
- Embedding fairness and transparency considerations

...

Romao, L., Oliveira, R., Alonso, S., and Kalinowski, M. **Agile Management for Machine Learning: A Systematic Mapping Study**. Euromicro Conference on Software Engineering and Advanced Applications (SEAA), 2025.

Polaris: A Framework for Trustworthy AI

Systematic literature review of the state of the practice (EASE 2023)

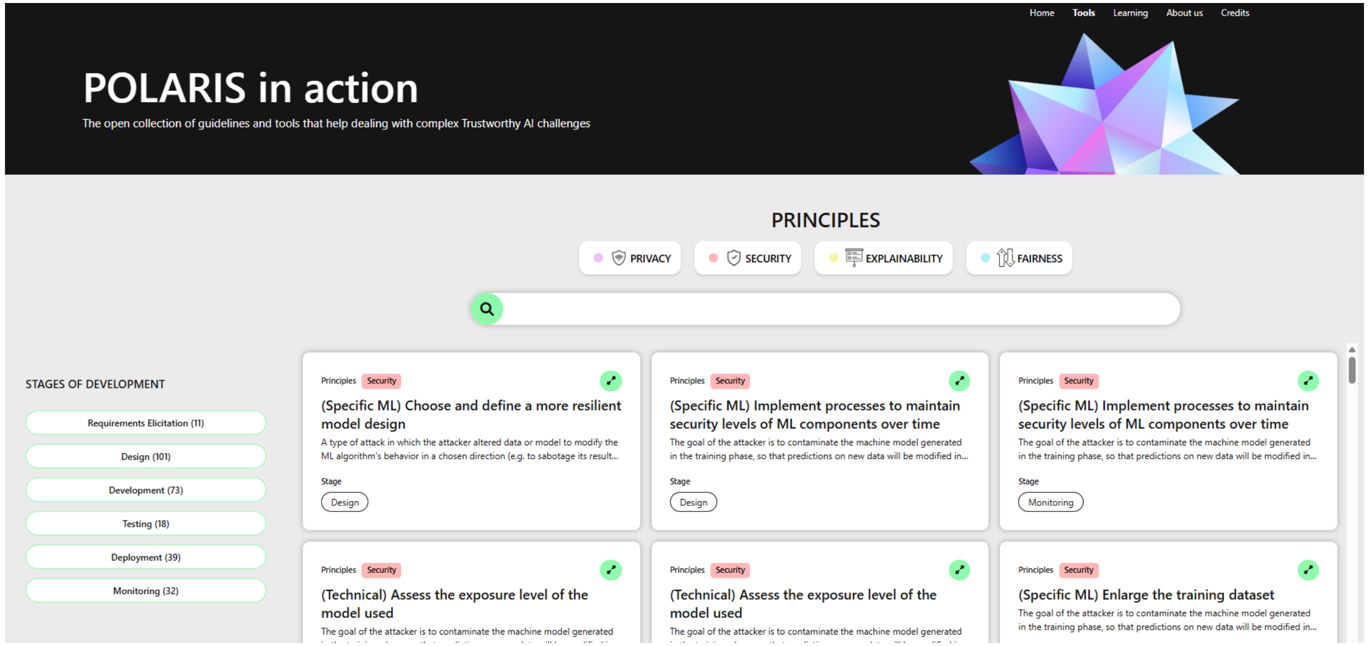
Survey with AI practitioners (EASE 2024)

POLARIS: A framework for Trustworthy AI
with actionable guidelines & tools (CAIN 2024)

Baldassarre, M.T., Gigante, D., Kalinowski, M. and Ragone, A., **POLARIS: A Framework to Guide the Development of Trustworthy AI Systems**. International Conference on AI Engineering – Software Engineering for AI (CAIN), 2024.

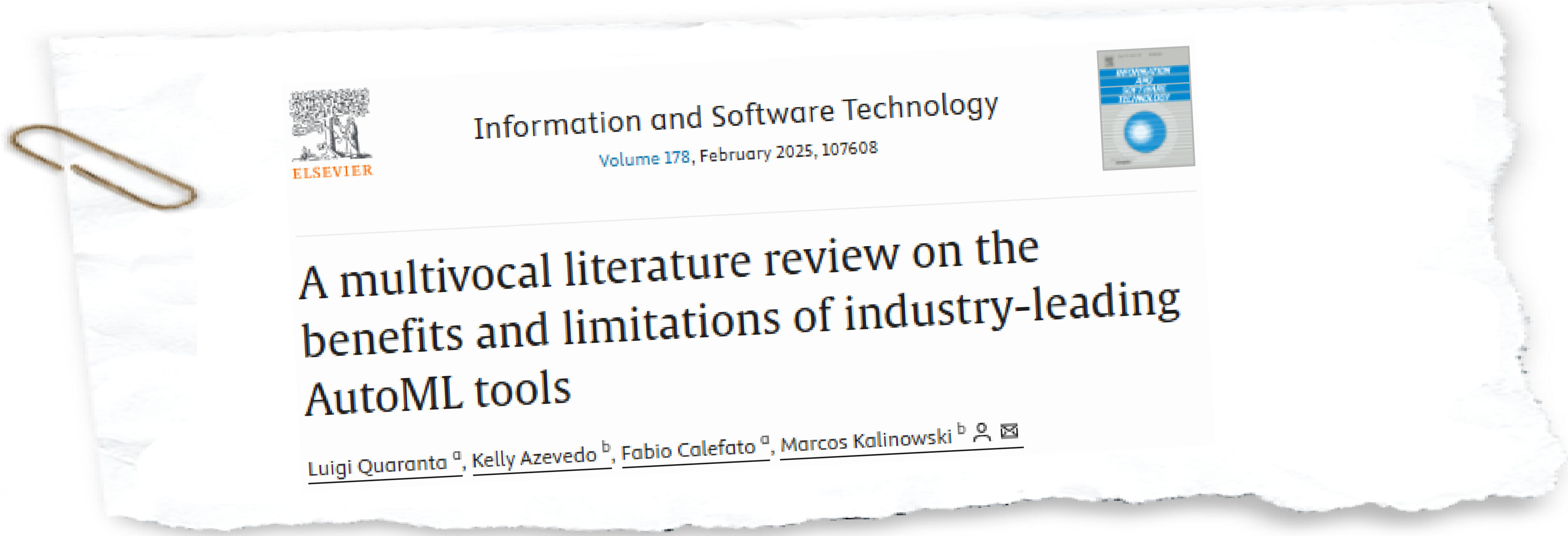
Polaris: A Framework for Trustworthy AI

https://polaris-app-5cc69.web.app/



Baldassarre, M.T., Gigante, D., Kalinowski, M. and Ragone, A., **POLARIS: A Framework to Guide the Development of Trustworthy AI Systems**. International Conference on AI Engineering – Software Engineering for AI (CAIN), 2024.

What about AutoML?


Information and Software Technology

Volume 178, February 2025, 107608

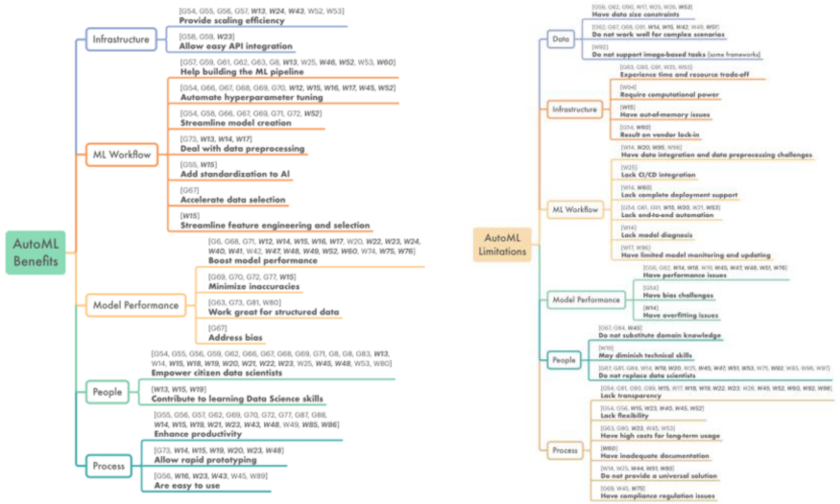
ELSEVIER

A multivocal literature review on the benefits and limitations of industry-leading AutoML tools

Luigi Quaranta [a], Kelly Azevedo [b], Fabio Calefato [a], Marcos Kalinowski [b]

Quaranta, L., Azevedo, K., Calefato, F. and Kalinowski, M., **A multivocal literature review on the benefits and limitations of industry-leading AutoML tools**. *Information and Software Technology*, vol, 178, 2025.

AutoML

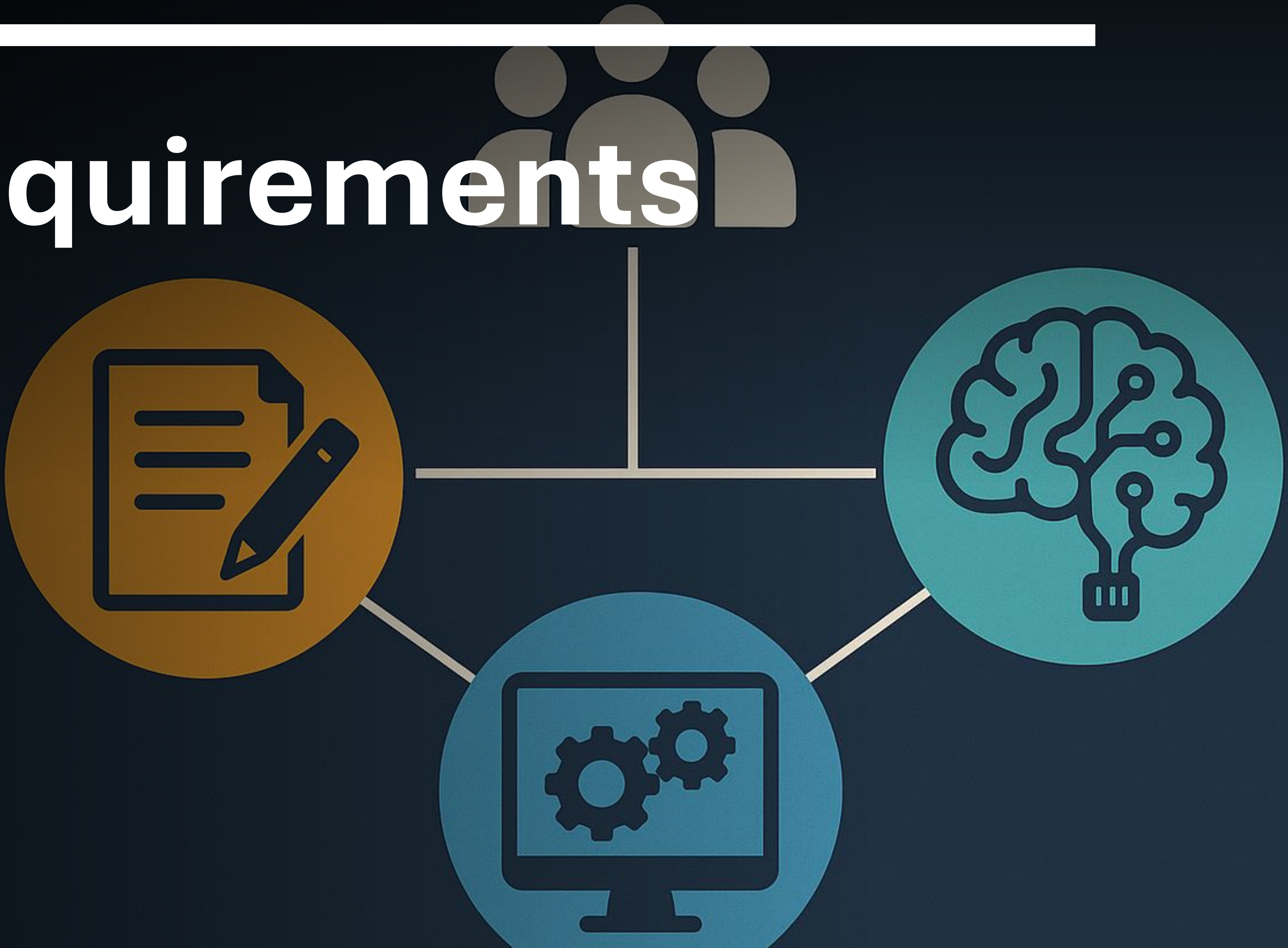


AutoML mainly supports technical feasibility assessments, but requirements are still key

AutoML tools can enhance the efficiency of machine learning by **complementing human expertise**

Skilled users are essential for effectively utilizing AutoML tools and applying their knowledge and insights

Requirements

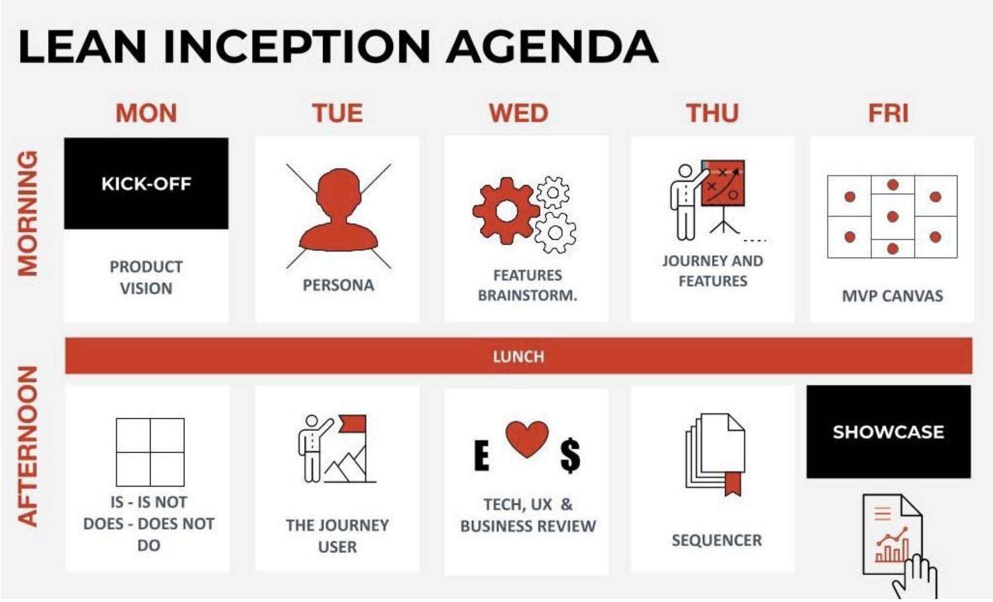


Ideation of ML-Enabled Systems

What about data?

Managing customer expectations?

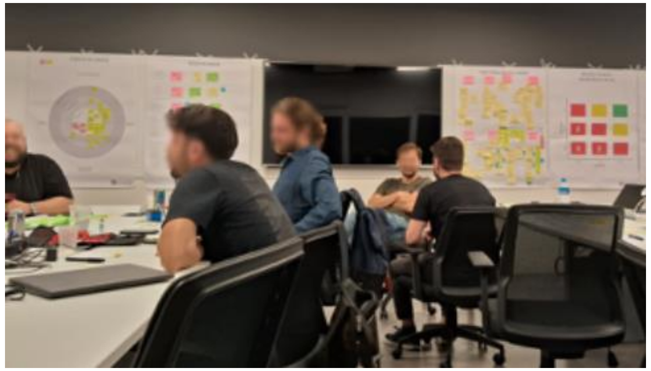
Aligning requirements with data?

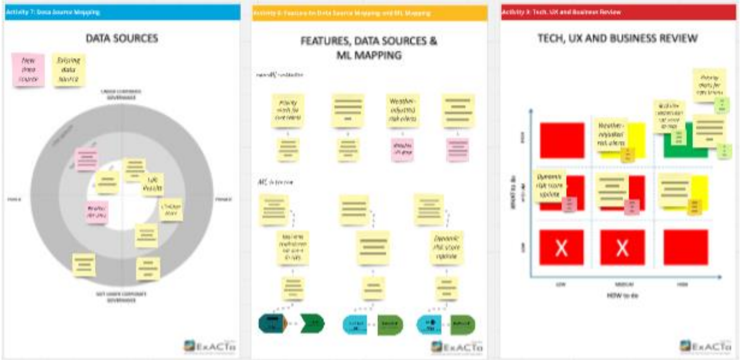


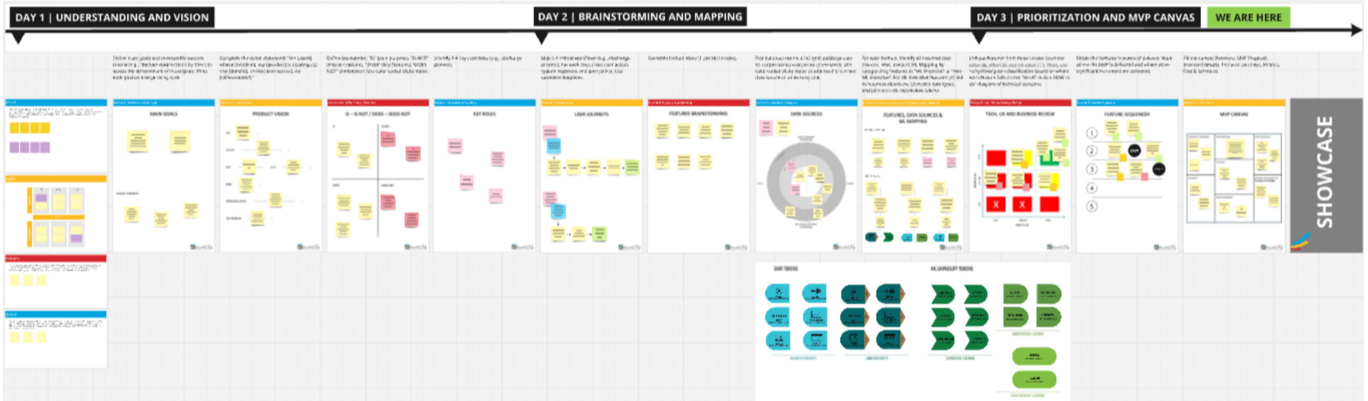
Define-ML

Extends Lean Inception with targeted activities: **Data Source Mapping**, **Feature-to-Data Source Mapping**, and **ML Mapping**

Helps to ground ideation in real data constraints, clarify ML feasibility, and align proposed features with business objectives



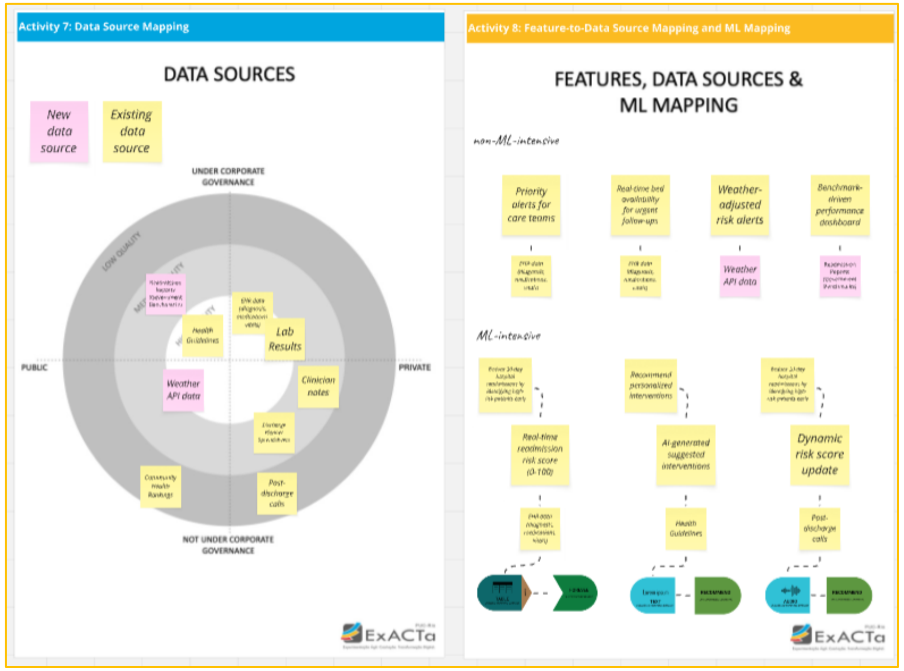


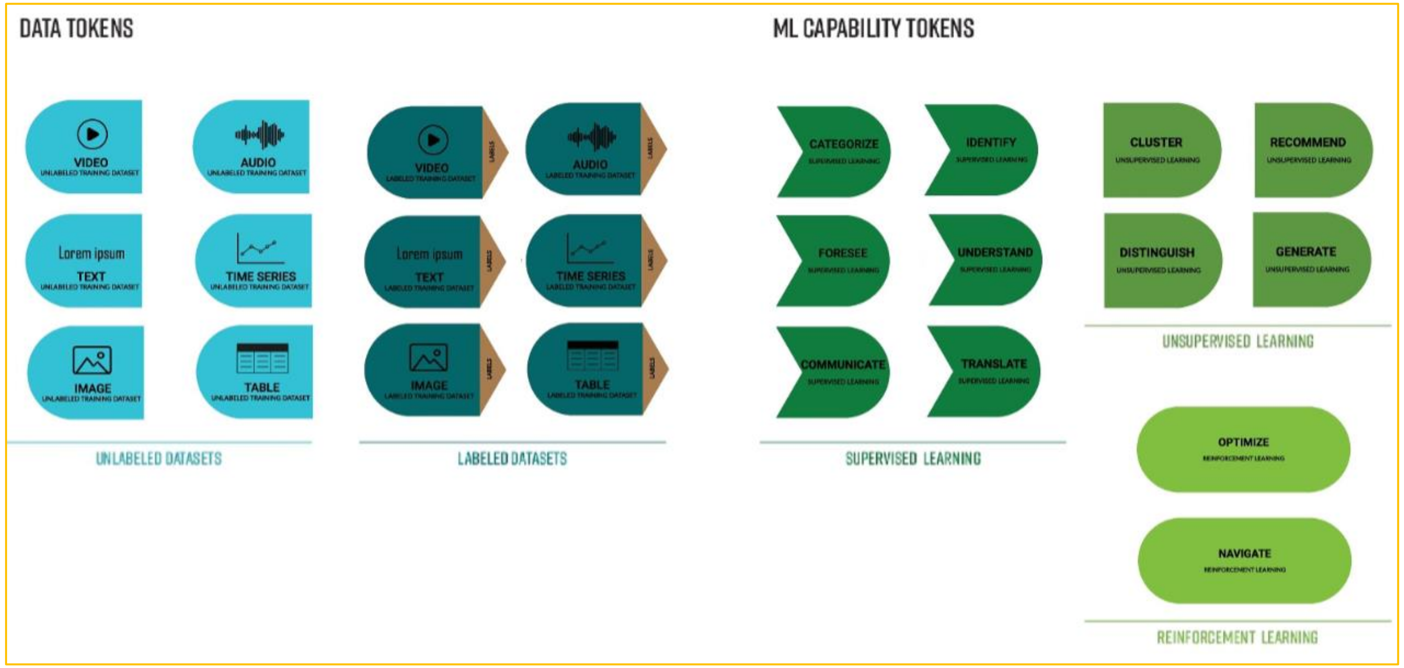


https://miro.com/miroverse/defineml-template/

Alonso, S., Alves, A.P.S., Romao, L., Lope, H., and Kalinowski, M., **Define-ML: An Approach to Ideate Machine Learning-Enabled Systems**, Euromicro Conference on Software Engineering and Advanced Applications (SEAA), 2025.

Define-ML





https://miro.com/miroverse/defineml-template/

Alonso, S., Alves, A.P.S., Romao, L., Lope, H., and Kalinowski, M., **Define-ML: An Approach to Ideate Machine Learning-Enabled Systems**, Euromicro Conference on Software Engineering and Advanced Applications (SEAA), 2025.

Requirements of ML-Enabled Systems

ELSEVIER

Journal of Systems and Software

Volume 213, July 2024, 112053

Identifying concerns when specifying machine learning-enabled systems: A perspective-based approach ☆

Hugo Villamizar[a], Marcos Kalinowski[a], Hélio Lopes[a], Daniel Mendez[b]

Villamizar, H., Kalinowski, M., Lopes, H. and Mendez, D., **Identifying concerns when specifying machine learning-enabled systems: A perspective-based approach**. *Journal of Systems and Software*, vol. 213, July 2024.

PerSpecML

ML-Enabled System Specification Approach

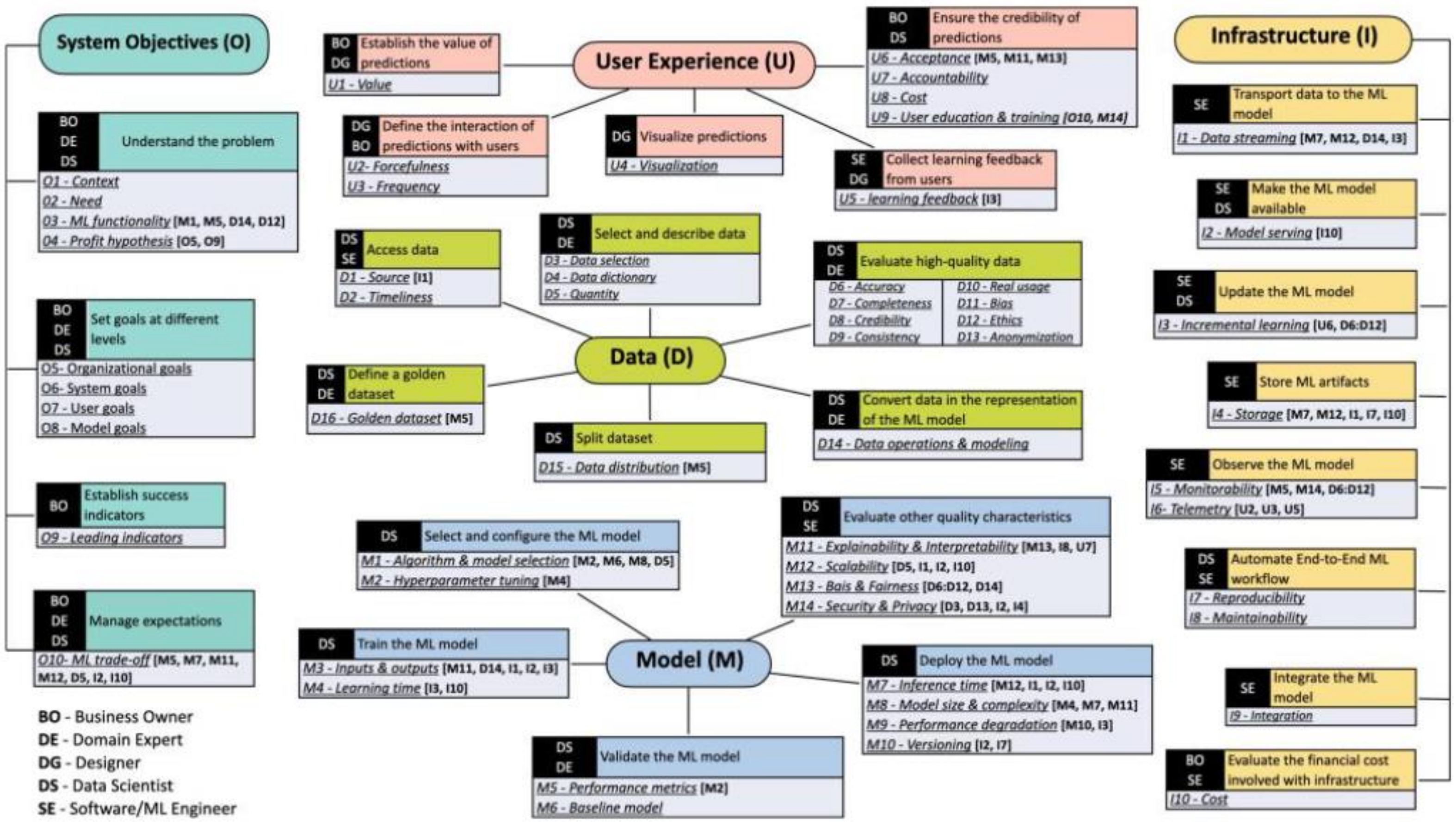
Based on analyzing **60 concerns** grouped into five perspectives: **objectives**, **user experience**, **infrastructure**, **model**, and **data**

Mediates the communication between **business owners**, **domain experts**, **designers**, **data scientists**, and **software/ML engineers**

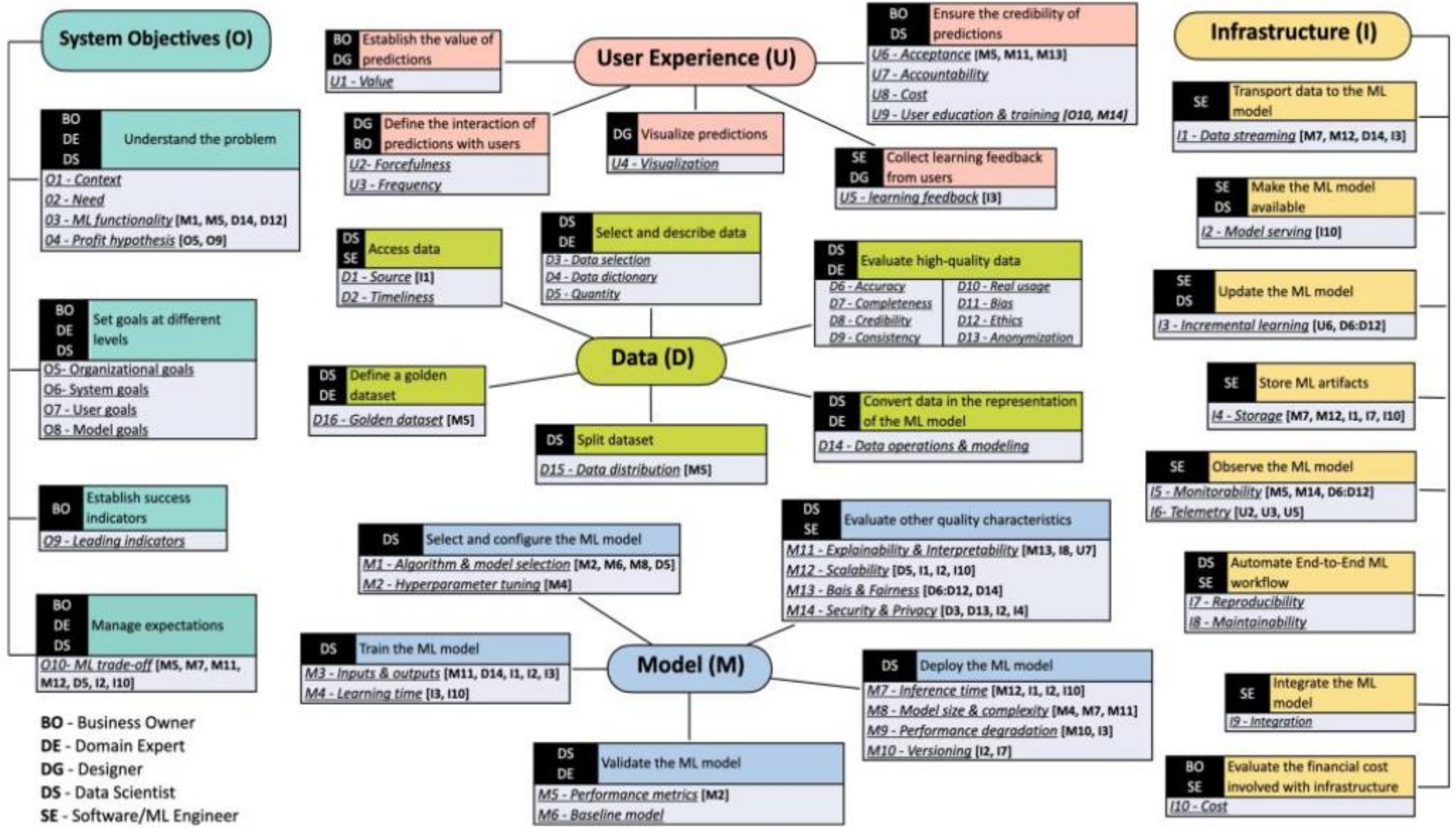
Villamizar, H., Kalinowski, M., Lopes, H. and Mendez, D., **Identifying concerns when specifying machine learning-enabled systems: A perspective-based approach**. *Journal of Systems and Software*, vol. 213, July 2024.

PerSpecML

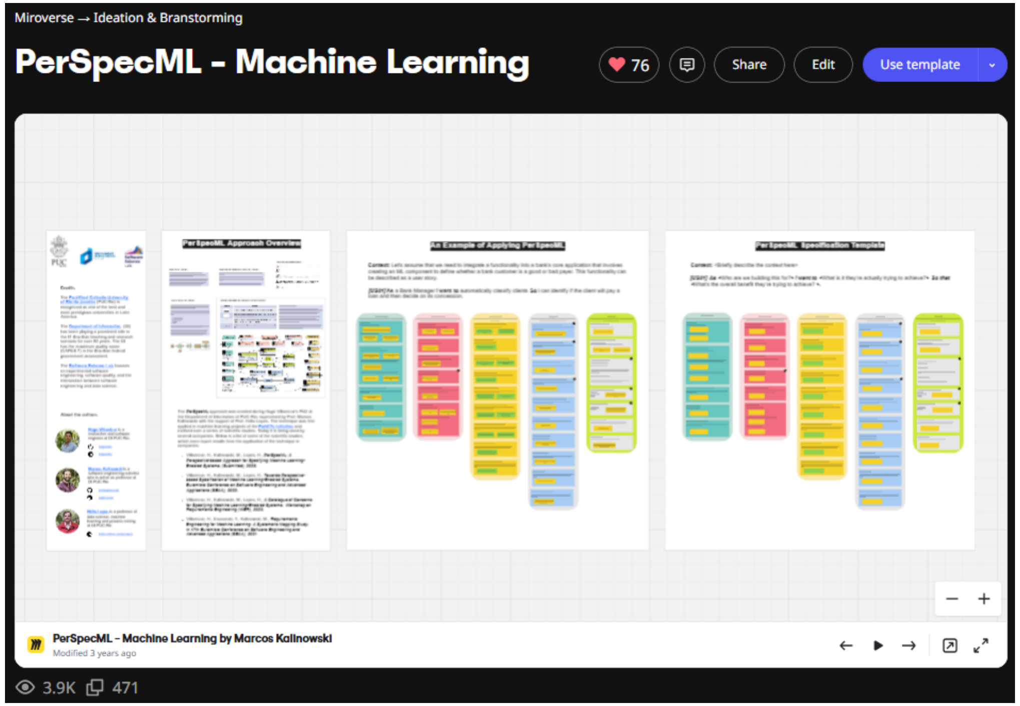
Task and Concern Diagram



Villamizar, H., Kalinowski, M., Lopes, H. and Mendez, D., **Identifying concerns when specifying machine learning-enabled systems: A perspective-based approach**. *Journal of Systems and Software*, vol. 213, July 2024.


System Objectives (O)
BO DE DS Understand the problem
O1 - Context
02 - Need
03 - ML functionality [M1, M5, D14, D12]
04 - Profit hypothesis [O5, O9]
BO DE DS Set goals at different levels
O5- Organizational goals
O6- System goals
O7 - User goals
O8 - Model goals
BO Establish success indicators
O9 - Leading indicators
BO DE DS Manage expectations
O10- ML trade-off [M5, M7, M11, M12, D5, I2, I10]
BO - Business Owner
DE - Domain Expert
DG - Designer
DS - Data Scientist
SE - Software/ML Engineer
User Experience (U)
BO DG Establish the value of predictions
U1 - Value
DG BO Define the interaction of predictions with users
U2- Forcefulness
U3 - Frequency
DG Visualize predictions
U4 - Visualization
BO DS Ensure the credibility of predictions
U6 - Acceptance [M5, M11, M13]
U7 - Accountability
U8 - Cost
U9 - User education & training [O10, M14]
SE DG Collect learning feedback from users
U5 - learning feedback [I3]
Data (D)
DS SE Access data
D1 - Source [I1]
D2 - Timeliness
DS DE Select and describe data
D3 - Data selection
D4 - Data dictionary
D5 - Quantity
DS DE Evaluate high-quality data
D6 - Accuracy
D7 - Completeness
D8 - Credibility
D9 - Consistency
D10 - Real usage
D11 - Bias
D12 - Ethics
D13 - Anonymization
DS DE Define a golden dataset
D16 - Golden dataset [M5]
DS Split dataset
D15 - Data distribution [M5]
DS DE Convert data in the representation of the ML model
D14 - Data operations & modeling
Model (M)
DS Select and configure the ML model
M1 - Algorithm & model selection [M2, M6, M8, D5]
M2 - Hyperparameter tuning [M4]
DS SE Evaluate other quality characteristics
M11 - Explainability & Interpretability [M13, I8, U7]
M12 - Scalability [D5, I1, I2, I10]
M13 - Bais & Fairness [D6:D12, D14]
M14 - Security & Privacy [D3, D13, I2, I4]
DS Train the ML model
M3 - Inputs & outputs [M11, D14, I1, I2, I3]
M4 - Learning time [I3, I10]
DS Deploy the ML model
M7 - Inference time [M12, I1, I2, I10]
M8 - Model size & complexity [M4, M7, M11]
M9 - Performance degradation [M10, I3]
M10 - Versioning [I2, I7]
DS DE Validate the ML model
M5 - Performance metrics [M2]
M6 - Baseline model
Infrastructure (I)
SE Transport data to the ML model
I1 - Data streaming [M7, M12, D14, I3]
SE DS Make the ML model available
I2 - Model serving [I10]
SE DS Update the ML model
I3 - Incremental learning [U6, D6:D12]
SE Store ML artifacts
I4 - Storage [M7, M12, I1, I7, I10]
SE Observe the ML model
I5 - Monitorability [M5, M14, D6:D12]
I6- Telemetry [U2, U3, U5]
DS SE Automate End-to-End ML workflow
I7 - Reproducibility
I8 - Maintainability
SE Integrate the ML model
I9 - Integration
BO SE Evaluate the financial cost involved with infrastructure
I10 - Cost

PerSpecML: Online Specification Template



https://miro.com/miroverse/perspecml-machine-learning/

Villamizar, H., Kalinowski, M., Lopes, H. and Mendez, D., **Identifying concerns when specifying machine learning-enabled systems: A perspective-based approach**. *Journal of Systems and Software*, vol. 213, July 2024.

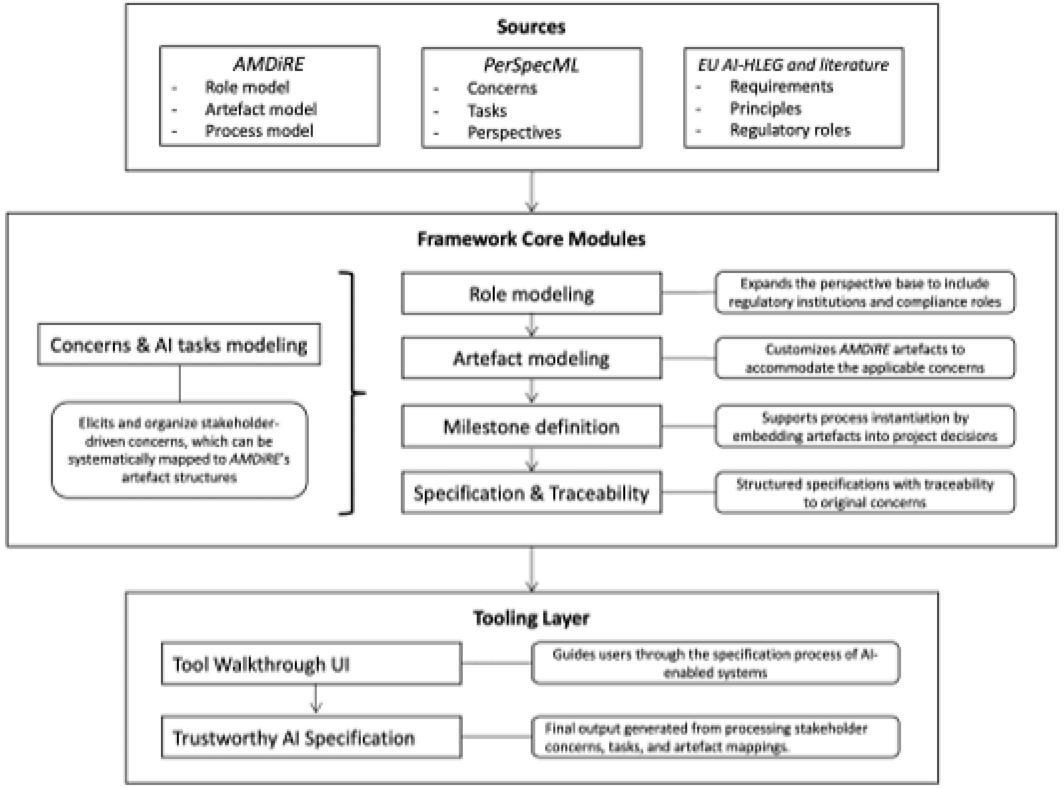
PerSpecML: Extensions

Generalizing beyond ML (FM-based Agents, RL, etc)

Extension of concerns focusing on trustworthy AI

Operationalization through artefacts

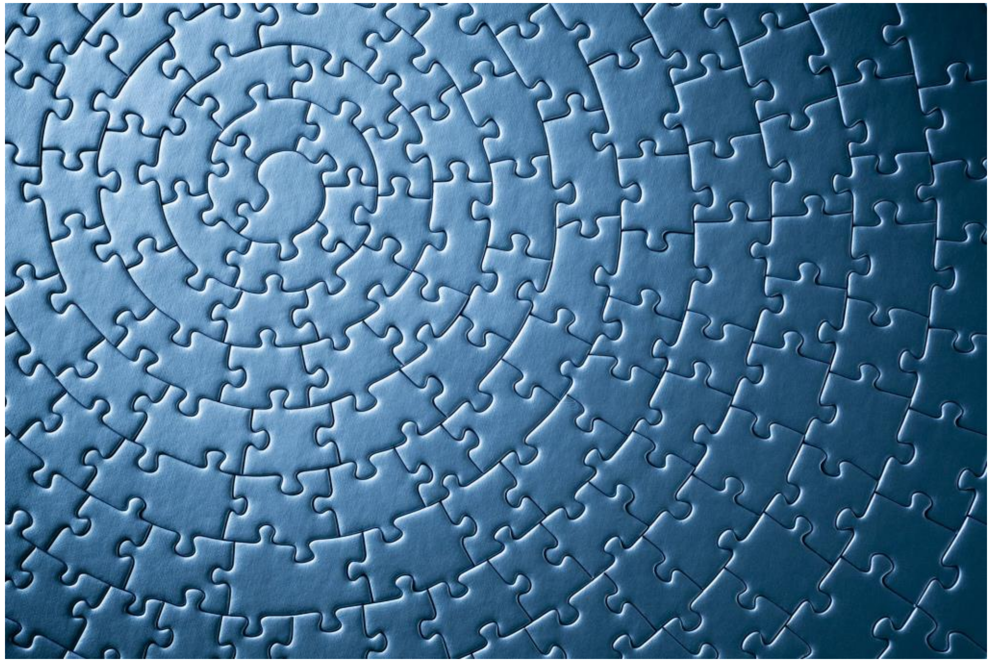
Providing domain knowledge



Villamizar, H., Mendez, D. and Kalinowski, M., 2025. **Towards a Framework for Operationalizing the Specification of Trustworthy AI Requirements**. Requirement Engineering for Trustworthy Artificial Intelligence (RETRAI@RE), 2025.

RE in Multi-Paradigm AI Engineering: The Road Ahead

From ML-Centric to Multi-Paradigm AI Engineering



Current Strengths and Limitations

Advances in AI Engineering have enabled robust ML systems.

The focus is still largely centered on ML.

Why Broaden the View?

Rapid advances in AI paradigms increase capability and software engineering complexity.

Individual paradigms have inherent limitations.

Intelligent systems can benefit from integrating perception, learning, reasoning, and decision-making.

Rise of Foundation Models and LLM-based Agents.

AI Paradigms

Machine Learning excels in pattern recognition but lacks explicit reasoning.

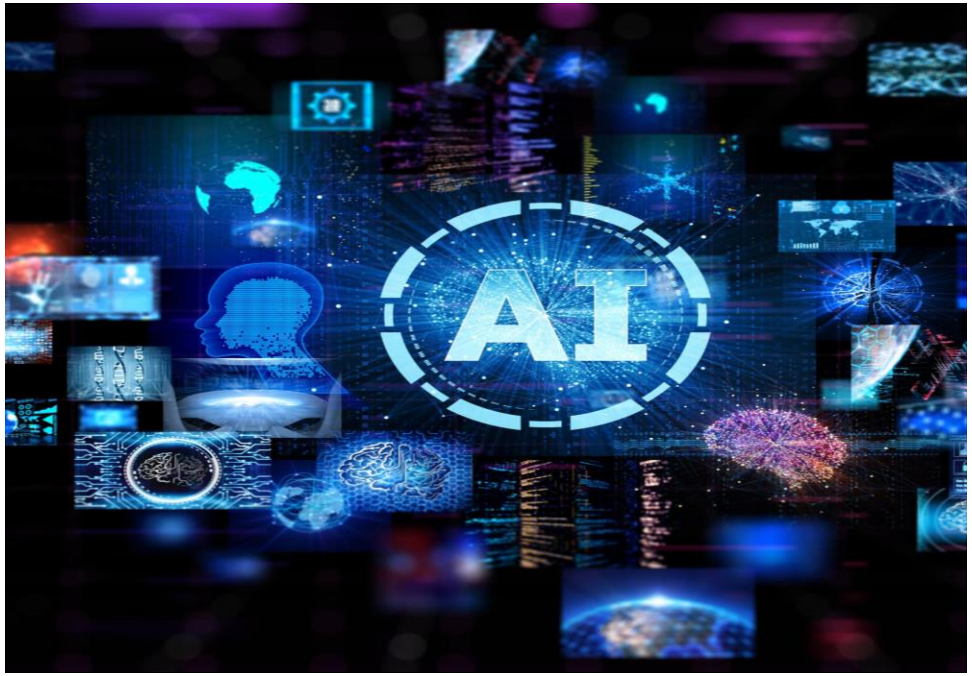
Reinforcement Learning offers adaptive decision-making but can lead to unpredictable behaviors.

Foundation Models generalize knowledge but face interpretability challenges.

Symbolic AI provides structure of readable logic and rules but lacks flexibility for complex problems.

Multi-Agent Systems enable collaboration but introduce emergent behaviors.

Rising Complexity, Growing Literacy Gap



Complexity Is Increasing

Foundation Models are now being used to perform reasoning, delegation, environment interaction, ...

Multi-paradigm AI systems are the new norm.

The Literacy Gap

AI adoption is widespread, but understanding is lagging.

AI literacy gap is widening, even among professionals.

Urgent need for accessible, cross-paradigm AI education.

Who will define requirements for multi-paradigm AI systems?

Combining Perception, Learning, Reasoning, and Decision-Making

Perception
Perception is the first step in understanding our environment.

Learning
Learning enhances our ability to adapt by incorporating new knowledge and experiences into our decision-making processes.

Reasoning
Reasoning allows individuals to analyze information and draw conclusions, which is crucial for effective problem-solving.

Decision-Making
Coherent decision-making integrates perception, learning, and reasoning to achieve **adaptable intelligence**.

Real-World Examples: Autonomous Vehicles

Paradigms Combined:

Machine Learning (for perception)

Symbolic AI (for rule-based traffic logic)

Reinforcement Learning (for dynamic decision-making)

Multi-Agent Systems (for interaction with other vehicles)

Tesla Autopilot and **Waymo’s self-driving systems** use deep learning to detect objects, symbolic logic for obeying traffic rules, RL to optimize behavior over time, and multi-agent reasoning to anticipate other drivers’ behavior.

Real-World Examples: Personal Assistants

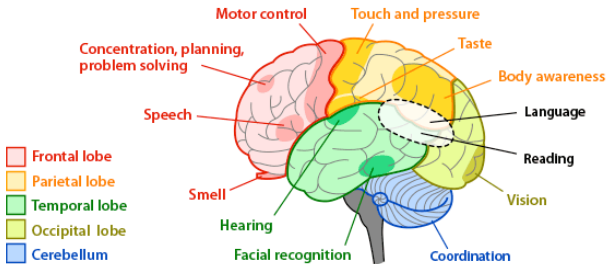
Paradigms Combined:

- Foundation Models (language understanding)
- Symbolic AI (context modeling, dialog trees)
- Machine Learning (voice recognition, sentiment analysis)
- Reinforcement Learning (dialog policy optimization)

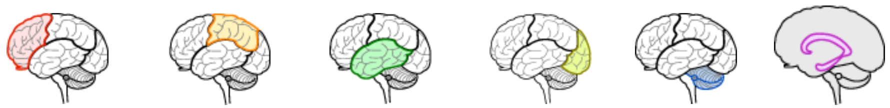
Amazon Alexa combines foundation models (e.g., AlexaTM) with symbolic dialog managers for intent handling and RL to improve user satisfaction over time through interaction.

What Exactly Are We Combining?

AI Paradigm	Brain Function	Explanation
Machine Learning	**Learning from experience**	Like how we improve through repetition and feedback
Reinforcement Learning	**Reward-based behavior adaptation**	Similar to how dopamine reinforces behavior when we succeed
Symbolic AI	**Logical reasoning and rule-following**	Mirrors how the brain uses explicit knowledge and structured logic
Foundation Models (NLP/CV)	**Language and vision processing**	Parallels the brain's natural language understanding and visual recognition
Multi-Agent Systems	**Social cognition and collaboration**	Reflects how we coordinate with others, plan, and divide tasks
Planning/Automated Reasoning	**Executive decision-making (Prefrontal Cortex)**	Comparable to how we plan steps to solve problems
Perception (Computer Vision, Speech)	**Sensory processing (Visual/Auditory Cortex)**	Equivalent to how we see and hear the world



High-level Components of Intelligence!



Intelligence Engineering: A New Chapter in AI Engineering?

A candidate definition (created for my CAIN 2025 keynote):

“**Intelligence Engineering** (within AI Engineering) is the discipline concerned with the systematic integration of multiple AI paradigms to design and realize intelligent behaviors in AI-enabled systems.”

Intelligence Engineering: What about RE?

Understanding of **"Intelligence Engineering"**, from a system's perspective, is needed.

RE aspects are still poorly understood across different AI paradigms and their combinations.

> How to **specify requirements involving intelligent behavior of multi-paradigm AI systems?**

Demand for AI literacy to define and govern intelligent behavior.

> How to **improve AI literacy of requirements engineers?**

AI Product Manager?

Research Roadmap: Requirements for Multi-Paradigm AI Systems

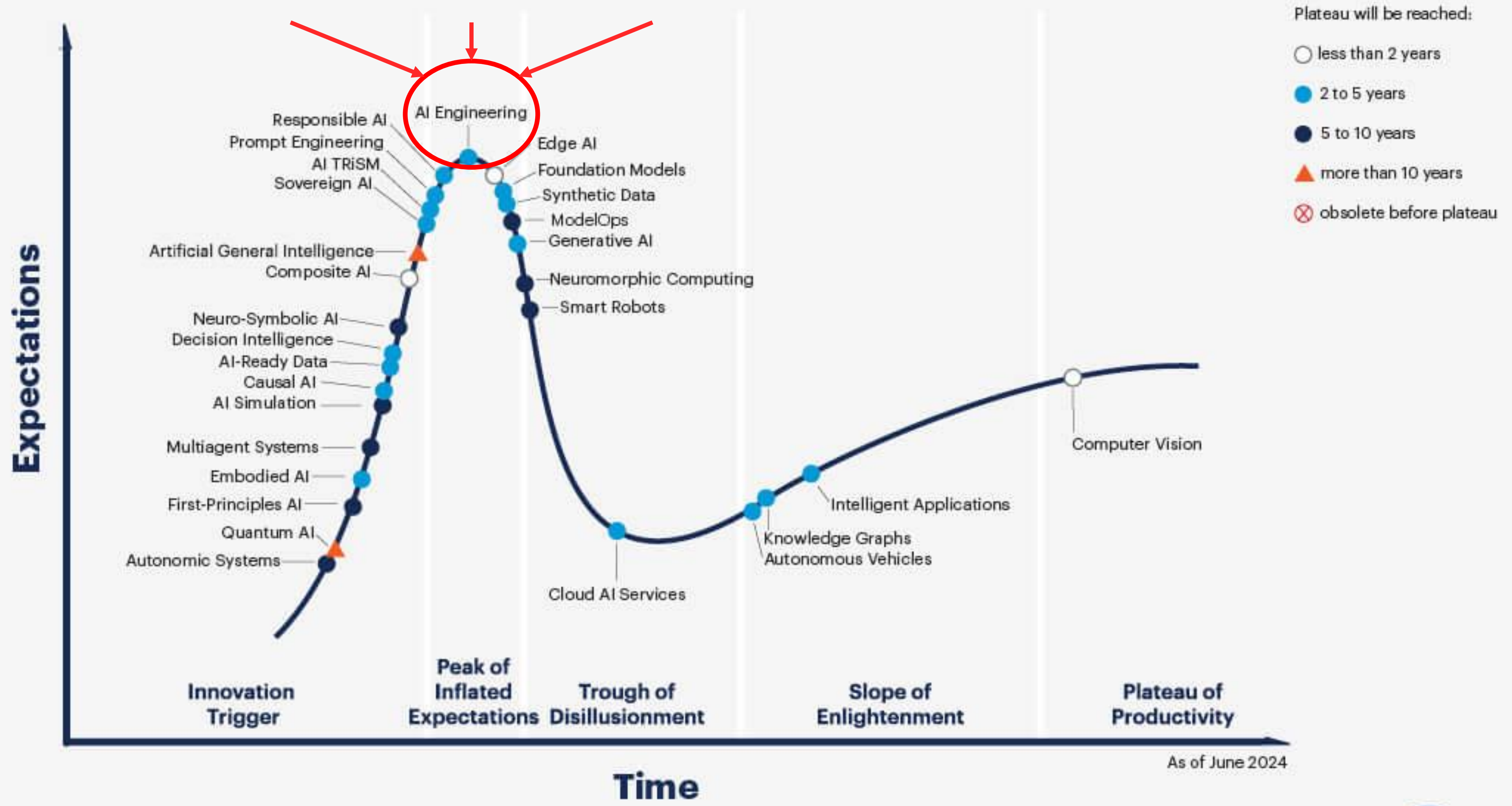
System Ideation
Methods for ideating intelligent systems across multiple AI paradigms.

Requirements Definition
Identify concerns when specifying requirements for different paradigms and their combinations.

Requirements Validation
Strategies to validate paradigm selections against system goals.

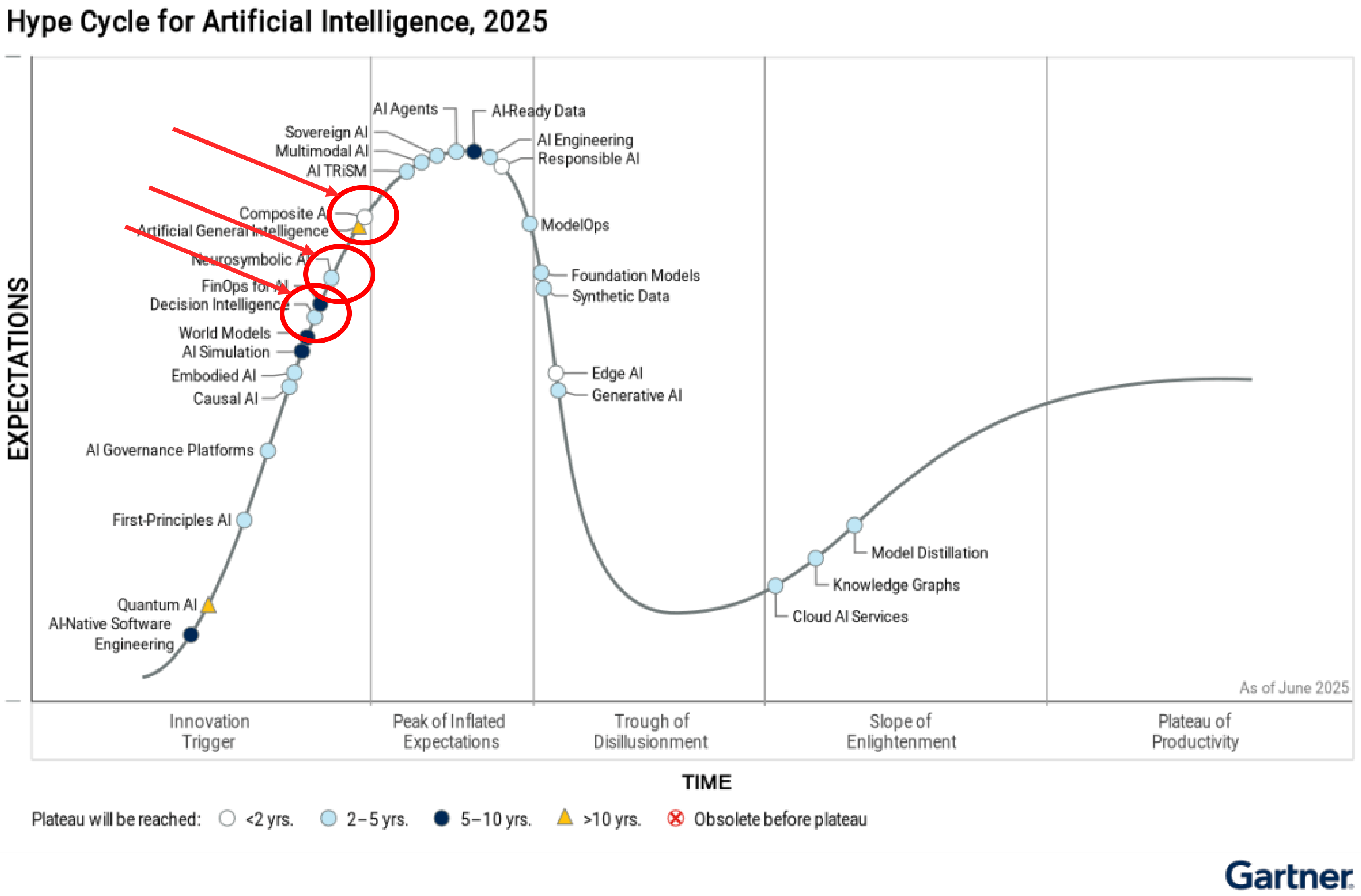
Requirements Management
Techniques for managing evolving requirements in multi-paradigm contexts.

Hype Cycle for Artificial Intelligence, 2024

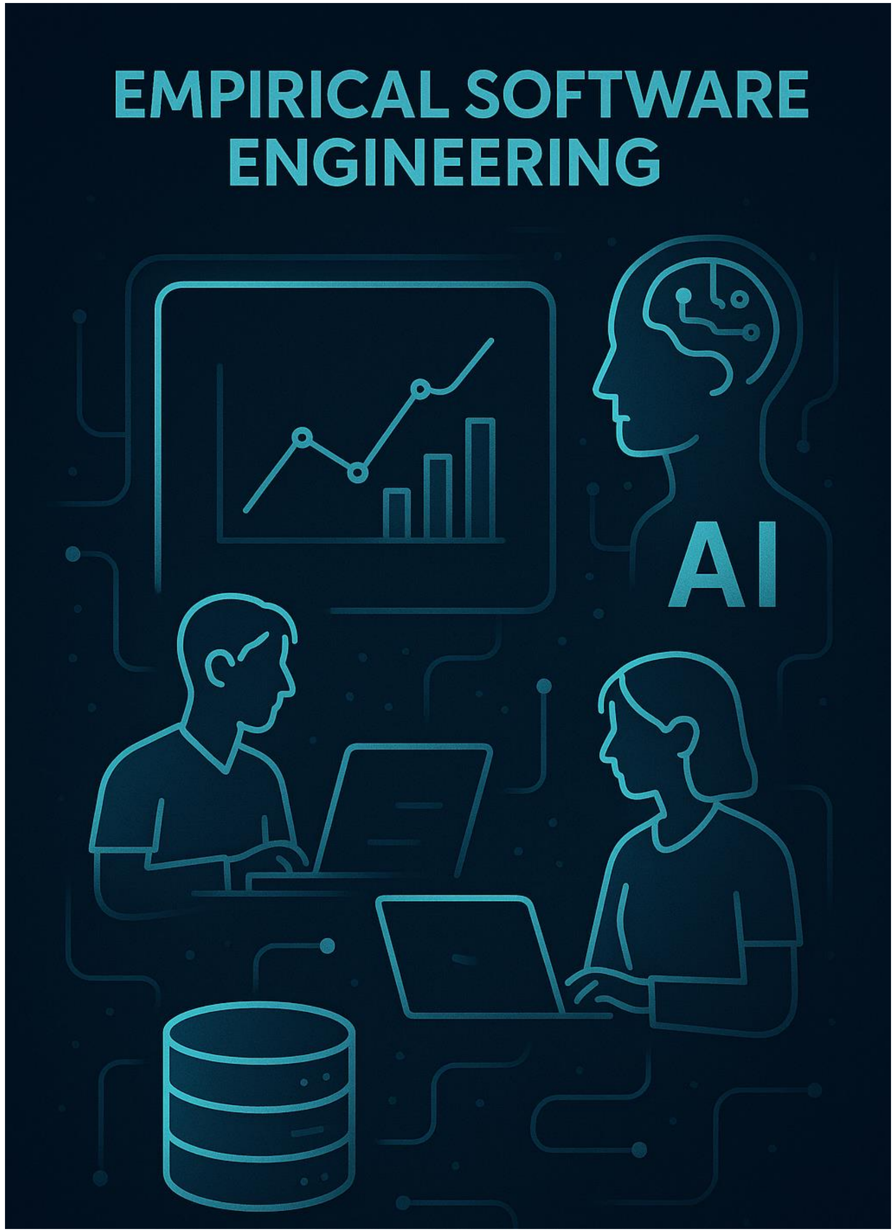


Source: Gartner
Commercial reuse requires approval from Gartner and must comply with the Gartner Content Compliance Policy on gartner.com.

Gartner



Composite AI: "combined application (or fusion) of different AI techniques to improve the efficiency of learning, broaden the level of knowledge representations, and ultimately solve a wider range of business problems in a more effective manner"



The Role of Empirical Software Engineering

At the *Peak of Inflated Expectations* excitement often outpaces proven results.

Scientific Evidence to Evolve Technologies

Empirical software engineering research can offer evidence on which technologies can support the responsible development of practical, trustworthy, scalable, and maintainable AI systems in real-world settings.


SCIENTIFIC EVIDENCE
SCIENTIFIC METHOD
EMPIRICAL
SOFTWARE ENGINEERING
INFLUENCER
BLOG POST
ANECDOTAL EVIDENCE
AI HYPE

Specifying Intelligence: RE for the Next Generation of AI Systems

Prof. Marcos Kalinowski, PUC-Rio

WER 2025, Rio de Janeiro, Brazil