

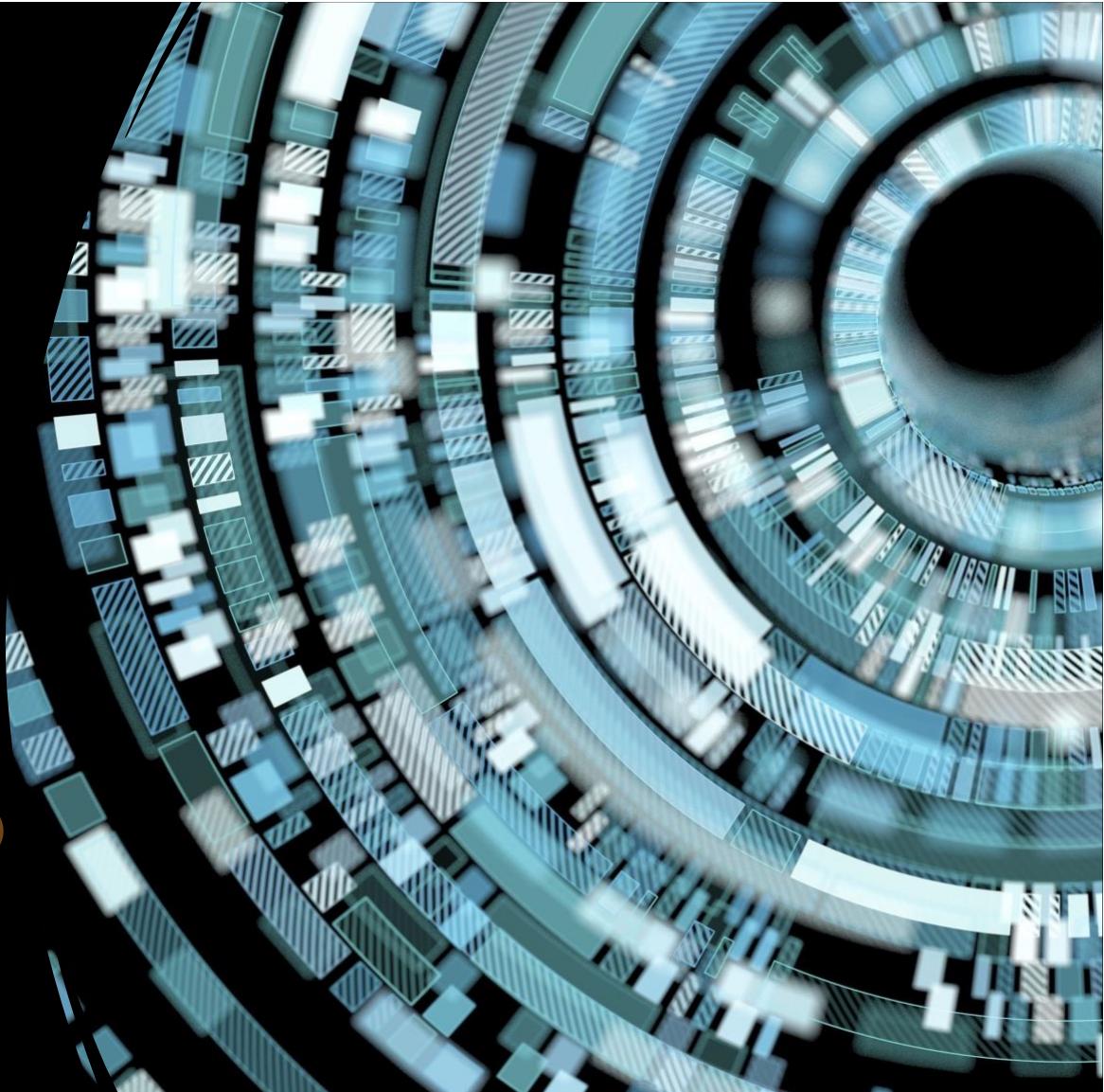
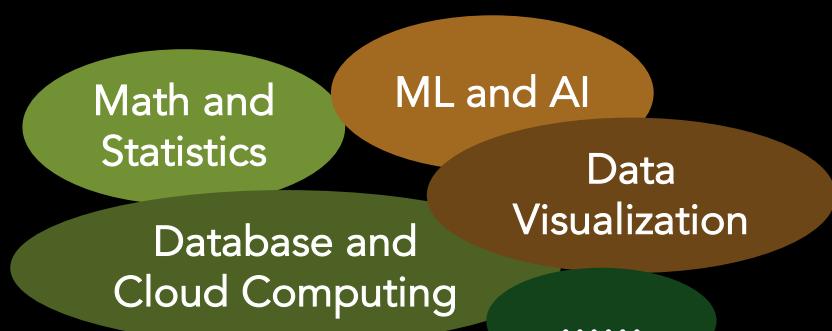
# Team and Collaboration

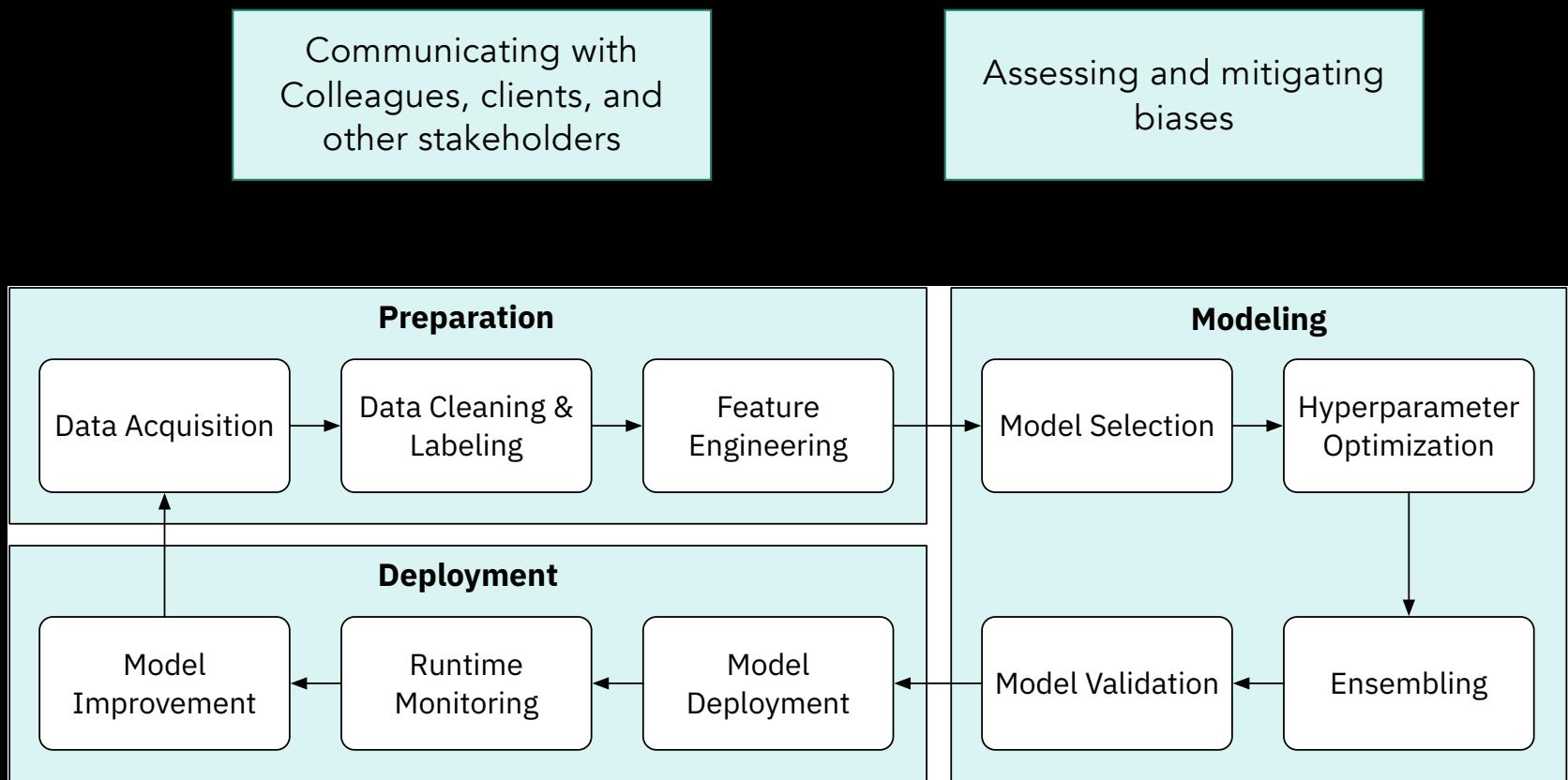
Jin Guo  
SOCS McGill University

# Data Scientists

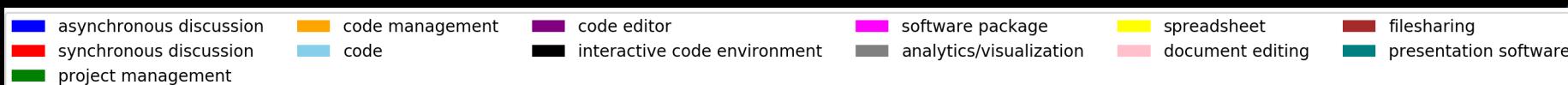
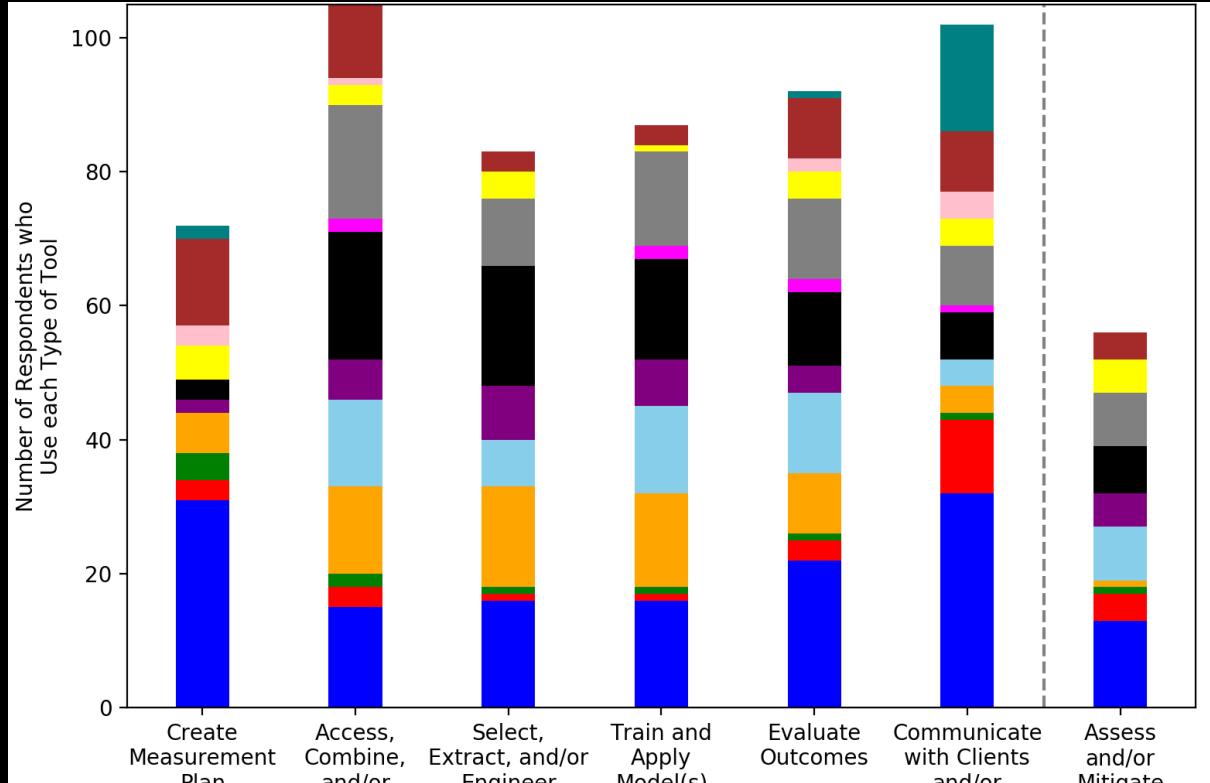
---

leverage modern machine learning techniques to identify insights from data





Wang, D. et al. (2019) 'Human-AI Collaboration in Data Science: Exploring Data Scientists' Perceptions of Automated AI', *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW), pp. 1–24. doi:[10.1145/3359313](https://doi.org/10.1145/3359313).

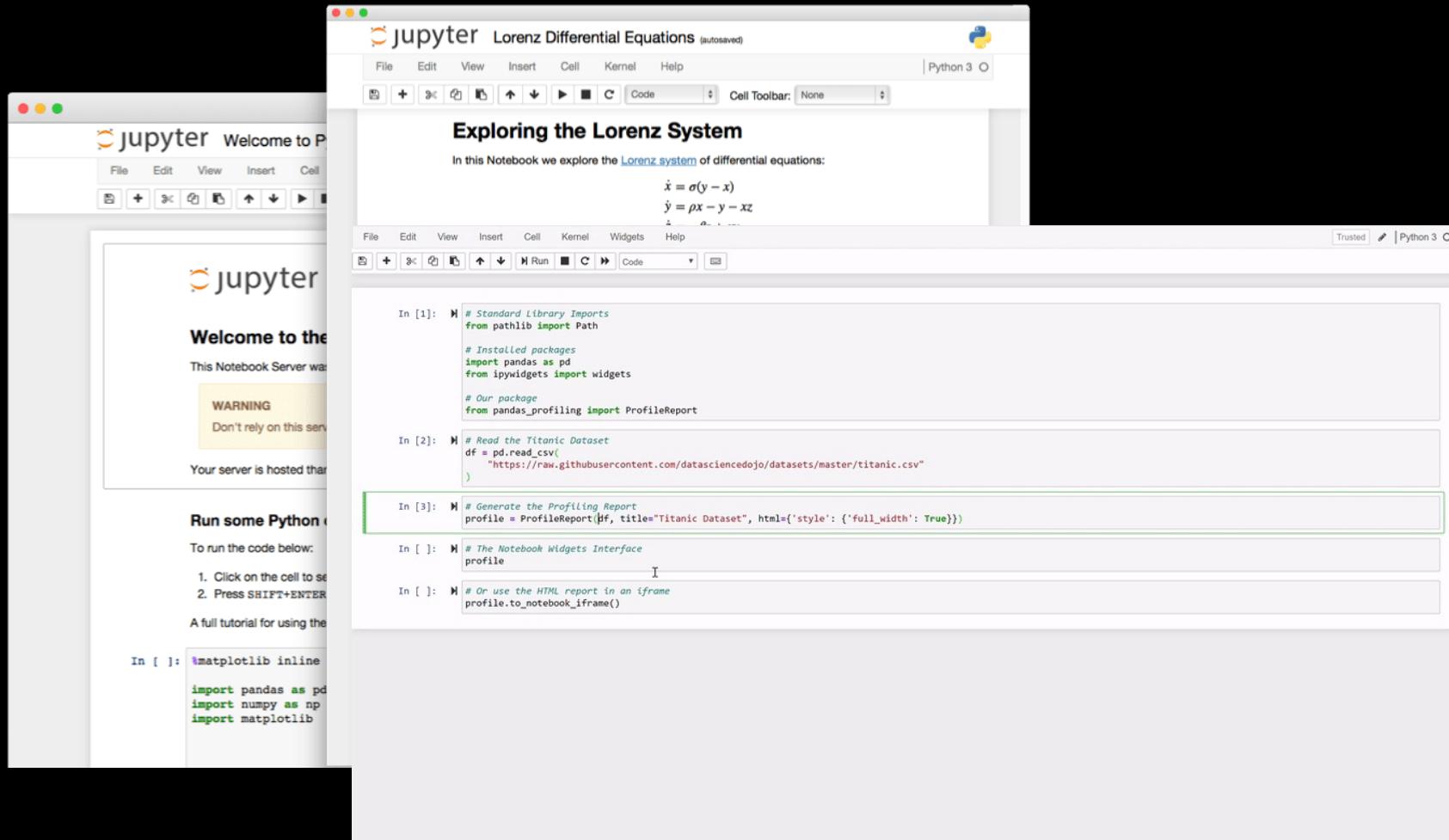


Zhang, A.X., Muller, M. and Wang, D. (2020) 'How do Data Science Workers Collaborate? Roles, Workflows, and Tools', *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW1), pp. 1–23. doi:[10.1145/3392826](https://doi.org/10.1145/3392826).

Table 2. Categories of data science tools and the number of times each tool was mentioned by respondents.

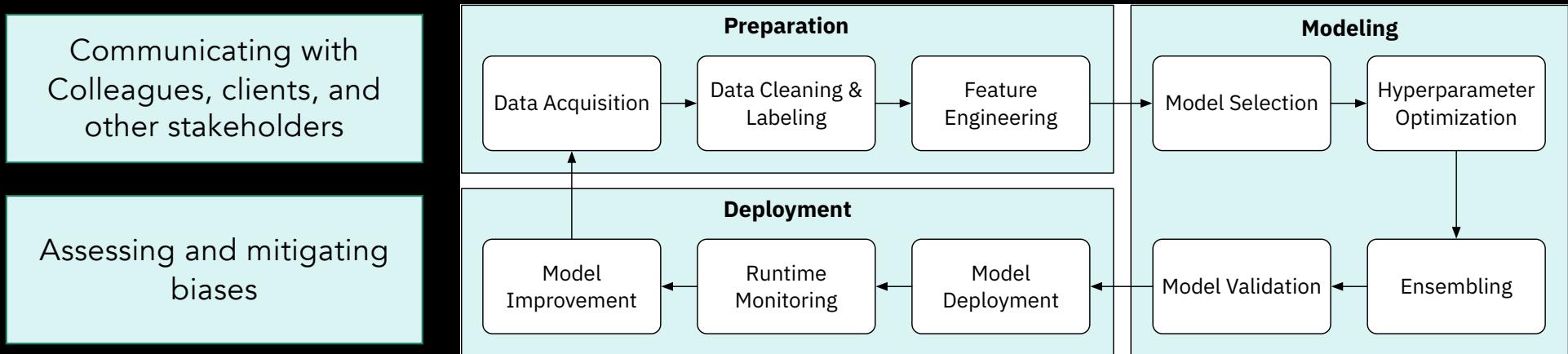
Tool Category	Tools Mentioned by Respondents (number of times mentioned)
asynchronous discussion	Slack (86), email (55), Microsoft Teams (1)
synchronous discussion	meeting (13), e-meeting (12), phone (1)
project management	Jira (8), ZenHub (2), Trello (1)
code management	GitHub (56) Git (5)
code	Python (42), R (9), Java (3), scripts (3)
code editor	visual Studio Code (11), PyCharm (11), RStudio (8), Eclipse (1), Atom
interactive code environment	Jupyter Notebook (16), SQL (6), terminal (4), Google Colab (4)
software package	Scikit Learn (5), Shiny App (2), Pandas (2)
analytics/visualization	SPSS (27), Watson Analytics (22), Cognos (7), ElasticSearch (4), Apache Spark (3), Graphana (2), Tableau (2), Logstash (2), Kibana (1)
spreadsheet	Microsoft Excel (22), spreadsheets (3), Google Sheets (1)
document editing	wiki (2), LaTeX (2), Microsoft Word (2), Dropbox Paper (2), Google Docs (1)
filesharing	Box (43), cloud (5), NFS (2), Dropbox (1), Filezilla (1)
presentation software	Microsoft Powerpoint (18), Prezi (1)

Note: **code** allows programmers to *write algorithms* for data science. **code editor** and **interactive code environment** provide a user experience for writing that code. **code management** is where the code may be stored and shared. By contrast, **analytics/visualization** provides “macro-level” tools that can invoke entire steps or modular actions in a data science pipeline.



<https://jupyter.org>

<https://github.com/pandas-profiling/pandas-profiling>



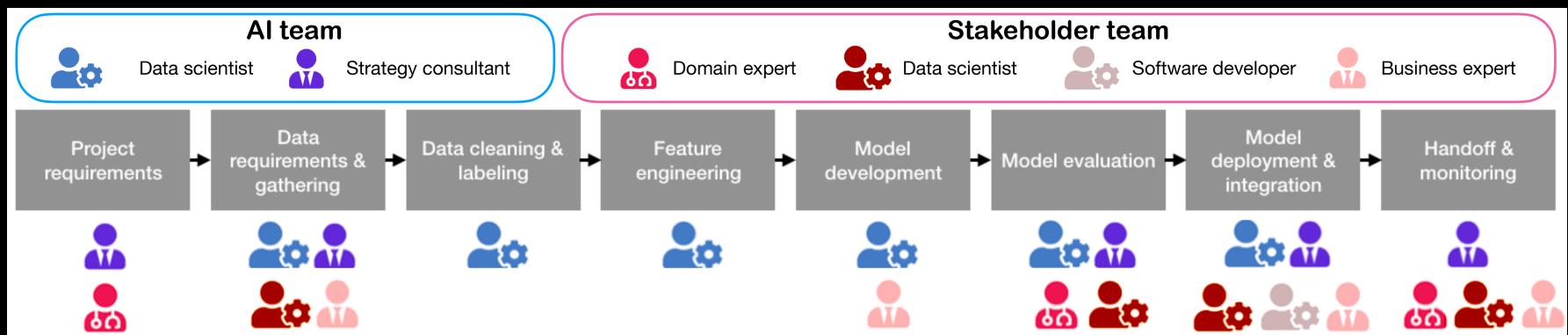
- Data Science for specialized industries and scientific discovery  
Work with domain experts for specific problems in health-care, insurance, banking, or other highly-specialized industries with sensitive data

Communication Gap

Define problem

Moving Target

- Data Science as a service  
Demonstrate or develop data science solutions, and educate clients



Piorkowski, D. et al. (2021) 'How AI Developers Overcome Communication Challenges in a Multidisciplinary Team: A Case Study', *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1), pp. 1–25. doi:[10.1145/3449205](https://doi.org/10.1145/3449205).

Table 1. A summary of the principles of how humans realize shared mental models [49]. This framework is helpful for describing collaborations activities in the domain of AI.

<b>Principle</b>	<b>Definition</b>
Consistency	Ability to maintain stability by resolving conflicts due to differing perceptions, differing knowledge states, asynchronous information, and missing updates.
Reactivity	Ability to effectively address unanticipated events or state changes by informing team members of the changes and adapting goals and plans to account for the new situations.
Proactivity	Ability to anticipate problems, bottlenecks and failures and take proactive actions, such as asking for clarification or offering assistance.
Coordination	Ability to work together via overall cooperative attitudes, such as establishing joint goals and plans, transparent task assignments, and truthful information sharing
Knowledge Stability	Ability to understand the staleness of information over time and adjust sampling rates according to their confidence in the information's validity.

Piorkowski, D. et al. (2021) 'How AI Developers Overcome Communication Challenges in a Multidisciplinary Team: A Case Study', *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1), pp. 1–25. doi:[10.1145/3449205](https://doi.org/10.1145/3449205).

# Activity

## Cross the Communication Gap

- Read and discuss Section 4 of the work "How AI Developers Overcome Communication Challenges in a Multidisciplinary Team: A Case Study"
- Identify concrete actions the AI developers perform
- List which principles it helps realize and identify which challenge it helps address.
- Use Miro to summarize the findings

# Best practice for AI development workflow

- Longer planning
- Shared Documentation effort
- Rolling from Smaller Pilot

More SE !

Piorkowski, D. et al. (2021) 'How AI Developers Overcome Communication Challenges in a Multidisciplinary Team: A Case Study', *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1), pp. 1–25. doi:[10.1145/3449205](https://doi.org/10.1145/3449205).

On next Tuesday:

Team and Collaboration by Martin

Quality Assessment