

Machine Learning in Production Fostering Interdisciplinary Teams

= (Process and Team Reflections)

One last crosscutting topic

Fundamentals of Engineering AI-Enabled Systems

Holistic system view: AI and non-AI components, pipelines, stakeholders, environment interactions, feedback loops

Requirements:

- System and model goals
- User requirements
- Environment assumptions
- Quality beyond accuracy
- Measurement
- Risk analysis
- Planning for mistakes

Architecture + design:

- Modeling tradeoffs
- Deployment architecture
- Data science pipelines
- Telemetry, monitoring
- Anticipating evolution
- Big data processing
- Human-AI design

Quality assurance:

- Model testing
- Data quality
- QA automation
- Testing in production
- Infrastructure quality
- Debugging

Operations:

- Continuous deployment
- Contin. experimentation
- Configuration mgmt.
- Monitoring
- Versioning
- Big data
- DevOps, MLOps

Teams and process: Data science vs software eng. workflows, interdisciplinary teams, collaboration points, technical debt

Responsible AI Engineering

Provenance,
versioning,
reproducibility

Safety

Security and
privacy

Fairness

Interpretability
and explainability

Transparency
and trust

Ethics, governance, regulation, compliance, organizational culture

Readings

Nahar, Nadia, Shurui Zhou, Grace Lewis, and Christian Kästner.
[**"Collaboration Challenges in Building ML-Enabled Systems: Communication, Documentation, Engineering, and Process."**](#) In International Conf. Software Engineering, 2022.

Learning Goals

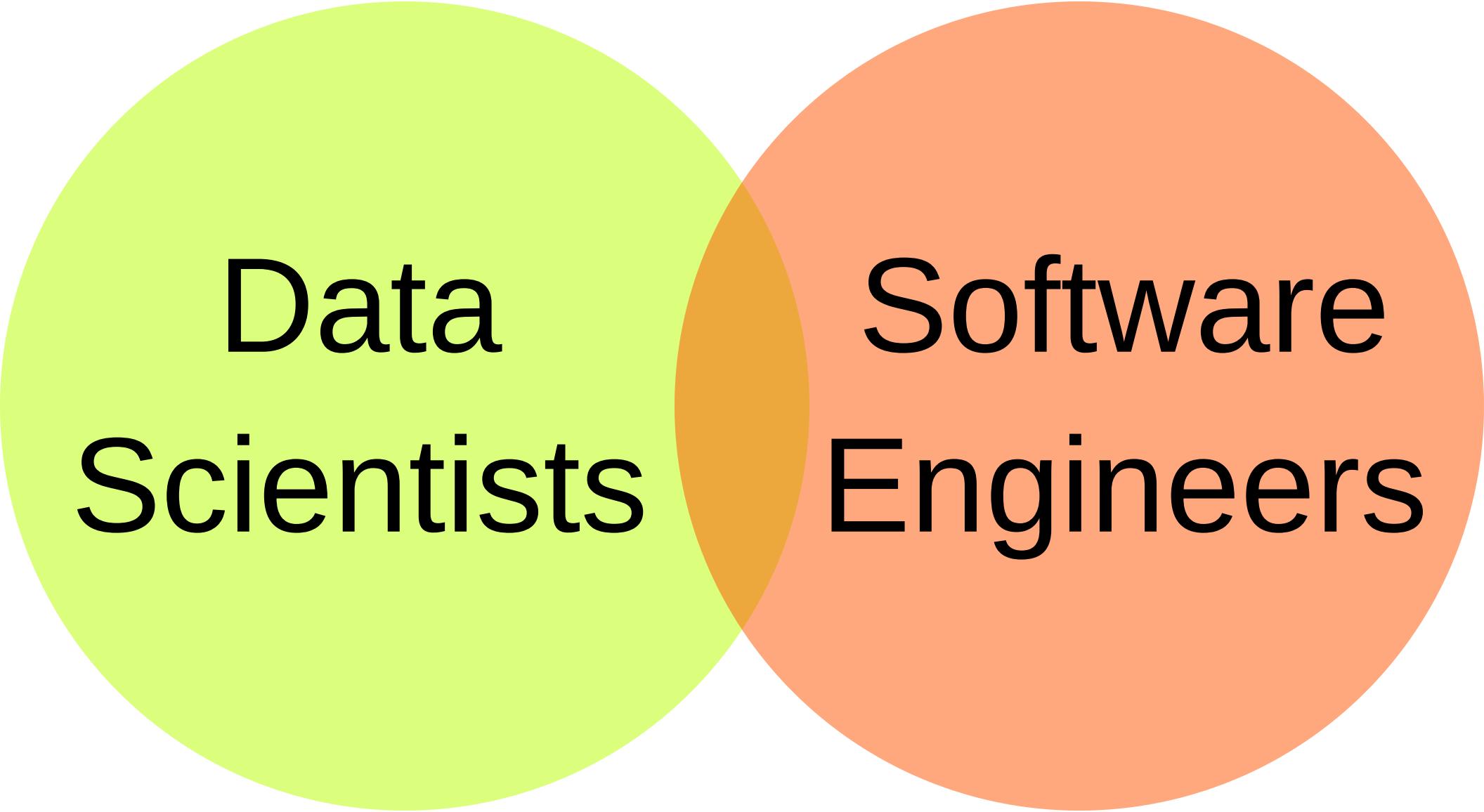
- Understand different roles in projects for AI-enabled systems
- Plan development activities in an inclusive fashion for participants in different roles
- Diagnose and address common teamwork issues
- Describe agile techniques to address common process and communication issues

Case Study: Depression Prognosis on Social Media



The Project

- Social media company of about 15000 employees, 500 developers and data scientists in US
- Use sentiment analysis on video data (and transcripts) to detect depression
- Planned interventions through recommending different content and showing ads for getting support, design for small group features
- Collaboration with mental health professionals and ML researches at top university



A Venn diagram consisting of two overlapping circles. The left circle is light green and contains the text "Data Scientists". The right circle is light orange and contains the text "Software Engineers". The overlapping area between the two circles is shaded in a darker orange.

**Data
Scientists**

**Software
Engineers**

Data scientist

- Often fixed dataset for training and evaluation (e.g., PBS interviews)
- Focused on accuracy
- Prototyping, often Jupyter notebooks or similar
- Expert in modeling techniques and feature engineering
- Model size, updateability, implementation stability typically does not matter

Software engineer

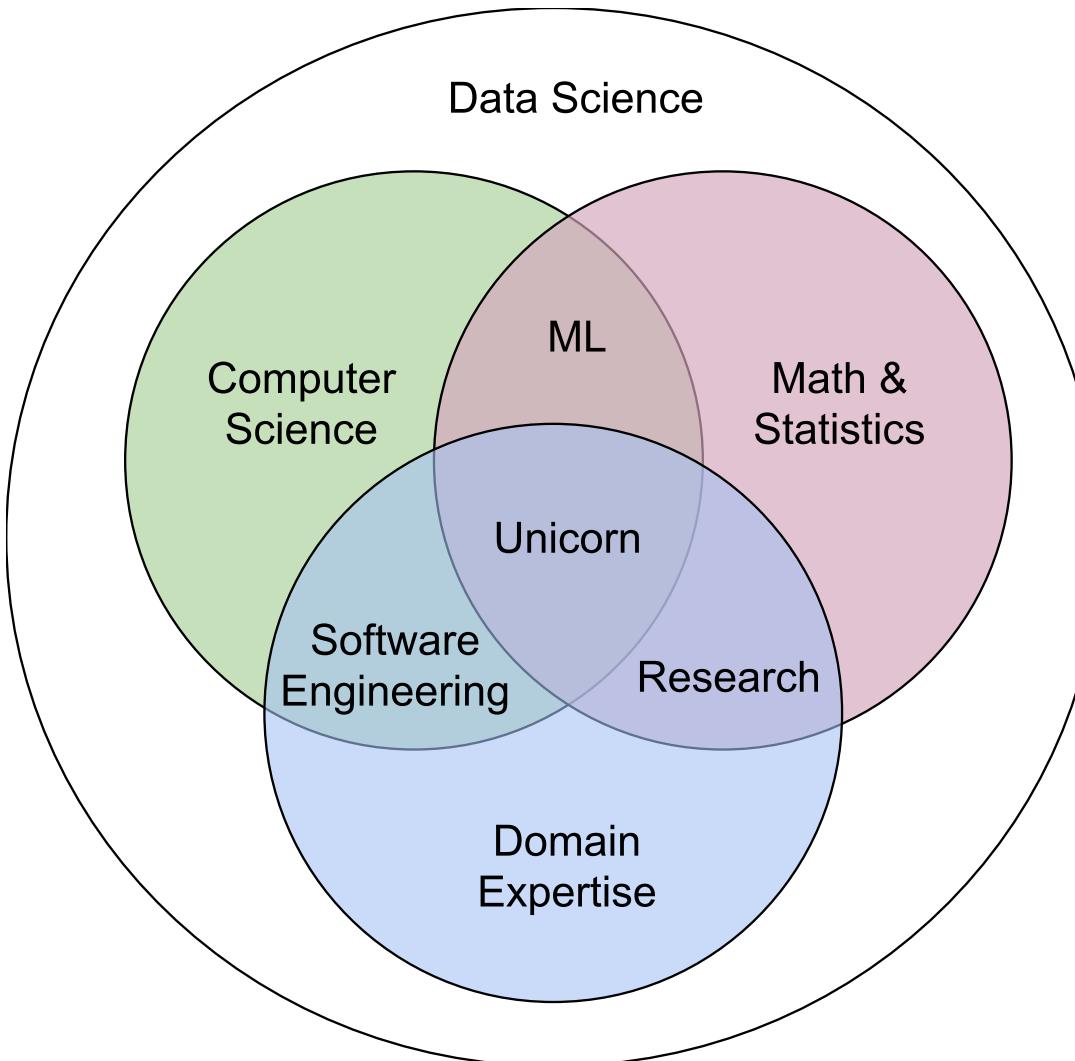
- Builds a product
- Concerned about cost, performance, stability, release time
- Identify quality through customer satisfaction
- Must scale solution, handle large amounts of data
- Detect and handle mistakes, preferably automatically
- Maintain, evolve, and extend the product over long periods
- Consider requirements for security, safety, fairness

Continuum of Skills

- Software Engineer
- Data Engineer
- Data Scientist
- Applied Scientist
- Research Scientist

Talk: Ryan Orban. [Bridging the Gap Between Data Science & Engineer: Building High-Performance Teams](#). 2016





By Steven Geringer, via Ryan Orban. [Bridging the Gap Between Data Science & Engineer: Building High-Performance Teams](#). 2016

Data Scientists At Microsoft

- Mostly analyzing product and customer data
- User engagement (which features users like and use, satisfaction, retention)
- Software productivity (bug prioritization, monitoring)
- Domain-specific problems (NLP quality, stock pricing, power prediction)
- Business intelligence (predicting investment, demand, sales)

Data Science Roles At Microsoft

- Polymath
- Data evangelist
- Data preparer
- Data shaper
- Data analyzer
- Platform builder
- 50/20% moonlighter
- Insight actors

Kim, Miryung, Thomas Zimmermann, Robert DeLine, and Andrew Begel. "[Data scientists in software teams: State of the art and challenges.](#)" IEEE Transactions on Software Engineering 44, no. 11 (2017): 1024-1038.

Many other Role Descriptions

- Data scientist
- Data analyst
- Data architect
- Data engineer
- Statistician
- Database administrator
- Business analyst
- Data and analytics manager



e.g. Martijn Theuwissen. [The different data science roles in the industry](#). 2015

Many other Role Descriptions

- Product Data Analyst (feature analysis)
- Business Intelligence, Analytics & Reporting (marketing)
- Modeling Analyst (financial forecasting)
- Machine Learning Engineer (user facing applications)
- Hybrid Data Engineer/Data Scientist (data pipelining)
- Hybrid Data Visualization Expert (communication, storytelling)
- Data Science Platforms & Tools Developer (supporting role)

Evolution of Data Science Roles



More or less engineering focus? More or less statistics focus? ...

Software Engineering Specializations

- Architectures
- Requirements engineers
- Testers
- Site reliability engineers
- Devops
- Safety
- Security
- UIX
- Distributed systems, cloud
- ...

Needed Roles in Depression Prognosis Projects?



Common other Roles in ML-Enabled Systems?

- Domain specialists
- Business, management, marketing
- Project management
- Designers, UI experts
- Operations
- Safety, security specialist
- Big data specialist
- Lawyers
- Social scientists, ethics
- ...

Interdisciplinary Teams

Unicorns -> Teams

- Domain experts
- Data scientists
- Software engineers
- Operators
- Business leaders

Necessity of Groups

- Division of labor
- Division of expertise (e.g., security expert, ML expert, data cleaning expert, database expert)

Team Issues discussed Today

- Process costs
- Groupthink
- Social loafing
- Multiple/conflicting goals

Team Issue: Process Costs

Case Studies

Disclaimer: All pictures represent abstract developer groups or products to give a sense of scale; they are not necessarily the developers of those products or developers at all.

How to structure teams?

Microblogging platform; 3 friends



How to structure teams?

Banking app; 15 developers and data analysts



How to structure teams?

Mobile game; 50ish developers?



How to structure teams?

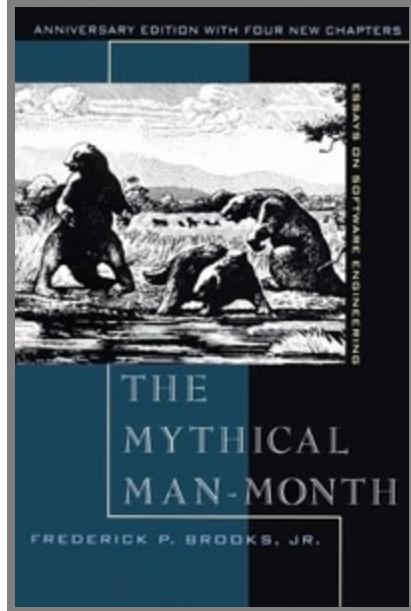
Mobile game; 200ish developers; distributed teams?

How to structure teams?

Self-driving cars; 1200 developers and data analysts



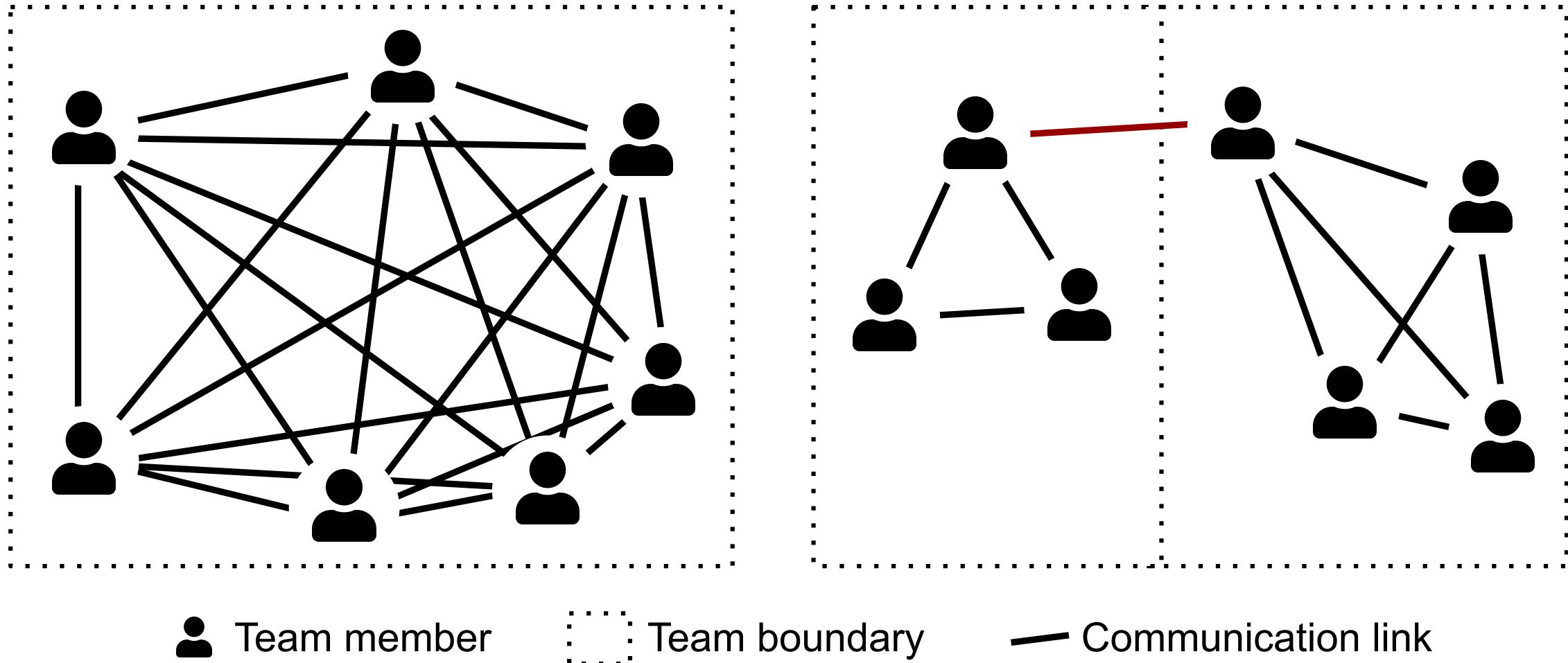
Mythical Man Month



Brooks's law: Adding manpower to a late software project makes it later

1975, describing experience at IBM developing OS/360

Process Costs



≡ $n(n - 1) / 2$ communication links within a team

Brook's Surgical Teams

- Chief programmer – most programming and initial documentation
- Support staff
 - Copilot: supports chief programmer in development tasks, represents team at meetings
 - Administrator: manages people, hardware and other resources
 - Editor: editing documentation
 - Two secretaries: one each for the administrator and editor
 - Program clerk: keeps records of source code and documentation
 - Toolsmith: builds specialized programming tools
 - Tester: develops and runs tests
 - Language lawyer: expert in programming languages, provides advice on producing optimal code.

Speaker notes

Would assume unicorns in today's context.



Microsoft's Small Team Practices

- Vision statement and milestones (2-4 month), no formal spec
- Feature selection, prioritized by market, assigned to milestones
- Modular architecture
- Allows small federated teams (Conway's law)
- Small teams of overlapping functional specialists

(Windows 95: 200 developers and testers, one of 250 products)

Microsoft's Feature Teams

- 3-8 developers (design, develop)
- 3-8 testers (validation, verification, usability, market analysis)
- 1 program manager (vision, schedule communication; leader, facilitator) – working on several features
- 1 product manager (marketing research, plan, betas)

Microsoft's Process

- "Synchronize and stabilize"
- For each milestone
 - 6-10 weeks feature development and continuous testing frequent merges, daily builds
 - 2-5 weeks integration and testing ("zero-bug release", external betas)
 - 2-5 weeks buffer

Agile Practices (e.g., Scrum)

- 7 ± 2 team members, collocated
- self managing
- Scrum master (potentially shared among 2-3 teams)
- Product owner / customer representative

Large teams (29 people) create around six times as many defects as small teams (3 people) and obviously burn through a lot more money. Yet, the large team appears to produce about the same amount of output in only an average of 12 days' less time. This is a truly astonishing finding, though it fits with my personal experience on projects over 35 years. -

Phillip Amour, 2006, CACM 49:9

Establish communication patterns

- Avoid overhead
 - Ensure reliability
 - Constraint latency
-
- e.g. Issue tracker vs email; online vs face to face

Establishing Interfaces

- When dividing work, need to agree on interface
- Common source of mismatch and friction
- **Examples?**
 - Team A uses data produced by Team B
 - Team C deploys model produced by team A
 - Team D uses model and needs to provide feedback to Team A
 - Team D waits for improvement/feature from model A
- Ideally interfaces are stable and well documented

Awareness

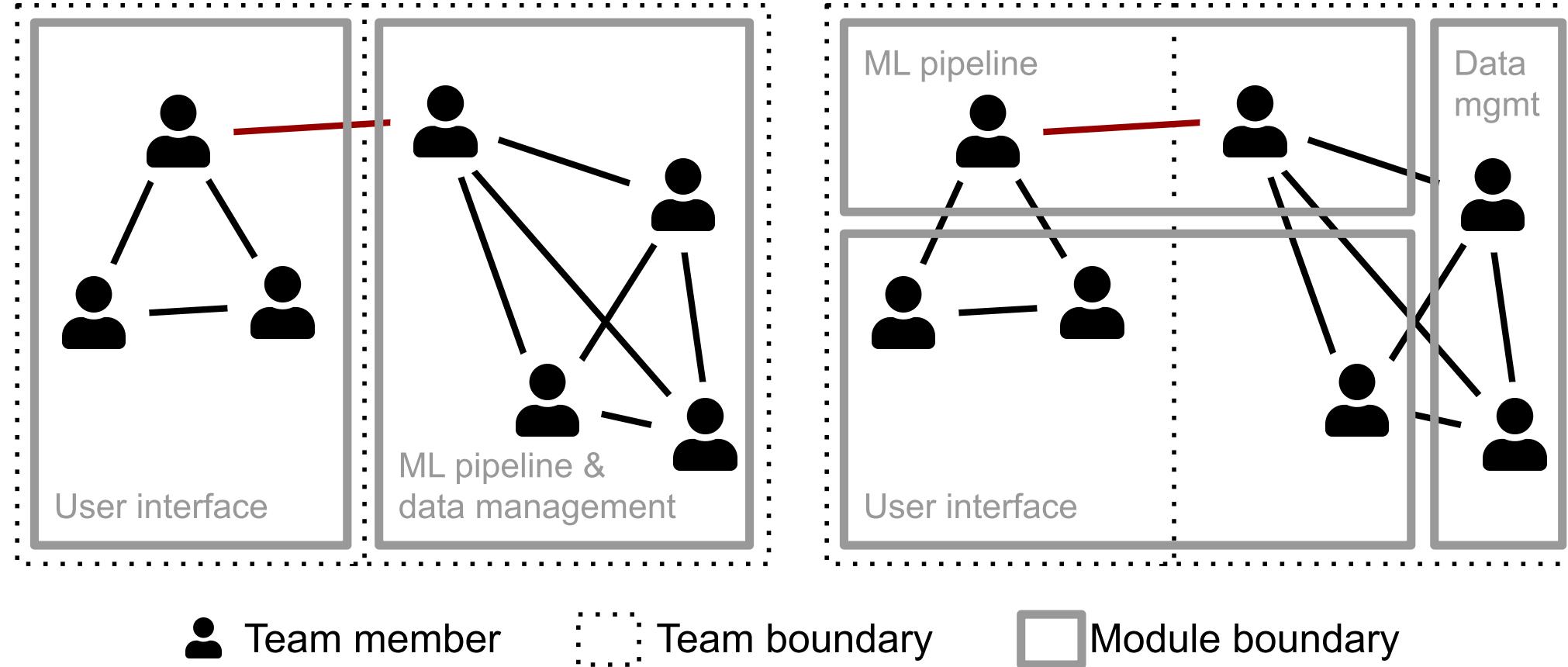
- Notifications
- Brook's documentation book
- Email to all
- Code reviews

Conway's Law

“Any organization that designs a system (defined broadly) will produce a design whose structure is a copy of the organization's communication structure.” – Mel Conway, 1967

“If you have four groups working on a compiler, you'll get a 4-pass compiler.”

Congruence



Structural congruence, Geographical congruence, Task congruence,
IRC communication congruence

Engineering Recommendations for Structuring AI-Enabled Systems

- Decompose the system
- Independent components (e.g. microservices)
- Isolate AI if possible
- Clear, stable interfaces, minimal coupling
- Monitoring to observe contracts and quality

Team Structure for Transcription Service?

the-changelog-318

← Dashboard | Quality: High ⓘ

Last saved a few seconds ago

... Share

00:00 ⓘ Offset 00:00 01:31:27

Play Back 5s 1x Volume

NOTES
Write your notes here

Speaker 5 ► 07:44

Yeah. So there's a slight story behind that. So back when I was in, uh, Undergrad, I wrote a program for myself to measure a, the amount of time I did data entry from my father's business and I was on windows at the time and there wasn't a function called time dot [inaudible] time, uh, which I needed to parse dates to get back to time, top of representation, uh, I figured out a way to do it and I gave it to what's called the python cookbook because it just seemed like something other people could use. So it was just trying to be helpful. Uh, subsequently I had to figure out how to make it work because I didn't really have to. Basically, it bothered me that you had to input all the locale information and I figured out how to do it over the subsequent months. And actually as a graduation gift from my Undergrad, the week following, I solved it and wrote it all out.

Speaker 5 ► 08:38

And I asked, uh, Alex Martelli, the editor of the Python Cookbook, which had published my original recipe, a, how do I get this into python? I think it might help

How did we do on your transcript? ⭐⭐⭐⭐⭐

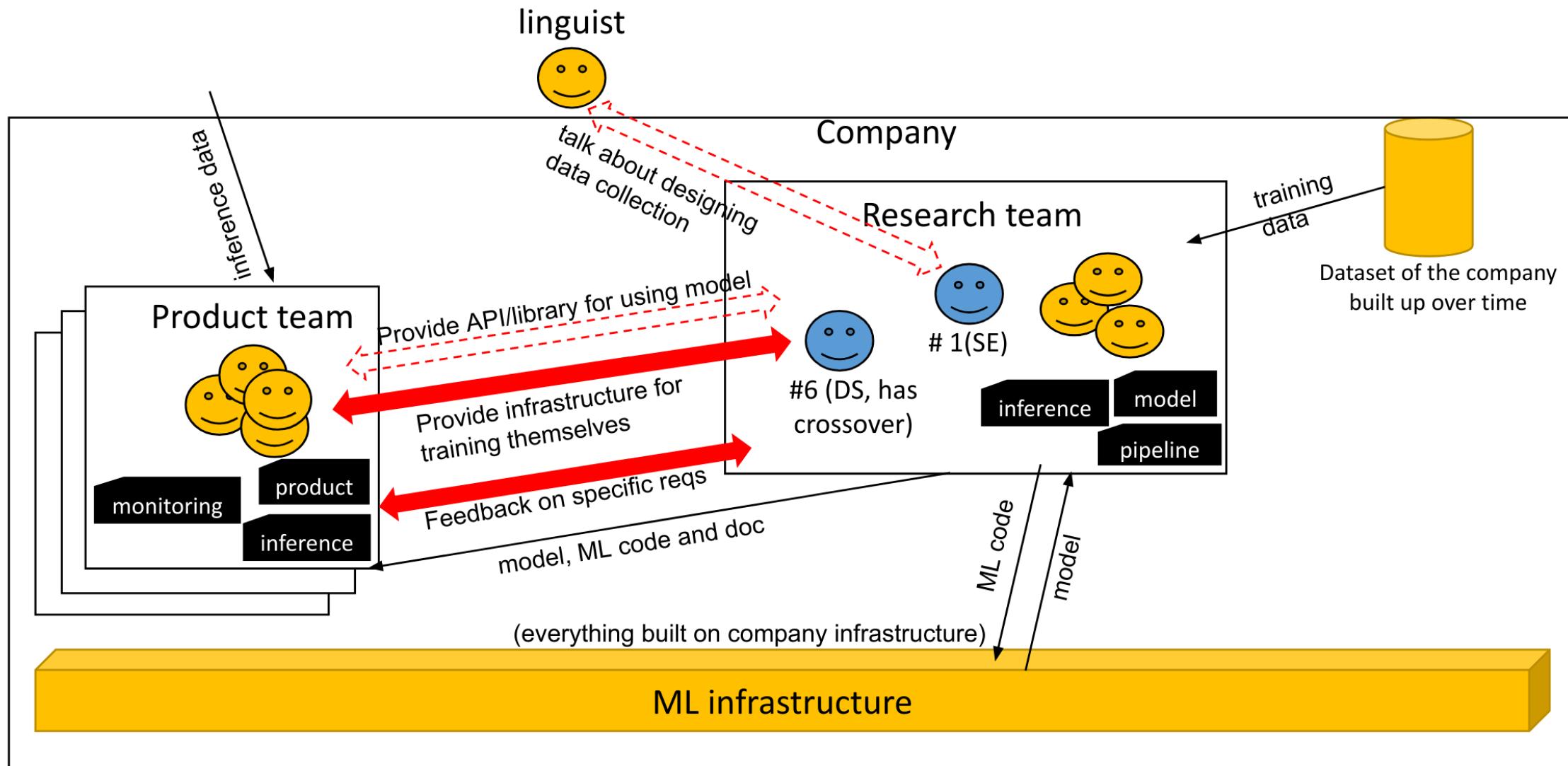
Breakout: Team Structure for Depression Prognosis

In groups, tagging team members, discuss and post in #lecture:

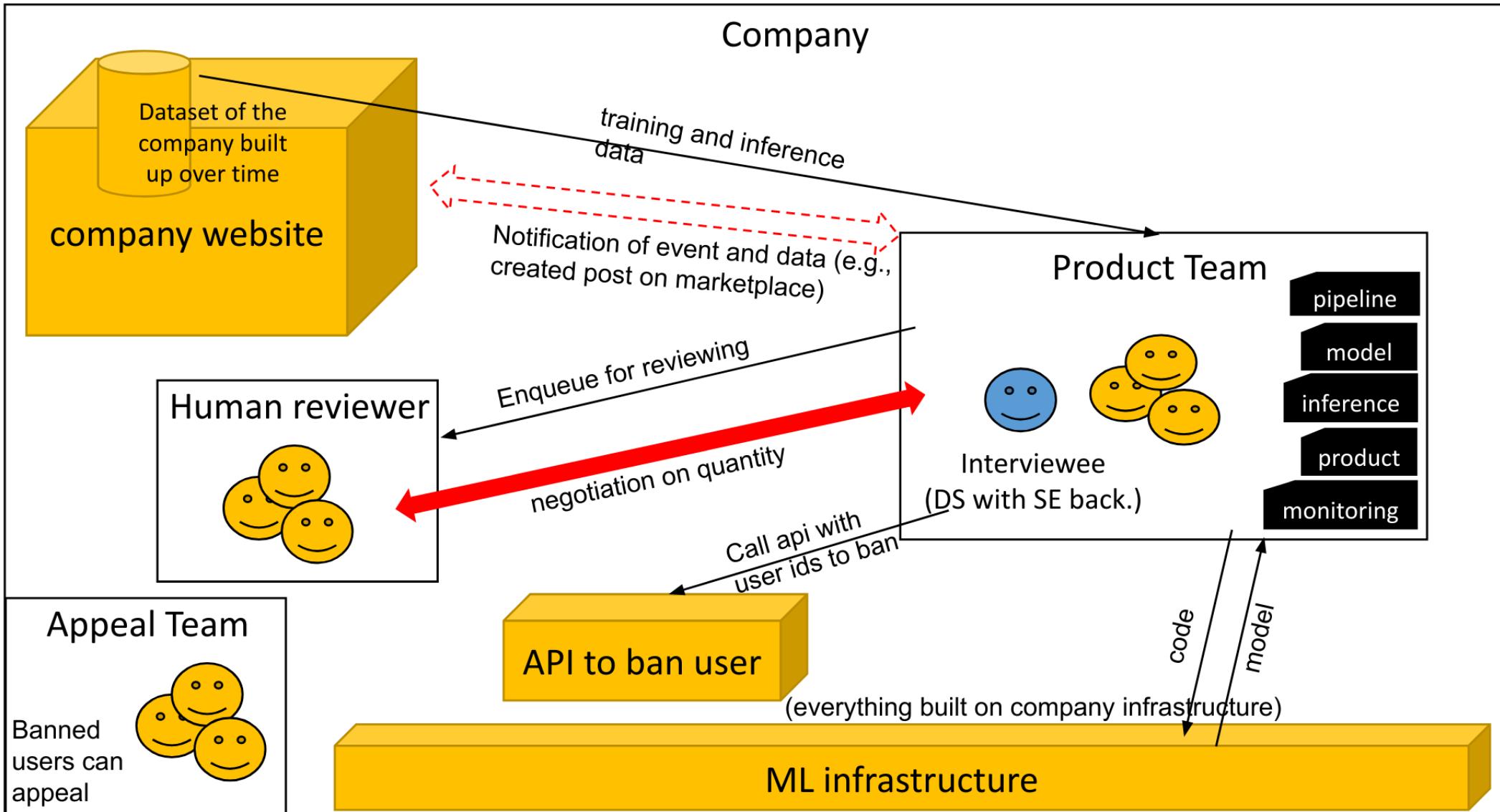
- How to decompose the work into teams?
- What roles to recruit for the teams

Story Time: Conflicts at the Interface between Teams

Research team in big tech company creates NLP-ML components used within various products in the company



Building product to detect scam on company platform



Common Challenge: Establishing Interfaces

- Formal vs informal agreements?
 - Service level agreements and automated enforcement?
 - Close collaboration vs siloed teams?
-
- Many concerns: prediction accuracy, generalization, execution time, scalability, data quality, data quantity, feedback latency, privacy, explainability, time estimation, ...
 - Formal agreements and enforcement expensive, slowing development? see technical debt

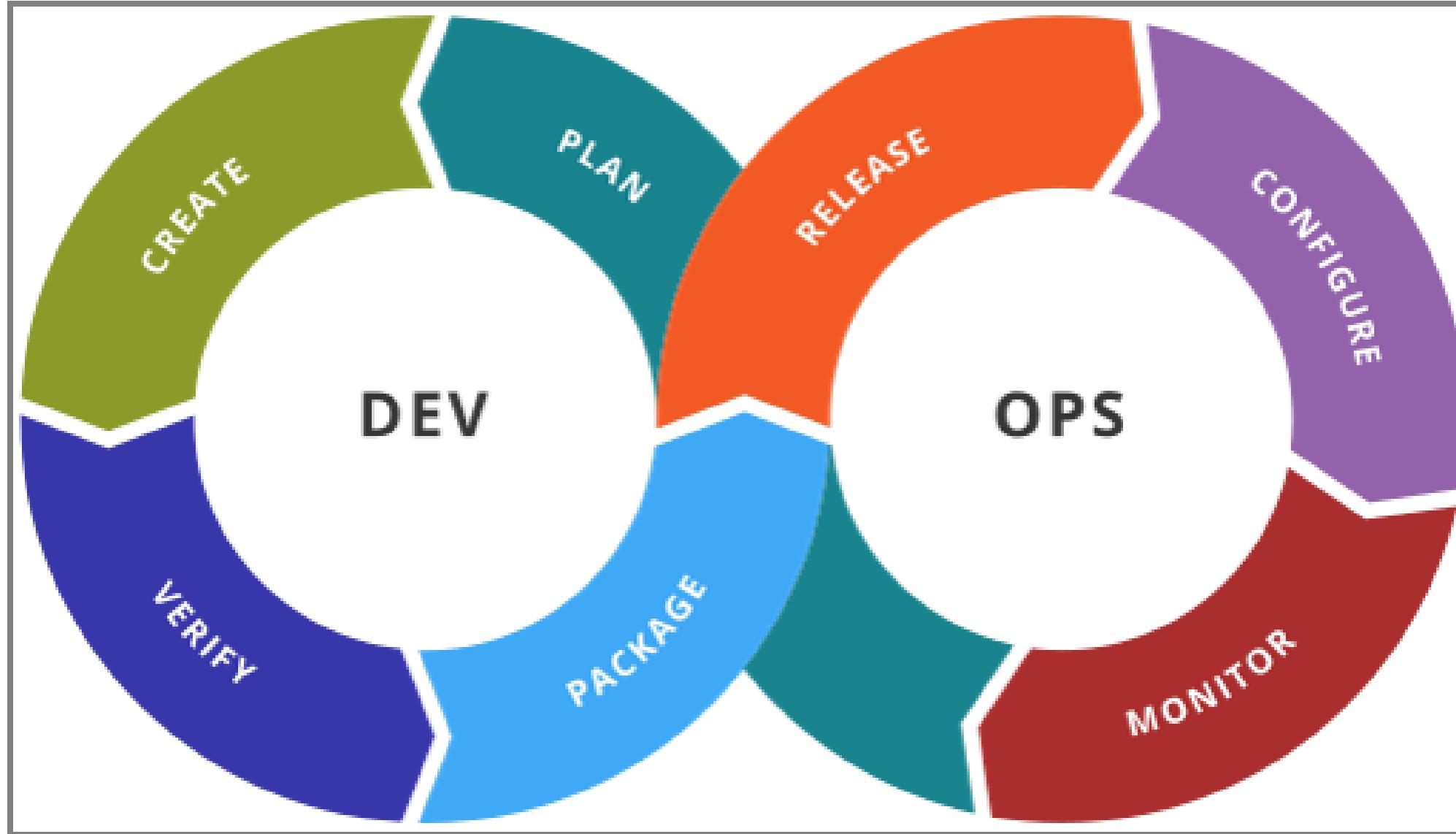
Common Collaboration Points

1. Understanding system requirements and ML capabilities
2. Understanding ML-specific requirements at the system level, reasoning about feedback loops
3. Project planning and architecture design
4. Data needs, data quality, data meaning
5. Documenting model output
6. Planning and monitoring for drift
7. Planning ML component QA (offline, online, monitoring)
8. Planning system QA (integration, interaction, safety, feedback loops)
9. Tool support for data scientists
10. From prototype to production (pipelines, versioning, operations, user interactions, ...)

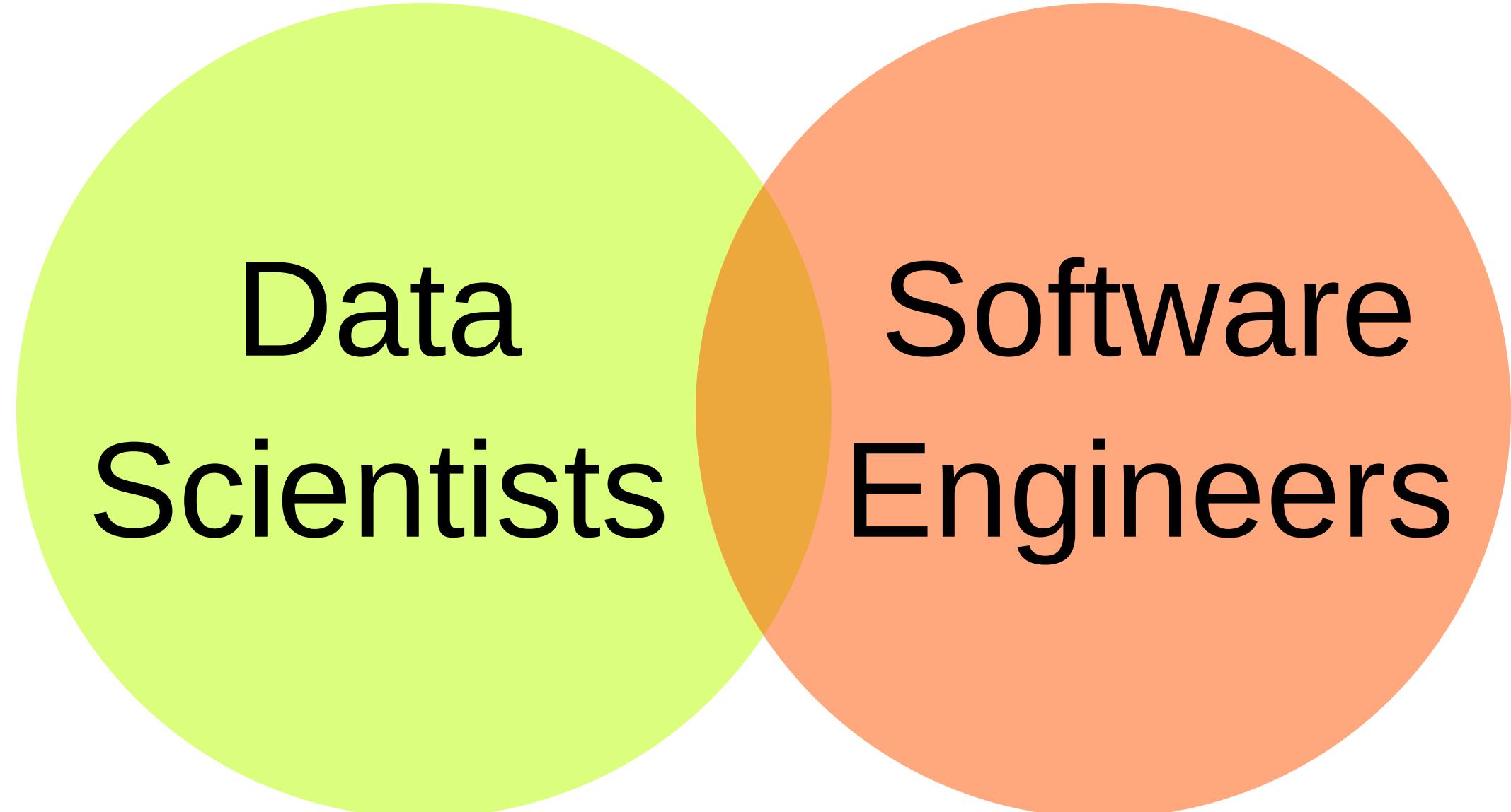
Team issues: Multiple/conflicting goals

(Organization of Interdisciplinary Teams)

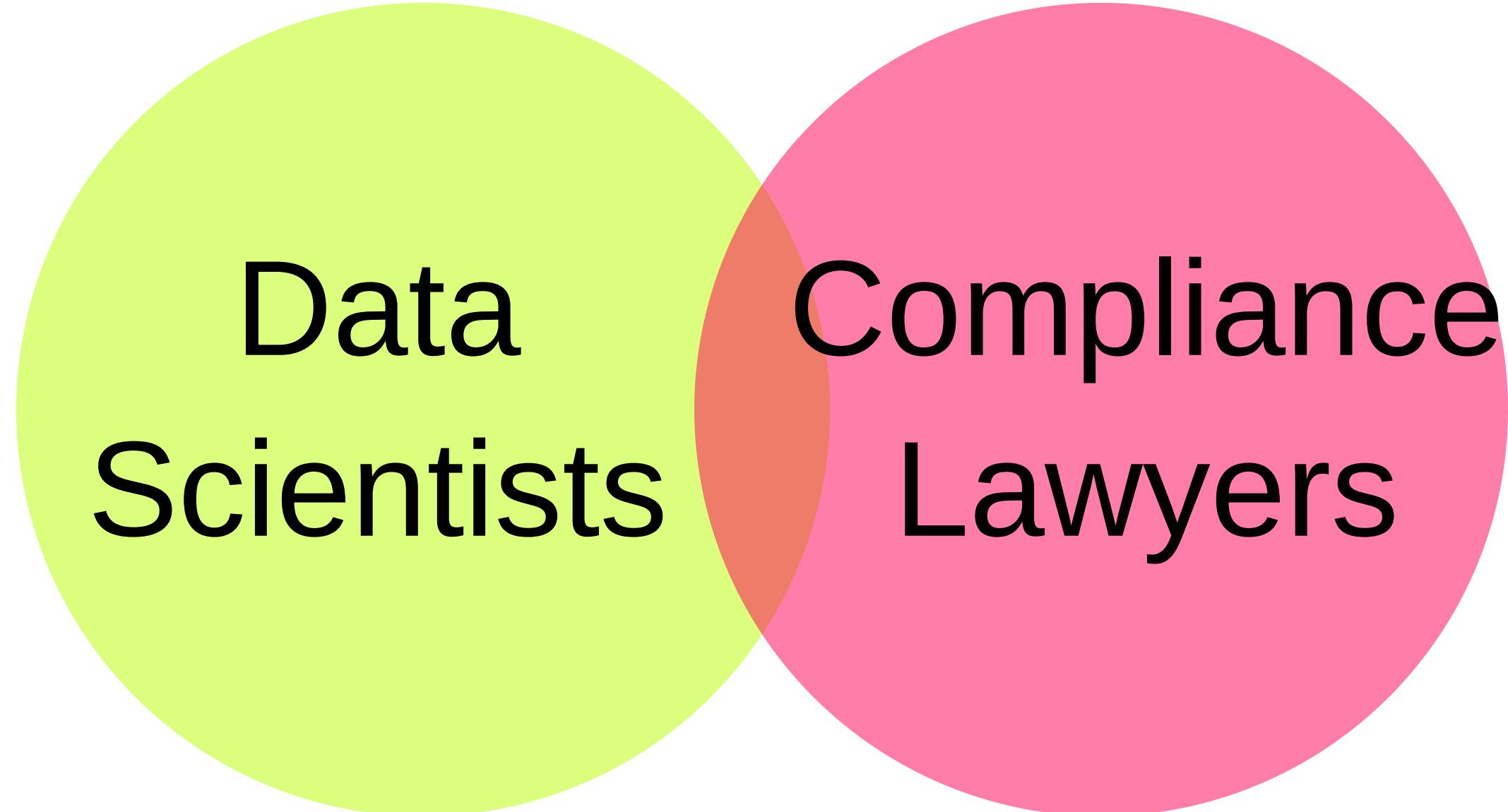
Conflicting Goals?



Conflicting Goals?



Conflicting Goals?



Conflicting Goals?



How to Address Goal Conflicts?



T-Shaped People

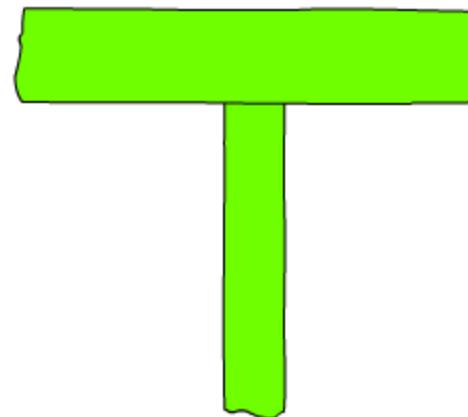
Broad-range generalist + Deep expertise



"I-shaped"
Expert at one thing



Generalist
Capable in a lot of things
but not expert in any



"T-shaped"
Capable in a lot of things
and expert in one of them

≡ Figure: Jason Yip. [Why T-shaped people?](#). 2018

T-Shaped People

Broad-range generalist + Deep expertise

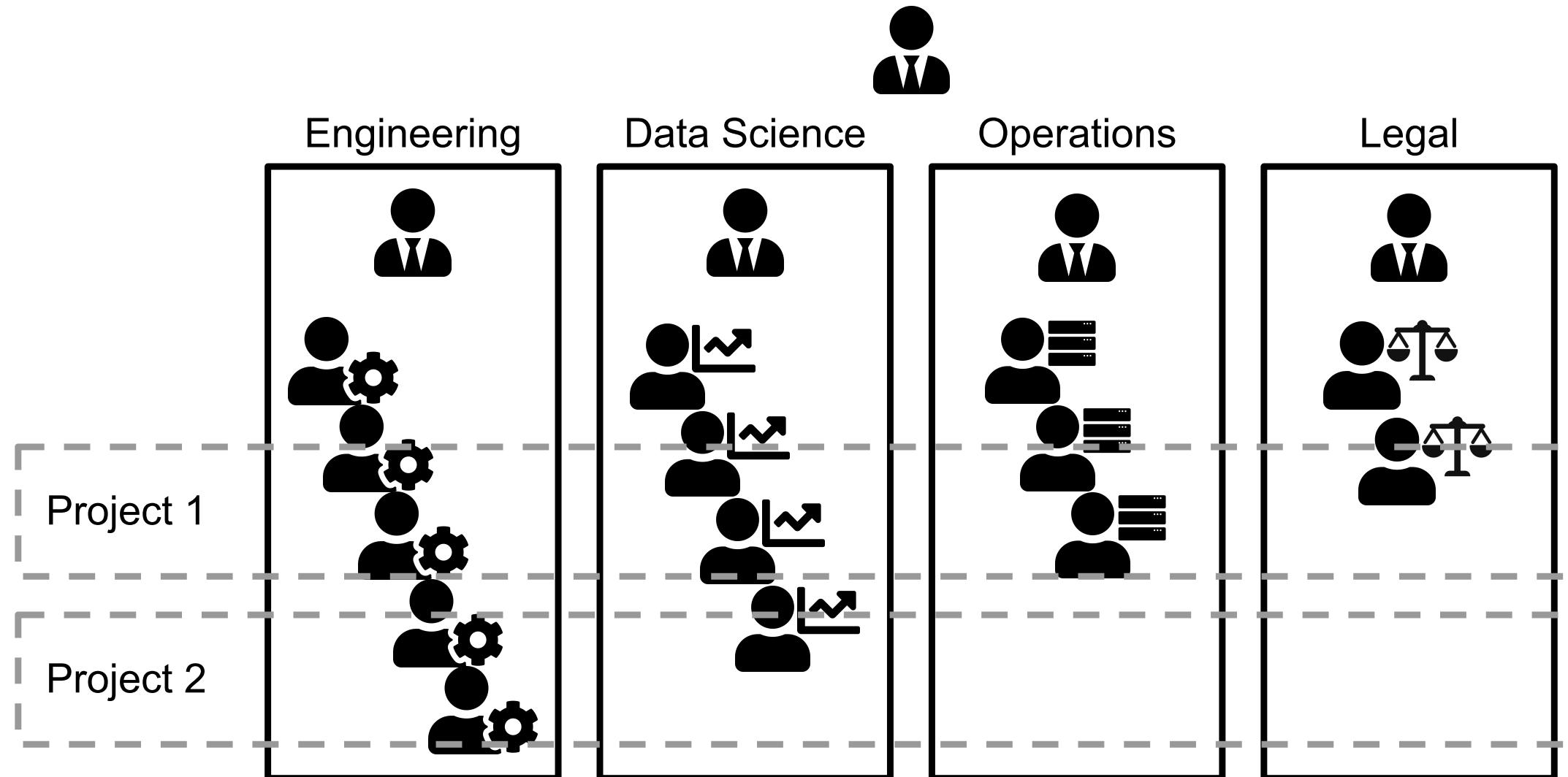
Example:

- Basic skills of software engineering, business, distributed computing, and communication
- Deep skills in deep neural networks (technique) and medical systems (domain)

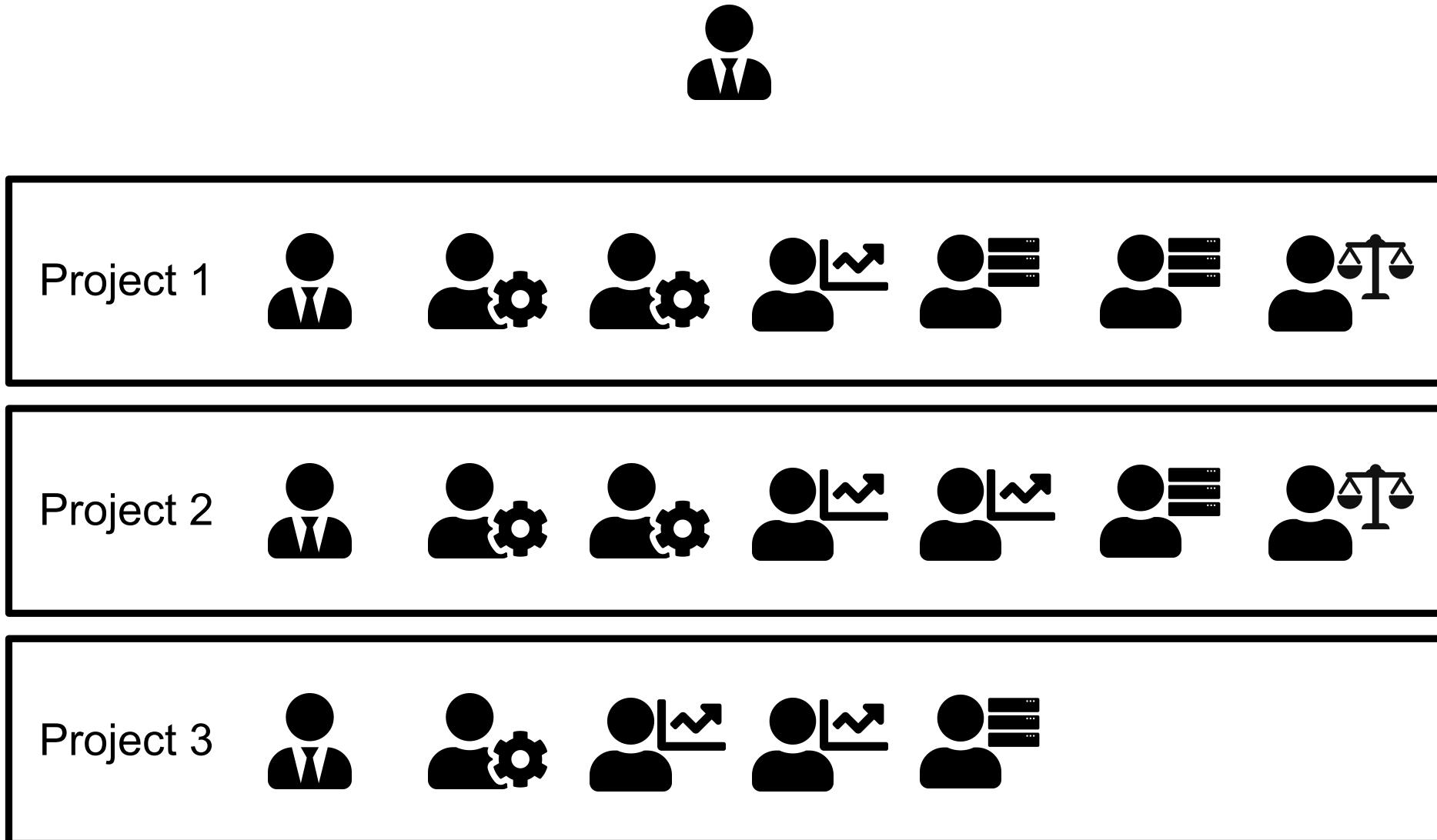
Team Composition

- Cover deep expertise in all important areas
- Aim for overlap in general skills
 - Fosters communication, same language

Matrix Organization



Project Organization



Case Study: Brøderbund

As the functional departments grew, staffing the heavily matrixed projects became more and more of a nightmare. To address this, the company reorganized itself into “Studios”, each with dedicated resources for each of the major functional areas reporting up to a Studio manager. Given direct responsibility for performance and compensation, Studio managers could allocate resources freely.

The Studios were able to exert more direct control on the projects and team members, but not without a cost. The major problem that emerged from Brøderbund’s Studio reorganization was that members of the various functional disciplines began to lose touch with their functional counterparts. Experience wasn’t shared as easily. Over time, duplicate effort began to appear.

Mantle, Mickey W., and Ron Lichty. [Managing the unmanageable: rules, tools, and insights for managing software people and teams](#). Addison-Wesley Professional, 2012.

Specialist Allocation (Organizational Architectures)

- Centralized: development teams consult with a core group of specialists when they need help
- Distributed: development teams hire specialists to be a first-class member of the team
- Weak Hybrid: centralized group of specialists and teams with critical applications hire specialists
- Strong Hybrid: centralized group of specialists and most teams also hire specialists

Tradeoffs?



Example: Security Roles

- Everyone: “security awareness” – buy into the process
- Developers: know the security capabilities of development tools and use them, know how to spot and avoid relevant, common vulnerabilities
- Managers: enable the use of security practices
- Security specialists: everything security

Allocation of Data Science/Software Engineering Expertise?



Commitment & Accountability

- Conflict is useful, expose all views
- Come to decision, commit to it
- Assign responsibilities
- Record decisions and commitments; make record available

Bell & Hart - 8 Causes of Conflict

- Conflicting resources.
- Conflicting styles.
- Conflicting perceptions.
- Conflicting goals.
- Conflicting pressures.
- Conflicting roles.
- Different personal values.
- Unpredictable policies.

Understanding causes helps design interventions. Examples?

Agile Techniques to Address Conflicting Goals?



Team issues: Groupthink



Groupthink

- Group minimizing conflict
- Avoid exploring alternatives
- Suppressing dissenting views
- Isolating from outside influences
- -> Irrational/dysfunctional decision making

Example: Time and Cost Estimation



Example: Use of Hype Technology

(agile, block chain, machine learning, devops, AIOps, ...)



Causes of Groupthink

- High group cohesiveness, homogeneity
- Structural faults (insulation, biased leadership, lack of methodological exploration)
- Situational context (stressful external threats, recent failures, moral dilemmas)

Symptoms

- Overestimation of ability: invulnerability, unquestioned belief in morality
- Closed-mindedness: ignore warnings, stereotyping; innovation averse Pressure toward uniformity: self-censorship, illusion of unanimity, ...



Diversity

*“Men and women have different viewpoints, ideas, and market insights, which enables better **problem solving**. A gender-diverse workforce provides easier **access to resources**, such as various sources of credit, multiple sources of information, and wider industry knowledge. A gender-diverse workforce allows the company to serve an **increasingly diverse customer base**. Gender diversity helps companies **attract and retain talented women**.”*

*“Cultural diversity leads to **process losses** through task conflict and decreased social integration, but to **process gains** through increased creativity and satisfaction.”*

Groupthink and AI-Enabled System Projects?



Groupthink and AI

- Need of AI
- Selection of learning method
- Fairness
- Safety requirements (e.g. Pitt delivery robot)
- Ethics

Mitigation Strategies

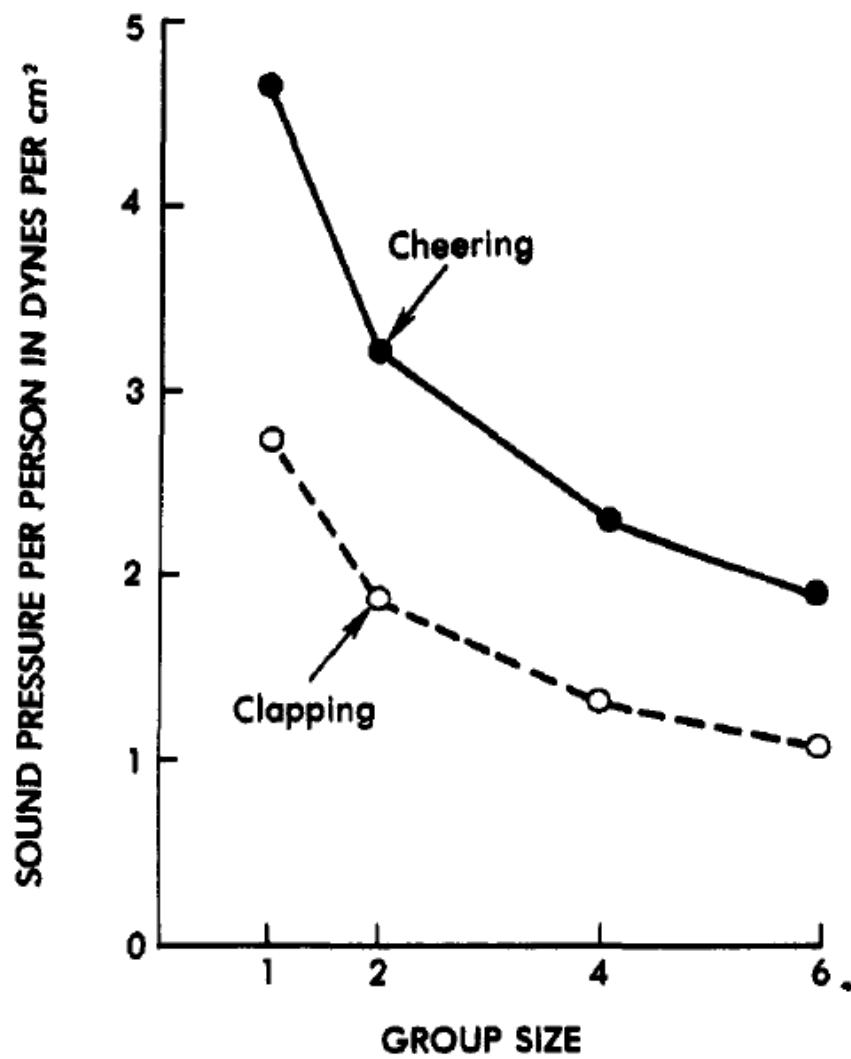


Mitigation Strategies

- Diversity in team composition
- Culture of open conflicts
- Appoint devil's advocate in discussions, moderate and rotate speaker order, leaders hide opinions in discussions
- Involve outside experts
- Always request a second solution
- Monitoring and process measurement
- Agile techniques as planning poker, on-site customer

Team issues: Social loafing





Latane, Bibb, Kipling Williams, and Stephen Harkins. "[Many hands make light the work: The causes and consequences of social loafing.](#)" Journal of personality and social psychology 37.6 (1979): 822.

Social Loafing

- People exerting less effort within a group
- Reasons
 - Diffusion of responsibility
 - Motivation
 - Dispensability of effort / missing recognition
 - Avoid pulling everybody / "sucker effect"
 - Submaximal goal setting
- “Evaluation potential, expectations of co-worker performance, task meaningfulness, and culture had especially strong influence”

Karau, Steven J., and Kipling D. Williams. "[Social loafing: A meta-analytic review and theoretical integration](#)." Journal of personality and social psychology 65.4 (1993): 681.

Mitigation Strategies



Mitigation Strategies

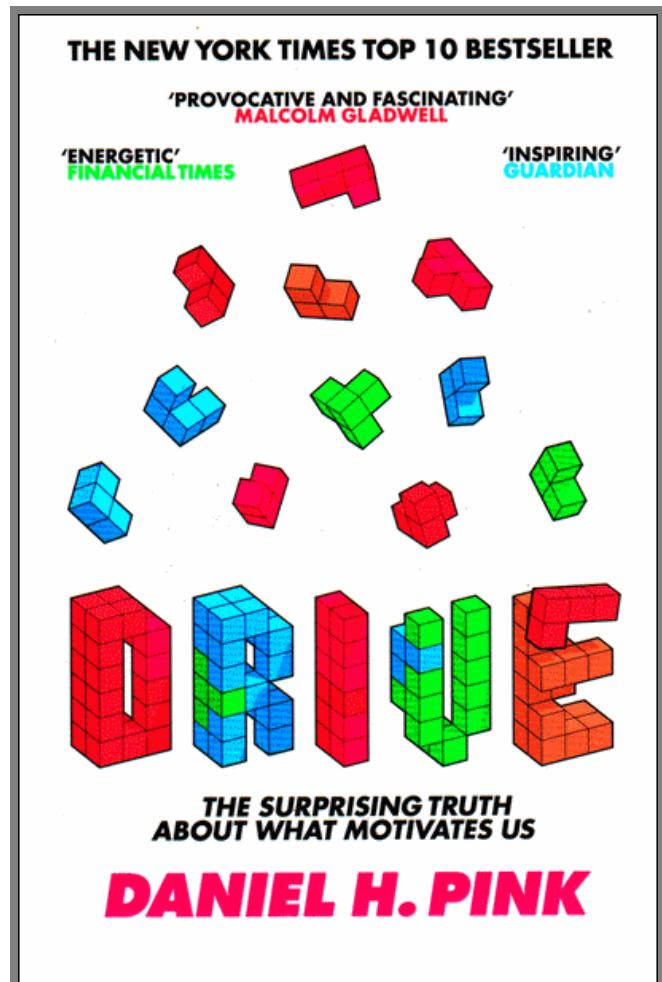
- Involve all team members, colocation
- Assign specific tasks with individual responsibility
 - Increase identifiability
 - Team contracts, measurement
- Provide choices in selecting tasks
- Promote involvement, challenge developers
- Reviews and feedback
- Team cohesion, team forming exercises
- Small teams

Responsibilities & Buy-In

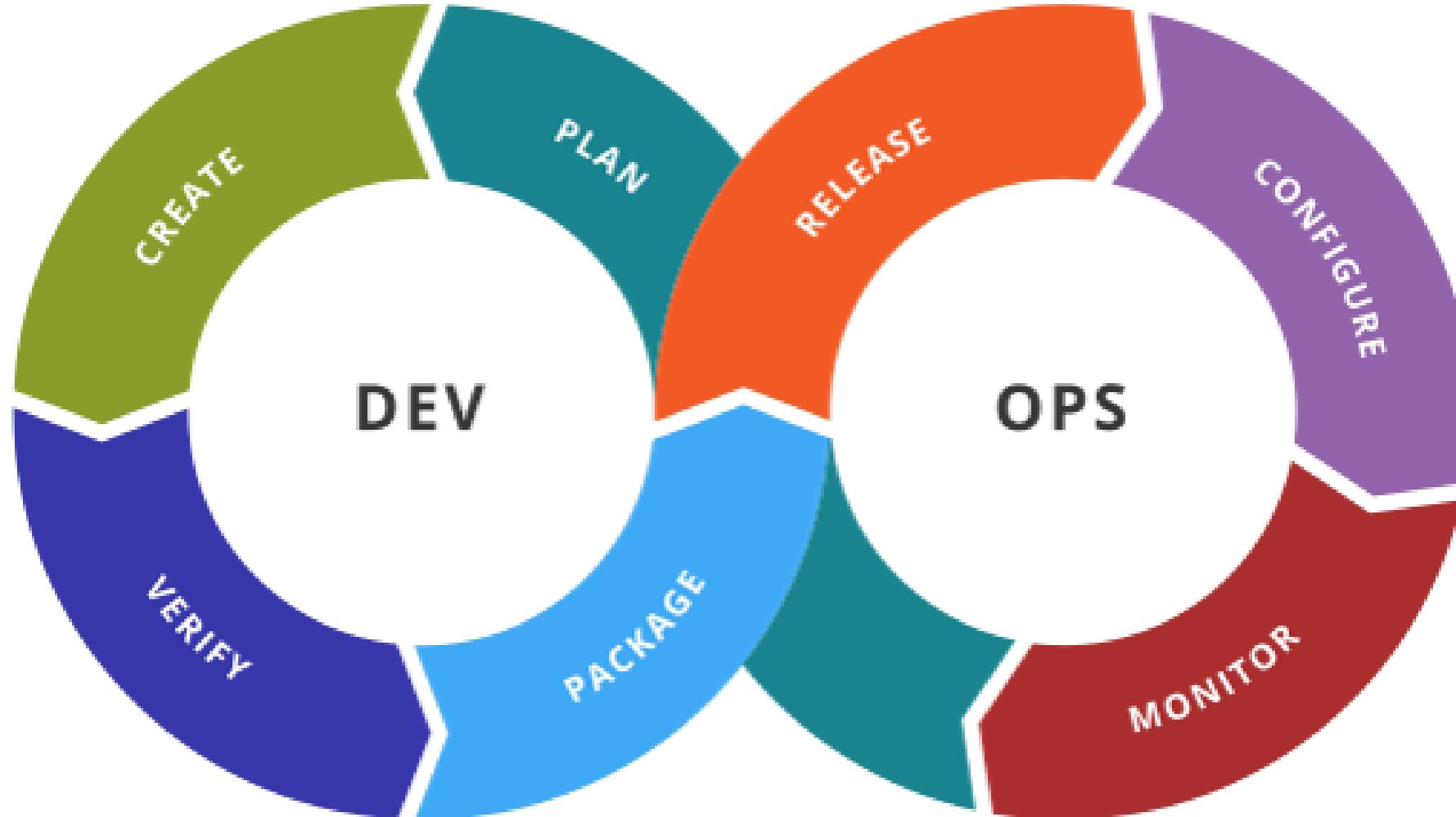
- Involve team members in decision making
- Assign responsibilities (ideally goals not tasks)
- Record decisions and commitments; make record available

Motivation

Autonomy * Mastery * Purpose



Learning from DevOps



DevOps: A culture of collaboration

- Overcome historic role and goal conflicts between developers and operators
- Joint planning for operations, joint responsibilities for testing and deployment
- Joint goals, joint vocabulary
- Joint tools (e.g., Docker, versioning, A/B testing, monitoring)
- Mutual benefits (faster releases, more telemetry, improved reliability, fewer conflicts)
- T-shaped professionals

Changing practices and culture is hard

- Ingrained "us vs them" and blame culture
- Inertia is hard to overcome ("this is how we always did things")
- Learning cost for new concepts and tools
- Extra effort for new practices (e.g., testing)
- Overwhelmed with current tasks, no time to learn/change
- Poor adoption may cause more costs than benefits

Working on Culture Change

- Bottom-up and top-down change possible
- Often introduced by individual advocates, convincing others
- Always requires supportive management
- Education helps generate buy-in
- Consultants can help with adoption and learning

- Demonstrate benefits in one small project, promote from there

Beyond DevOps

Summary

- Team dysfunctions well studied
- Know the signs, know the interventions
- Small teams, crossfunctional teams
 - Deliberately create teams, respect congruence, define interfaces
 - Hire T-shaped developers
- Create awareness and accountability

Further Readings

- Brooks Jr, Frederick P. [The mythical man-month: essays on software engineering](#). Pearson Education, 1995.
- DeMarco, Tom, and Tim Lister. [Peopleware: productive projects and teams](#). Addison-Wesley, 2013.
- Mantle, Mickey W., and Ron Lichty. [Managing the unmanageable: rules, tools, and insights for managing software people and teams](#). Addison-Wesley Professional, 2019.
- Lencioni, Patrick. "[The five dysfunctions of a team: A Leadership Fable.](#)" Jossey-Bass (2002).
- Rakova, Bogdana, Jingying Yang, Henriette Cramer, and Rumman Chowdhury. "[Where responsible AI meets reality: Practitioner perspectives on enablers for shifting organizational practices.](#)" Proceedings of the ACM on Human-Computer Interaction 5, no. CSCW1 (2021): 1-23.
- Luz, Welder Pinheiro, Gustavo Pinto, and Rodrigo Bonifácio. "[Adopting DevOps in the real world: A theory, a model, and a case study.](#)" Journal of Systems and Software 157 (2019): 110384.
- Sambasivan, Nithya, Shivani Kapania, Hannah Highfill, Diana Akrong, Praveen Paritosh, and Lora M. Aroyo. "["Everyone wants to do the model work, not the data work": Data Cascades in High-Stakes AI](#)". In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, pp. 1-15. 2021.