Team & Collaboration

Presentation for COMP 599: Designing and Building Intelligent Systems

1. **Motivation**
2. **Tools**
3. **Roles**
4. **Activity**
5. **Examples**

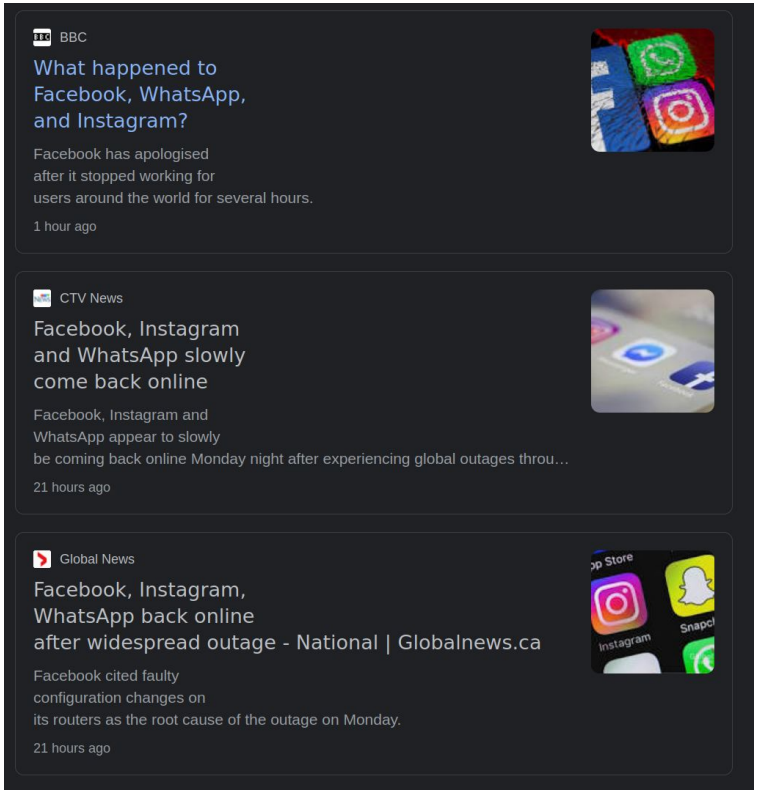
Martin Pömsl
October 5th, 2021

Motivation

Collaboration is all about communication.

Yesterday, Facebook experienced problems that disrupted the communication between

- Users and services
- Employees and servers
- Employees in different physical locations
- Employees and electronic on-site locks



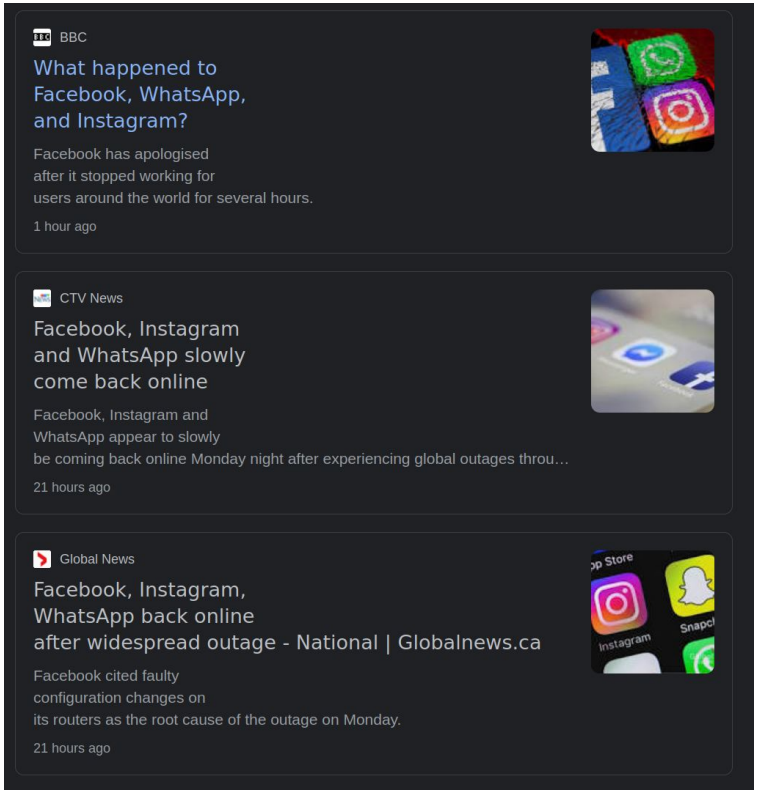
Source: https://9to5mac.com/2021/10/05/facebook-outage-cause-mistake/

Motivation

Choose communication tools wisely:

- If you have too many tools, attention gets divided and communication becomes ineffective

- If you have too few tools, you become reliant on them and they might not be ideal for every purpose of communication



Tools

Table 2. Reported Technologies in Use: Types, Tools, Uses, and Types of Common Ground

Types of Technologies	Example Tools	Uses	Types of Common Ground
Co-Editing Systems	Google Docs, Google Sheets	- teach, tell and explain tabular data sets, DS methods	Content
		- brainstorm, compile and compare different perspectives	Content
		- summarize progress, document decision rationales	Content
		- plan schedule, moderate progress	Process
		- specify work division	Process
Communication Systems	Emails	- exchange meeting notes and align quick decisions	Content
		- schedule meetings and exchange meeting agendas	Process
	Skype, Zoom	- present and discuss findings	Content
		- make decisions, discuss and align goals and protocols	Content, Process
		- exchange quick questions and clarifications	Content
	Slack	- propose and brainstorm ideas	Content
Co-Creation Systems with Version Control	GitHub	- conduct collaborative analytics	Content, Process
		- curate and share work processes	Content, Process
Repository Systems	PRO-ACT database	- curate and share data sets	Content
Expertise Systems	MeSH terms	- query domain-specific knowledge	Content

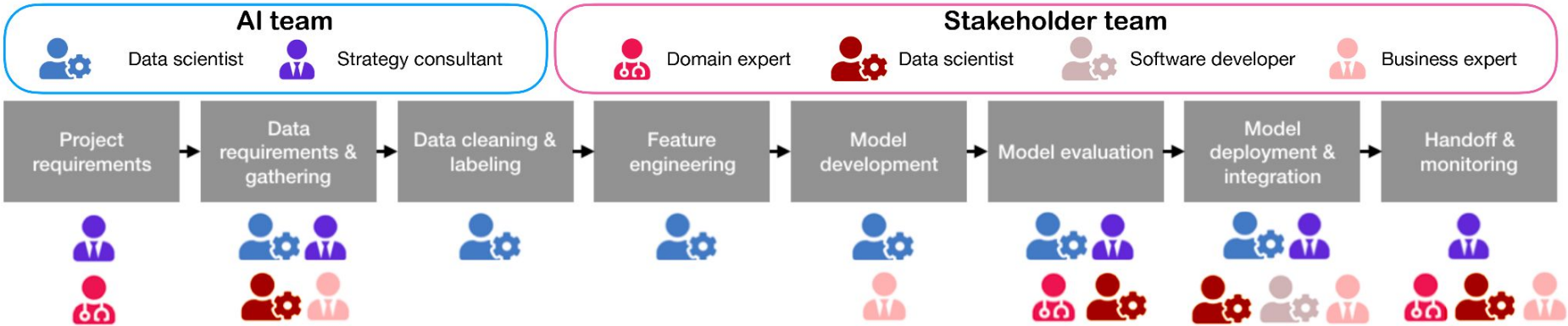
Yaoli Mao, Dakuo Wang, Michael Muller, Kush R. Varshney, Ioana Baldini, Casey Dugan, and Aleksandra Mojsilović. 2019. How Data Scientists Work Together With Domain Experts in Scientific Collaborations: To Find The Right Answer Or To Ask The Right Question? *Proc. ACM Hum.-Comput. Interact.* 3, GROUP, Article 237 (December 2019), 23 pages. DOI: https://doi.org/10.1145/3361118. Table 2.

Tools

Involved roles	Corresponding phase	Tools	Purpose of the conversation
	Project requirement	Internal data warehosuse	Interactive Q&A session. Domain experts explain concepts from raw data (e.g., meaning of each column)
		Powerpoint	Education session. To bridge the knowledge gap and laying the groundwork for upcoming collaboration
	Model development	Powerpoint	Sync up and report the progress update. Sharing results of different configuration of models
		Github, Cloud contents management tool	Sharing high-level details of how the model works and how it maps to the business problem
		Github, Word documents	Sharing algorithm and error details of how the model works
	Model evaluation	Powerpoint, Excel, Mural	Finalizing the model. Include additional metrics such as user acceptance testing
	Model deployment		Sharing road maps. At this point, stakeholders understand various metrics in ML and confusion matrix

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. Figure 3.

Roles



- Data Scientists
- Strategy Consultants
- Domain Experts
- Business Experts
- Software Developers

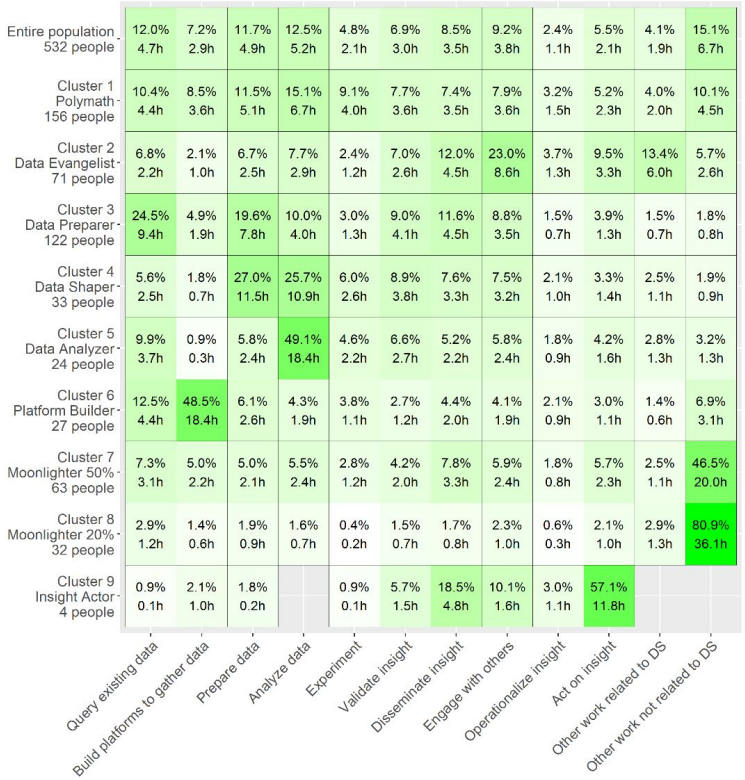
→ Understand roles to communicate

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. Figure 2.

Roles

What is a Data Scientist?

- Survey of 532 data scientists at Microsoft
- Nine distinct role clusters of activities
- Limitation: Data scientists are self-identified



M. Kim, T. Zimmermann, R. DeLine and A. Begel, "Data Scientists in Software Teams: State of the Art and Challenges," in *IEEE Transactions on Software Engineering*, vol. 44, no. 11, pp. 1024-1038, 1 Nov. 2018, doi: 10.1109/TSE.2017.2754374. Table 1.

Activity

- Collaborate as a group (whole class) to write a simple Python script
- Two phases:
  - Phase 1 (5 minutes): Without any communication (no speaking, no Slack)
  - Phase 2 (5 minutes): With communication (Slack: `#team-collaboration-activity`)
- Constraints:
  - Each person can only write two lines of code!

Activity

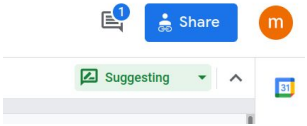
Task: Find arguments x, y ∈ [-500, 500] for the eggholder function that minimize it!

1. Open Google Doc

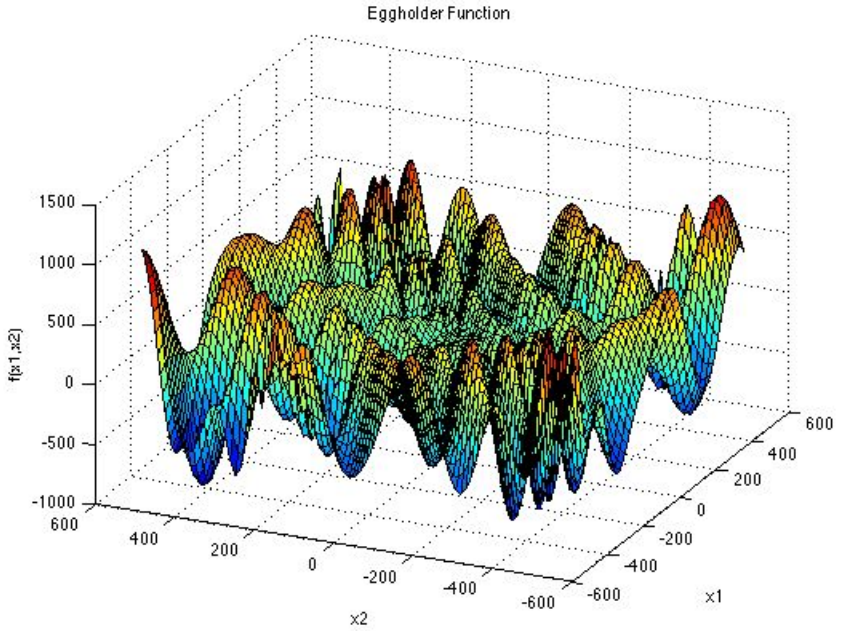
   (link in `#team-collaboration-activity`)

2. Turn on “Suggesting”

   in top right corner



3. Write your two lines of code!



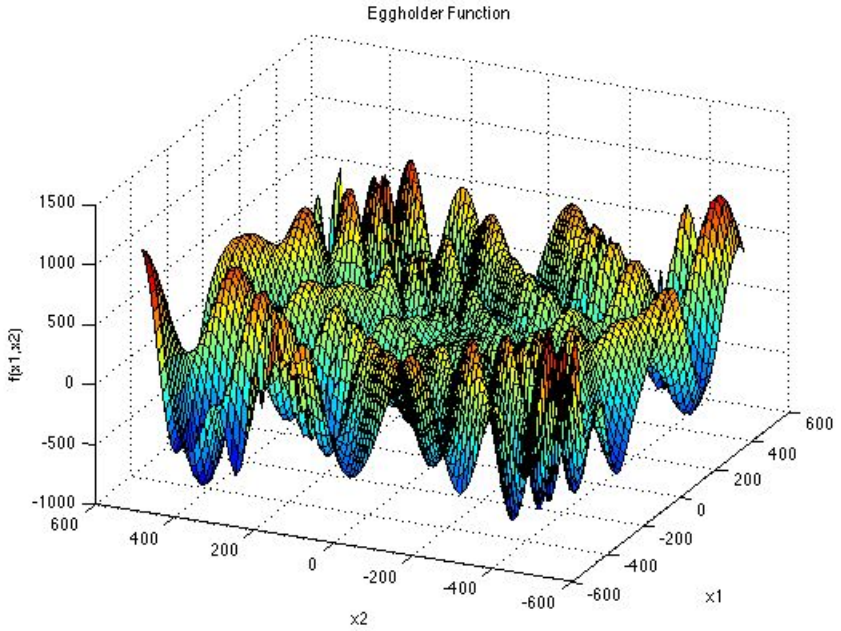
https://www.sfu.ca/~ssurjano/egg.html

Activity: Solution I

Task: Find arguments x, y ∈ [-500, 500] for the eggholder function that minimize it!

```
def minimize_eggholder():

    best_x, best_y = -500, -500
    best_res = 0

    for x in range(-500, 500):
        for y in range(-500, 500):

            res = eggholder(x, y)

            if res < best_res:

                best_res = res
                best_x, best_y = x, y

    print(best_x, best_y, best_res)
```

Output: 482.00 433.00 -950.86

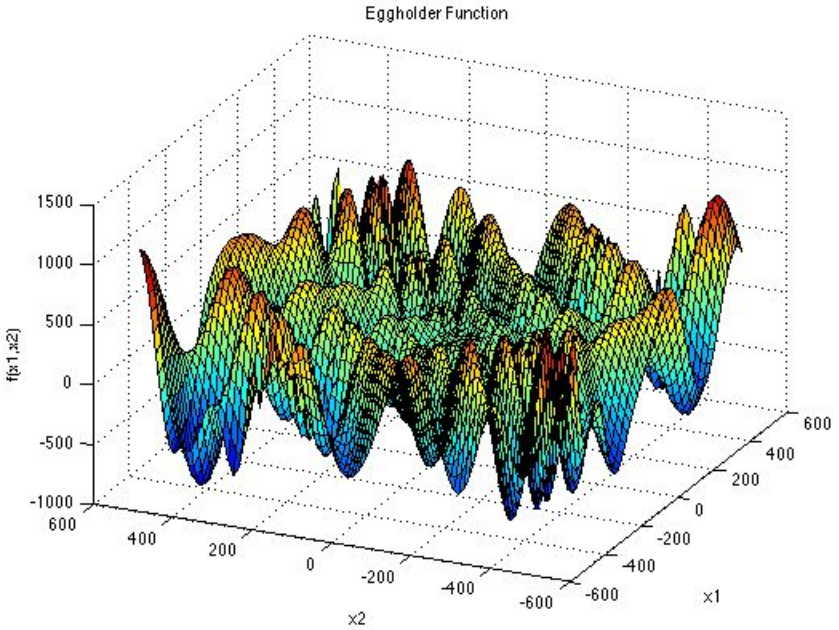


https://www.sfu.ca/~ssurjano/egg.html

Activity: Solution II

Task: Find arguments x, y ∈ [-500, 500] for the eggholder function that minimize it!

```
def minimize_eggholder():

    from scipy import optimize

    bounds = [(-500, 500), (-500, 500)]

    arr_func = lambda a: eggholder(a[0], a[1])

    res = optimize.dual_annealing(arr_func, bounds)

    print(*res.x, res.fun)
```



Output: 482.35 432.87 -956.91

https://www.sfu.ca/~ssurjano/egg.html

Example: AirBnb

- **Roles:** Central data science team with sub-teams facing engineers, marketers
- **Communication:** Initially direct conversation (7 people), now focus on cross-team
- **Use Cases:**
  - Predict Location Relevance
  - Suggest Search Queries

Source: https://venturebeat.com/2015/06/30/how-we-scaled-data-science-to-all-sides-of-airbnb-over-5-years-of-hypergrowth/

Example: Uber

- **Roles:** Product, specialist, research, ML platforms teams
- **Communication:** Annual internal ML conference, reading groups
- **Use Cases:**
  - Predict Demand Localization
  - Uber Eats Suggestions
  - ETA Prediction

Source: https://eng.uber.com/scaling-michelangelo/

Example: LinkedIn

- **Roles:** Data science is full product team with blurry distinctions
- **Communication:** Small teams should sit together, regular contact across teams
- **Use Cases:**
  - People You May Know (PYMK)
  - Career Explorer
  - Who's Viewed My Profile

Source: https://www.datascienceassn.org/sites/default/files/Building%20Data%20Science%20Teams.pdf

Any Questions?

Thanks for Participating!

Martin Pömsl
October 5th, 2021