# Lab 7: How to Devise Research Questions, AKA Choose Your Own XAI Adventure

Jon Dodge

2.22.24

Sources:
*Case Study Research: Design and Methods*
by Robert Yin
*Guide to Advanced Empirical Software Engineering*
edited by Forrest Shull et. al

# Goals

In this session:

1. Brainstorm a shortlist of XAI-oriented RQs (≈10m)
2. Refine your RQs based on question type (≈10m)
3. Refine your RQs to be win-win (≈5m)
4. Identify methods and data sources for your RQ (≈10m)

On your own (or if you finish early):

1. Find 3 pieces of related work.
2. If someone else already answered your RQ:
   Refine it further OR Pick another RQ from your list and repeat step 1

# What to turn in

Form groups of 2-3 students, we will discuss work as a group, but each of you is responsible to submit work **individually**.

Please keep the output of all activities in a single document (one per *person*, not one per *team*) so we can see your work as we forge a good RQ together.

For this reason, if you have an idea and abandon it, we would prefer you *strike it out instead of deleting it.*

# Activity 1: Brainstorm RQs (≈10m)

1. Individually, write down 3-5 RQs (or at least topics you are interested to study).

Examples:

- What information do shoutcasters seek to generate explanations, and where do they find it? [1]
- How do end users think machine-learned programs make decisions? [3]
- Does Explanatory Debugging help users personalize a classifier more efficiently than instance labeling? [2]

You may find it helpful to refer to your reflections or to what you wrote in your "Introduce yourself" submission.

# Activity 2: RQ Types (≈10m)

Consult with 1-2 neighbors:

1. Take turns verbalizing an RQ from your list (try to get through at least 3 from each member)
2. As a group, determine what kind of question it is and write down what you decide for each question.
3. If you wish your question had different types, change the questions!

In case you forgot the types:

- Exploratory questions
- Base-rate questions
- Relationship questions
- Causality questions
- Design questions

# Making RQs "Win-Win"

Results can be positive or negative.
Let's guarantee an **interesting** result.

For example:

- **Proposed RQ**: Does X occur?
  - Do we have a paper if X does not occur?
  - **Improved RQ**: To what extent does X occur?

Now we describe *reasons for* and *situations featuring* the discrepancy in expectation.

# Activity 3: Win-Win RQs (≈5m)

Consult with 1-2 neighbors, again taking turns working on an RQ from your list together:

- Is this RQ Win-Win?
  - If not, work together to change it!
- If you imagine any data in the world to exist, identify some sufficient data sources for the RQs.

# What will you accept as evidence?

You need data sources that can address your RQs.

Examples:

- What information do shoutcasters seek to generate explanations, and where do they find it?
  - **Method** - Fly on the Wall
  - **Data source** - Navigations in the video
- Does Explanatory Debugging help users personalize a classifier *more efficiently* than instance labeling?
  - **Method** - Instrumenting Systems
  - **Data source** - Accuracy data sampled at regular intervals as the participants work

# Method "Menu" Source: Shull page 11

- Inquisitive
  - Brainstorming and focus group
  - Interviews
  - Questionnaires
  - *Conceptual modeling
  - *Work Diaries
  - Think Aloud Sessions
  - *Shadowing and Observation (third person observation)
  - *Participant observation (first person observation, i.e. joining the team)
- Observational (indirect)
  - Instrumenting systems
  - Fly on the wall
- Observational (independent)
  - *Analysis of work databases
  - Analysis of tool use logs
  - *Documentation analysis
  - Static and dynamic analysis

# Activity 4: Choosing Methods + Data Sources (≈10m)

Consult with 1-2 neighbors, again taking turns working on an RQ from your list together:

1. Identify some methods that might be appropriate to answer this RQ.
2. If you imagine any data in the world to exist, identify some sufficient data sources for the RQs.
3. Last, narrow down to data that actually exist, however imperfect.

# On your own:

1. Find 3 pieces of related work.
2. If someone else already answered your RQ:
   Refine it further
   - OR -
   Pick another RQ from your list and repeat step 1
3. Choose 1-2 methods that would answer the RQ from the list on the following slide.
4. What is the best real-world data source that is analogous to the data source you provided in Activity 3 for this RQ?

As before, show your work as you go!

# References I


Jonathan Dodge, Sean Penney, Claudia Hilderbrand, Andrew Anderson, Logan Simpson, and Margaret Burnett.
How the experts do it: Assessing and explaining agent behaviors in real-time strategy games.
In *ACM Conference on Human Factors in Computing Systems*, page To Appear, 2018.

Todd Kulesza, Margaret Burnett, Weng-Keen Wong, and Simone Stumpf.
Principles of explanatory debugging to personalize interactive machine learning.
In *ACM International Conference on Intelligent User Interfaces*, pages 126–137. ACM, 2015.

# References II

Todd Kulesza, Simone Stumpf, Margaret Burnett, Weng-Keen Wong, Yann Riche, Travis Moore, Ian Oberst, Amber Shinsel, and Kevin McIntosh.
Explanatory debugging: Supporting end-user debugging of machine-learned programs.
In *Visual Languages and Human-Centric Computing (VL/HCC), 2010 IEEE Symposium on*, pages 41–48. IEEE, 2010.