Human-Centered Computing
Research-Group

Freie Universität Berlin

«**Exercise Human-Centered Data Science**»
Session #11

February 2nd, 2021

Agenda

1. **A5** ✓

2. **A7 - Ethics**

   » What are the (new) topics/concept(s)/methods introduced in this guest lecture?

   » Which concepts or ideas presented in the guest lectures do you find useful for your data science practice?

   » Any open questions?

Group Work - Part 1

What are the (new) topics/concept(s)/methods introduced in this guest lecture?

- Work in groups:
  - https://pad.spline.inf.fu-berlin.de/hNuC5cyMGS
  - 15 Minutes

Group Work - Part 2

Which concepts or ideas presented in the guest lectures do you find useful for your data science practice? **How can you apply it in your work/life/daily-business as a e.g. programmer or data scientist or other?**
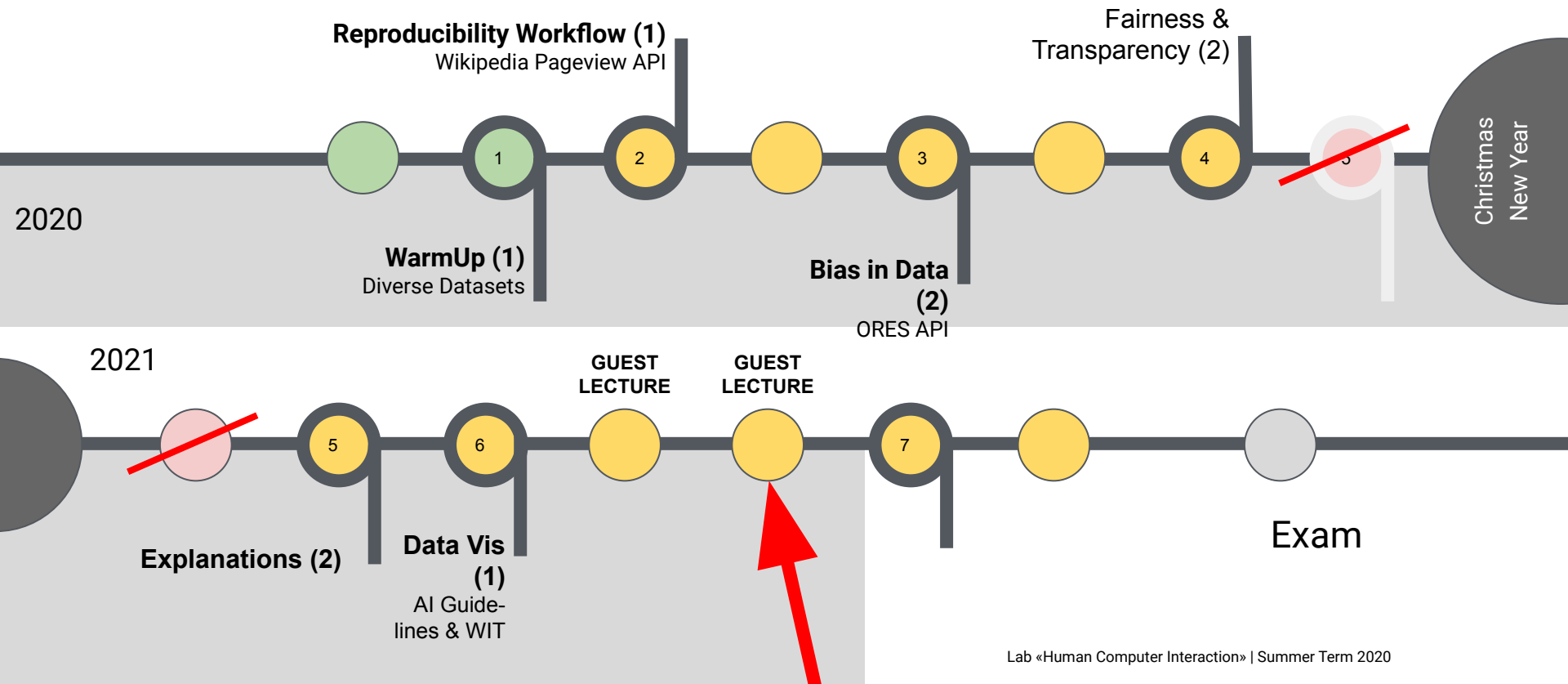
- Work in groups:
  - https://pad.spline.inf.fu-berlin.de/hNuC5cyMGS
  - 15 Minutes

Group Work - Part 3

Is there anything that is unanswered or you want to know more about?

- Work <u>individually</u>:
  - https://pad.spline.inf.fu-berlin.de/hNuC5cyMGS
  - 10 Minutes

Schedule

Reproducibility Workflow (1)
Wikipedia Pageview API

Fairness & Transparency (2)

2020

WarmUp (1)
Diverse Datasets

Bias in Data (2)
ORES API

1 2 3 4 5

Christmas New Year

2021

5 6 7

GUEST LECTURE

GUEST LECTURE

Explanations (2)

Data Vis (1)
AI Guide-lines & WIT

Exam

Questions

Data Science Ethics Checklist

`ethics checklist` `deon`

A. Data Collection

☐ **A.1 Informed consent**: If there are human subjects, have they given informed consent, where subjects affirmatively opt-in and have a clear understanding of the data uses to which they consent?

☐ **A.2 Collection bias**: Have we considered sources of bias that could be introduced during data collection and survey design and taken steps to mitigate those?

☐ **A.3 Limit PII exposure**: Have we considered ways to minimize exposure of personally identifiable information (PII) for example through anonymization or not collecting information that isn't relevant for analysis?

☐ **A.4 Downstream bias mitigation**: Have we considered ways to enable testing downstream results for biased outcomes (e.g., collecting data on protected group status like race or gender)?

B. Data Storage

☐ **B.1 Data security**: Do we have a plan to protect and secure data (e.g., encryption at rest and in transit, access controls on internal users and third parties, access logs, and up-to-date software)?

☐ **B.2 Right to be forgotten**: Do we have a mechanism through which an individual can request their personal information be removed?

☐ **B.3 Data retention plan**: Is there a schedule or plan to delete the data after it is no longer needed?