

# Introduction to social media data analysis within social data science

Max Pellert

*University of Konstanz*

Social Media Data Analysis



## Max Pellert

Assistant Professor  
University of Mannheim

Download CV as PDF

Professor for Social and Behavioural Data Science (interim, W2) at the University of Konstanz

Assistant Professor (Business School of the University of Mannheim)

I worked in industry at SONY Computer Science Laboratories in Rome, Italy



## Max Pellert

Assistant Professor  
University of Mannheim

Download CV as PDF

PhD from the Complexity Science Hub Vienna and the Medical University of Vienna in Computational Social Science

Studies in Psychology and History and Philosophy of Science

Msc in Cognitive Science and Bsc in Economics (both University of Vienna)

# (Some) Research interests

- Computational Social Science
- Digital traces
- Affective expression in text
- Natural Language Processing
- Collective emotions
- Belief updating
- **Psychometrics of AI**

Pellert, M., Lechner, C. M., Wagner, C., Rammstedt, B., & Strohmaier, M. (2024). AI Psychometrics: Assessing the Psychological Profiles of Large Language Models Through Psychometric Inventories. *Perspectives on Psychological Science*. <https://doi.org/10.1177/17456916231214460>

# About the professorship



Website: [dgarcia.eu](http://dgarcia.eu)

Twitter: [@dgarcia\\_eu](https://twitter.com/dgarcia_eu)

Github: [dgarcia-eu](https://github.com/dgarcia-eu)

Email: [david.garcia@uni-konstanz.de](mailto:david.garcia@uni-konstanz.de)

- Professor for Social and Behavioral Data Science, University of Konstanz
- Faculty member of the Complexity Science Hub Vienna
- Privatdozent at ETH Zurich and TU Graz

# David Garcia

- BSc in Informatics at Universidad Autonoma de Madrid (UAM), Madrid, Spain
- MSc in Computer Science (theory of computation) at ETH Zurich
- PhD and Habilitation at MTEC department at ETH Zurich
- Previously: PI at the Medical University of Vienna and Complexity Science Hub Vienna
- Previously: Professor for Computational and Behavioral Social Sciences at TU Graz



# Computational Social Science & Social Data Science

University of Konstanz, TU Graz, and Complexity Science Hub Vienna

Multidisciplinary team: psychologists, sociologists, CS, math



David Garcia



Hannah Metzler



Jana Lasser



Segun Aroyehun



Petar Jerčić



VIENNA SCIENCE  
AND TECHNOLOGY FUND



Anna Di Natale



Alina Herderich



Apeksha Shetty



Emma Fraxanet



Joao Pinheiro Neto



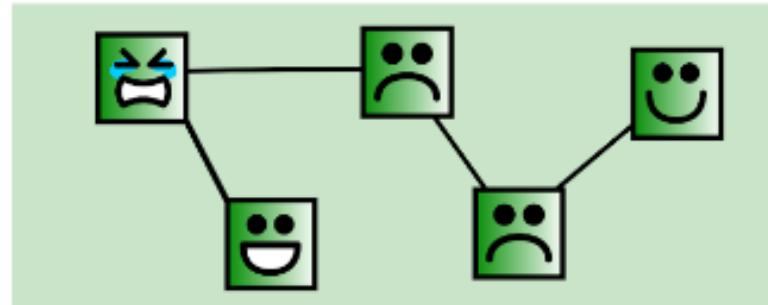
European Research Council

Established by the European Commission

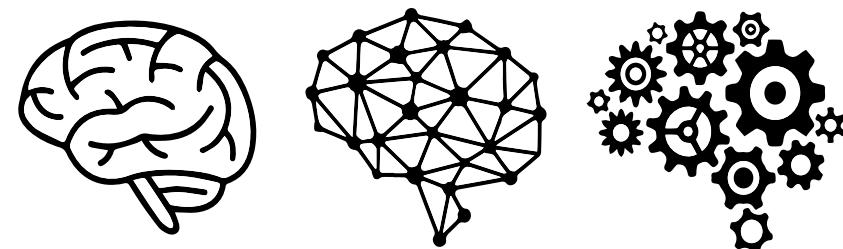
# Research lines



Methods in CSS

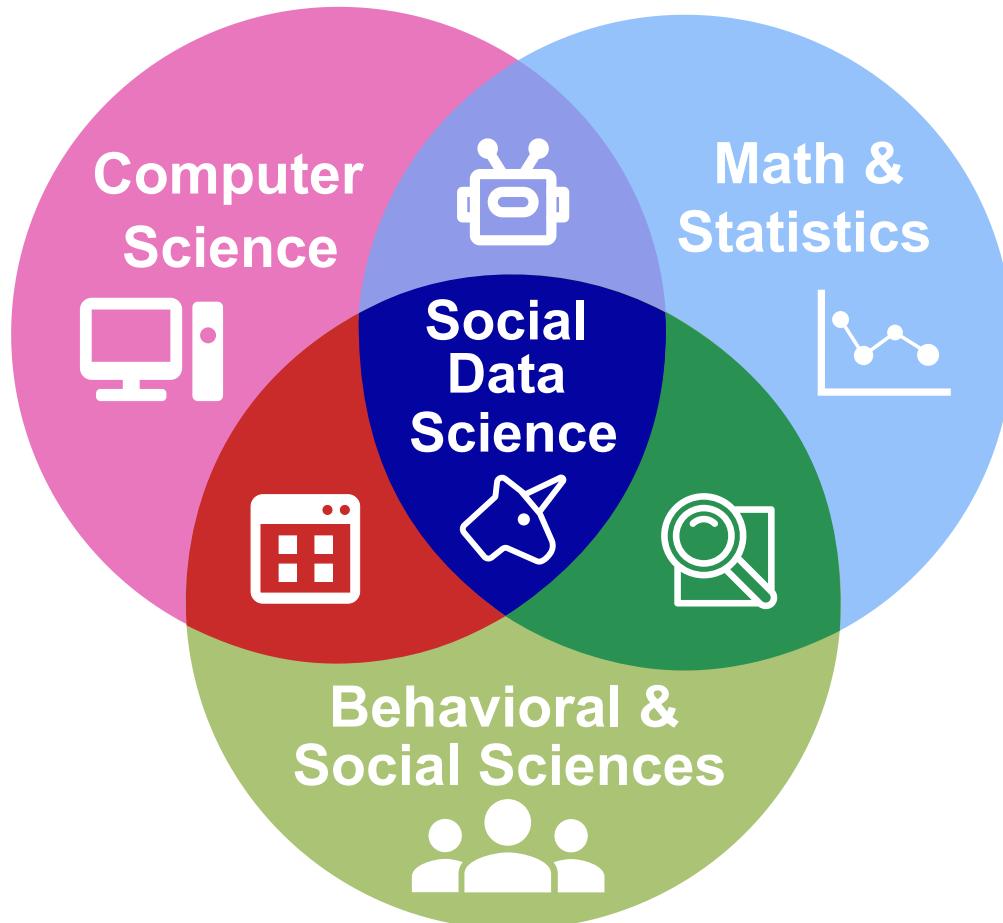


Computational Affective Science



Information Disorder

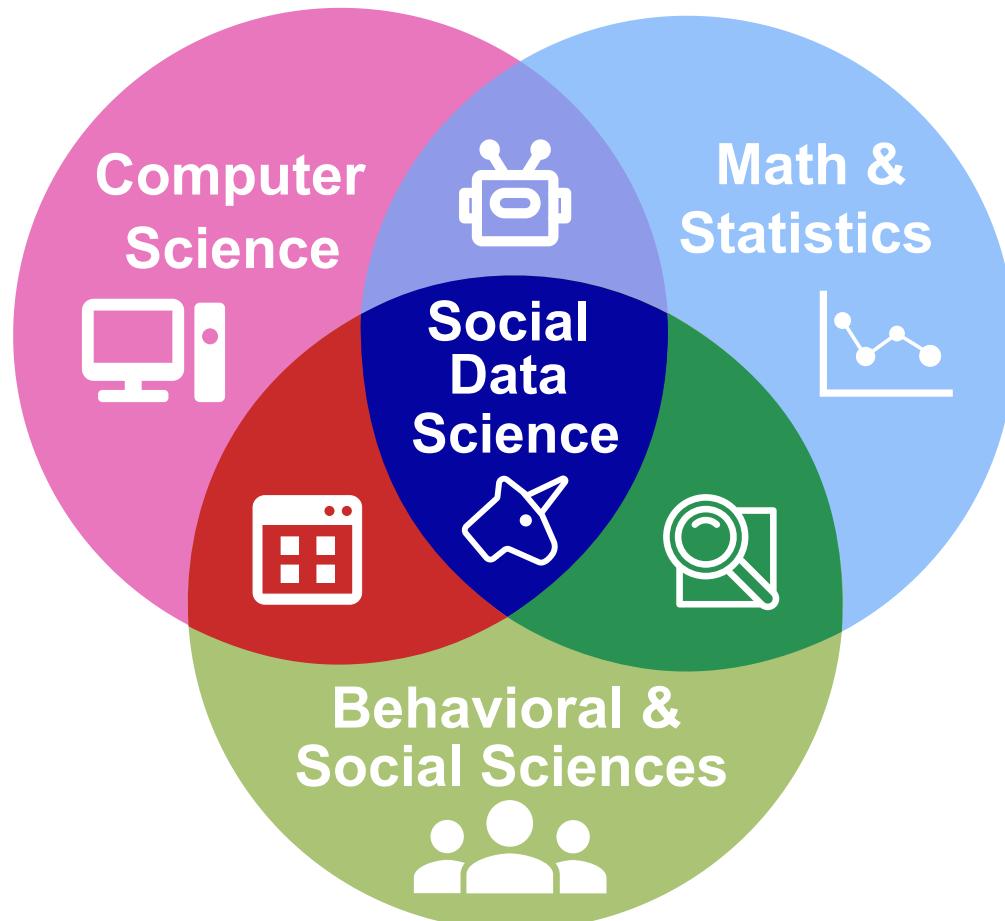
# Developing and connecting methods in Data Science



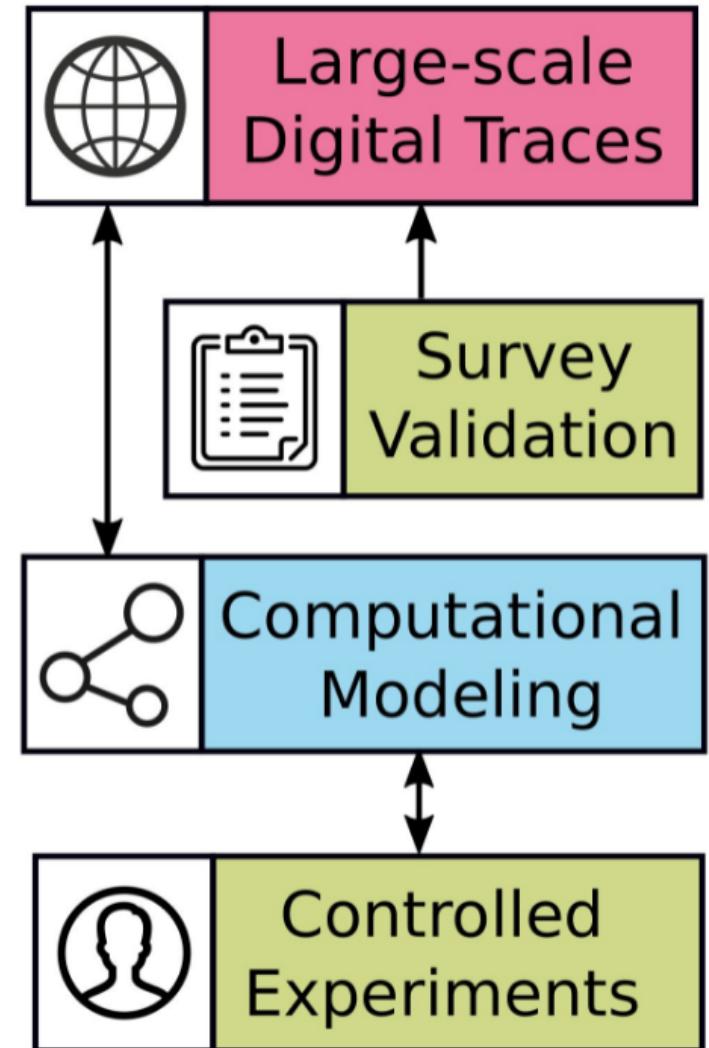
Statistics: Making sound inferences from too little data

Data Science: Making meaningful inferences from too much data

# Developing and connecting methods in Data Science



Statistics: Making sound inferences from too little data  
Data Science: Making meaningful inferences from too much data



# Who are you?

- SEDS students: year 1
- SEDS students: year 2
- SEDS students: year 3+
- Other programmes?

# Who are you?

- SEDS students: year 1
- SEDS students: year 2
- SEDS students: year 3+
- Other programmes?

**And more importantly: What do you expect from this course?**

# Who are you?

- SEDS students: year 1
- SEDS students: year 2
- SEDS students: year 3+
- Other programmes?

**And more importantly: What do you expect from this course?**

**Side question: Which social media do you use? What for?**

# Course objectives

After this course, you should be able to:

- understand a wide variety of techniques to retrieve **digital trace data from online platforms**
- store, process, and summarize online data for **quantitative analysis**
- perform **statistical analyses** to test hypotheses, derive insights, and formulate predictions
- interpret the results of data analysis with respect to testable **theories of human behavior**
- understand the **limitations** of observational data analysis with respect to data volume, statistical power, and external validity

# Course topics

## Block 1: Introduction

- Introduction to social media data analysis within social data science
- Algorithms and digital traces: The case of Google trends
- Ethics and privacy in social media data analysis
- Exercise 1: Future orientation and search

## Block 2: Social dynamics

- Social trends and the Simmel effect
- Social impact theory and its application to social media
- Exercise 2: Testing the division of impact hypothesis in social media

# Course topics

## Block 3: Text in social media

- Dictionary methods in social media data analysis
- Supervised social media text analysis
- The measurement of meaning in social media
- Exercise 3: Validating sentiment analysis for social media text

## Block 4: Networks in social media

- Introduction to social networks and the friendship paradox
- Centrality measures in online social networks
- Communities and assortativity in social media
- Exercise 4: Political assortativity on Twitter

# Exercises and practical sessions

- Exercise practical sessions are provided over the semester
- Time and place: Thursdays at 10:00, room D431
- Your tutors will be Claire and Elena, they will co-teach tutorials with me
- Assignments will be released on Wednesday evening or the latest Thursdays before the tutorial
- The deadline is before a tutorial a few weeks later
- Assignment submissions through Github as in ICSS
- **Assignments count 40% of the final grade of the course (10% each)**
  - Exercise 1: Future orientation and search
  - Exercise 2: Testing the division of impact hypothesis in social media
  - Exercise 3: Validating sentiment analysis for social media text
  - Exercise 4: Political assortativity on Twitter

# Learning materials and evaluation

- Github materials from last year: <https://github.com/dgarcia-eu/SocialMediaDataAnalysis>
  - The frontpage contains links to lecture slides
  - It is populated with written handouts
- Course Ilias for lecture: TBA
- Course Ilias for tutorials: TBA
- Final evaluation based on:
  - Project (60%)
  - Exercise assignments (40%) - each with its own grade worth 10%
  - **Questions?**

# Related courses

- **Introduction to Computation for the Social Sciences**
  - Previous semester prerequisite. You should have taken it, otherwise talk to other students to learn the how to of that course, especially exercises
- **Current Topics and Projects in Social Data Science**
  - Additional seminar to discuss current research and develop projects building on this course and ICSS
  - Thursdays after tutorial in the same room
- **Network Science of Online Interaction**
  - Seminar with exercises focusing on network science methods
  - Includes introduction to network science and data analysis projects
  - Good alternative to this course
- Other courses: Colloquium as a preparation for a Master Thesis, Data Science Projects, and Computational Modelling of Social Systems seminar during the Winter Semester

# What is social media?

Information and Communication Technologies that enable users to share content with each other and with larger audiences.

Social media are old but changed a lot recently. One example:

**VATIAM AEDILEM FURUNCULI ROGANT**

(The petty thieves ask for Vatia as aedile)

Graffito on the wall of a building in Pompeii

How do our social media differ from a writing on a street wall?

# What can we measure in social media?

"The idea that anything as subtle and complex as all the manifestations of changes in temperature could be measured and quantified on a single numerical scale was scoffed at as impossible, even by the leading philosophers of the sixteenth century."

—Arthur Jensen, Bias in Mental Testing, 1980

- Measuring meaning through text seemed very hard but has been improving substantially over time
- Tracking the connections between Billions of people seemed impossible and now it is a boring reality
- What now might feel subjective might be quantitative science in the future

# What can we use social media data for?

"The plan was to connect these devices to a network –it would ride on the existing TV networks– so that the total national happiness at any moment in time could be determined. The algedonic meter, as the device was called (from the Greek algos, "pain," and hedone, "pleasure"), would measure only raw pleasure-or-pain reactions to show whether government policies were working."

–Evgeny Morozov, The New Yorker, 2014

"More is different"

–Paul Anderson, 1972

# What is the aim of Social Data Science?

The aim of Social Data Science is  
**the Quantitative Understanding of Human Behavior.**

- **Quantitative:** As opposed to qualitative or descriptive, we aim for robust findings grounded in strong evidence that can be quantified.
- **Understanding:** Not just predicting, we want to be able to generalize and combine knowledge, and even to motivate interventions or policies.
- **Human:** We will not study particles or objects. Measurement validity and ethics will be a challenge.
- **Behavior:** Observable changes, structures, dynamics, and patterns; not just stories or theories

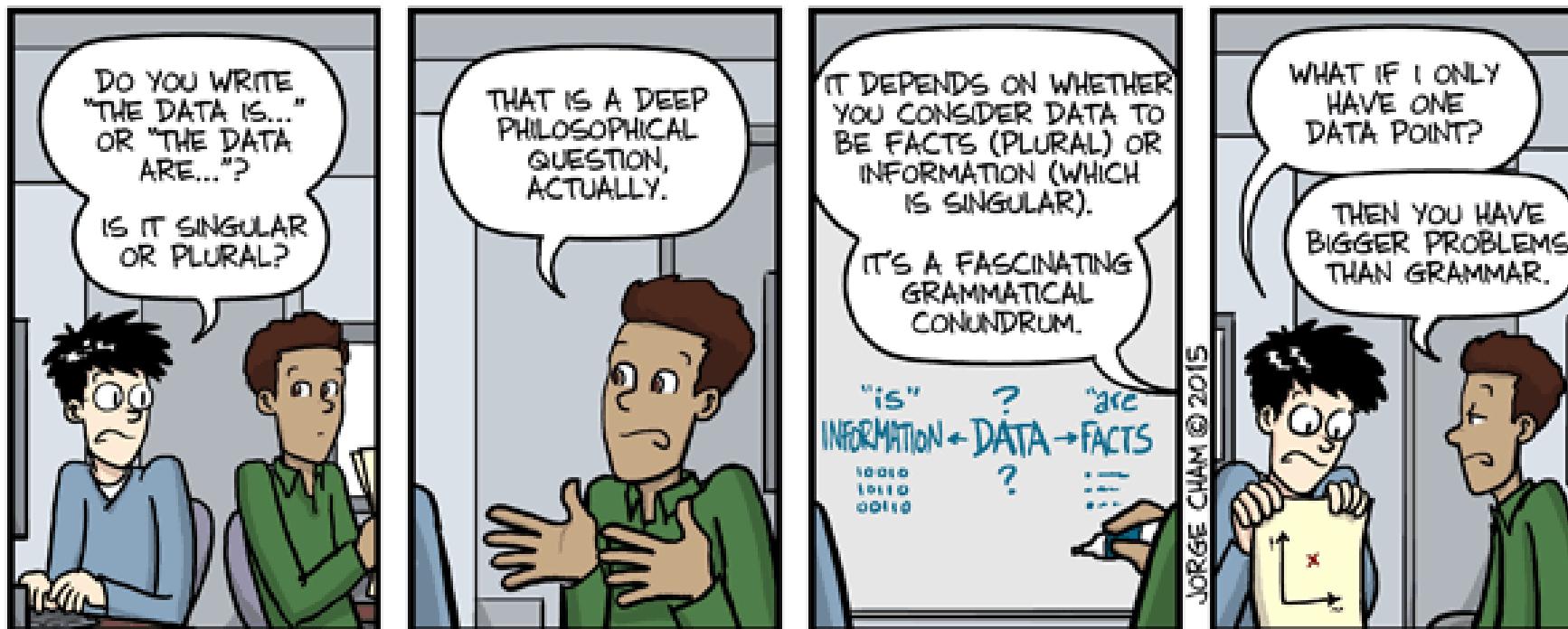
# How are we going to do it?

Retrieving, processing, analyzing, and interpreting social media data.



# What is data?

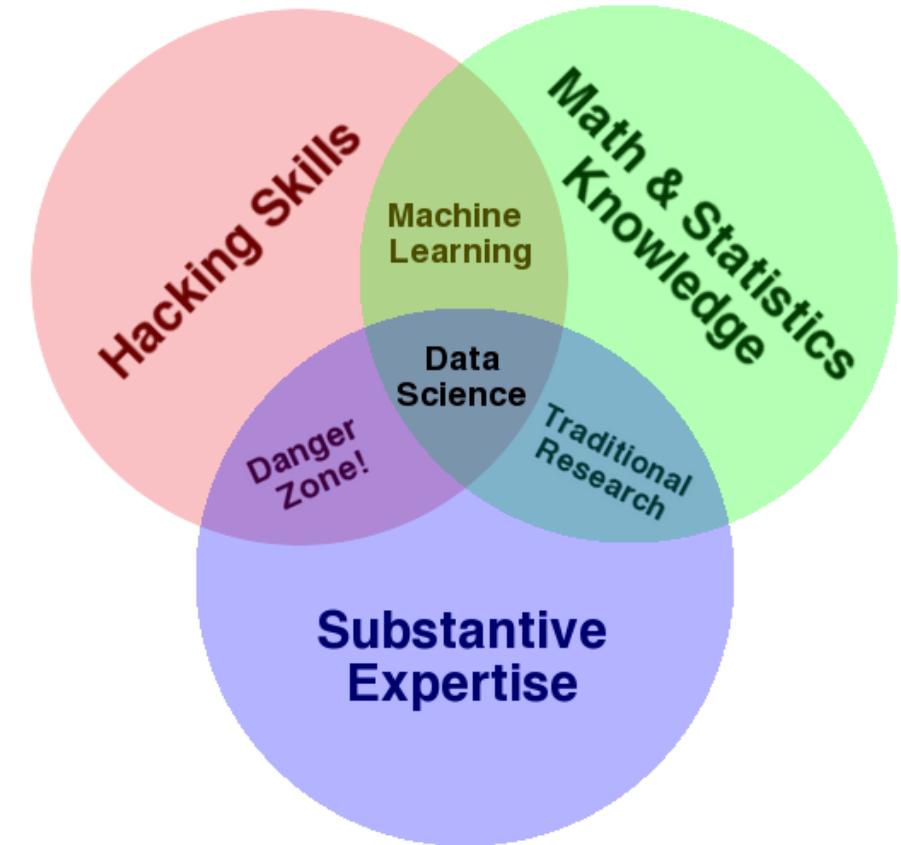
- | Data: Facts in the form of stored and transmittable information.
- *Data* is the plural of the Latin word *datum*. *Data* means "given (things)".
- Data is given to us, it is not fabricated nor simulated.



# What is Social Data Science?

**Social Data Science** is the application of Data Science to study human behavior and social interaction.

- Importance of combining methods, synthesis, and communication
- Gathering digital traces with Computer Science techniques
- Analysis interpretation with respect to what is known in the relevant Social Sciences



**Interdisciplinarity:** learning from the mistakes of one discipline to improve the others - many examples in this course

# Why digital traces?

Data from digital traces have six properties that can complement other traditional data sources in the social sciences:

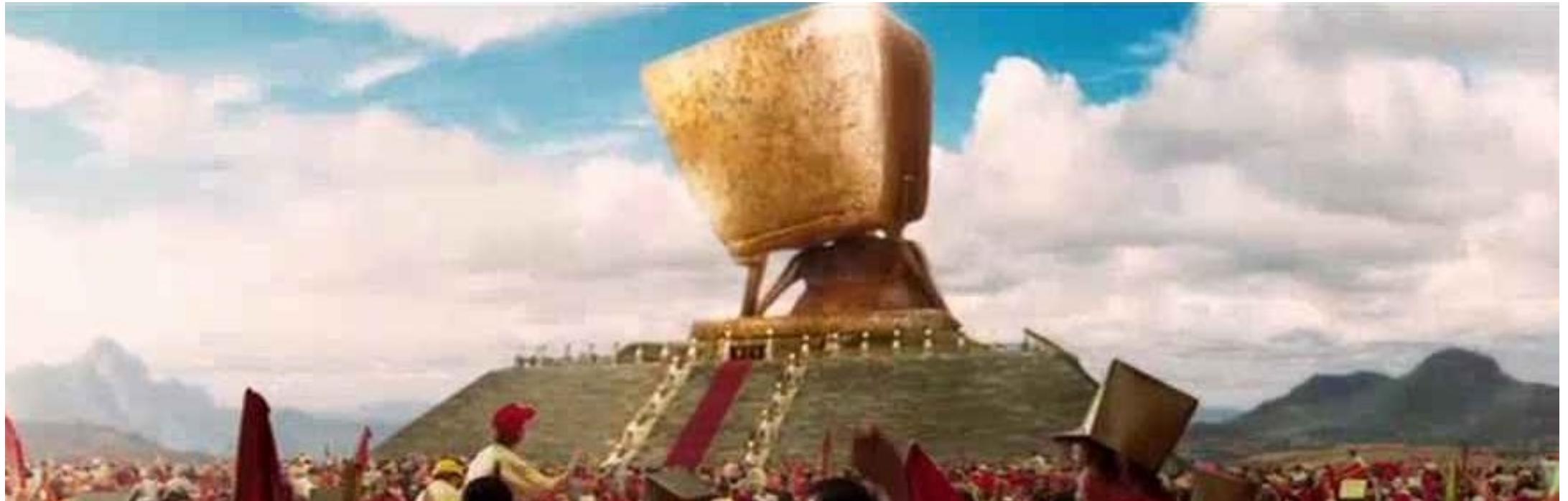
- **Big Data:** Observing large amounts of humans across demographics
- **Fast Data:** Quantifying aspects of human behavior in real time
- **Long Data:** Retrieving longitudinal data and at various timescales
- **Deep Data:** Gathering persistent information on individuals
- **Mixed Data:** Combining heterogeneous datasources and unstructured data
- **Strange Data:** Locating small subcommunities or deviant behavior

# What are the limitations of digital traces?

With the great potential of digital traces, also come great challenges:

- **Platform biases:** caused by their design, algorithms, social bots, etc
- **Data gatekeepers:** Not everyone can access some data sources
- **Performative behavior:** Talking online is not the same than offline.
- **Representativity issues:** Not everyone leaves digital traces.
- **Observational data:** Testing causal mechanisms is not straightforward.
- **The data deluge:** Too much data enables black-box predictions that can be useful but limit our understanding.

# The importance of questions in Social Data Science



Deep Thought, from the movie version of The Hitchhiker's Guide to the Galaxy.

**Understanding our questions is a prerequisite to understanding their answers.**

# Summary

- In this course, you will learn how to retrieve, analyze, and interpret social media data
- Our aim is aligned with Social Data Science: quantitative understanding of human behavior
- Course contents include social dynamics, text analysis, social networks
- Emphasis on critique and understanding of limitations and potentials
- Join us for the first tutorial on Thursday at 10:00!
- Questions?