POSTCAST SCRIPT

INTRODUCTION

Hello and welcome to CAISS notes, a series of short introductions to the world of Computational Social Science, or CSS for short, brought to you by the CAISS hub.

Find out more about us at the end of the podcast.

Today you have me, Matt Asher, to explain all about perhaps the most fundamental question of Computational social science:

What is it?

DEFINITIONS OF CSS

As with most fundamental questions in research, the answer is either it depends, or it covers everything.

When the term was originally coined, it was a description of the state of the research space as much as anything else.

 Lazer et al, noted in 2009 that

"(a) computational social science is emerging that leverages the capacity to collect and analyze data with an unprecedented breadth and depth and scale." (1)

which leave the definition pretty broad, and focused on what we might call *big data* for social sciences.

The definition has evolved since then, and will often depend on who you talk to, especially if they come from different background.

For a top-level definition, the one at Wikipedia is as good as any, where CSS is

"an interdisciplinary academic sub-field concerned with computational approaches to the social sciences." (2)

though I would perhaps disagree with the 'sub-field' as CSS has an important role to play in multiple existing fields

Commonality of definitions

Two things that most definitions of computational social science agree on is that it is about

1) social data

 and

2) computers,

though CSS is very much more than just large repositories of digital data and the computational methods used to analyse them.

IMPORATNCE OF CSS

Theories of human behaviour are often developed as discrete explanations with little or no interaction with other theoretical understandings, particularly between disciplines.

However, modelling of complex systems demands the 'blending' of these theories in order to provide a complete, or at least more complete, picture of the multitude of factors which affect human behaviour. For example, within defence, we are striving to holistically consider the individual, societal and community factors which are all important drivers when looking at participation in, and solutions to, conflict.

Any serious consideration of CSS must recognise that the social science expertise must be introduced at all stage of the data collection and analysis, from thoughts of how the data collection will be affected by regional social values, to the implications of adjusting for bias in the processing techniques.

An integrated approach from across specialities and also across disciplines, for example, defence, industry, and academia, can bring together existing and varied expertise to attempt to answer relevant CSS questions.

We will go into more details about specific uses of CSS in later CAISSnotes, but for now is sufficient to note that there have recently been a number of high-profile media cases where assessment tools used in official capacities, such as for predicting recidivism (8), have been shown to create sexist (3,4), ableist (5), or racial (6) bias.

Such examples demonstrate how tools devised to process social data can come with serious consequence if social theory is not included in the application of data processing techniques.

Basically, computer scientists can forget that they are processing data on real people and need to understand how to handle it.

As we advance into the digital age this type of data will continue to grow exponentially and become more integrated in the lives of society, making it necessary for social science disciplines to take it into account. A true collaboration between social scientists and data scientists will not only produce methodologically stronger research, but it will also make it possible to study things neither could study alone.

Wrap up

In conclusion, CSS should be thought of as a synthesis of different fields, from social sciences and psychology, through statical data analysis to computer engineering and even including legal and ethical studies, pulling these divers approaches together to new and different ways of thinking about data in social situations to answer questions that are relevant to society today and in the future.

OUTRODUCTION

Thank you for listening to the end. I hope this short explanation had deepened your knowledge and make you hungry for more. I would like to thank George, Sophie, Grace, Joe & Glen for all their work to make this possible.

This podcast is created by the Computational and AI in Social Sciences Hub, a collaboration between Dstl, Lancaster University and the Alan Turing Institute, to raise awareness of the interdisciplinary nature of data analysis in the modern world, bringing together experts from technical, social, psychological and other backgrounds. For more details, find our website by searching for CAISS hub or follow us on social on Twiitter (X) @CAISS_uk, that's C,A,I,S,S,underscore,U,K , and on linked in as CAISS.

{Beep Out}

CITATIONS

1. Lazer, *et al.*, Social science. Computational social science. *Science (New York, NY)* **323**, 721–723 (2009).

2. https://en.wikipedia.org/wiki/Computational_social_science

3. Dressel, J., & Farid, H. (2018). The accuracy, fairness, and limits of predicting recidivism. *Science advances*, *4*(1), eaao5580.

4. Dastin, J. (2018). Amazon scraps secret AI recruiting tool that showed bias against women. UK Reuters, 10.10.2018.

5. Dastin, J. (2022). Amazon scraps secret AI recruiting tool that showed bias against women. In *Ethics of data and analytics* (pp. 296-299). Auerbach Publications. Alternate Link

6. Job hunting for neurodivergent people: "AI recruitment means I've got zero chance." BBC.co.uk.