



LES RENCONTRES DE STATISTIQUE APPLIQUÉE

Brainpower Meets Machine Power: Chat-GPT synergies with social scientists

Vendredi 12 janvier 2024

Ined (9h-12h45, salle Alfred Sauvy)



EUROPEAN DOCTORAL
SCHOOL OF
DEMOGRAPHY

Service des Méthodes Statistiques (Ined) in collaboration with the European Doctoral School of Demography (EDSD)

Ined : Campus Condorcet, 9, cours des Humanités, Aubervilliers

The development of artificial intelligence techniques has numerous consequences for our activities. However, the rapid pace of adaptation is such that individuals often feel helpless and uncertain in the face of resulting technological changes. In response to this formidable challenge and its implications for our work, we are organizing a session featuring experts in this field.

B. Crabbé (University of Paris) will give an introductory presentation recounting the history of these contributions. K. Smaili (Loria, Lorraine University) will detail the mechanics inherent in the ChatGPT engine. D. Schwab (Université Grenoble Alpes University) will present the general principles and questions raised by OpenAI's large language models. Finally, Paul Gay (Pau University and Paris-Saclay University) will focus on the environmental aspects related to the development of these technologies. E. Dagorn (Ined) will ensure the follow-up of the presentations and initiate focused discussions during the conclusion.

Free on registration <https://statapp.site.ined.fr/fr/2024/ia/>

Speakers and abstracts

9:20 • Benoît Crabé (Paris Cité University)

The origins of large language models: an historical perspective on computational language modeling

This talk will provide elements on the history of language modeling with the aim to shed light on the nature of current large language models. We start from the original insights of A. Markov, C. Shannon and W. Weaver and contrast them with the insights of N. Chomsky and symbolic AI. We will provide elements explaining the failure of symbolic AI for language and the generalization of statistical methods for language modeling. The last step of this evolution is the so-called deep learning paradigm that gave birth to current Large Language Models.

10:05 • Kamel Smaili (Lorraine University - Loria)

Let's explore transformer together

In this presentation, we will provide a detailed explanation of the Transformer architecture, which forms the foundation for state-of-the-art generative AI systems.

We will delve into the intricacies of both the encoder and the decoder, emphasizing the significance of positional encoding and multi-head attention. The presentation will be enriched with illustrative examples. Regarding the decoder, we will highlight the distinction between the two types of multi-head attention employed within it.

11:05 • Didier Schwab (Grenoble Alpes University)

From gpt to chatgpt: general principles and issues raised by openai's large language models

In this presentation, we explain what we know about chatGPT, with what data it is learned, what it really does and what its technical and ethical limits are.

11:50 • Paul Gay (Pau University and Paris-Saclay University)

Integrating environmental impact of ai in a data center

This presentation is an introduction to the environmental impact of machine learning. It will describe the different impacts from the digital sector : carbon and water footprint, metal consumption, ecotoxicity, etc. The main findings and orders of magnitude from the state of the art will be provided as well the current limits and uncertainty in the estimation of these impacts. It will be shown that these impacts can be analysed at different scopes : a program, a computing device, an data science project in a data center, the use of an open source model by a community... Long term impacts will be also covered through the concepts of the Jevons effects and other indirect effects. The last part will give entry points to available tools and methods that data scientists can use to measure and reduce the impact of their projects. These include measuring the energy consumption of a piece of software, and considering larger and more general impact thanks to life cycle analysis methodologies.

12:35 • Etienne Dagorn (Ined)

Discussion and conclusion

Information and presentations of past sessions on <https://statapp.site.ined.fr/>