efl | the Data Science Institute

Democratization of Data Science & AI

ANNUAL CONFERENCE 2022 OF EFL – THE DATA SCIENCE INSTITUTE

efl – the Data Science Institute
October 11, 2022 at the Casino Building of the Goethe University Frankfurt

14:30	REGISTRATION & COFFEE
15:00	WELCOME
15:10	EXAMPLE-DRIVEN DATA CLEANING Ziawasch Abdejan
15:40	CURRENT DEVELOPMENTS IN FOUNDATION MODELS Markus Schmitz
16:10	KEYNOTE: DESIGNING AI SYSTEMS WITH EVERYONE IN MIND Fernanda Viégas & Martin Wattenberg
16:50	LAUDATIO: 20 YEARS EFL Joachim Würmeling
17:10	REFRESHMENT BREAK
17:40	CHALLENGES IN DATA DEMOCRATIZATION AT BUNDESBANK Stefan Bender
18:10	DATA CONSIDERATIONS FOR RESPONSIBLE DATA-DRIVEN SYSTEMS Asia Biega
18:40	GET TOGETHER

DZ BANK Gruppe

DEUTSCHE BÖRSE GROUP

finanz informatik

FACTSET

AMERICAN EXPRESS

Deutsche Leasing

GOETHE UNIVERSITÄT FRANKFURT AM MAIN

TECHNISCHE UNIVERSITÄT DARMSTADT

Democratization of Data Science & AI

ANNUAL CONFERENCE 2022 OF EFL – THE DATA SCIENCE INSTITUTE

efl – the Data Science Institute
October 11, 2022 at the Casino Building of the Goethe University Frankfurt

EXAMPLE-DRIVEN DATA CLEANING

Prof. Dr. Ziawasch Abdejan
Research Director, Leibniz University Hannover and
Visiting Academic at Amazon

Data is being produced in large volumes and many shapes and there is an increasing interest in applying novel data science methods to extract new insights. Often, these data science applications are carried out by researchers who lack computer science and systems expertise. This is aggravated by the fact that research data is often the combination of from multiple sources, such as third-party archives, open data, and experiments carried out by many different people. The heterogeneity of these sources requires integration and curation efforts, which is only feasible through heavy human intervention. To facilitate this process, one has to increase its automation through scalable algorithms for the encapsulation of the complex integration and curation steps. In this talk, I will outline our contributions towards more declarative and example-based data integration and curation steps with a special focus on data cleaning.

Democratization of Data Science & AI

ANNUAL CONFERENCE 2022 OF EFL – THE DATA SCIENCE INSTITUTE

efl – the Data Science Institute
October 11, 2022 at the Casino Building of the Goethe University Frankfurt

CURRENT DEVELOPMENTS IN FOUNDATION MODELS

Dr. Markus Schmitz
AI Consulting, Aleph Alpha GmbH

Large Language Models (LLMs) have been a breakthrough discovery in AI research, especially in the field of Natural Language Processing. These amazing capabilities are fuelled by enormous model sizes that pose unprecedented possibilities for working with information in natural language. Applications and system built atop this technology can facilitate understanding, cross the boundaries of language through semantics and bridge gaps in communication and education to increase democratization of information. While many developments and applications developed on top of this technology is still mainly functional, these advancements can fuel many more applications that improve access and participation for everybody.

DZ BANK Gruppe

DEUTSCHE BÖRSE GROUP

finanz informatik

FACTSET

AMERICAN EXPRESS

Deutsche Leasing

GOETHE UNIVERSITÄT FRANKFURT AM MAIN

TECHNISCHE UNIVERSITÄT DARMSTADT

Democratization of Data Science & AI

ANNUAL CONFERENCE 2022 OF EFL – THE DATA SCIENCE INSTITUTE

efl – the Data Science Institute
October 11, 2022 at the Casino Building of the Goethe University Frankfurt

KEYNOTE: DESIGNING AI SYSTEMS WITH EVERYONE IN MIND

Prof. Dr. Fernanda Viégas & Prof. Dr. Martin Wattenberg
Principal Scientists, Google Research & Gordon McKay
Professors, Harvard University

How should people relate to artificial intelligence technology? Is it a tool to be used, a partner to be consulted, or perhaps a source of inspiration and awe? As technology advances, choosing useful human/AI relationship framings will become an increasingly important question for designers, technologists and users. We'll discuss a series of research projects–ranging from data visualizations and tools for medical practitioners to guidelines for designers–that illustrate how AI can play each of these roles. By providing users with a diversity of engagement possibilities, we hope to develop more responsible and effective ways to construct, use and evaluate this technology.

DZ BANK Gruppe

DEUTSCHE BÖRSE GROUP

finanz informatik

FACTSET

AMERICAN EXPRESS

Deutsche Leasing

GOETHE UNIVERSITÄT FRANKFURT AM MAIN

TECHNISCHE UNIVERSITÄT DARMSTADT

Democratization of Data Science & AI

ANNUAL CONFERENCE 2022 OF EFL – THE DATA SCIENCE INSTITUTE

efl – the Data Science Institute
October 11, 2022 at the Casino Building of the Goethe University Frankfurt

CHALLENGES IN DATA DEMOCRATIZATION AT BUNDESBANK

Prof. Dr. Stefan Bender
Head of Research Data and Service Center, Deutsche Bundesbank

Realizing the full analytical potential of AI / ML algorithms requires access to detailed micro data at the e.g. individual bank, enterprise or security level. However, providing access to this micro data is quite a complex endeavor. The challenge comes from protecting the identity of the reporting entities while simultaneously allowing data users to study distinct features of these entities, e.g. effects on an individual person's health or a company's business model. In order to master this balancing act between increased confidentiality and the high analytical value of micro data, data providers need to implement statistical, organizational or technical measures to prevent the disclosure of sensitive information. The Research Data and Service Centre (RDSC) of the Deutsche Bundesbank provides standardized access to selected microdata collected by the Deutsche Bundesbank in accordance with its statutory mandate to be used in independent scientific research projects. Within the applicable legal framework and data protection regulations, Stefan Bender (joined with Jannick Blaschke and Christian Hirsch) argue that institutions wanting to share microdata efficiently should approach this from the perspective of a cost-benefit analysis. This requires, first, minimizing the costs for researcher and data provider along the three dimensions: process, documentation, and automation of micro data. Examples include high-quality data on the data (i.e. meta data) but also on access (annodata) and usage of the micro data. Second, institutions interested in efficient data sharing also need to aim to maximize the benefit of data sharing. Examples include Amazon-like data recommendation, a data network or a dataset impact factor. Summarizing, this presentation and the examples herein provide some insides into how information can be transformed into knowledge for the wealth of society. Furthermore, first ideas on how data and information can be accessed beyond independent research are discussed.

Democratization of Data Science & AI

ANNUAL CONFERENCE 2022 OF EFL – THE DATA SCIENCE INSTITUTE

efl – the Data Science Institute
October 11, 2022 at the Casino Building of the Goethe University Frankfurt

DATA CONSIDERATIONS FOR RESPONSIBLE DATA-DRIVEN SYSTEMS

Dr. Asia Biega
Max Planck Institute for Security and Privacy (MPI-SP)

Data-driven systems collect, process and generate data from user interactions. To ensure these processes are responsible, we constrain them with a variety of social, legal, and ethical norms. In this talk, I will discuss several considerations for responsible data practice. I will show how these concepts can be operationalized as well as highlighting the computational and normative challenges that arise when these guiding principles are implemented in practice.