Data literacy

Eric Solano, Ph.D.

During my extensive career I have published theses, dissertations, peer reviewed papers, conference papers and technical papers. Most of these papers were based on data driven research projects, and contributed to innovative ways to structure data and to build models ranging from operations research, mathematical modeling, systems dynamics modeling, agent-based modeling, stochastic models, discrete time models, etc. They contributed to the collective improvement of knowledge in data modeling and analysis by pushing the limits of technology at the time.

Some of these articles continue having an impact on the data analytics community as they are reviewed and cited by others. My Google Scholar account reports, for example, more than one hundred and fifty citations for one of my papers: "Life cycle based solid waste management". This publication presented seminal research with big data and complex mathematical optimization models to support decision making. The impact of this research was and still is seen in many areas: for example, revenue generation from recovered materials and energy; and cost savings through the selection of optimal sets of technologies and processes. I received awards for some of these publications.

In more recent years, I have mentored college students, graduate students and interns. The common denominator of all these training courses and mentorships are data driven projects. I have contributed to data literacy within my data science team, the extended IT group and the business partners. I have conducted multiple trainings at Continental, including the use of Business Intelligence tools like tableau and PowerBI; the design of databases in Microsoft SQL Server, Redshift, Cassandra and PosgreSQL; the design and deployment of machine learning models using python, sciPy, scikit-learn, pyTorch and other libraries; the design and deployment of distributed computing applications using EMR cluster, Apache Spark, S3, HDFS, etc.

More recently, I have been part of a team exploring options to implement Large Language Models (LLM) at Continental. We call this team Conti-GPT supported by Continental's AI lab and from a handful of participants we grew to hundreds of technical and non-technical staff members participating in weekly calls from all over the world. We provided customized generative AI solutions available at Continental: usage of closed-source models, such as GPT-4 or Claude, via providers like Microsoft or AWS; usage of open-source models, such as **Llama3** or **Mistral**. We have highlighted the pros and cons of these options. With strictly confidential data, our own infrastructure might be a requirement and open source is the only option. With public, internal or confidential data, using a provider is probably OK and "pay as you go" is probably more cost-efficient. Available options are Microsoft Azure and GPT4 or Amazon Bedrock and Claude AI. Colleagues have used these options to generate very interesting uses cases, for example, in the creation of internal regulatory documents in the natural languages from countries where Continental has a presence.