

# Reflections on Data Science 2023

Ludek Cizinsky, Lukas Rasocha, Gergo Gyori, Katalin Literati-Dobos

{luci,lukr,gegy,klit}@itu.dk

## Table of Contents

- #1. Lead author: [Ludek Cizinsky], Calling BS on "Alena Schillerova's tweet claiming unrealistic growth in gas and electricity prices that people in Czechia had to pay." (April, 2023) 2**
- #2. Lead author: [Ludek Cizinsky], Calling BS on "A few more years and the Poles will be richer than most Western European nations in terms of real domestic economic activity." (May, 2023) 3**
- #3. Lead author: [Lukas Rasocha], Calling BS on "Steve Milloy's tweet claiming that global warming due to C02 emmission is a hoax." (January, 2023) 4**
- #4. Lead author: [Lukas Rasocha], Calling BS on "Dr. Simone Gold's tweet claiming 'The same number of athletes died in the last TWO years as compared to a prior 38 years.'" (January, 2023) 5**

**#1. Lead author: [Ludek Cizinsky], Calling BS on "Alena Schillerova's tweet claiming unrealistic growth in gas and electricity prices that people in Czechia had to pay." (April, 2023)**

Alena Schillerova, a member of the Czech parliament, claims in her tweet that average natural gas price that people in Czechia had to pay in the second half of 2022 was by 231 % higher compared to the average natural gas price in the second half of 2021. She then also makes the same claim for electricity price which grew by 97 %. I call both of these claims BS and I believe they are a result of Schillerova's **confirmation bias**. The fact is that Schillerova is the face of the main opposing party to the current Czech government. Naturally, any statistic that would claim that Czechia is not doing so great under current government is aligned with her existing beliefs and as such she might not look for alternative explanations. What could further convince her is the "fancy" visualizations from Eurostat, the [source](#) of her statistics.

The key problem with Schillerova's data for the natural gas and electricity prices in Czechia is that they are based on the prices the new customers have to pay when they want to start using particular energy provider. In other words, these are not the actual prices paid by Czech households in which case Schillerova's claims would be valid. ([source](#)) Unfortunately, there is no law that requires Czech energy companies to report the actual paid prices. However, since these numbers drew a lot of media attention, I was able to find statements from two Czech energy companies regarding the growth in actual prices paid ([source](#)), which I believe refute Schillerova's claims. When it comes to gas prices, CEZ, which held a 12 % market share in 2021 ([source](#)), reported only an 87 % growth. For electricity prices, EON, which held a 15 % market share in 2021 ([source](#)), states that the prices for their consumers in fact dropped by 43 %. While these do not apply to the whole consumer base in Czechia (given the providers' market shares), both companies must stay competitive as well as profitable therefore it is reasonable to assume that similar amount of growth or decrease in prices would be reported by the other providers. In summary, Schillerova's confirmation bias likely led her to omit proper due diligence on what kind of prices Eurostat uses for their computations. It is fair to assume that she would not post such tweet after the proper due diligence since the change in prices is far less shocking (see above). The clear con of Schillerova's tweet is that it spreads panic among people. On the other hand, it seems to now speed up the process of coming up with legislation that requires energy companies to report the actually paid prices.

**#2. Lead author: [Ludek Cizinsky], Calling BS on "A few more years and the Poles will be richer than most Western European nations in terms of real domestic economic activity." (May, 2023)**

In the start of his tweet, Foubert correctly states that Poland is catching up with France when it comes to Gross Domestic Product (GPD) at Purchasing Power Parity (PPP) per capita, i.e., gross domestic product which accounts for living costs in the given country. I verified the stated numbers using the World Bank's [database](#). However, I call BS on his second statement in which he claims that *in a few more years Poles will be richer than most Western European nations in terms of real domestic economic activity* which I plan to refute using **Fermi estimation**. I first decided to investigate when will Poland catch up with France since Foubert is using it as a reference in the tweet. I used again the World Bank's [database](#) to compute average year to year growth rate of GDP at PPP per capita in the past two pre COVID-19 years, i.e., I compared 2019 vs 2018 and 2018 vs 2017. This yielded a rough estimate of 0.08 for Poland and 0.06 for France. I then made a simplifying assumption that both nations will keep this growth rate for the given number of years. Therefore, to compute GDP at PPP per capita in  $N$  years, I used the following formula:  $s * (1 + gr)^N$  where  $s$  is the starting value of GDP at PPP per capita and  $gr$  is the growth rate. This led me to a conclusion that it would take roughly 16 years for Poland to overtake France which clearly refutes the few years claim of Foubert. To further support my estimate and given the availability of World Bank's [data](#), I have done more fine grained estimate of growth using past 10 years of pre COVID years (2010 to 2019) for nations which are geographically Western European nations, see table 1a. I decided to exclude Ireland which had growth rate higher than Poland, as such Poland would never catch up given my methodology. I then used my formula to compute the number of years it would take Poland to overtake given nation in GDP at PPP per capita (see table 1b). For most of the nations, it would take more than a decade which is in line with my initial estimate, thus providing further evidence that refutes Foubert's statement.

Country	Growth
Romania	0.072
<b>Poland</b>	<b>0.062</b>
Czechia	0.048
Germany	0.046
Belgium	0.040
Austria	0.039
France	0.039
Portugal	0.035
United Kingdom	0.035
Luxembourg	0.033
Switzerland	0.033
Netherlands	0.031
Spain	0.030

(a) Mean Year-to-Year Growth of GDP at PPP per capita between 2010 and 2019.

Country	Overtake in (years)
Portugal	0
Spain	3
United Kingdom	12
France	14
Netherlands	18
Austria	21
Belgium	22
Switzerland	26
Germany	28
Luxembourg	46

(b) How long would it take Poland to overtake the given country in GDP at PPP per capita

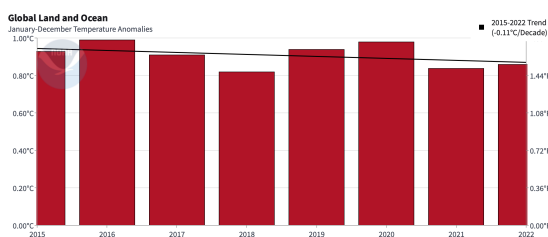
Table 1: Data in both tables were obtained using World Bank's [database](#) and the whole computation procedure is computed in this [notebook](#).

### #3. Lead author: [Lukas Rasocha], Calling BS on "Steve Milloy's tweet claiming that global warming due to C02 emission is a hoax." (January, 2023)

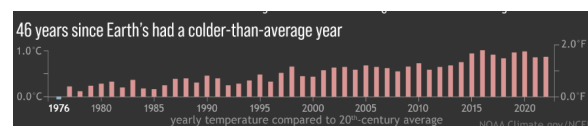
Steve Milloy, the author of the [tweet](#), argues that global warming caused by CO2 emissions is a hoax, based on a figure comparing yearly temperature to the 20th-century average. He includes a regression line to support his claim of a cooling trend and a declining temperature rate. I call BS his claim that global warming due to C02 emission is a hoax. The reason for this is that he is **lying with his visualization**.

Figure 1 presents both Milloy's figure and the original figure from the [NOAA's press release](#), contradicting his assertions. First, the original figure clearly demonstrates that it has been 46 years since the average yearly temperature was cooler than the 20th-century average. Furthermore, it highlights that the years 2014 to 2022 rank among the ten warmest years. Finally, Professor of geosciences at Princeton University Gabriel Vecchi explains in one of the [articles](#) that climate patterns such as [El Niño and La Niña](#) can cause temporary surface warming and cooling, resulting in fluctuations in the warming rate over short periods. Consequently, cherry-picking specific timeframes to portray a trend in warming/cooling is an invalid approach. Thus, Milloy's distortion of the visualization by selectively zooming in on a small portion of the original figure, is obscuring the overall long-term warming trend. Finally, another visualization technique that distorts reader's perception is the regression line which human brain naturally extends further and consequently can lead the reader to the wrong conclusion that subsequent years will follow the same trajectory.

Global warming is an ongoing and significant issue, with substantial efforts already undertaken. However, much work remains, and politicians play a critical role in driving change to mitigate global warming. To facilitate the implementation of measures by politicians, it is essential to ensure the public is well-informed. Unfortunately, tweets like the one by Steve Milloy only deepen divisions among people's opinions instead of fostering unity. As a consequence, progress towards achieving zero CO2 emissions is hindered, slowing down potential advancements.



(a) Steve Milloy's figure used in his [tweet](#).



(b) Original figure from [NOAA's press release](#).

Figure 1: For both figures, the x-axis captures time in years and the y-axis captures the difference between average temperature in the given year and average temperature of the whole 20th century. The only difference between the two figures is the time span which was selectively chosen by Milloy to be the years between 2015 and 2022. From such perspective, one can see a cooling trend which is further supported by the regression line. However, as argued above, the original NOAA's visualization shows the long term warming trend and as such disputes Milloy's claim.

**#4. Lead author: [Lukas Rasocha], Calling BS on "Dr. Simone Gold's tweet claiming 'The same number of athletes died in the last TWO years as compared to a prior 38 years.'" (January, 2023)**

In her Twitter thread, Dr. Simone Gold, an experienced emergency physician, begins by stating that prior to 2020, athletes being incapacitated or dropping dead was not a common occurrence. However, she expresses concern that such incidents are now happening with alarming frequency. To support her claim, Dr. Gold backs her claim by two statistics coming from the [letter](#) to the editor of an immunology journal. The first statistic ([source](#)) states that between 1966 and 2004 (a span of 38 years), a total of 1101 athletes died due to various heart conditions. The second statistic, sourced from another article ([source](#)), indicates that from 2021 until the beginning of 2023, the same number of athletes died. To summarise, I call BS on Dr. Gold claim that *The same number of athletes died in the last TWO years as compared to a prior 38 years* due to **unfair comparison** being made since she compares counts of two different groups of people.

Firstly, the first source focuses specifically on athletes under the age of 35, whereas the second source encompasses athletes of all age groups. Secondly, there are disparities in how the two sources define an athlete. The definition provided in the second source is more ambiguous and encompasses a broader spectrum of individuals who engage in athletic activities. Thirdly, it is important to note that the first source only includes athletes who died from various heart conditions, while the second source takes into account all types of deaths among athletes. Consequently, the two counts of the athletes are not the same, contrary to what Dr. Gold wrote. As a result, drawing any meaningful conclusions based on the comparison of these two numbers is not valid due to the inherent unfairness in the comparison. Although Dr. Gold does not explicitly state it in her tweet, the context, such as the year range for the second statistic starting in 2021, suggests that she is attempting to link the "increased" death rate among athletes to vaccination. Such a tweet supports the opinion that vaccines are dangerous and advocates for avoiding them. However, promoting vaccine hesitancy in this manner can hinder the progress of recovering from the global pandemic and returning to the "normal world".

The presentation by Group A, B, E, and M in the course "Reflections on Data Science" discussed the energy consumption of machine learning (ML) and its societal implications. The groups highlighted that the increased demand for artificial intelligence (AI) has led to a significant increase in greenhouse gas emissions, with training a single AI model emitting as much carbon as five cars in their lifetimes. By 2025, AI is estimated to account for 5 % of global electricity consumption.

The groups discussed the energy consumption of different AI technologies, such as ChatGPT's electricity consumption, data storage, and the training and iteration of datasets, leading to larger models that require more energy. They also provided several tools for estimating the carbon footprint of deep learning and highlighted the importance of reducing energy consumption and tracking and reporting costs, just like other metrics.

Furthermore, the groups explored the societal implications of AI's energy consumption, including environmental, ethical, and regulatory implications, with the need for solutions such as using renewable energy, developing more efficient technology, and buying carbon offsets. They also emphasized the importance of ethical safeguards, policies, and assessment protocols to mitigate the impact of AI's energy consumption on the environment and society.

In conclusion, the groups concluded that the incorporation of large machine learning models into existing products elevates the user experience, but there needs to be a debate on whether it justifies the large energy consumption. The adoption of AI and ML in emerging markets raises environmental challenges, and researchers and policymakers need to consider the environmental impact of AI. They highlighted the impending energy and resource strain that is likely to occur and the need for solutions to address this issue.