ML-based Election Result Forecasting

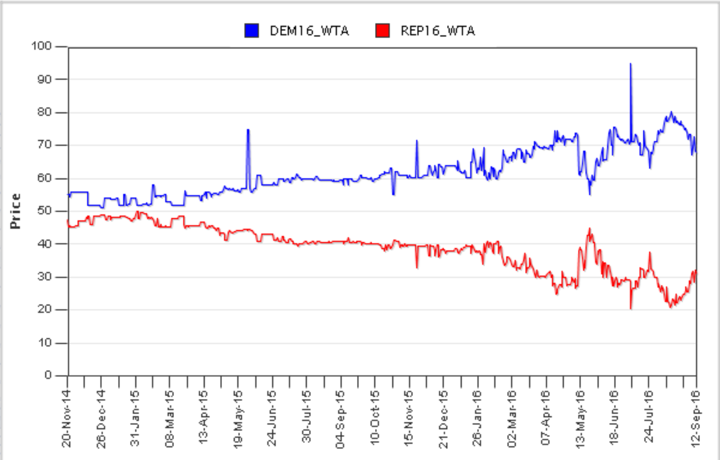
G. Chatziparaskevas, Prof. Ioannis Pitas
Aristotle University of Thessaloniki
pitas@csd.auth.gr
www.aiia.csd.auth.gr

Version 1.0

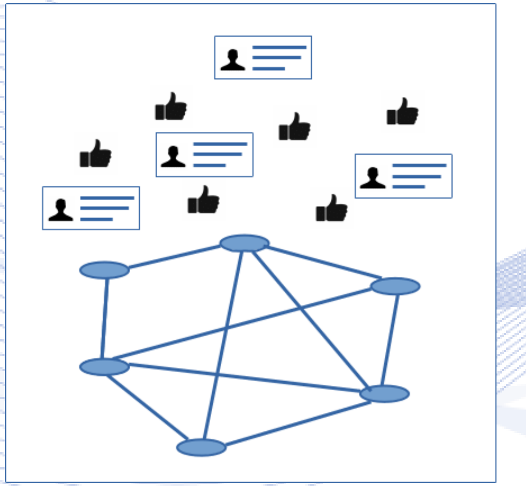
Artificial Intelligence &
Information Analysis Lab

Forecasting election results: General Approaches

1. Traditional: Polls, Prediction markets.
2. ML-based: Opinion mining from social media, online newspaper comments etc.



Source: Iowa Electronic Markets



Forecasting election results from tweets

1. Use a large number of tweets posted in the pre-election period.
2. Identify the tweets referring to political parties.
3. Assign a sentiment value to each tweet.
4. Produce metrics that can be used as proxy of the public opinion.
5. Apply a forecasting model to the metrics.

Forecasting election results from tweets: Metrics extraction 1/3

1. Aggregate tweets per day.
2. Find the number of tweets referring to i-th party with positive (p_i) and negative (n_i) sentiment.
3. Compute daily metrics to be used as proxies for vote intention.
4. Trivial approach: Positive tweets share of i-th party:

$$S_i = \frac{p_i}{\sum_{j=1}^{N_{parties}} p_j}$$

- Drawback: Tweets with negative sentiment not used yet more numerous.

Artificial Intelligence & Information Analysis Lab

Forecasting election results from tweets: Metrics extraction 2/3

Target: ***Take advantage of both positive and negative tweets.***

The idea:

1) Count the negative tweets for one party (n_i) as supportive for the other and vice versa.
2) Use the share of all the tweets referring to two parties that are supportive for each party of the pair to measure the preference of one party over the other.

Consider two parties i, j:

$$S_i = \frac{p_i + n_j}{Total_{i,j}},\ S_j = \frac{p_j + n_i}{Total_{i,j}},\ Total_{i,j} = p_i + n_i + p_j + n_j$$

Forecasting election results from tweets: Metrics extraction 3/3

Idea: ***Use their pairwise differences instead indicating preference of one party over the other***:

1) For each pair or parties i, j compute the difference of their metrics:

$$d_{ij} = S_i - S_j.$$

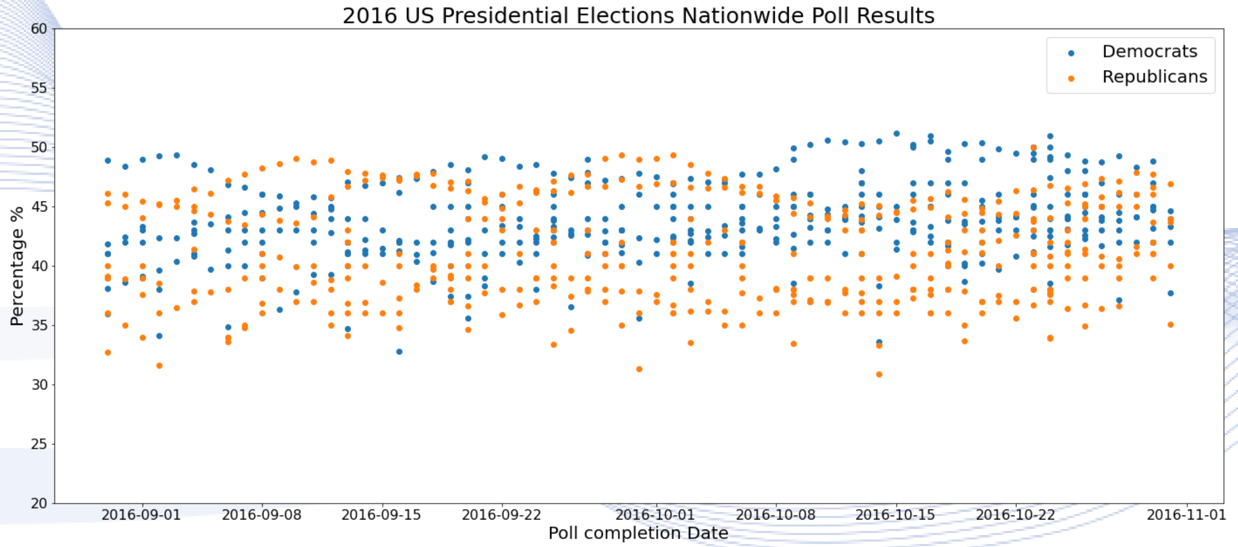
2) Apply the forecasting model to the differences d_{ij}.

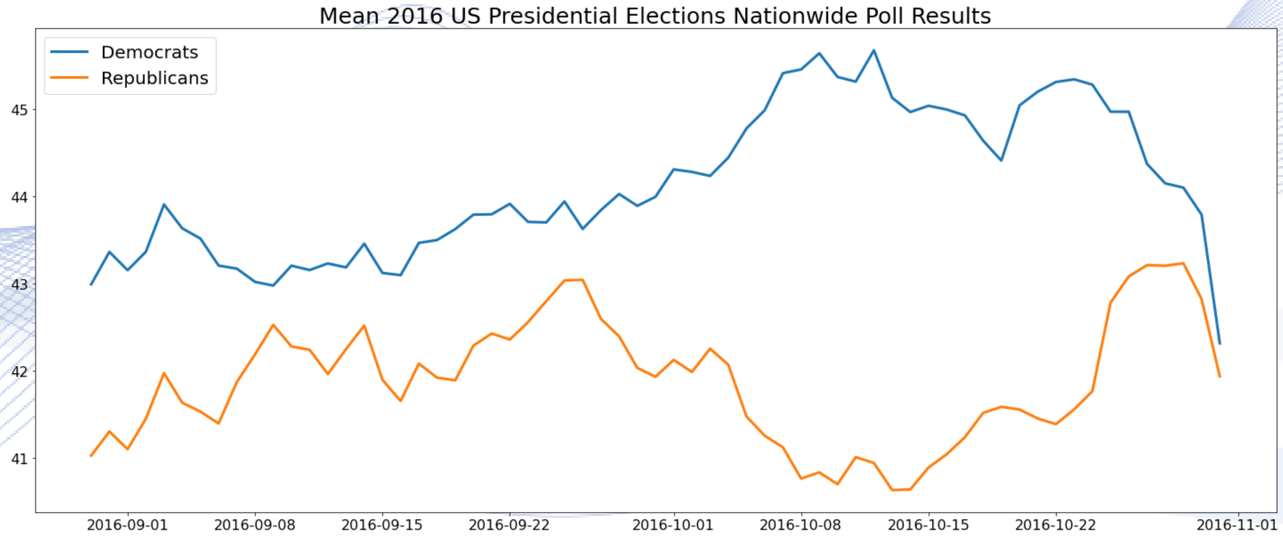
3) MDS can be applied in the end to extract the ordinality of the metrics / ranking of the parties.

Artificial Intelligence & Information Analysis Lab

2016 US Presidential Elections Polls

1. Use nation-wide poll results during pre-election period.
2. Isolate the main contenders (Democrats & Republicans).
3. Extract mean result per day averaging the results of polls being active at that day.

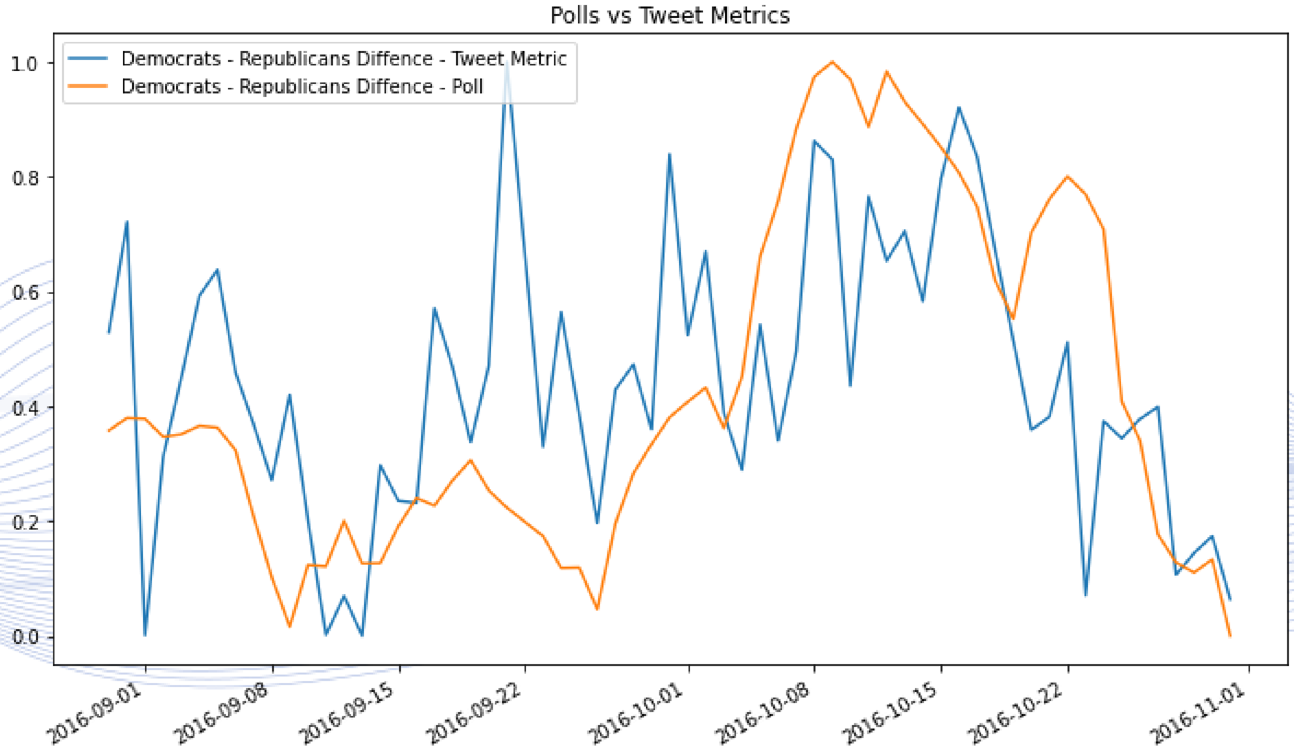




Artificial Intelligence & Information Analysis Lab

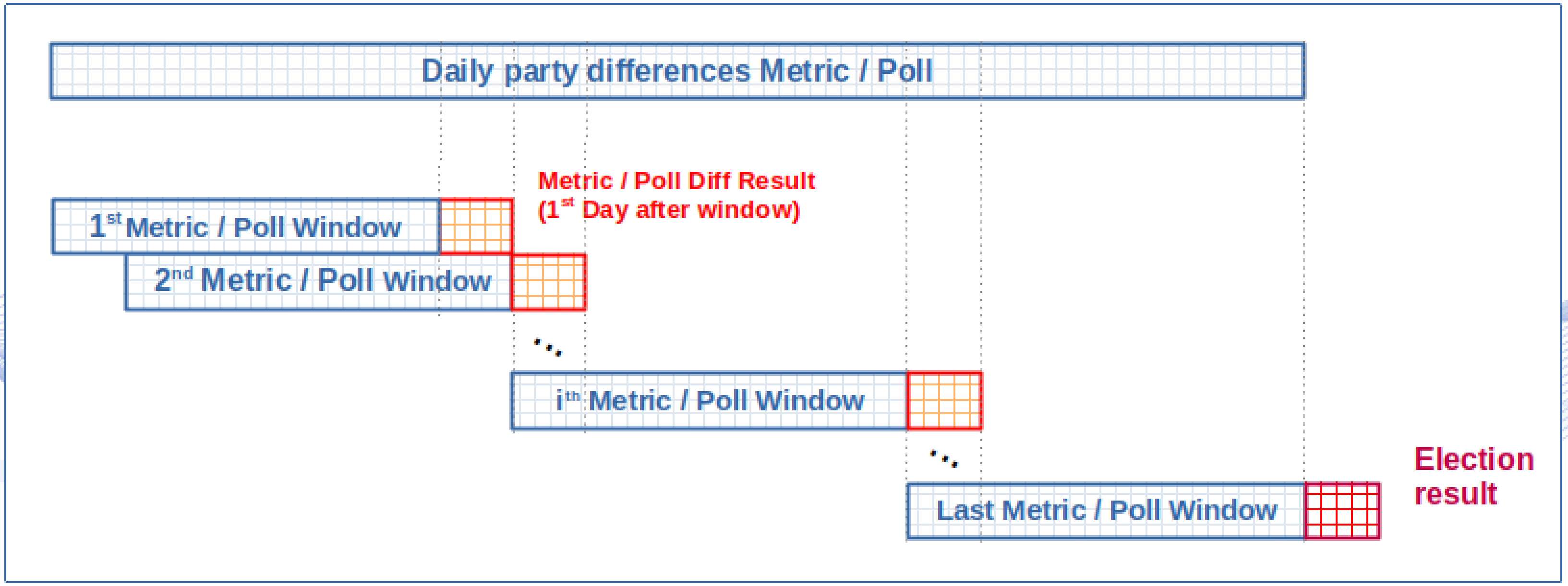
2016 US Presidential Elections
Tweet Metric vs Poll

1. Normalized Difference between Democrats and Republicans as obtained from Polls and Tweet Metric:
2. Correlation is evident!



Forecasting election results from tweets: Time-series Formulation

Strategy: Rolling window single-step ahead forecasting.
Input: Window of daily party metric and poll differentials.
Output: Poll differential on the day after each input window - Election result differential estimate for the last input window.



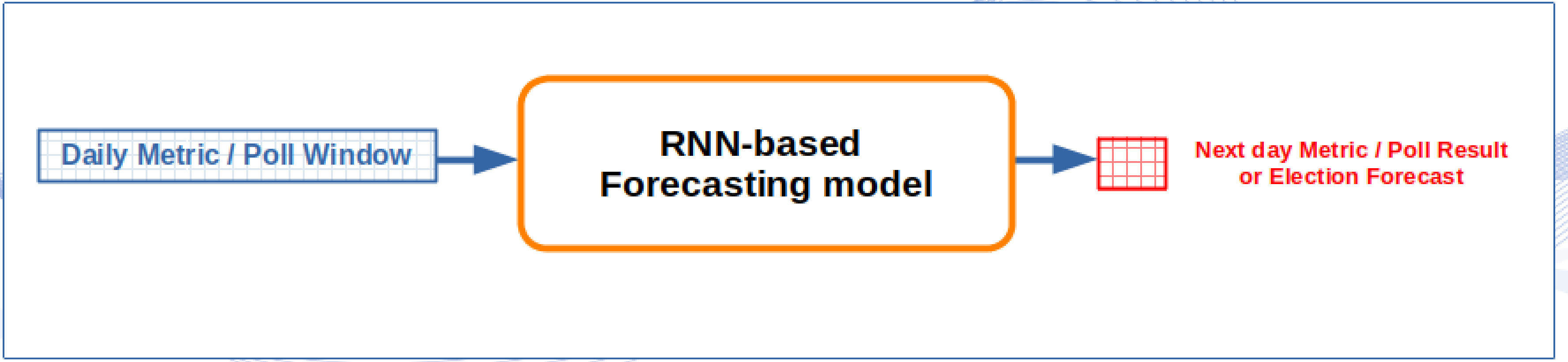
Forecasting election results from tweets: datasets

Tweets dataset:
https://www.kaggle.com/paulrohan2020/2016-usa-presidential-election-tweets61m-rows

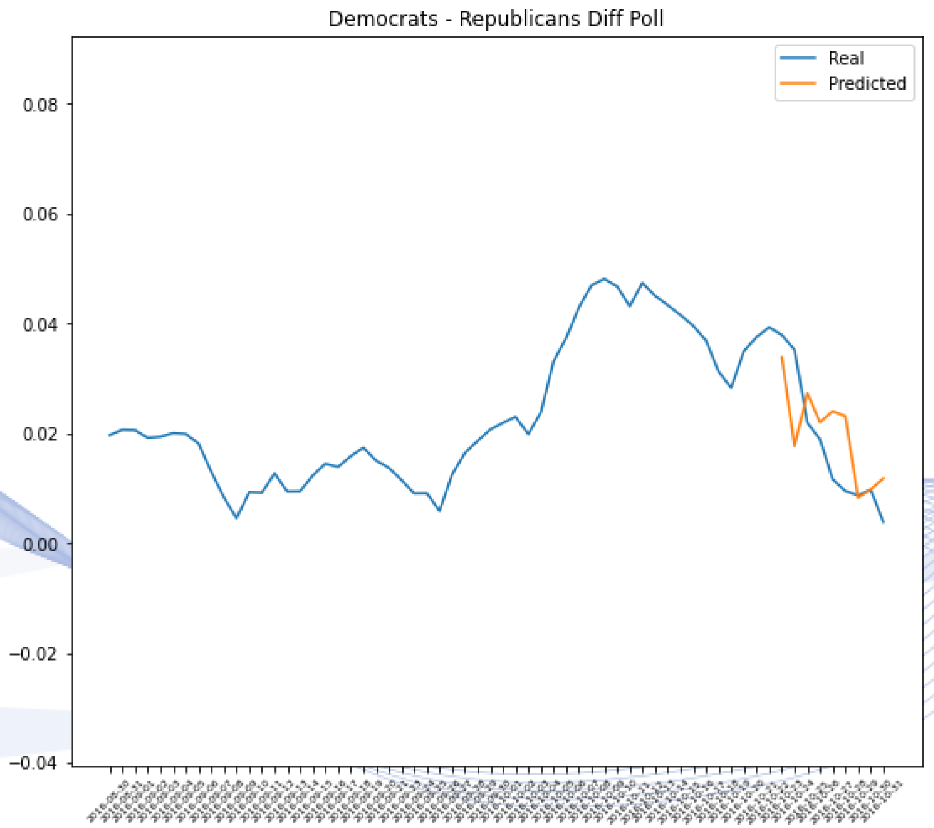
Poll dataset:
https://github.com/EugeneYilia/data_analysis_2

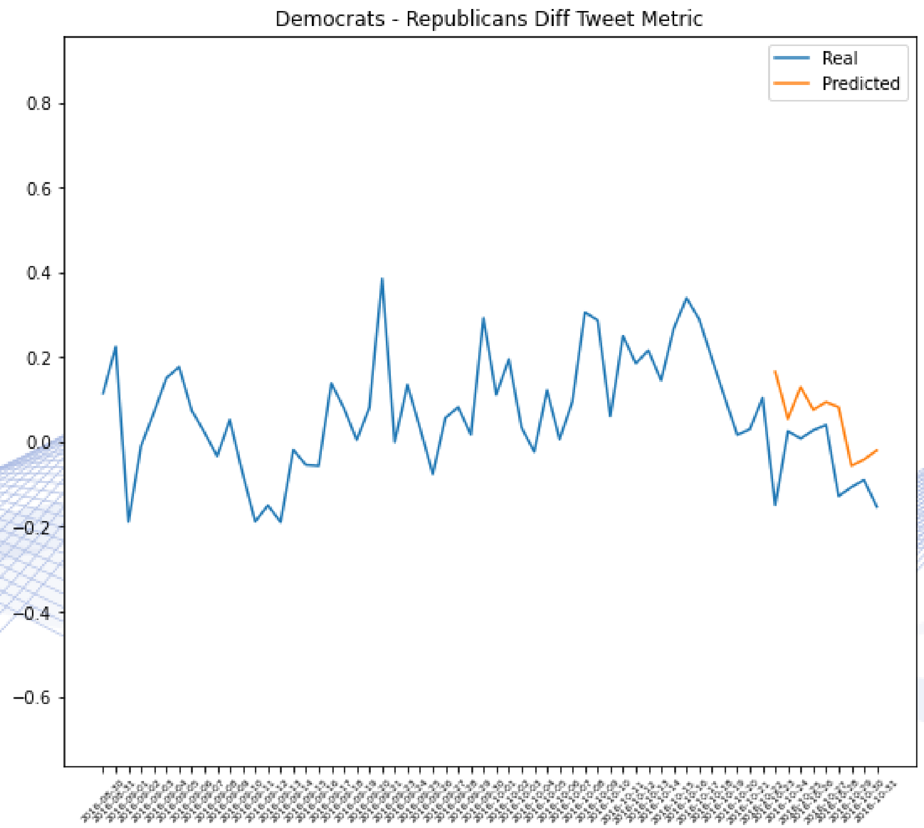
Forecasting election results from tweets: Forecasting Model

Model: RNN-based producing single-time-step forecast.
Training – Evaluation:
- Dataset split: 80-20 Train / Validation – Test split.
- Training: 5-fold cross-validation.
- Hyper-parameter optimization with “SMAC” method.



2016 US Presidential elections Poll and Tweet Metric Forecasts





Artificial Intelligence & Information Analysis Lab

2016 US Presidential Election actual Result Forecast

	Democrats vs Republicans Difference	Democrats vs Republicans Difference (basis points)	Winner
Actual Election Result	-0.021000	-2	Republicans
Poll Forecast	0.003840	0.3	Democrats
Tweet Metric Forecast	-0.098890	-9	Republicans
Mean of Poll & Tweet Metric Forecasts	−0,047525	-4	Republicans

Main References

Tweets dataset:
https://www.kaggle.com/paulrohan2020/2016-usa-presidential-election-tweets61m-rows

Poll dataset:
https://github.com/EugeneYilia/data_analysis_2

Bibliography

[BOV2018] Bovet, A., Morone, F. & Makse, H.A. Validation of Twitter opinion trends with national polling aggregates: Hillary Clinton vs Donald Trump. Sci Rep 8, 8673 (2018).

[HEW2021] H. Hewamalage, C. Bergmeir, and K. Bandara, “Recurrent neural networks for time series forecasting: Current status and future directions,” *International Journal of Forecasting*, vol. 37, no. 1, pp. 388–427, 2021.

Bibliography

[1] I. Pitas, “Artificial Intelligence Science and Society Part A: Introduction to AI Science and Information Technology“, Amazon/Kindle Direct Publishing, 2022,

https://www.amazon.com/dp/9609156460?ref_=pe_3052080_397514860

[2] I. Pitas, “Artificial Intelligence Science and Society Part B: AI Science, Mind and Humans“, Amazon/Kindle Direct Publishing, 2022, https://www.amazon.com/dp/9609156479?ref_=pe_3052080_397514860

[3] I. Pitas, “Artificial Intelligence Science and Society Part C: AI Science and Society”, Amazon/Kindle Direct Publishing, 2022, https://www.amazon.com/dp/9609156487?ref_=pe_3052080_397514860

[4] I. Pitas, “Artificial Intelligence Science and Society Part D: AI Science and the Environment“, Amazon/Kindle Direct Publishing, 2022,

https://www.amazon.com/dp/9609156495?ref_=pe_3052080_397514860

Q & A

Thank you very much for your attention!

More material in
http://icarus.csd.auth.gr/cvml-web-lecture-series/

Contact: Prof. I. Pitas
pitas@csd.auth.gr