

Systems: needed more
than ever.
Partly due to AI

Jon Crowcroft, 29.11.2023

<https://www.cst.cam.ac.uk/people/jac22>

Stacks and Stacks of Latency

- “No-one ever got a speedup from adding another layer”, J.Ousterhout
- Baseline your code so when you move from Toy to Dawn, don't slow down

McSherry, Isard, Murray: “Scalability, but at What Cost?”

<https://www.usenix.org/system/files/conference/hotos15/hotos15-paper-mcsherry.pdf>

Learning, through Unlearning, without Catastrophic Forgetting

- Suppose we need to unlearn one thing (gdpr etc etc)?
- Do we delete that item from input and retrain from scratch?
 - Obviously not, so what can we do?
 - Well how about we run differential privacy (while training) to determine whether an input item made a statistically significant difference in the *training data*?
 - As a side effect, get privacy:-)
 - Or we could do Shapley Values or Integrated Gradients to see whether an item actually resulted in a significant change to the *model*?
 - As a side effect, get explainability:-)

Libraries are your friend...

- Re-inventing wheels –
 - e.g. distributed ML SGD, re-invent de-synch
 - Its stochastic, so you can do random stuff 😊
 - Federated (aggregators) – re-invent clustering
 - Its federated, like peer-to-peer filesharing 😊
- Metrics, please – e.g. “accuracy” – not hallucinations 😊
 - Perplexity is not your friend
 - Empirical may be too late...

Conclusion

- Systems people aren't going to out of work any day soon...
- Good systems design is not just post hoc optimization
- Software archeology will uncover evidence of entire cities being levelled to the ground but the occasional jewel.



- A small pot from 6 layers down in Meggido(*) (approx. 6000 years old):-
- *a.k.a. Armageddon...