GETTING AI TO FORGET YOU

RAYMOND BROWN – NITRD CO-CHAIR PRIVACY R&D IWG

These comments are my own and not official representation of NSA
Forgetting is a tricky business, both for humans and for artificial intelligence. For humans, forgetting is more than just a failure to remember; it’s an active process that helps the brain take in new information and make decisions more effectively.

AI’s lasting memory raises not only technical issues, but concerns related to privacy, law and ethics. Imagine if your household robot witnessed you having a sneaky cigarette despite what you tell your doctor and insurance company? What if your TV heard you smother your spouse?
This is not advocation for AI to be changed so we can conduct crimes!
WHILE FORGETTING IS SOMETHING THAT MIGHT CAUSE US FRUSTRATION, IT IS THE WAY WE FORGET THAT MAKES PEOPLE STILL SUPERIOR TO AI.

Machine learning algorithms are poor at knowing when to keep old information and when to discard outdated information.

“Connectionist AI” (AI that often uses neural networks modelled on the structure of the brain) faces several problems related to “forgetting”. These include over-fitting, which is when a learning machine stores overly detailed information from past experiences, hindering its ability to generalize and predict future events.

Another problem is “Catastrophic forgetting”. Researchers are trying to build artificial neural networks that can appropriately adjust to new information without abruptly forgetting what they learned before.

Sometimes the neurons of an artificial neural network adopt undesirable activation patterns early in the learning process, damaging the future learning ability of the AI.

An alternative approach to storing memories in robots is symbolic memory representations where knowledge is represented by logical facts (“birds fly”, “Tweety is a bird”, so therefore, “Tweety can fly”). These highly structured human-created representations can be easily deleted, just like deleting a file on a computer.
HENCE...

• IT'S NOT JUST THE MECHANICS OF MAKING AI FORGET THAT WE NEED TO PUSH RESEARCH ON, BUT WHEN AND WHY IT SHOULD FORGET THAT NEEDS TO BE ADDITIONALLY STUDIED AND ADVANCED.
HOW IS THE FIELD APPROACHED TODAY

Leading techniques:
- Homomorphic Encryption
- Differential Privacy

Does anyone see the irony here?
Concretely, we may frame the problem of data deletion in machine learning as follows. Suppose a statistical model is trained on \((N)\) datapoints. For example, the model could be trained to perform disease diagnosis from data collected from \((N)\) patients. To delete the data sampled from the \((i)\)-th patient from our trained model, we would like to update it such that it becomes independent of sample \((i)\), and looks as if it had been trained on the remaining \((N - 1)\) patients. A naive approach to satisfy the requested deletion would be to retrain the model from scratch on the data from the remaining \((N - 1)\) patients.
• For many applications, this is not a tractable solution – the costs (in time, computation, and energy) for training many machine learning models can be quite high. Large scale algorithms can take weeks to train and consume large amounts of electricity and other resources. Hence, we posit that efficient data deletion is a fundamental data management operation for machine learning models and AI systems, just like in relational databases or other classical data structures.
DATA DELETION IS NOT PRIVACY

• **Algorithms** that support efficient deletion do not have to be private, and algorithms that are private do not have to support efficient deletion.

• **An operation** that outputs the entire dataset in the clear could support data deletion, whereas such an operation is certainly not private. In this sense, the challenge of data deletion only arises in the presence of computational limitations.

• **Privacy**, on the other hand, presents statistical challenges, even in the absence of any computational limitations. However, data deletion has direct connections and consequences in data privacy and security, which we explore in more detail in Appendix A.
ENSEMBLE: A RADICAL APPROACH TO DELETION WHILE RETAINING LEGACY MEMORIES
ENSEMBLE: A RADICAL APPROACH TO DELETION WHILE RETAINING LEGACY MEMORIES
ENSEMBLE: A RADICAL APPROACH TO DELETION WHILE RETAINING LEGACY MEMORIES
QUESTIONS?

• It's possible that our brains and distinctly human processes, like forgetting, hold the map to creating strong artificial intelligence, but scientists are collectively still figuring out how to read the directions.

HTTP://NAFTALI-TISHBY.MYSTRIKINGLY.COM/