All.Net Analyst Report and Newsletter

Welcome to our Analyst Report and Newsletter

Why Federated models in AI leak secrets

I was at a meeting last week discussing how AI models might be used to support archives for efficiency, effectiveness, etc. while maintaining the requirements for those archives to provide reliable and authentic records and maintain confidentiality of relevant record data elements.

Someone across the room indicated that they were going to federate models, not records themselves, so that you could effectively get results across multiple models without leaking content. I countered that the models were going to leak content, and they indicated that indeed federating models and not the content would not leak information.

If you disagree, we'll meet in the corner during the break

So just before the next break, I indicated to all present that anybody who thought Federated models would not leak content could meet me in the corner of the room and we would settle it. There was a general Oooo in the room as it sounded something like meeting me in the alley outside the bar.

A simple question

So we met there and I asked a simple question:

> Will the Federated model give different results from the individual models?
>
> ***Them: Yes***
>
> Me: Then it leaks information.
>
> ***Them: But I don't know how I could get that information out***

The next morning

That evening, I decided to pose a thought experiment to the group, which I did the next morning. Here's how it goes:

- Let's make a large language AI model of all the works of literature and other content ever written before 1940. As comprehensive as you like. That's the first model.
- Now let's make another model of only classified information about nuclear weapons.
- Now federate the models and start asking questions about anything involving nuclear weapons.

Obviously, since there were no nuclear weapons or literature about it before 1940, anything you get will be from the classified documents.

- Now federate a model from the confidential information about disabilities from a university that started keeping those records in in the 1960s. Start asking about those issues. And so on...

Conclusions

Folks who don't really know security make lots of big mistakes. Get an expert please.

