CRAIG ([00:00](https://www.temi.com/editor/t/ZUCSrneVBFHP9wAsgc6D0fIbtzjwR_bd7SUbxpWJpkvfl4l1-FlKAwVWaUwXvaGwLI8ocq_8r2fxP1AIqfvxA2Fu1tQ?loadFrom=DocumentDeeplink&ts=0.09)):

Hi, I'm Craig Smith and this is Eye on AI.

CRAIG ([00:13](https://www.temi.com/editor/t/ZUCSrneVBFHP9wAsgc6D0fIbtzjwR_bd7SUbxpWJpkvfl4l1-FlKAwVWaUwXvaGwLI8ocq_8r2fxP1AIqfvxA2Fu1tQ?loadFrom=DocumentDeeplink&ts=13.89)):

COVID-19 has swept across the world was startling speed, but with equally startling speed, the machine learning community has responded. By the end of March, initiatives had sprung up around the world bringing the power of machine learning to bear on tracking the virus and finding a cure. One of the quickest to organize was in Montreal where Turring award winner Yoshua Bengio and his team at Mila, together with the University of Montreal, sprang into action with a task force to tackle the problem. This week I speak with Irina Rish, a professor at the University of Montreal and a Mila academic member, who is helping head that task force. She talked about where the efforts currently stand and where they expect to go in the weeks and months ahead. It's early days as Dr. Rish makes clear, but her team's efforts provide a welcome bright spot in a gloomy situation. You can follow their progress and find ways to participate at their website: AI-against-COVID.ca. I hope you find the conversation as important as I did.

CRAIG ([01:37](https://www.temi.com/editor/t/ZUCSrneVBFHP9wAsgc6D0fIbtzjwR_bd7SUbxpWJpkvfl4l1-FlKAwVWaUwXvaGwLI8ocq_8r2fxP1AIqfvxA2Fu1tQ?loadFrom=DocumentDeeplink&ts=97.24)):

What I like to do is start by asking who you are, where you came from, how you got involved in machine learning, and then we'll go from there.

IRINA ([01:46](https://www.temi.com/editor/t/ZUCSrneVBFHP9wAsgc6D0fIbtzjwR_bd7SUbxpWJpkvfl4l1-FlKAwVWaUwXvaGwLI8ocq_8r2fxP1AIqfvxA2Fu1tQ?loadFrom=DocumentDeeplink&ts=106.9)):

So I graduated from University of California. Irvine and was doing artificial intelligence as my main subject. I was mainly working on probabilistic models, patient networks and inference and then machine learning. So I kind of worked on machine learning for a long time. Then I was a researcher at IBM, TJ Watson research in New York. At IBM, I had quite a bit of experience with biological data and particularly neural imaging, neuroscience data, applying machine learning to data analysis, and also kind of sitting on two chairs, trying to also apply neuroscience ideas to better artificial intelligence algorithms. Anyways, there's this intersection between machine learning and biology. So I joined the University of Montreal and Mila, the Quebec AI Institute, in October - formerly Montreal Institute for Learning Algorithms, we have probably close to 400 people by now - teaching classes there on AI and different aspects of machine learning.

IRINA ([02:55](https://www.temi.com/editor/t/ZUCSrneVBFHP9wAsgc6D0fIbtzjwR_bd7SUbxpWJpkvfl4l1-FlKAwVWaUwXvaGwLI8ocq_8r2fxP1AIqfvxA2Fu1tQ?loadFrom=DocumentDeeplink&ts=175.67)):

And then the whole situation started developing with COVID. Actually, interestingly, it's students at Mila who first started alerting everyone. They created even a Slack channel on coronavirus, when people were still kind of not taking it seriously. Basically when adults were taking it seriously, kids actually started noticing and alarming adults, which was quite interesting. So, in a couple of weeks we all of course realized what's going on. Mila definitely adopted the social distancing and started working from home since I think March 12th. So basically, then the situation started developing. As we know in the US, things were developing much faster than anticipated. In Montreal things are not yet developing as fast, but still it's still an exponential curve. Just maybe there's a bit different exponent, which is probably because people indeed seem to follow social distancing recommendations. But given what's going on, a very natural idea that I started discussing with students and professors and with Yoshua Bengio ideas, and put together some form of taskforce about AI against coronavirus and try to see what people, data analysis, skill machine learning, AI in general can do if they team up with biologists and epidemiologists.

IRINA ([04:27](https://www.temi.com/editor/t/ZUCSrneVBFHP9wAsgc6D0fIbtzjwR_bd7SUbxpWJpkvfl4l1-FlKAwVWaUwXvaGwLI8ocq_8r2fxP1AIqfvxA2Fu1tQ?loadFrom=DocumentDeeplink&ts=267.19)):

Then for anyone who actually studies pandemics in general and this particular virus specifically, the challenge is finding data sets. For example if you want, data sets on various omics like proteomics, genomics and transcriptomics and so on, it's not readily available yet. But they are kind of being created as we speak. So more than a week already, almost two weeks, we were just working on maybe logistics initially trying to put together websites jointly with people in Montreal, at universities, colleagues in IBM, research colleagues in Germany. We have website actually listing all our activities now. It's live. There are collaborative spaces. There is a open Slack channel which was created specifically for these collaborative efforts between Mila and other places across the globe. Pretty much it's open Slack and people keep joining in that Slack. We have multiple projects and people trying to cluster around that. At the same time, there is a collaborative side more generally, not just with Mila, that was created by Mila students and professors in Germany. Everything is outlined now on one central website that we are just putting together. It's up and running now. It's called AI-against-COVID.ca.

IRINA ([06:06](https://www.temi.com/editor/t/ZUCSrneVBFHP9wAsgc6D0fIbtzjwR_bd7SUbxpWJpkvfl4l1-FlKAwVWaUwXvaGwLI8ocq_8r2fxP1AIqfvxA2Fu1tQ?loadFrom=DocumentDeeplink&ts=366.76)):

Various efforts, not just at Mila and collaborators of Mila, but also at Vector Institute and University of Toronto and also University of Alberta and more generally CIFAR, which is a Canadian organization supporting AI research. So this website is basically keeping track of various efforts at all three organizations. More specifically the Mila website keeps track of all the projects and this direction that are taking place right now. I personally kind of more involved in two projects that really to data analysis or omics data jointly with my colleagues at Mila and our colleagues at other institutions. So there are multiple clusters of collaborators around this direction of biological data or mixed data and also clinical data.

IRINA ([07:04](https://www.temi.com/editor/t/ZUCSrneVBFHP9wAsgc6D0fIbtzjwR_bd7SUbxpWJpkvfl4l1-FlKAwVWaUwXvaGwLI8ocq_8r2fxP1AIqfvxA2Fu1tQ?loadFrom=DocumentDeeplink&ts=424.15)):

There is a new clinical trial that is starting now here in Montreal trying another potential drug and it's supposed to be like an accelerated trial because they are really trying to get data soon as maybe June. There are activities also in the direction of not necessarily biological data analysis, but other relevant data sets. For example, tracking progression of disease. Somewhat similar to what John Hopkins university is doing. They have this map.

IRINA ([07:38](https://www.temi.com/editor/t/ZUCSrneVBFHP9wAsgc6D0fIbtzjwR_bd7SUbxpWJpkvfl4l1-FlKAwVWaUwXvaGwLI8ocq_8r2fxP1AIqfvxA2Fu1tQ?loadFrom=DocumentDeeplink&ts=458.8)):

We have ideas about, again on the webpage, about how to maybe use data collected in peer to peer fashion, say using apps so that you can have an app that constantly updates your risk, your probability of getting infected depending on where you go, whom you communicate with and possibly other factors. And this type of app, the high level description was put online couple of days ago by your Yoshua Bengio in his blog. There is a big team trying to kind of work on this app and make it functional as soon as possible. I'm also involved there. There are also multiple areas of machine learning from like good old graphical probabilistic models and probability of [inaudible]. There is algorithms [inaudible] propagation to epidemiological models. It's all basically being created as we speak and the directions may change, but that's kind of one umbrella topic of having this personal guardian angel that can use the data and can be useful in the shorter term, while we are working on potentially novel drugs or trying drug repurposing and just doing simulations testing. What kind of existing drugs may be helpful with this specific virus.

INTERLUDE ([08:58](https://www.temi.com/editor/t/ZUCSrneVBFHP9wAsgc6D0fIbtzjwR_bd7SUbxpWJpkvfl4l1-FlKAwVWaUwXvaGwLI8ocq_8r2fxP1AIqfvxA2Fu1tQ?loadFrom=DocumentDeeplink&ts=538.831)):

CHORD

CRAIG ([09:04](https://www.temi.com/editor/t/ZUCSrneVBFHP9wAsgc6D0fIbtzjwR_bd7SUbxpWJpkvfl4l1-FlKAwVWaUwXvaGwLI8ocq_8r2fxP1AIqfvxA2Fu1tQ?loadFrom=DocumentDeeplink&ts=544.01)):

The task force you're, you're one of the leaders of this task force, began two weeks ago?

IRINA ([09:11](https://www.temi.com/editor/t/ZUCSrneVBFHP9wAsgc6D0fIbtzjwR_bd7SUbxpWJpkvfl4l1-FlKAwVWaUwXvaGwLI8ocq_8r2fxP1AIqfvxA2Fu1tQ?loadFrom=DocumentDeeplink&ts=551.08)):

Yeah, I would say yeah, definitely. This is all the end of this week. I think it really kind of all started basically the weekend when the Mila was closed. That was I guess March 12th.

CRAIG ([09:24](https://www.temi.com/editor/t/ZUCSrneVBFHP9wAsgc6D0fIbtzjwR_bd7SUbxpWJpkvfl4l1-FlKAwVWaUwXvaGwLI8ocq_8r2fxP1AIqfvxA2Fu1tQ?loadFrom=DocumentDeeplink&ts=564.2)):

On this app that Yoshua's talking about. It sounds similar in a way to what they've done in China.

IRINA ([09:31](https://www.temi.com/editor/t/ZUCSrneVBFHP9wAsgc6D0fIbtzjwR_bd7SUbxpWJpkvfl4l1-FlKAwVWaUwXvaGwLI8ocq_8r2fxP1AIqfvxA2Fu1tQ?loadFrom=DocumentDeeplink&ts=571.16)):

And Israel.

CRAIG ([09:32](https://www.temi.com/editor/t/ZUCSrneVBFHP9wAsgc6D0fIbtzjwR_bd7SUbxpWJpkvfl4l1-FlKAwVWaUwXvaGwLI8ocq_8r2fxP1AIqfvxA2Fu1tQ?loadFrom=DocumentDeeplink&ts=572.27)):

Oh is that right? Is Israel doing the same? Where in China, they're mining Alibaba payment app data to track where people have been and then they score according to your exposure to known disease carriers and you're given a color coded label and actually the government then is enforcing increasingly restrictive levels of quarantine on you depending on your, your risk factor. But this would be kind of a personal version of that, is that right?

IRINA ([10:06](https://www.temi.com/editor/t/ZUCSrneVBFHP9wAsgc6D0fIbtzjwR_bd7SUbxpWJpkvfl4l1-FlKAwVWaUwXvaGwLI8ocq_8r2fxP1AIqfvxA2Fu1tQ?loadFrom=DocumentDeeplink&ts=606.77)):

Exactly. The main difference, I know there is lots of concerns. When I was posting this information on Twitter and yoshua posted this on Facebook, there was very mixed reaction because people are afraid. Even though we say the main purpose of this is to be peer to peer. This is not being sent - any individual information - is not being sent to government. That's the main difference. It will be federated learning. Government or any authority organizations may have some summary statistics but they not going to have your personal data. That's the main difference. That's why we do not want to deploy the existing apps of different kinds and the reason is because our thinking was that okay, eventually if we implement this hammer and dance strategy, right, and after the hammer we will have to start reopening businesses and going around life in a more or less normal way.

IRINA ([11:07](https://www.temi.com/editor/t/ZUCSrneVBFHP9wAsgc6D0fIbtzjwR_bd7SUbxpWJpkvfl4l1-FlKAwVWaUwXvaGwLI8ocq_8r2fxP1AIqfvxA2Fu1tQ?loadFrom=DocumentDeeplink&ts=667.4)):

But we'll have to be very cautious. We'll be in the dance stage. You want to be more informed about the situation, about the chances about the danger over particular setting which you are in. And presumably this information will allow you to avoid complete lockdown, complete social distancing. You still can be in safe situations, some kind of real time data analysis to allow you to evaluate that. So that's, that's the main idea. So that's type of an app that may not necessarily be used like next week because we have much more pressing issues next week and in the next two weeks and now probably there is no way but social distancing. But hopefully it will not last for too long. And when we, at least in the phase when we start getting out of the house, we will be more informed and can cautiously, instead of just binary, social distancing-no-social distancing, something in between which is more important, more situation dependent. So that's the idea. So we kind of try to balance the economic factors versus safety and saving lives.

INTERLUDE ([12:17](https://www.temi.com/editor/t/ZUCSrneVBFHP9wAsgc6D0fIbtzjwR_bd7SUbxpWJpkvfl4l1-FlKAwVWaUwXvaGwLI8ocq_8r2fxP1AIqfvxA2Fu1tQ?loadFrom=DocumentDeeplink&ts=737.17)):

CHORD

CRAIG ([12:20](https://www.temi.com/editor/t/ZUCSrneVBFHP9wAsgc6D0fIbtzjwR_bd7SUbxpWJpkvfl4l1-FlKAwVWaUwXvaGwLI8ocq_8r2fxP1AIqfvxA2Fu1tQ?loadFrom=DocumentDeeplink&ts=740.37)):

On the data analysis side, can you explain a little bit of what specifically you're looking at?

IRINA ([12:25](https://www.temi.com/editor/t/ZUCSrneVBFHP9wAsgc6D0fIbtzjwR_bd7SUbxpWJpkvfl4l1-FlKAwVWaUwXvaGwLI8ocq_8r2fxP1AIqfvxA2Fu1tQ?loadFrom=DocumentDeeplink&ts=745.16)):

Okay, so there are drugs everybody talks about. There are alternatives also being tried and they seem to be promising. And that's a joint work with Julie Hussin from University of Montreal as I mentioned, and several other people, more like enabler type of collaborative spaces. So it starts with this AI against COVID and then the peer to peer AI based tracing that I mentioned or essentially it might be a family of apps at the end of the day. And this whole thing that I was talking about, drug development, multi omic profiling and viral evolution studies. It relates to data collection in that accelerated trial that I mentioned. So this is a double blind clinical study called Co-Corona at the Montreal Heart Institute and they just announced that they are launching the Canadian clinical study and they want to determine short term treatment for COVID. So there again, they will be collecting the data and we will be helping them analyze this data, try to figure out both genetic or other factors that will determine whether that particular drug will be more helpful for patients. So you can have a better idea about more personalized treatments.

CRAIG ([13:41](https://www.temi.com/editor/t/ZUCSrneVBFHP9wAsgc6D0fIbtzjwR_bd7SUbxpWJpkvfl4l1-FlKAwVWaUwXvaGwLI8ocq_8r2fxP1AIqfvxA2Fu1tQ?loadFrom=DocumentDeeplink&ts=821.16)):

The AI application or the machine learning application in that case in the clinical trial is what exactly?

IRINA ([13:47](https://www.temi.com/editor/t/ZUCSrneVBFHP9wAsgc6D0fIbtzjwR_bd7SUbxpWJpkvfl4l1-FlKAwVWaUwXvaGwLI8ocq_8r2fxP1AIqfvxA2Fu1tQ?loadFrom=DocumentDeeplink&ts=827.95)):

You want to detect, interactions with viral genomic sequences and human genome. You want to analyze genomic data, you probably going to have other type of omics like proteomics. So you basically trying to see what variables in collected data are most relevant to the outcome, to positive outcomes. So when does it work, when doesn't it work, so who is going to benefit most from this treatment, hopefully. We're going to use various approaches including neural networks and other machine learning approaches and basically going to train them on large amount of sequences in the genomic sequence is available from various subjects, various patients also from different regions because there are mutations in the virus depending on location and so on. So this seems to be like quite a varying type of a virus.

CRAIG ([14:44](https://www.temi.com/editor/t/ZUCSrneVBFHP9wAsgc6D0fIbtzjwR_bd7SUbxpWJpkvfl4l1-FlKAwVWaUwXvaGwLI8ocq_8r2fxP1AIqfvxA2Fu1tQ?loadFrom=DocumentDeeplink&ts=884.35)):

When you say proteomic and transcriptomic, can you explain what that means? That data?

IRINA ([14:51](https://www.temi.com/editor/t/ZUCSrneVBFHP9wAsgc6D0fIbtzjwR_bd7SUbxpWJpkvfl4l1-FlKAwVWaUwXvaGwLI8ocq_8r2fxP1AIqfvxA2Fu1tQ?loadFrom=DocumentDeeplink&ts=891.26)):

Okay, so basically with genomic, that kind of more classical type of data, basically you want to know what proteins are involved in the functioning of this virus, how it operates, how it functions. I just look at this as high dimensional data and I'm trying to find correlations or even better causal relationships between different aspects of the data and the outcome, which is how much it helps to say improve symptoms or whether it helps or not to eventually get rid of them and how fast. So you're looking at data coming out of trials or you're looking at large data sets to try and find correlations. So data will be coming out of those trials because we can decide what to measure. But you are right that besides this particular trial there are, that's a bit difficult to get access and that's what we are working on, there are apparently other data sets either existing or being collected. Most of them being collected.

IRINA ([15:52](https://www.temi.com/editor/t/ZUCSrneVBFHP9wAsgc6D0fIbtzjwR_bd7SUbxpWJpkvfl4l1-FlKAwVWaUwXvaGwLI8ocq_8r2fxP1AIqfvxA2Fu1tQ?loadFrom=DocumentDeeplink&ts=952.85)):

There is this large World Health Organization initiative on collecting data for the most promising treatments. Again, there are other initiatives, but the problem, as I said, it's still most of the time in the collection stage, so the data are to be collected and that's a bit of a problem. You cannot just start applying machine learning right away. But on the other hand you can decide what to measure. And as I was saying, the variability comes from different aspects of each patient. You can have equally healthy patients, young patients and one develops severe symptoms and may not make it and another recovers. There was this article about two young doctors in China which was quite striking, which makes you wonder how come you have such a huge variance in their disease progression, trajectory and outcome given seemingly similar other factors. So what are other factors that may be explaining this huge variance and that's one of the motivations. For that to try and to understand how the virus works in the body and what are those critical things that may affect the results

INTERLUDE ([17:08](https://www.temi.com/editor/t/ZUCSrneVBFHP9wAsgc6D0fIbtzjwR_bd7SUbxpWJpkvfl4l1-FlKAwVWaUwXvaGwLI8ocq_8r2fxP1AIqfvxA2Fu1tQ?loadFrom=DocumentDeeplink&ts=1028.52)):

MUSIC

CRAIG ([17:32](https://www.temi.com/editor/t/ZUCSrneVBFHP9wAsgc6D0fIbtzjwR_bd7SUbxpWJpkvfl4l1-FlKAwVWaUwXvaGwLI8ocq_8r2fxP1AIqfvxA2Fu1tQ?loadFrom=DocumentDeeplink&ts=1052.64)):

The data sets that you are trying to access or that you're collecting and the data that's coming out of the clinical trial include what data points?

IRINA ([17:42](https://www.temi.com/editor/t/ZUCSrneVBFHP9wAsgc6D0fIbtzjwR_bd7SUbxpWJpkvfl4l1-FlKAwVWaUwXvaGwLI8ocq_8r2fxP1AIqfvxA2Fu1tQ?loadFrom=DocumentDeeplink&ts=1062.9)):

So for example, a whole genome data set, in particular we are planning to use machine learning, particularly deep networks or other techniques to, you can have this timestamp whole genome data set and you can try to make use of that. You can also actually train on large amounts of sequences available from various geographic regions for example, and you could try to predict point mutations in RNA sequences and so on and so forth. You're trying to understand what's going on at that level.

CRAIG ([18:20](https://www.temi.com/editor/t/ZUCSrneVBFHP9wAsgc6D0fIbtzjwR_bd7SUbxpWJpkvfl4l1-FlKAwVWaUwXvaGwLI8ocq_8r2fxP1AIqfvxA2Fu1tQ?loadFrom=DocumentDeeplink&ts=1100.42)):

And that that's the genome of the virus?

IRINA ([18:23](https://www.temi.com/editor/t/ZUCSrneVBFHP9wAsgc6D0fIbtzjwR_bd7SUbxpWJpkvfl4l1-FlKAwVWaUwXvaGwLI8ocq_8r2fxP1AIqfvxA2Fu1tQ?loadFrom=DocumentDeeplink&ts=1103.44)):

Both. Well you have the genetic information of the person and the virus and you want to see how the two interact. The virus is one thing kind of different subject is the variability differences between people. That's also important thing and that's an interesting question that is part of this whole family of projects. As we know, typically kids, infants are usually as vulnerable as much older adults in many diseases, but with this particular virus it's not a U-shaped curve. By U shape, I mean that the youngest and the oldest are most vulnerable. Now the youngest seem to be most of the time not as vulnerable as we already know, like younger people and especially kids, most of them, yes, there were some kids who died, but majority seem to have much easier symptoms, if any. And that's a good question because you may want to study and try to understand what is the difference in the virus interaction in their bodies in particular in their lungs.

IRINA ([19:31](https://www.temi.com/editor/t/ZUCSrneVBFHP9wAsgc6D0fIbtzjwR_bd7SUbxpWJpkvfl4l1-FlKAwVWaUwXvaGwLI8ocq_8r2fxP1AIqfvxA2Fu1tQ?loadFrom=DocumentDeeplink&ts=1171.27)):

Why it is strikingly different from older population, what is involved. It's not the typical thing for all viruses, not even for old coronaviruses. Maybe if you understand why, maybe it can give you some idea how to make older people more resilient. Again, it's all hypothesizing, but there are these key interesting and kind of different properties of this particular virus and how it behaves that you really need to understand to try to think about what kind of treatment might be best. Again, it's a little bit like searching and the dark room for a black cat and you're not even sure the cat is there, but you keep searching.

CRAIG ([20:17](https://www.temi.com/editor/t/ZUCSrneVBFHP9wAsgc6D0fIbtzjwR_bd7SUbxpWJpkvfl4l1-FlKAwVWaUwXvaGwLI8ocq_8r2fxP1AIqfvxA2Fu1tQ?loadFrom=DocumentDeeplink&ts=1217.29)):

the way you would do that. Just on the algorithmic level, you have these two databases, you have one database of viral genomes in their variations and then you have the data base of patient genomes and their variations and do you do just a pattern matching? You're trying to find correlations between them and then once you find the correlation you can look into which genes or gene families are different in patients that have worse or better outcomes?

IRINA ([20:50](https://www.temi.com/editor/t/ZUCSrneVBFHP9wAsgc6D0fIbtzjwR_bd7SUbxpWJpkvfl4l1-FlKAwVWaUwXvaGwLI8ocq_8r2fxP1AIqfvxA2Fu1tQ?loadFrom=DocumentDeeplink&ts=1250.83)):

Of course what you just described. It's a very simple approach that first comes to mind and definitely should be tried but there are kind of - the interesting and powerful property of this multivariate machine learning methods is if you plug in the data, basically each long vector of different variables describing those sequences and their variations for the virus, for the person. Also the RNA data, if you have some available, if you have all this information you can build multivariate predictive models like think about how we used to classify images of cats versus dogs and you may use something like a deep neural net for that. But you can use something much simpler. You can use like a linear regression model or linear classifier or you can make it more sophisticated and go all the way to nonlinear multi level such as deep networks and at the end of the day all this success in applications of AI for example to image recognition was that you might be able to train your model to achieve very high accuracy of say image recognition. And this is instead of images you might be predicting whether the treatment was successful or not in a particular person and then instead of just simple pairwise correlation saying that, okay, so there is a particular variable here and particular genes there were correlated with a positive outcome.

IRINA ([22:23](https://www.temi.com/editor/t/ZUCSrneVBFHP9wAsgc6D0fIbtzjwR_bd7SUbxpWJpkvfl4l1-FlKAwVWaUwXvaGwLI8ocq_8r2fxP1AIqfvxA2Fu1tQ?loadFrom=DocumentDeeplink&ts=1343.02)):

You say, no, I look now at these groups of variables. So basically just like with image processing, you would say, I can look at this whole pattern, multivariate pattern or just like in playing game of Go, you can say, well this particular pattern is predictive that I am in a good state of the game versus a bad one. Something like that. The point is that with those more sophisticated methods, you will be able to discover those multi-variate patterns and say if this and that, but not that and not that all together, then you have a higher chance of benefiting from the drug. And sometimes in general, and not just this application I'm coming from, for example, applications where you use similar techniques for brain imaging data analysis, you may not be able to say much by looking at individual correlation of this [inaudible] in the brain or this particular gene with the outcome. You only start seeing very accurate predictions when you look at those subsets, groups or patterns, and that's the power of multivariate predictive modeling.

IRINA ([23:35](https://www.temi.com/editor/t/ZUCSrneVBFHP9wAsgc6D0fIbtzjwR_bd7SUbxpWJpkvfl4l1-FlKAwVWaUwXvaGwLI8ocq_8r2fxP1AIqfvxA2Fu1tQ?loadFrom=DocumentDeeplink&ts=1415.53)):

You may need to see the whole picture and to see the whole picture with multiple things involved simultaneoulsy, you need those multivariate predictive models of various flavors. With deep learning being one successful lately.

CRAIG ([23:50](https://www.temi.com/editor/t/ZUCSrneVBFHP9wAsgc6D0fIbtzjwR_bd7SUbxpWJpkvfl4l1-FlKAwVWaUwXvaGwLI8ocq_8r2fxP1AIqfvxA2Fu1tQ?loadFrom=DocumentDeeplink&ts=1430.21)):

It's really a question of collecting the right data. Is that right?

IRINA ([23:54](https://www.temi.com/editor/t/ZUCSrneVBFHP9wAsgc6D0fIbtzjwR_bd7SUbxpWJpkvfl4l1-FlKAwVWaUwXvaGwLI8ocq_8r2fxP1AIqfvxA2Fu1tQ?loadFrom=DocumentDeeplink&ts=1434.11)):

Deep learning has shown to be successful, but there are simple methods. There are different methods, especially when the data are small, relatively small. Of course we're going to try deep learning but not only deep learning. And then I was mentioning very large scale data in terms of number of involved variables, like again back to the brain application because it just came to my mind. We had a specific case where you have about 80,000 variables corresponding to [inaudible], or areas in the brain and you have only 40 subjects. And you're trying to classify which subjects are actually going to develop schizophrenia. And yes, people have such data.

IRINA ([24:37](https://www.temi.com/editor/t/ZUCSrneVBFHP9wAsgc6D0fIbtzjwR_bd7SUbxpWJpkvfl4l1-FlKAwVWaUwXvaGwLI8ocq_8r2fxP1AIqfvxA2Fu1tQ?loadFrom=DocumentDeeplink&ts=1477.61)):

Amazingly you can learn and generalized to unseen examples or new subjects from such a small population of just 40 people despite the fact that your data dimensionalities are humongous, like dozens of thousands of variables and you can do it with methods even like good old classical support vector machines or good old classical fast logistic regression. But they may be quite effective because they have what is called regularization incorporated in them, which basically means if you have many, many degrees of freedom but not enough samples but you really restrict the set of your models, you may still be able to learn something useful. And that's a well known thing in statistics, regularization. Then, the same thing is applied to deep learning but since models are much larger and more complicated and more powerful, they need more data, it's the same trade off how to impose constraints on those networks so they will be able to learn with probably smaller data.

IRINA ([25:49](https://www.temi.com/editor/t/ZUCSrneVBFHP9wAsgc6D0fIbtzjwR_bd7SUbxpWJpkvfl4l1-FlKAwVWaUwXvaGwLI8ocq_8r2fxP1AIqfvxA2Fu1tQ?loadFrom=DocumentDeeplink&ts=1549.37)):

What I'm saying, we can try and we will try various, a range of methods including deep learning but not limited to it.

CRAIG ([25:57](https://www.temi.com/editor/t/ZUCSrneVBFHP9wAsgc6D0fIbtzjwR_bd7SUbxpWJpkvfl4l1-FlKAwVWaUwXvaGwLI8ocq_8r2fxP1AIqfvxA2Fu1tQ?loadFrom=DocumentDeeplink&ts=1557.77)):

How large are the data sets that you've collected so far? As I mentioned, the dataset that we will be looking at is not yet even there and we're still searching for data sets online and I assume that potentially the data that we might be getting from a hospital here, which particularly focuses on younger adults and kids. Luckily in their age range, we didn't have too many cases, dozens or something. It's kind of good news. It might be not as good for data collection, but we probably will also look at data collected in other regions, China and Italy and New York. I would like to be able to have methods that if needed, can learn even from dozens of patients and do not have to wait for many, many, many thousands, although hundreds and thousands of data points would be definitely better.

IRINA ([26:54](https://www.temi.com/editor/t/ZUCSrneVBFHP9wAsgc6D0fIbtzjwR_bd7SUbxpWJpkvfl4l1-FlKAwVWaUwXvaGwLI8ocq_8r2fxP1AIqfvxA2Fu1tQ?loadFrom=DocumentDeeplink&ts=1614.84)):

But that is the reason I'm talking about like severely constrained regularized methods or maybe simpler methods than deep learning is we'll need to try what can work with potentially smaller data.

CRAIG ([27:07](https://www.temi.com/editor/t/ZUCSrneVBFHP9wAsgc6D0fIbtzjwR_bd7SUbxpWJpkvfl4l1-FlKAwVWaUwXvaGwLI8ocq_8r2fxP1AIqfvxA2Fu1tQ?loadFrom=DocumentDeeplink&ts=1627.47)):

In terms of collecting data from other countries, is there at this point a repository that countries are contributing data to for researchers to draw from?

IRINA ([27:20](https://www.temi.com/editor/t/ZUCSrneVBFHP9wAsgc6D0fIbtzjwR_bd7SUbxpWJpkvfl4l1-FlKAwVWaUwXvaGwLI8ocq_8r2fxP1AIqfvxA2Fu1tQ?loadFrom=DocumentDeeplink&ts=1640.55)):

There are multiple initiatives and things kind of keep changing as we speak, so I would say yes. We have linked to data sets as they are being collected. Also our AI-against-COVID.ca was collecting various things for last week. It's just like too many things happening in parallel as you can imagine, but they are being collected. Most of them are still under construction. There are many more datasets of different kinds which do not necessarily involve the genomics proteomics. There are data sets about disease progression and so on, like John Hopkins, Those data are easily available or there are 19 datasets where thousands of scientific papers on the topic. And the AI problem. there would be to use natural language processing to mine the papers for more information.

IRINA ([28:15](https://www.temi.com/editor/t/ZUCSrneVBFHP9wAsgc6D0fIbtzjwR_bd7SUbxpWJpkvfl4l1-FlKAwVWaUwXvaGwLI8ocq_8r2fxP1AIqfvxA2Fu1tQ?loadFrom=DocumentDeeplink&ts=1695.631)):

But this is a different type. These data sets are more easily available while the medical data sets are still, I would say, most of the time under construction, but I'm pretty sure within a week or so there will be more and plus I'm also searching and discovering more of such data sets as we go.

CRAIG ([28:35](https://www.temi.com/editor/t/ZUCSrneVBFHP9wAsgc6D0fIbtzjwR_bd7SUbxpWJpkvfl4l1-FlKAwVWaUwXvaGwLI8ocq_8r2fxP1AIqfvxA2Fu1tQ?loadFrom=DocumentDeeplink&ts=1715.22)):

Sasha Luccioni, I don't know if I'm pronouncing her name working with UN global pulse. Can you talk about that effort to map the landscape of papers.

IRINA ([28:45](https://www.temi.com/editor/t/ZUCSrneVBFHP9wAsgc6D0fIbtzjwR_bd7SUbxpWJpkvfl4l1-FlKAwVWaUwXvaGwLI8ocq_8r2fxP1AIqfvxA2Fu1tQ?loadFrom=DocumentDeeplink&ts=1725.18)):

They just published two days ago, I think it was out on the 23rd, they put together a high level survey paper about how AI can help the situation, trying to outline different types of activities and what can you do. She also put together this list of data sets. Of course it's incomplete and things change every day and that's what we try to maintain on the webpage. But on the webpage, in the data sets section, you can find that. So Sasha is a post-doc with Yoshua Bengio so she, she actually really put together quite quickly that survey paper on how AI could help COVID. What I would like this webpage to do is to provide surveys of recent developments like that one on a regular basis.

INTERLUDE ([29:35](https://www.temi.com/editor/t/ZUCSrneVBFHP9wAsgc6D0fIbtzjwR_bd7SUbxpWJpkvfl4l1-FlKAwVWaUwXvaGwLI8ocq_8r2fxP1AIqfvxA2Fu1tQ?loadFrom=DocumentDeeplink&ts=1775.13)):

MUSIC

CRAIG ([30:15](https://www.temi.com/editor/t/ZUCSrneVBFHP9wAsgc6D0fIbtzjwR_bd7SUbxpWJpkvfl4l1-FlKAwVWaUwXvaGwLI8ocq_8r2fxP1AIqfvxA2Fu1tQ?loadFrom=DocumentDeeplink&ts=1815.58)):

China and Italy for example, are they putting genomic data online?

IRINA ([30:20](https://www.temi.com/editor/t/ZUCSrneVBFHP9wAsgc6D0fIbtzjwR_bd7SUbxpWJpkvfl4l1-FlKAwVWaUwXvaGwLI8ocq_8r2fxP1AIqfvxA2Fu1tQ?loadFrom=DocumentDeeplink&ts=1820.3)):

I'm not sure by the way that if such data exists, they will be immediately publicly available because it's medical data and you can imagine, as usual, medical data is sensitive. So I think that more reliable. Is direct connection with colleagues who are working on those cases and collecting the data than waiting for absolutely publicly available data set.

CRAIG ([30:46](https://www.temi.com/editor/t/ZUCSrneVBFHP9wAsgc6D0fIbtzjwR_bd7SUbxpWJpkvfl4l1-FlKAwVWaUwXvaGwLI8ocq_8r2fxP1AIqfvxA2Fu1tQ?loadFrom=DocumentDeeplink&ts=1846.15)):

You mentioned federated learning for the personal app. Could you put up a federated learning model that people could download all over the world and train on their individual data sets without having to pool data?

IRINA ([30:59](https://www.temi.com/editor/t/ZUCSrneVBFHP9wAsgc6D0fIbtzjwR_bd7SUbxpWJpkvfl4l1-FlKAwVWaUwXvaGwLI8ocq_8r2fxP1AIqfvxA2Fu1tQ?loadFrom=DocumentDeeplink&ts=1859.91)):

That would be ideal. I'm pretty sure some people are thinking about that. I haven't seen yet any calls for collaboration of this type. Again, there are so many things happening. I would say to search for such calls for collaboration, it's a really good idea to include as a project in this task force to ask people working or planning to work on such data to organize ourselves. Federated learning is very powerful, but you need to organize it.

CRAIG ([31:32](https://www.temi.com/editor/t/ZUCSrneVBFHP9wAsgc6D0fIbtzjwR_bd7SUbxpWJpkvfl4l1-FlKAwVWaUwXvaGwLI8ocq_8r2fxP1AIqfvxA2Fu1tQ?loadFrom=DocumentDeeplink&ts=1892.58)):

Yeah, of course.

IRINA ([31:33](https://www.temi.com/editor/t/ZUCSrneVBFHP9wAsgc6D0fIbtzjwR_bd7SUbxpWJpkvfl4l1-FlKAwVWaUwXvaGwLI8ocq_8r2fxP1AIqfvxA2Fu1tQ?loadFrom=DocumentDeeplink&ts=1893.331)):

Peer to peer federated approaches to various aspects and various types of data in this domain, that's what we should do. But logistically it's always a nightmare. Last two weeks, we've had quite a bit of logistical nightmares and I think we are not the only ones.

CRAIG ([31:51](https://www.temi.com/editor/t/ZUCSrneVBFHP9wAsgc6D0fIbtzjwR_bd7SUbxpWJpkvfl4l1-FlKAwVWaUwXvaGwLI8ocq_8r2fxP1AIqfvxA2Fu1tQ?loadFrom=DocumentDeeplink&ts=1911.37)):

Yeah. Well this is interesting because this is really the first time since the rise of machine learning or really since the rise of more traditional AI methods that there's been a global effort to, to solve a problem and maybe structures will come out of this that will allow the sharing of data or the sharing of federated learning models going forward. So Mila is doing this. How many centers do you think there are around the world that are working in the way that Mila is working and how do people find you or is there a clearing house somewhere that you're watching and where you're listed along with, I don't know, a hundred other projects?

IRINA ([32:36](https://www.temi.com/editor/t/ZUCSrneVBFHP9wAsgc6D0fIbtzjwR_bd7SUbxpWJpkvfl4l1-FlKAwVWaUwXvaGwLI8ocq_8r2fxP1AIqfvxA2Fu1tQ?loadFrom=DocumentDeeplink&ts=1956.52)):

That's what I was mentioning in the beginning that we put together this all Canadian website. Then there are collaborative spaces. Stanford even put together already a course about AI and COVID. And Stanford, just announced, a virtual conference on AI for solving COVID problems on April 1st. I just signed up. I mean the emails are coming in as we speak and things are being posted as we speak so a week ago, I think there were, from what I've seen besides ours, there were like three maybe other collaborative spaces online and so even just like keeping track of trying to do similar things is important and we need to do that. But I think it's kind of also growing exponentially. Just like the virus, which is a good thing.

CRAIG ([33:29](https://www.temi.com/editor/t/ZUCSrneVBFHP9wAsgc6D0fIbtzjwR_bd7SUbxpWJpkvfl4l1-FlKAwVWaUwXvaGwLI8ocq_8r2fxP1AIqfvxA2Fu1tQ?loadFrom=DocumentDeeplink&ts=2009.34)):

This is in North America. Is there much of those kinds of collaborative spaces appearing in Europe or Asia?

IRINA ([33:36](https://www.temi.com/editor/t/ZUCSrneVBFHP9wAsgc6D0fIbtzjwR_bd7SUbxpWJpkvfl4l1-FlKAwVWaUwXvaGwLI8ocq_8r2fxP1AIqfvxA2Fu1tQ?loadFrom=DocumentDeeplink&ts=2016.76)):

The one that I got involved in first was actually it was in Germany, although the webpage is in English. To be honest, I don't know but I think there will be and it's a good point to keep our eye on that because I'm trying to make sure we collect not just information about scientific projects and analysis projects and data sets and also papers like research papers, events like the AI conference at Stanford. And I would like information about related projects of different kinds, including those maybe not English. So trying to make sure that our landing page before we branch into particular organizations and same with Mila is doing it. So on the landing page for this AI-against-COVID.ca will be a quickly updated source of all the recent developments across the world. This is a big challenge because there are exponentially many streams of information, but if we are smart about involving students and students are extremely good at finding information. Initially, I remember I was a bit skeptical and was saying that maybe you should focus on your research instead of wasting time and a week later, I was saying thank you.

CRAIG ([34:58](https://www.temi.com/editor/t/ZUCSrneVBFHP9wAsgc6D0fIbtzjwR_bd7SUbxpWJpkvfl4l1-FlKAwVWaUwXvaGwLI8ocq_8r2fxP1AIqfvxA2Fu1tQ?loadFrom=DocumentDeeplink&ts=2098.64)):

How hopeful are you that this global effort to apply different kinds of data analysis to data sets, how hopeful are you that it will come up within a reasonable period of time with some useful action plans?

IRINA ([35:14](https://www.temi.com/editor/t/ZUCSrneVBFHP9wAsgc6D0fIbtzjwR_bd7SUbxpWJpkvfl4l1-FlKAwVWaUwXvaGwLI8ocq_8r2fxP1AIqfvxA2Fu1tQ?loadFrom=DocumentDeeplink&ts=2114.84)):

I have no choice but to be hopeful. There is no plan B so I'm hopeful. We have to move fast and we have to move smart about organizing people, but I prefer not to think about what happens if it doesn't work.

CRAIG ([35:40](https://www.temi.com/editor/t/ZUCSrneVBFHP9wAsgc6D0fIbtzjwR_bd7SUbxpWJpkvfl4l1-FlKAwVWaUwXvaGwLI8ocq_8r2fxP1AIqfvxA2Fu1tQ?loadFrom=DocumentDeeplink&ts=2140.17)):

That's it. For this week's podcast, I want to commend Irina and her colleagues in Montreal and the all machine learning practitioners around the world for their efforts to combat the virus. These are strange times but in many ways hopeful times. It's heartening to see the machine learning community come together at such a swift pace. If you want to learn more about these efforts, you can find a link to a transcript of this show on our website. Eys on AI. And I encourage everyone to check out the website, a AI-against-COVID.ca and get involved.

CRAIG ([36:23](https://www.temi.com/editor/t/ZUCSrneVBFHP9wAsgc6D0fIbtzjwR_bd7SUbxpWJpkvfl4l1-FlKAwVWaUwXvaGwLI8ocq_8r2fxP1AIqfvxA2Fu1tQ?loadFrom=DocumentDeeplink&ts=2183.79)):

Be safe. And remember, the singularity may not be near, but AI is about to change your world. Pay attention

Speaker 2 ([36:47](https://www.temi.com/editor/t/ZUCSrneVBFHP9wAsgc6D0fIbtzjwR_bd7SUbxpWJpkvfl4l1-FlKAwVWaUwXvaGwLI8ocq_8r2fxP1AIqfvxA2Fu1tQ?loadFrom=DocumentDeeplink&ts=2207.42)):

[inaudible].