WEBVTT

NOTE duration: ”00:24:56.746”

NOTE Confidence: 0.86740375

00:00:00.080 --> 00:00:02.159 So next up, we have,

NOTE Confidence: 0.86740375

00:00:02.560 --> 00:00:03.780 doctor Deep Jaylee,

NOTE Confidence: 0.96079326

00:00:04.880 --> 00:00:05.380 who's,

NOTE Confidence: 0.99491256

00:00:05.759 --> 00:00:06.799 we're lucky to have him

NOTE Confidence: 0.99491256

00:00:06.799 --> 00:00:07.040 from,

NOTE Confidence: 0.8248296

00:00:07.680 --> 00:00:09.280 Apple machine learning where he's,

NOTE Confidence: 0.9935884

00:00:09.920 --> 00:00:11.039 lead leads a team of

NOTE Confidence: 0.9935884

00:00:11.039 --> 00:00:13.299 researchers working on fundamental techniques

NOTE Confidence: 0.9935884

00:00:13.360 --> 00:00:14.340 for machine learning.

NOTE Confidence: 0.96508133

00:00:14.934 --> 00:00:16.215 I know Deep for some

NOTE Confidence: 0.96508133

00:00:16.215 --> 00:00:17.015 time, and I know that

NOTE Confidence: 0.96508133

00:00:17.015 --> 00:00:18.375 he's been actually working on

NOTE Confidence: 0.96508133

00:00:18.375 --> 00:00:20.215 deep learning, before deep learning

NOTE Confidence: 0.96508133

00:00:20.215 --> 00:00:21.255 was cool a long time

NOTE Confidence: 0.96508133

00:00:21.255 --> 00:00:21.755 ago.

NOTE Confidence: 0.9582097

00:00:22.215 --> 00:00:24.295 So he actually did, his

NOTE Confidence: 0.9582097

00:00:24.295 --> 00:00:26.134 PhD under the supervision of

NOTE Confidence: 0.9582097

00:00:26.134 --> 00:00:27.335 Jeff Hinton, who just got

NOTE Confidence: 0.9582097

00:00:27.335 --> 00:00:28.455 a Nobel Prize, you may

NOTE Confidence: 0.9582097

00:00:28.455 --> 00:00:29.980 have heard, in the foundational

NOTE Confidence: 0.9582097

00:00:29.980 --> 00:00:31.260 days of deep learning. And

NOTE Confidence: 0.9582097

00:00:31.260 --> 00:00:32.720 then he joined Google Brain.

NOTE Confidence: 0.9962205

00:00:33.180 --> 00:00:34.460 He worked on deep learning

NOTE Confidence: 0.9962205

00:00:34.460 --> 00:00:35.280 models for

NOTE Confidence: 0.9464793

00:00:35.580 --> 00:00:37.340 sequences, and then he also,

NOTE Confidence: 0.9464793

00:00:37.580 --> 00:00:38.479 worked at various,

NOTE Confidence: 0.83981866

00:00:39.340 --> 00:00:40.880 places such as in media,

NOTE Confidence: 0.88259125

00:00:41.739 --> 00:00:43.575 Google Brain Robotics. And then

NOTE Confidence: 0.88259125

00:00:43.655 --> 00:00:44.854 somehow he got to finance

NOTE Confidence: 0.88259125

00:00:44.854 --> 00:00:45.575 for a bit, D. E.

NOTE Confidence: 0.88259125

00:00:45.575 --> 00:00:47.415 Shaw, and then, also before

NOTE Confidence: 0.88259125

00:00:47.415 --> 00:00:49.575 that, international labs. So we're

NOTE Confidence: 0.88259125

00:00:49.575 --> 00:00:50.534 gonna hear, I think, from

NOTE Confidence: 0.88259125

00:00:50.534 --> 00:00:52.475 Deepan the latest in generative

NOTE Confidence: 0.88259125

00:00:52.615 --> 00:00:53.115 models.

NOTE Confidence: 0.9937703

00:00:56.295 --> 00:00:57.355 Thank you, John.

NOTE Confidence: 0.84077704

00:00:59.320 --> 00:00:59.820 Well,

NOTE Confidence: 0.9794217

00:01:00.600 --> 00:01:01.640 I'll be presenting a lot

NOTE Confidence: 0.9794217

00:01:01.640 --> 00:01:02.840 of work with my colleagues

NOTE Confidence: 0.9794217

00:01:02.840 --> 00:01:04.200 at Apple. I have to

NOTE Confidence: 0.9794217

00:01:04.200 --> 00:01:05.560 say, this is a little

NOTE Confidence: 0.9794217

00:01:05.560 --> 00:01:06.920 different from my usual talk

NOTE Confidence: 0.9794217

00:01:06.920 --> 00:01:08.280 where I get into the

NOTE Confidence: 0.9794217

00:01:08.280 --> 00:01:10.200 nitty gritty of, machine learning

NOTE Confidence: 0.9794217

00:01:10.200 --> 00:01:12.380 and why one slight variation

NOTE Confidence: 0.9794217

00:01:12.440 --> 00:01:14.104 is more important than the

NOTE Confidence: 0.9794217

00:01:14.104 --> 00:01:14.604 other.

NOTE Confidence: 0.9750039

00:01:15.145 --> 00:01:15.645 Instead,

NOTE Confidence: 0.9960849

00:01:15.944 --> 00:01:17.225 for today, I thought maybe

NOTE Confidence: 0.9960849

00:01:17.225 --> 00:01:19.084 I would touch base on

NOTE Confidence: 0.9960849

00:01:19.384 --> 00:01:20.985 what I thought were three

NOTE Confidence: 0.9960849

00:01:20.985 --> 00:01:22.284 essential things that,

NOTE Confidence: 0.9969192

00:01:23.145 --> 00:01:24.685 people working in life sciences

NOTE Confidence: 0.9969192

00:01:24.744 --> 00:01:25.564 should think about

NOTE Confidence: 0.99412423

00:01:26.330 --> 00:01:27.770 in advances in machine learning

NOTE Confidence: 0.99412423

00:01:27.770 --> 00:01:28.990 as they would be relevant

NOTE Confidence: 0.9983893

00:01:29.290 --> 00:01:30.250 to them when they're looking

NOTE Confidence: 0.9983893

00:01:30.250 --> 00:01:30.830 at data.

NOTE Confidence: 0.99715424

00:01:32.410 --> 00:01:34.090 Okay. So what are these

NOTE Confidence: 0.99715424

00:01:34.090 --> 00:01:35.530 three things that I have

NOTE Confidence: 0.99715424

00:01:35.530 --> 00:01:36.490 in mind? Well, the first

NOTE Confidence: 0.99715424

00:01:36.490 --> 00:01:36.990 one,

NOTE Confidence: 0.972235

00:01:37.610 --> 00:01:38.590 is that recently,

NOTE Confidence: 0.98850167

00:01:38.970 --> 00:01:40.650 neural networks have got really

NOTE Confidence: 0.98850167

00:01:40.650 --> 00:01:42.715 good at trying to embed

NOTE Confidence: 0.9978884

00:01:43.175 --> 00:01:44.555 various kinds of data,

NOTE Confidence: 0.9996989

00:01:45.175 --> 00:01:46.475 into a vector representation.

NOTE Confidence: 0.99133164

00:01:46.855 --> 00:01:47.495 You know, if you get

NOTE Confidence: 0.99133164

00:01:47.495 --> 00:01:48.935 some data, you first need

NOTE Confidence: 0.99133164

00:01:48.935 --> 00:01:50.055 to convert that into a

NOTE Confidence: 0.99133164

00:01:50.055 --> 00:01:51.015 form that you can work

NOTE Confidence: 0.99133164

00:01:51.015 --> 00:01:52.635 with in statistical models.

NOTE Confidence: 0.9789339

00:01:53.095 --> 00:01:54.135 And so this is a

NOTE Confidence: 0.9789339

00:01:54.135 --> 00:01:55.955 requirement. And in the past,

NOTE Confidence: 0.9789339

00:01:56.295 --> 00:01:56.850 you know,

NOTE Confidence: 0.9798089

00:01:57.409 --> 00:01:58.530 people had some ways to

NOTE Confidence: 0.9798089

00:01:58.530 --> 00:01:59.729 do it. And recently, there's

NOTE Confidence: 0.9798089

00:01:59.729 --> 00:02:00.450 been a lot of progress

NOTE Confidence: 0.9798089

00:02:00.450 --> 00:02:01.330 in this, so I wanna

NOTE Confidence: 0.9798089

00:02:01.330 --> 00:02:02.850 touch that touch upon that

NOTE Confidence: 0.9798089

00:02:02.850 --> 00:02:03.590 a little bit.

NOTE Confidence: 0.9955533

00:02:04.210 --> 00:02:05.430 I should also say,

NOTE Confidence: 0.99941206

00:02:06.530 --> 00:02:08.470 we can fit these representations

NOTE Confidence: 0.99579495

00:02:08.850 --> 00:02:10.130 very well. No one really

NOTE Confidence: 0.99579495

00:02:10.130 --> 00:02:10.950 knows why,

NOTE Confidence: 0.9960451

00:02:11.410 --> 00:02:12.470 but we can.

NOTE Confidence: 0.7925912

00:02:13.385 --> 00:02:13.885 And,

NOTE Confidence: 0.8788985

00:02:15.224 --> 00:02:17.565 recently, generated models themselves have,

NOTE Confidence: 0.9696978

00:02:18.584 --> 00:02:20.264 become really powerful. So we

NOTE Confidence: 0.9696978

00:02:20.264 --> 00:02:22.584 have really an uncanny ability

NOTE Confidence: 0.9696978

00:02:22.584 --> 00:02:24.364 to generate data now that

NOTE Confidence: 0.9987009

00:02:25.130 --> 00:02:25.610 doesn't

NOTE Confidence: 0.9533761

00:02:26.250 --> 00:02:27.770 that really surprises me every

NOTE Confidence: 0.9533761

00:02:27.770 --> 00:02:30.010 day, from, you know, generated

NOTE Confidence: 0.9533761

00:02:30.010 --> 00:02:31.310 models of text to

NOTE Confidence: 0.9001668

00:02:31.690 --> 00:02:32.669 models of images.

NOTE Confidence: 0.96143454

00:02:32.970 --> 00:02:34.730 In that regard, I'll touch

NOTE Confidence: 0.96143454

00:02:34.730 --> 00:02:35.770 on, like, the two main

NOTE Confidence: 0.96143454

00:02:35.770 --> 00:02:36.270 techniques,

NOTE Confidence: 0.9713913

00:02:36.650 --> 00:02:38.750 autoregressive models and diffusion models

NOTE Confidence: 0.9713913

00:02:38.889 --> 00:02:40.510 today and how they work.

NOTE Confidence: 0.96528107

00:02:41.505 --> 00:02:42.785 And I should also highlight

NOTE Confidence: 0.96528107

00:02:42.785 --> 00:02:44.145 that now we can do

NOTE Confidence: 0.96528107

00:02:44.145 --> 00:02:45.905 this across modalities. So not

NOTE Confidence: 0.96528107

00:02:45.905 --> 00:02:47.025 only is it just a

NOTE Confidence: 0.96528107

00:02:47.025 --> 00:02:48.145 model for text or just

NOTE Confidence: 0.96528107

00:02:48.145 --> 00:02:49.365 a model for images,

NOTE Confidence: 0.96656656

00:02:49.905 --> 00:02:51.105 but instead, we can build

NOTE Confidence: 0.96656656

00:02:51.105 --> 00:02:52.085 models that,

NOTE Confidence: 0.9559004

00:02:52.625 --> 00:02:54.480 work across all of them.

NOTE Confidence: 0.9525898

00:02:55.280 --> 00:02:56.560 I should say everything's gonna

NOTE Confidence: 0.9525898

00:02:56.560 --> 00:02:57.200 be on a really high

NOTE Confidence: 0.9525898

00:02:57.200 --> 00:02:58.080 level, but if you want

NOTE Confidence: 0.9525898

00:02:58.080 --> 00:02:59.360 to get into nitty gritties,

NOTE Confidence: 0.9525898

00:02:59.360 --> 00:03:00.000 we can,

NOTE Confidence: 0.9984657

00:03:00.319 --> 00:03:01.860 touch base after the talk.

NOTE Confidence: 0.94609773

00:03:03.040 --> 00:03:03.519 And so,

NOTE Confidence: 0.9990627

00:03:05.519 --> 00:03:07.299 I will end by one

NOTE Confidence: 0.9888308

00:03:07.915 --> 00:03:09.055 little vignette on

NOTE Confidence: 0.9472133

00:03:09.435 --> 00:03:11.595 doing conformer predictions with diffusion

NOTE Confidence: 0.9472133

00:03:11.595 --> 00:03:12.955 models. So you're given a,

NOTE Confidence: 0.9403915

00:03:13.514 --> 00:03:14.955 compound. You want to predict

NOTE Confidence: 0.9403915

00:03:14.955 --> 00:03:15.615 what the,

NOTE Confidence: 0.9709051

00:03:16.075 --> 00:03:17.855 structure of that compound is.

NOTE Confidence: 0.9709051

00:03:17.915 --> 00:03:19.595 And, you know, everybody's seen

NOTE Confidence: 0.9709051

00:03:19.595 --> 00:03:20.095 AlphaFold.

NOTE Confidence: 0.91292906

00:03:21.200 --> 00:03:22.819 There's specific methodologies,

NOTE Confidence: 0.9478802

00:03:23.360 --> 00:03:24.880 the early methods in AlphaFold

NOTE Confidence: 0.9478802

00:03:24.880 --> 00:03:25.919 used, which use a lot

NOTE Confidence: 0.9478802

00:03:25.919 --> 00:03:27.680 of information like multiple sequence

NOTE Confidence: 0.9478802

00:03:27.680 --> 00:03:28.980 alignments and so on.

NOTE Confidence: 0.96829486

00:03:29.360 --> 00:03:30.480 But now our techniques are

NOTE Confidence: 0.96829486

00:03:30.480 --> 00:03:31.840 getting powerful enough where you

NOTE Confidence: 0.96829486

00:03:31.840 --> 00:03:33.760 can do things have initial

NOTE Confidence: 0.96829486

00:03:33.760 --> 00:03:35.440 without that much information. And

NOTE Confidence: 0.96829486

00:03:35.440 --> 00:03:36.515 so I think this is

NOTE Confidence: 0.96829486

00:03:36.515 --> 00:03:37.175 an interesting

NOTE Confidence: 0.97139764

00:03:37.715 --> 00:03:38.615 approach to,

NOTE Confidence: 0.9223178

00:03:39.315 --> 00:03:40.435 to highlight, and I think

NOTE Confidence: 0.9223178

00:03:40.435 --> 00:03:42.035 latest alpha fold three also

NOTE Confidence: 0.9223178

00:03:42.035 --> 00:03:43.095 works on diffusion.

NOTE Confidence: 0.99624777

00:03:44.515 --> 00:03:46.455 So there's some commonality there.

NOTE Confidence: 0.98113465

00:03:48.115 --> 00:03:49.155 Okay. So,

NOTE Confidence: 0.98390925

00:03:49.555 --> 00:03:51.390 how does, how does this

NOTE Confidence: 0.99972415

00:03:52.110 --> 00:03:53.730 embedding of data into

NOTE Confidence: 0.96314895

00:03:54.270 --> 00:03:55.730 any space work? So,

NOTE Confidence: 0.99950266

00:03:56.030 --> 00:03:56.530 traditionally,

NOTE Confidence: 0.9960791

00:03:56.910 --> 00:03:58.209 we think of this as

NOTE Confidence: 0.9960791

00:03:58.270 --> 00:04:00.030 representation learning. You're given some

NOTE Confidence: 0.9960791

00:04:00.030 --> 00:04:00.530 data.

NOTE Confidence: 0.9823809

00:04:00.910 --> 00:04:02.110 Before you can do anything

NOTE Confidence: 0.9823809

00:04:02.110 --> 00:04:03.150 about it, you want to

NOTE Confidence: 0.9823809

00:04:03.150 --> 00:04:04.475 first convert that into a

NOTE Confidence: 0.9823809

00:04:04.475 --> 00:04:06.315 usable form by embedding it

NOTE Confidence: 0.9823809

00:04:06.315 --> 00:04:08.235 into some vector space with

NOTE Confidence: 0.9823809

00:04:08.235 --> 00:04:08.975 n dimensions.

NOTE Confidence: 0.944022

00:04:09.995 --> 00:04:11.035 And then you plug that

NOTE Confidence: 0.944022

00:04:11.035 --> 00:04:12.555 into a statistical model, and

NOTE Confidence: 0.944022

00:04:12.555 --> 00:04:13.915 you can do things like

NOTE Confidence: 0.944022

00:04:13.915 --> 00:04:15.195 make predictions on it or

NOTE Confidence: 0.944022

00:04:15.195 --> 00:04:16.875 maybe do unsupervised learning like

NOTE Confidence: 0.944022

00:04:16.875 --> 00:04:18.635 clustering that we just saw

NOTE Confidence: 0.944022

00:04:18.635 --> 00:04:19.135 previously.

NOTE Confidence: 0.940194

00:04:20.630 --> 00:04:21.370 And so,

NOTE Confidence: 0.99309945

00:04:22.150 --> 00:04:23.669 what people used to do

NOTE Confidence: 0.99309945

00:04:23.669 --> 00:04:24.410 in the past,

NOTE Confidence: 0.9479278

00:04:25.270 --> 00:04:27.029 was there was a specific

NOTE Confidence: 0.9479278

00:04:27.029 --> 00:04:28.550 technique for every modality. You

NOTE Confidence: 0.9479278

00:04:28.550 --> 00:04:29.750 had images. You would use

NOTE Confidence: 0.9479278

00:04:29.750 --> 00:04:31.770 two d convolutions, your convolutional

NOTE Confidence: 0.9946903

00:04:32.175 --> 00:04:32.574 models.

NOTE Confidence: 0.9938262

00:04:33.375 --> 00:04:34.914 For text, you would embed

NOTE Confidence: 0.9868399

00:04:35.214 --> 00:04:37.154 each text into a little,

NOTE Confidence: 0.99318224

00:04:38.014 --> 00:04:40.034 descriptor for it. For waveforms,

NOTE Confidence: 0.99318224

00:04:40.175 --> 00:04:42.335 you might convert, the waveform

NOTE Confidence: 0.99318224

00:04:42.335 --> 00:04:43.714 into spectral representations

NOTE Confidence: 0.98420733

00:04:44.095 --> 00:04:45.794 and then embed those into

NOTE Confidence: 0.8894292

00:04:46.175 --> 00:04:47.955 a big size vector space.

NOTE Confidence: 0.8960921

00:04:50.040 --> 00:04:50.940 And so,

NOTE Confidence: 0.9763082

00:04:51.880 --> 00:04:53.960 you know, they what because

NOTE Confidence: 0.9763082

00:04:53.960 --> 00:04:55.400 of this limitation, what happened

NOTE Confidence: 0.9763082

00:04:55.400 --> 00:04:56.699 is that everybody had,

NOTE Confidence: 0.99223924

00:04:57.560 --> 00:04:59.080 models for different kinds of

NOTE Confidence: 0.99223924

00:04:59.080 --> 00:05:00.120 data, and they were all

NOTE Confidence: 0.99223924

00:05:00.120 --> 00:05:00.620 separated.

NOTE Confidence: 0.9791212

00:05:01.435 --> 00:05:02.955 And then over time, people

NOTE Confidence: 0.9791212

00:05:02.955 --> 00:05:03.455 decided,

NOTE Confidence: 0.9919666

00:05:03.755 --> 00:05:05.035 well, let's try and embed

NOTE Confidence: 0.9919666

00:05:05.035 --> 00:05:06.555 different modalities into the same

NOTE Confidence: 0.9919666

00:05:06.555 --> 00:05:07.055 space.

NOTE Confidence: 0.9841268

00:05:07.514 --> 00:05:08.875 And once they're in the

NOTE Confidence: 0.9841268

00:05:08.875 --> 00:05:10.315 same space, we'll just combine

NOTE Confidence: 0.9841268

00:05:10.315 --> 00:05:11.835 them by little things like

NOTE Confidence: 0.9841268

00:05:11.835 --> 00:05:13.595 adding the representations from different

NOTE Confidence: 0.9841268

00:05:13.595 --> 00:05:14.495 spaces together

NOTE Confidence: 0.99603015

00:05:14.960 --> 00:05:16.080 or maybe even putting a

NOTE Confidence: 0.99603015

00:05:16.080 --> 00:05:17.120 small neural net on top

NOTE Confidence: 0.99603015

00:05:17.120 --> 00:05:17.700 of that.

NOTE Confidence: 0.99346393

00:05:18.800 --> 00:05:20.820 But that was really quite

NOTE Confidence: 0.99346393

00:05:20.880 --> 00:05:21.380 inflexible,

NOTE Confidence: 0.97771955

00:05:22.480 --> 00:05:24.160 in that, the kinds of

NOTE Confidence: 0.97771955

00:05:24.160 --> 00:05:25.200 changes you could make to

NOTE Confidence: 0.97771955

00:05:25.200 --> 00:05:27.220 the representation were limited. And,

NOTE Confidence: 0.9503719

00:05:28.000 --> 00:05:28.960 what you did at,

NOTE Confidence: 0.978797

00:05:29.585 --> 00:05:30.625 times when you wanted to

NOTE Confidence: 0.978797

00:05:30.625 --> 00:05:31.825 use the model was pretty

NOTE Confidence: 0.978797

00:05:31.825 --> 00:05:32.705 much how you trained it.

NOTE Confidence: 0.978797

00:05:32.705 --> 00:05:33.665 If you had two things

NOTE Confidence: 0.978797

00:05:33.665 --> 00:05:35.185 going in during training, you

NOTE Confidence: 0.978797

00:05:35.185 --> 00:05:36.325 only use two things,

NOTE Confidence: 0.9616071

00:05:37.505 --> 00:05:39.505 during evaluation. So, like, if

NOTE Confidence: 0.9616071

00:05:39.505 --> 00:05:40.545 you had images and text,

NOTE Confidence: 0.9616071

00:05:40.545 --> 00:05:41.745 you could just use images

NOTE Confidence: 0.9616071

00:05:41.745 --> 00:05:43.345 and text and no other

NOTE Confidence: 0.9616071

00:05:43.345 --> 00:05:43.845 combinations.

NOTE Confidence: 0.9682243

00:05:45.029 --> 00:05:46.389 And so in twenty twenty

NOTE Confidence: 0.9682243

00:05:46.389 --> 00:05:47.669 seven, all this changed with

NOTE Confidence: 0.9682243

00:05:47.669 --> 00:05:48.330 the attention,

NOTE Confidence: 0.9968327

00:05:49.190 --> 00:05:50.169 models paper.

NOTE Confidence: 0.98066217

00:05:50.629 --> 00:05:51.449 It was really,

NOTE Confidence: 0.98299843

00:05:51.990 --> 00:05:53.750 a breakthrough paper, which has

NOTE Confidence: 0.98299843

00:05:53.750 --> 00:05:55.830 had its implications in various

NOTE Confidence: 0.98299843

00:05:55.830 --> 00:05:56.330 forms.

NOTE Confidence: 0.71730655

00:05:56.654 --> 00:05:57.154 So

NOTE Confidence: 0.9213064

00:05:58.495 --> 00:05:59.855 the I the basic idea

NOTE Confidence: 0.9213064

00:05:59.855 --> 00:06:01.154 is you can take,

NOTE Confidence: 0.97762746

00:06:01.855 --> 00:06:03.535 embeddings of different data, and

NOTE Confidence: 0.97762746

00:06:03.535 --> 00:06:04.035 then

NOTE Confidence: 0.98797077

00:06:04.415 --> 00:06:05.955 you can combine these embeddings

NOTE Confidence: 0.98797077

00:06:06.014 --> 00:06:07.535 by choosing what's important. So

NOTE Confidence: 0.98797077

00:06:07.535 --> 00:06:09.075 there's this notion of attention.

NOTE Confidence: 0.98797077

00:06:09.294 --> 00:06:10.095 I won't go into the

NOTE Confidence: 0.98797077

00:06:10.095 --> 00:06:11.235 details necessarily,

NOTE Confidence: 0.9808657

00:06:11.759 --> 00:06:13.120 on that, but there's this

NOTE Confidence: 0.9808657

00:06:13.120 --> 00:06:14.240 notion of looking at your

NOTE Confidence: 0.9808657

00:06:14.240 --> 00:06:14.740 data,

NOTE Confidence: 0.98815113

00:06:15.199 --> 00:06:16.400 and looking at different parts

NOTE Confidence: 0.98815113

00:06:16.400 --> 00:06:17.620 of it and choosing,

NOTE Confidence: 0.9882102

00:06:18.400 --> 00:06:19.139 those parts,

NOTE Confidence: 0.99383694

00:06:19.599 --> 00:06:21.460 if it seems relevant to,

NOTE Confidence: 0.99383694

00:06:21.759 --> 00:06:22.740 the model itself.

NOTE Confidence: 0.98453283

00:06:23.120 --> 00:06:24.720 And this is all learned

NOTE Confidence: 0.98453283

00:06:24.720 --> 00:06:26.324 during training itself of the

NOTE Confidence: 0.98453283

00:06:26.324 --> 00:06:27.764 model, so this attention is

NOTE Confidence: 0.98453283

00:06:27.764 --> 00:06:28.964 not baked in beforehand. You

NOTE Confidence: 0.98453283

00:06:28.964 --> 00:06:29.764 just learn how to do

NOTE Confidence: 0.98453283

00:06:29.764 --> 00:06:30.964 it, as part of the

NOTE Confidence: 0.98453283

00:06:30.964 --> 00:06:31.464 training.

NOTE Confidence: 0.9455217

00:06:34.164 --> 00:06:34.664 So,

NOTE Confidence: 0.990575

00:06:35.845 --> 00:06:37.205 what this also offers is

NOTE Confidence: 0.990575

00:06:37.205 --> 00:06:39.145 this really interesting ability to

NOTE Confidence: 0.990575

00:06:39.310 --> 00:06:40.589 change how you embed data

NOTE Confidence: 0.990575

00:06:40.589 --> 00:06:42.130 into your model. So instead

NOTE Confidence: 0.9457501

00:06:42.509 --> 00:06:43.250 of just,

NOTE Confidence: 0.9997985

00:06:44.190 --> 00:06:45.250 using your traditional

NOTE Confidence: 0.9149885

00:06:45.630 --> 00:06:46.610 way of approaching,

NOTE Confidence: 0.97693294

00:06:47.710 --> 00:06:49.389 embeddings where you just put

NOTE Confidence: 0.97693294

00:06:49.389 --> 00:06:50.830 your data in through some

NOTE Confidence: 0.97693294

00:06:50.830 --> 00:06:51.330 prebaked

NOTE Confidence: 0.9988715

00:06:51.710 --> 00:06:52.830 model, what you can do

NOTE Confidence: 0.9988715

00:06:52.830 --> 00:06:53.650 is you can

NOTE Confidence: 0.9834541

00:06:54.495 --> 00:06:56.335 apply these attention models to

NOTE Confidence: 0.9834541

00:06:56.335 --> 00:06:57.935 sort of compress your data

NOTE Confidence: 0.9834541

00:06:57.935 --> 00:06:59.395 into a fixed size representation.

NOTE Confidence: 0.9834541

00:06:59.455 --> 00:07:00.895 So here's an example for

NOTE Confidence: 0.9834541

00:07:00.895 --> 00:07:02.335 images. You can take an

NOTE Confidence: 0.9834541

00:07:02.335 --> 00:07:04.275 image, split it into patches,

NOTE Confidence: 0.9834541

00:07:04.415 --> 00:07:05.615 and then learn an attention

NOTE Confidence: 0.9834541

00:07:05.615 --> 00:07:07.055 model on tap, which kind

NOTE Confidence: 0.9834541

00:07:07.055 --> 00:07:08.575 of compresses the whole image

NOTE Confidence: 0.9834541

00:07:08.575 --> 00:07:09.715 down to a single

NOTE Confidence: 0.9856396

00:07:10.099 --> 00:07:11.060 vector, and now you can

NOTE Confidence: 0.9856396

00:07:11.060 --> 00:07:12.660 use that for anything else

NOTE Confidence: 0.9856396

00:07:12.660 --> 00:07:13.300 you want to do with

NOTE Confidence: 0.9856396

00:07:13.300 --> 00:07:13.800 it.

NOTE Confidence: 0.99936116

00:07:16.419 --> 00:07:16.919 Furthermore,

NOTE Confidence: 0.97257316

00:07:17.460 --> 00:07:19.300 what's really interesting is now

NOTE Confidence: 0.97257316

00:07:19.300 --> 00:07:20.180 you can do this for

NOTE Confidence: 0.97257316

00:07:20.180 --> 00:07:21.860 various modalities across time. You

NOTE Confidence: 0.97257316

00:07:21.860 --> 00:07:22.759 can have images.

NOTE Confidence: 0.98781484

00:07:23.164 --> 00:07:24.604 You can have text. You

NOTE Confidence: 0.98781484

00:07:24.604 --> 00:07:25.645 can have videos. You can

NOTE Confidence: 0.98781484

00:07:25.645 --> 00:07:27.164 have sound. They all get

NOTE Confidence: 0.98781484

00:07:27.164 --> 00:07:28.525 embedded into the same space,

NOTE Confidence: 0.98781484

00:07:28.525 --> 00:07:29.245 and you can,

NOTE Confidence: 0.9778141

00:07:30.525 --> 00:07:31.645 compress them down to the

NOTE Confidence: 0.9778141

00:07:31.645 --> 00:07:32.685 same format. So you can

NOTE Confidence: 0.9778141

00:07:32.685 --> 00:07:33.965 do things like apply it

NOTE Confidence: 0.9778141

00:07:33.965 --> 00:07:35.645 to different sentences of different

NOTE Confidence: 0.9778141

00:07:35.645 --> 00:07:36.845 lengths. So you don't have

NOTE Confidence: 0.9778141

00:07:36.845 --> 00:07:38.940 to worry about periodicity or

NOTE Confidence: 0.9778141

00:07:38.940 --> 00:07:40.300 or the fact that everything

NOTE Confidence: 0.9778141

00:07:40.300 --> 00:07:40.940 has to be on the

NOTE Confidence: 0.9778141

00:07:40.940 --> 00:07:42.880 same length, with this device,

NOTE Confidence: 0.9992147

00:07:43.419 --> 00:07:44.539 and you can do this

NOTE Confidence: 0.9992147

00:07:44.539 --> 00:07:46.160 across different data types.

NOTE Confidence: 0.9905334

00:07:47.419 --> 00:07:48.160 And so,

NOTE Confidence: 0.95703995

00:07:48.620 --> 00:07:49.580 you know, it's a really

NOTE Confidence: 0.95703995

00:07:49.580 --> 00:07:51.064 powerful tool, and I, you

NOTE Confidence: 0.95703995

00:07:51.064 --> 00:07:52.185 know, I wanted to highlight

NOTE Confidence: 0.95703995

00:07:52.185 --> 00:07:53.224 that today because I think,

NOTE Confidence: 0.95703995

00:07:53.224 --> 00:07:54.425 you know, if you're dealing

NOTE Confidence: 0.95703995

00:07:54.425 --> 00:07:56.664 with multivariate data, you can,

NOTE Confidence: 0.95703995

00:07:56.664 --> 00:07:58.104 over time, think about clever

NOTE Confidence: 0.95703995

00:07:58.104 --> 00:07:58.604 techniques,

NOTE Confidence: 0.9917822

00:07:59.064 --> 00:08:00.504 on how to combine them

NOTE Confidence: 0.9917822

00:08:00.504 --> 00:08:01.625 together. And a lot of

NOTE Confidence: 0.9917822

00:08:01.625 --> 00:08:02.125 ingenuity

NOTE Confidence: 0.9240962

00:08:02.504 --> 00:08:03.784 that's gone into things like

NOTE Confidence: 0.9240962

00:08:03.784 --> 00:08:04.824 alpha bold and stuff is

NOTE Confidence: 0.9240962

00:08:04.824 --> 00:08:06.044 about how do you combine,

NOTE Confidence: 0.97064024

00:08:06.720 --> 00:08:08.319 various, data that goes in

NOTE Confidence: 0.97064024

00:08:08.319 --> 00:08:08.819 there.

NOTE Confidence: 0.9897727

00:08:09.280 --> 00:08:10.560 So it requires some experience,

NOTE Confidence: 0.9897727

00:08:10.560 --> 00:08:11.440 but I think, you know,

NOTE Confidence: 0.9897727

00:08:11.440 --> 00:08:12.720 just with a little tweaking,

NOTE Confidence: 0.9897727

00:08:12.960 --> 00:08:13.840 you get pretty good at

NOTE Confidence: 0.9897727

00:08:13.840 --> 00:08:14.340 it.

NOTE Confidence: 0.9360766

00:08:16.400 --> 00:08:18.160 Okay. So switching to generative

NOTE Confidence: 0.9360766

00:08:18.160 --> 00:08:18.660 models.

NOTE Confidence: 0.9735073

00:08:19.895 --> 00:08:20.775 Once you have an embedding,

NOTE Confidence: 0.9735073

00:08:20.775 --> 00:08:21.975 you can do sort of

NOTE Confidence: 0.9735073

00:08:21.975 --> 00:08:23.275 generative models of data.

NOTE Confidence: 0.991169

00:08:23.815 --> 00:08:25.575 What's a generative model? A

NOTE Confidence: 0.991169

00:08:25.575 --> 00:08:26.775 generative model is a model

NOTE Confidence: 0.991169

00:08:26.775 --> 00:08:28.535 that allows, by definition, to

NOTE Confidence: 0.991169

00:08:28.535 --> 00:08:30.215 generate new data of that

NOTE Confidence: 0.991169

00:08:30.215 --> 00:08:30.715 modality.

NOTE Confidence: 0.81144845

00:08:31.895 --> 00:08:32.395 Additionally,

NOTE Confidence: 0.997471

00:08:32.850 --> 00:08:34.470 it can help you quantify

NOTE Confidence: 0.997471

00:08:34.610 --> 00:08:36.529 whether something you're seeing has

NOTE Confidence: 0.997471

00:08:36.529 --> 00:08:38.230 high probability or low probability,

NOTE Confidence: 0.9947732

00:08:39.329 --> 00:08:40.529 so that you can do

NOTE Confidence: 0.9947732

00:08:40.529 --> 00:08:41.190 other things,

NOTE Confidence: 0.9453179

00:08:42.050 --> 00:08:43.490 with that probability as in

NOTE Confidence: 0.9453179

00:08:43.490 --> 00:08:45.190 build tools on top of,

NOTE Confidence: 0.99960065

00:08:45.730 --> 00:08:47.110 those measures itself.

NOTE Confidence: 0.973057

00:08:48.154 --> 00:08:49.995 There's a wide variety of,

NOTE Confidence: 0.9813427

00:08:50.795 --> 00:08:52.495 techniques for generative models,

NOTE Confidence: 0.97441506

00:08:53.514 --> 00:08:55.035 that the machine learning community

NOTE Confidence: 0.97441506

00:08:55.035 --> 00:08:56.634 has built over time, but

NOTE Confidence: 0.97441506

00:08:56.634 --> 00:08:58.314 I'll basically just be talking

NOTE Confidence: 0.97441506

00:08:58.314 --> 00:08:59.134 about autoregressive

NOTE Confidence: 0.9916067

00:08:59.434 --> 00:09:01.595 models and diffusion models, which

NOTE Confidence: 0.9916067

00:09:01.595 --> 00:09:03.600 are really the mainstay of,

NOTE Confidence: 0.99830574

00:09:04.220 --> 00:09:05.820 of the models today. You're

NOTE Confidence: 0.99830574

00:09:05.820 --> 00:09:07.279 quite familiar with them.

NOTE Confidence: 0.9160248

00:09:08.220 --> 00:09:10.380 Autoregressive models are an example

NOTE Confidence: 0.9160248

00:09:10.380 --> 00:09:11.199 would be CHATCPT,

NOTE Confidence: 0.90412205

00:09:12.940 --> 00:09:14.160 for diffusion models,

NOTE Confidence: 0.99603665

00:09:14.540 --> 00:09:15.980 you know, something like stable

NOTE Confidence: 0.99603665

00:09:15.980 --> 00:09:17.679 diffusion for image generation

NOTE Confidence: 0.96388674

00:09:18.205 --> 00:09:19.105 is an example.

NOTE Confidence: 0.9845737

00:09:21.645 --> 00:09:22.545 Okay. So,

NOTE Confidence: 0.9953401

00:09:23.165 --> 00:09:24.705 with autoregressive models,

NOTE Confidence: 0.98527426

00:09:26.365 --> 00:09:27.565 the goal is to build

NOTE Confidence: 0.98527426

00:09:27.565 --> 00:09:28.225 a model,

NOTE Confidence: 0.9962593

00:09:29.245 --> 00:09:30.524 where you get a probability

NOTE Confidence: 0.9962593

00:09:30.524 --> 00:09:31.825 for any data point.

NOTE Confidence: 0.9878708

00:09:32.220 --> 00:09:32.459 And,

NOTE Confidence: 0.95433795

00:09:33.339 --> 00:09:34.540 the way we we do

NOTE Confidence: 0.95433795

00:09:34.540 --> 00:09:35.980 this with autoregressive models is

NOTE Confidence: 0.95433795

00:09:35.980 --> 00:09:37.500 to convert high dimensional data

NOTE Confidence: 0.95433795

00:09:37.500 --> 00:09:38.559 into a sequence

NOTE Confidence: 0.96518314

00:09:39.100 --> 00:09:40.699 and then measure the probability

NOTE Confidence: 0.96518314

00:09:40.699 --> 00:09:42.139 of the sequence using the

NOTE Confidence: 0.96518314

00:09:42.139 --> 00:09:44.320 chain rule of conditional probability,

NOTE Confidence: 0.96518314

00:09:44.459 --> 00:09:45.600 which is just basically

NOTE Confidence: 0.9940352

00:09:46.735 --> 00:09:48.195 multiply the probabilities

NOTE Confidence: 0.9452506

00:09:48.735 --> 00:09:49.235 of,

NOTE Confidence: 0.9831361

00:09:49.695 --> 00:09:51.635 one variable given the rest.

NOTE Confidence: 0.9831361

00:09:51.775 --> 00:09:52.815 I think the details are

NOTE Confidence: 0.9831361

00:09:52.815 --> 00:09:53.934 not too important. I'll try

NOTE Confidence: 0.9831361

00:09:53.934 --> 00:09:55.535 and highlight with this sort

NOTE Confidence: 0.9831361

00:09:55.535 --> 00:09:56.675 of example here.

NOTE Confidence: 0.96374416

00:09:57.135 --> 00:09:58.015 Let's say you have the

NOTE Confidence: 0.96374416

00:09:58.015 --> 00:09:59.135 web and you want to

NOTE Confidence: 0.96374416

00:09:59.135 --> 00:10:00.495 build a generative model of

NOTE Confidence: 0.96374416

00:10:00.495 --> 00:10:00.995 text.

NOTE Confidence: 0.99962986

00:10:01.829 --> 00:10:02.790 What you would do is

NOTE Confidence: 0.99962986

00:10:02.790 --> 00:10:04.329 you would take the entire

NOTE Confidence: 0.9417427

00:10:04.630 --> 00:10:06.309 dataset of text and convert

NOTE Confidence: 0.9417427

00:10:06.309 --> 00:10:08.410 it to input output pairs

NOTE Confidence: 0.9606176

00:10:09.829 --> 00:10:11.350 of of the type x

NOTE Confidence: 0.9606176

00:10:11.350 --> 00:10:12.470 and y. So you're given

NOTE Confidence: 0.9606176

00:10:12.470 --> 00:10:14.069 some data x and you

NOTE Confidence: 0.9606176

00:10:14.069 --> 00:10:15.295 want to predict y. Know,

NOTE Confidence: 0.9606176

00:10:15.295 --> 00:10:16.895 you're familiar with sort of

NOTE Confidence: 0.9606176

00:10:16.895 --> 00:10:18.575 regression or logistic regression. It's

NOTE Confidence: 0.9606176

00:10:18.575 --> 00:10:19.915 the same sort of,

NOTE Confidence: 0.97986174

00:10:20.415 --> 00:10:22.335 technique. You're basically just trying

NOTE Confidence: 0.97986174

00:10:22.335 --> 00:10:22.995 to predict,

NOTE Confidence: 0.98455364

00:10:23.615 --> 00:10:25.295 some some target given some

NOTE Confidence: 0.98455364

00:10:25.295 --> 00:10:25.795 input.

NOTE Confidence: 0.9953063

00:10:26.175 --> 00:10:27.455 And so with with the

NOTE Confidence: 0.9953063

00:10:27.455 --> 00:10:28.495 web, what you would do

NOTE Confidence: 0.9953063

00:10:28.495 --> 00:10:29.455 is you would just take

NOTE Confidence: 0.9953063

00:10:29.455 --> 00:10:30.275 all the prefixes.

NOTE Confidence: 0.9826467

00:10:31.050 --> 00:10:31.550 So,

NOTE Confidence: 0.8988892

00:10:32.410 --> 00:10:33.770 you got a Shakespearean verse

NOTE Confidence: 0.8988892

00:10:33.770 --> 00:10:34.990 here, and you could say,

NOTE Confidence: 0.8988892

00:10:35.130 --> 00:10:35.850 to be or not to

NOTE Confidence: 0.8988892

00:10:35.850 --> 00:10:37.290 be. So you convert that

NOTE Confidence: 0.8988892

00:10:37.290 --> 00:10:38.350 into data examples.

NOTE Confidence: 0.8964758

00:10:38.809 --> 00:10:41.050 Empty start, first word being

NOTE Confidence: 0.8964758

00:10:41.050 --> 00:10:41.550 two,

NOTE Confidence: 0.9859921

00:10:42.010 --> 00:10:43.690 and then x being two,

NOTE Confidence: 0.9859921

00:10:43.690 --> 00:10:44.890 and the next word is

NOTE Confidence: 0.9859921

00:10:44.890 --> 00:10:45.390 b.

NOTE Confidence: 0.94860864

00:10:45.855 --> 00:10:47.695 That's another data example. And

NOTE Confidence: 0.94860864

00:10:47.695 --> 00:10:48.575 then to b is an

NOTE Confidence: 0.94860864

00:10:48.575 --> 00:10:49.934 in an input for another

NOTE Confidence: 0.94860864

00:10:49.934 --> 00:10:50.834 one, and

NOTE Confidence: 0.83322877

00:10:51.295 --> 00:10:52.355 r is the target,

NOTE Confidence: 0.94911176

00:10:53.135 --> 00:10:54.415 word. And so you can

NOTE Confidence: 0.94911176

00:10:54.415 --> 00:10:56.415 convert the entire dataset entire

NOTE Confidence: 0.94911176

00:10:56.415 --> 00:10:57.955 web into such a database,

NOTE Confidence: 0.94911176

00:10:58.095 --> 00:10:58.595 and

NOTE Confidence: 0.99007714

00:10:59.059 --> 00:11:00.339 you are now learning a

NOTE Confidence: 0.99007714

00:11:00.339 --> 00:11:01.959 model that learns to predict,

NOTE Confidence: 0.9947223

00:11:02.660 --> 00:11:04.339 the next word given whatever

NOTE Confidence: 0.9947223

00:11:04.339 --> 00:11:05.320 context it is.

NOTE Confidence: 0.9926897

00:11:06.339 --> 00:11:07.620 And so, you know, this

NOTE Confidence: 0.9926897

00:11:07.620 --> 00:11:08.360 is the workhorse

NOTE Confidence: 0.9559631

00:11:08.820 --> 00:11:10.660 of how, current models like

NOTE Confidence: 0.9559631

00:11:10.660 --> 00:11:12.579 ChatDPD work. You just take

NOTE Confidence: 0.9559631

00:11:12.579 --> 00:11:13.860 the web and just do

NOTE Confidence: 0.9559631

00:11:13.860 --> 00:11:14.360 next

NOTE Confidence: 0.9912101

00:11:14.754 --> 00:11:15.574 token prediction,

NOTE Confidence: 0.9499623

00:11:16.274 --> 00:11:17.954 as it were. And then

NOTE Confidence: 0.9499623

00:11:17.954 --> 00:11:18.834 when you want to run

NOTE Confidence: 0.9499623

00:11:18.834 --> 00:11:20.355 the model, what you do

NOTE Confidence: 0.9499623

00:11:20.355 --> 00:11:22.035 is you feed in some

NOTE Confidence: 0.9499623

00:11:22.035 --> 00:11:24.035 context such as what it,

NOTE Confidence: 0.9499623

00:11:24.274 --> 00:11:25.654 why is the sky blue,

NOTE Confidence: 0.9499623

00:11:25.795 --> 00:11:26.595 and then you let the

NOTE Confidence: 0.9499623

00:11:26.595 --> 00:11:28.434 model generate the next, word,

NOTE Confidence: 0.9499623

00:11:28.434 --> 00:11:29.559 which which it has already

NOTE Confidence: 0.9499623

00:11:29.559 --> 00:11:31.640 learned from its model, and

NOTE Confidence: 0.9499623

00:11:31.640 --> 00:11:32.839 then you take that word,

NOTE Confidence: 0.9499623

00:11:32.839 --> 00:11:34.040 for example, in this case,

NOTE Confidence: 0.9499623

00:11:34.040 --> 00:11:34.540 because,

NOTE Confidence: 0.99696064

00:11:35.000 --> 00:11:36.220 and you feed that

NOTE Confidence: 0.9994837

00:11:36.520 --> 00:11:37.820 word back in the next

NOTE Confidence: 0.55545586

00:11:38.279 --> 00:11:38.779 sentence,

NOTE Confidence: 0.93831074

00:11:39.160 --> 00:11:40.360 why is the sky blue

NOTE Confidence: 0.93831074

00:11:40.360 --> 00:11:41.640 because, and then you have

NOTE Confidence: 0.93831074

00:11:41.640 --> 00:11:43.020 it predict the next word,

NOTE Confidence: 0.93831074

00:11:43.160 --> 00:11:44.140 because of,

NOTE Confidence: 0.9392017

00:11:45.695 --> 00:11:47.535 Raleigh, scattering, and so and

NOTE Confidence: 0.9392017

00:11:47.535 --> 00:11:48.975 so on. So you basically

NOTE Confidence: 0.9392017

00:11:48.975 --> 00:11:50.495 run the model during inference,

NOTE Confidence: 0.9392017

00:11:50.495 --> 00:11:52.014 and it it just generates

NOTE Confidence: 0.9392017

00:11:52.014 --> 00:11:52.995 text, and that,

NOTE Confidence: 0.8744515

00:11:53.695 --> 00:11:54.355 that generated,

NOTE Confidence: 0.998186

00:11:55.214 --> 00:11:56.035 output is,

NOTE Confidence: 0.9990994

00:11:56.574 --> 00:11:57.875 you know, what you see.

NOTE Confidence: 0.9575497

00:11:59.850 --> 00:12:00.890 So how do you apply

NOTE Confidence: 0.9575497

00:12:00.890 --> 00:12:02.490 this to other modalities other

NOTE Confidence: 0.9575497

00:12:02.490 --> 00:12:04.250 than text? It's quite clear

NOTE Confidence: 0.9575497

00:12:04.250 --> 00:12:05.070 for text.

NOTE Confidence: 0.9768272

00:12:05.450 --> 00:12:06.570 It's a discrete data, and

NOTE Confidence: 0.9768272

00:12:06.570 --> 00:12:07.690 these models work really well

NOTE Confidence: 0.9768272

00:12:07.690 --> 00:12:09.290 for discrete data. They don't

NOTE Confidence: 0.9768272

00:12:09.290 --> 00:12:10.410 really work so well in

NOTE Confidence: 0.9768272

00:12:10.410 --> 00:12:12.110 continuous regression space.

NOTE Confidence: 0.97010106

00:12:12.809 --> 00:12:14.410 So it's easy to apply

NOTE Confidence: 0.97010106

00:12:14.410 --> 00:12:16.464 just to things like protein

NOTE Confidence: 0.97010106

00:12:16.464 --> 00:12:18.005 sequences, amino acids,

NOTE Confidence: 0.9714646

00:12:18.785 --> 00:12:20.065 and so on that are

NOTE Confidence: 0.9714646

00:12:20.065 --> 00:12:22.065 naturally discrete. It's a little

NOTE Confidence: 0.9714646

00:12:22.065 --> 00:12:24.565 trickier for, high dimensional data,

NOTE Confidence: 0.96896863

00:12:25.904 --> 00:12:27.445 which are not like strings.

NOTE Confidence: 0.96896863

00:12:27.505 --> 00:12:29.240 So when it highlights

NOTE Confidence: 0.9737557

00:12:29.860 --> 00:12:31.059 how how people do this,

NOTE Confidence: 0.9737557

00:12:31.540 --> 00:12:32.820 what you end up doing

NOTE Confidence: 0.9737557

00:12:32.820 --> 00:12:34.580 is building a model which

NOTE Confidence: 0.9737557

00:12:34.580 --> 00:12:35.860 first encodes your data into

NOTE Confidence: 0.9737557

00:12:35.860 --> 00:12:36.980 a sequence of,

NOTE Confidence: 0.92654634

00:12:38.100 --> 00:12:38.660 of tokens,

NOTE Confidence: 0.20023243

00:12:39.140 --> 00:12:39.640 dispute

NOTE Confidence: 0.93720037

00:12:40.660 --> 00:12:41.160 tokens,

NOTE Confidence: 0.9979974

00:12:41.620 --> 00:12:42.360 and then

NOTE Confidence: 0.9896786

00:12:42.765 --> 00:12:44.305 you learn another model.

NOTE Confidence: 0.9763159

00:12:44.765 --> 00:12:46.125 Typically, actually, you learn both

NOTE Confidence: 0.9763159

00:12:46.125 --> 00:12:47.265 of these models together.

NOTE Confidence: 0.94430864

00:12:47.645 --> 00:12:48.605 And the other model is

NOTE Confidence: 0.94430864

00:12:48.605 --> 00:12:50.845 called reconstruction model, which takes

NOTE Confidence: 0.94430864

00:12:50.845 --> 00:12:51.345 in,

NOTE Confidence: 0.9988439

00:12:52.205 --> 00:12:53.725 the output tokens and converts

NOTE Confidence: 0.9988439

00:12:53.725 --> 00:12:54.845 it back to the data

NOTE Confidence: 0.9988439

00:12:54.845 --> 00:12:55.345 itself.

NOTE Confidence: 0.9334738

00:12:56.010 --> 00:12:57.370 And once you have this,

NOTE Confidence: 0.9334738

00:12:57.370 --> 00:12:58.570 you can convert your entire

NOTE Confidence: 0.9334738

00:12:58.570 --> 00:13:00.010 data into the sequence of

NOTE Confidence: 0.9334738

00:13:00.010 --> 00:13:00.510 tokens,

NOTE Confidence: 0.8152515

00:13:01.929 --> 00:13:03.230 and then learn the autoregressive

NOTE Confidence: 0.98194706

00:13:03.690 --> 00:13:05.050 model on that sequence of

NOTE Confidence: 0.98194706

00:13:05.050 --> 00:13:06.650 tokens by just predicting the

NOTE Confidence: 0.98194706

00:13:06.650 --> 00:13:08.510 next token given the history

NOTE Confidence: 0.998178

00:13:08.890 --> 00:13:09.790 of its tokens.

NOTE Confidence: 0.7742119

00:13:10.650 --> 00:13:10.809 And,

NOTE Confidence: 0.9991715

00:13:12.464 --> 00:13:13.904 you can now generate this

NOTE Confidence: 0.9991715

00:13:13.904 --> 00:13:15.125 new kind of data

NOTE Confidence: 0.9734298

00:13:15.505 --> 00:13:17.425 by running the autoregressive model,

NOTE Confidence: 0.9734298

00:13:17.425 --> 00:13:19.505 generating some sequences, and then

NOTE Confidence: 0.9734298

00:13:19.505 --> 00:13:21.205 converting that to real data.

NOTE Confidence: 0.9666131

00:13:22.945 --> 00:13:24.225 Here's an example on how

NOTE Confidence: 0.9666131

00:13:24.225 --> 00:13:25.345 you might apply this to

NOTE Confidence: 0.9666131

00:13:25.345 --> 00:13:26.725 modality like speech.

NOTE Confidence: 0.94226086

00:13:28.100 --> 00:13:29.779 So you have speech is

NOTE Confidence: 0.94226086

00:13:29.779 --> 00:13:30.760 really just waveforms.

NOTE Confidence: 0.93279773

00:13:31.860 --> 00:13:33.059 And so to be able

NOTE Confidence: 0.93279773

00:13:33.059 --> 00:13:34.420 to build a model actually,

NOTE Confidence: 0.93279773

00:13:34.420 --> 00:13:36.340 you were really really wanted

NOTE Confidence: 0.93279773

00:13:36.340 --> 00:13:36.980 to do it. You could

NOTE Confidence: 0.93279773

00:13:36.980 --> 00:13:38.440 just model speech directly,

NOTE Confidence: 0.9765533

00:13:38.900 --> 00:13:40.279 and people have done that.

NOTE Confidence: 0.981973

00:13:41.495 --> 00:13:43.335 But it's harder to, deal

NOTE Confidence: 0.981973

00:13:43.335 --> 00:13:45.015 with that because speech happens

NOTE Confidence: 0.981973

00:13:45.015 --> 00:13:45.895 at a very fast rate,

NOTE Confidence: 0.981973

00:13:45.895 --> 00:13:46.775 so the data would be

NOTE Confidence: 0.981973

00:13:46.775 --> 00:13:48.535 just too much. So, typically,

NOTE Confidence: 0.981973

00:13:48.535 --> 00:13:49.575 what people will do now

NOTE Confidence: 0.981973

00:13:49.575 --> 00:13:50.695 is convert speech into a

NOTE Confidence: 0.981973

00:13:50.695 --> 00:13:51.755 spectral representation,

NOTE Confidence: 0.9363853

00:13:52.695 --> 00:13:54.135 by just taking windows of

NOTE Confidence: 0.9363853

00:13:54.135 --> 00:13:54.955 speech and,

NOTE Confidence: 0.97367334

00:13:55.360 --> 00:13:56.960 computing a Fourier spectrum in

NOTE Confidence: 0.97367334

00:13:56.960 --> 00:13:58.800 it. So this original waveform

NOTE Confidence: 0.97367334

00:13:58.800 --> 00:14:00.960 is converted to a frequency

NOTE Confidence: 0.97367334

00:14:00.960 --> 00:14:02.900 diagram over time showing how

NOTE Confidence: 0.9692934

00:14:03.280 --> 00:14:04.020 the sound,

NOTE Confidence: 0.99793005

00:14:04.480 --> 00:14:06.559 is is distributing energy over

NOTE Confidence: 0.99793005

00:14:06.559 --> 00:14:08.400 these different frequencies on on

NOTE Confidence: 0.99793005

00:14:08.400 --> 00:14:09.220 y axis.

NOTE Confidence: 0.9845158

00:14:11.834 --> 00:14:13.675 And once once you've converted

NOTE Confidence: 0.9845158

00:14:13.675 --> 00:14:15.355 that to this, format, you

NOTE Confidence: 0.9845158

00:14:15.355 --> 00:14:16.714 can learn an inverse model

NOTE Confidence: 0.9845158

00:14:16.714 --> 00:14:18.075 called the vocoder, which will

NOTE Confidence: 0.9845158

00:14:18.075 --> 00:14:18.975 just generate,

NOTE Confidence: 0.98965985

00:14:19.834 --> 00:14:21.535 the raw waveform from

NOTE Confidence: 0.92968524

00:14:22.740 --> 00:14:23.240 from,

NOTE Confidence: 0.99695516

00:14:23.780 --> 00:14:24.280 this

NOTE Confidence: 0.77232134

00:14:24.740 --> 00:14:25.640 coded speech.

NOTE Confidence: 0.9596989

00:14:26.340 --> 00:14:27.000 Now, unfortunately,

NOTE Confidence: 0.7789766

00:14:27.380 --> 00:14:27.880 the,

NOTE Confidence: 0.9638625

00:14:28.980 --> 00:14:30.100 spectrum on the right hand

NOTE Confidence: 0.9638625

00:14:30.100 --> 00:14:31.640 side is it's still continuous.

NOTE Confidence: 0.9638625

00:14:31.700 --> 00:14:33.540 It's not discrete, and so

NOTE Confidence: 0.9638625

00:14:33.540 --> 00:14:34.580 it's hard to embed it

NOTE Confidence: 0.9638625

00:14:34.580 --> 00:14:35.080 into,

NOTE Confidence: 0.966589

00:14:35.460 --> 00:14:36.920 an autoregressive model.

NOTE Confidence: 0.9789751

00:14:37.455 --> 00:14:38.815 So what what you can

NOTE Confidence: 0.9789751

00:14:38.815 --> 00:14:40.175 do is simply just take

NOTE Confidence: 0.9789751

00:14:40.175 --> 00:14:42.255 that data and tokenize it

NOTE Confidence: 0.9789751

00:14:42.255 --> 00:14:44.175 by discretizing the data. So

NOTE Confidence: 0.9789751

00:14:44.175 --> 00:14:46.035 just round divide by

NOTE Confidence: 0.9619739

00:14:46.415 --> 00:14:48.015 take off the minimum, divide

NOTE Confidence: 0.9619739

00:14:48.015 --> 00:14:49.855 by maximum, and just convert

NOTE Confidence: 0.9619739

00:14:49.855 --> 00:14:50.575 it into a range of

NOTE Confidence: 0.9619739

00:14:50.575 --> 00:14:51.855 numbers between zero and some

NOTE Confidence: 0.9619739

00:14:51.855 --> 00:14:52.835 maximum bin.

NOTE Confidence: 0.9940748

00:14:53.290 --> 00:14:54.170 And now you have a

NOTE Confidence: 0.9940748

00:14:54.170 --> 00:14:54.670 discretized

NOTE Confidence: 0.95988053

00:14:54.970 --> 00:14:56.750 version. You can convert back,

NOTE Confidence: 0.9949306

00:14:57.690 --> 00:14:58.890 to the original as well

NOTE Confidence: 0.9949306

00:14:58.890 --> 00:14:59.790 by just mapping,

NOTE Confidence: 0.99894947

00:15:00.330 --> 00:15:00.830 the

NOTE Confidence: 0.9706952

00:15:01.210 --> 00:15:03.370 continuous values to, the discrete

NOTE Confidence: 0.9706952

00:15:03.370 --> 00:15:04.990 values back to continuous codes.

NOTE Confidence: 0.9781245

00:15:05.545 --> 00:15:06.425 And so now you have

NOTE Confidence: 0.9781245

00:15:06.425 --> 00:15:07.625 this machinery by which you

NOTE Confidence: 0.9781245

00:15:07.625 --> 00:15:09.385 can take this continuous data,

NOTE Confidence: 0.9781245

00:15:09.385 --> 00:15:10.745 convert it to tokens, and

NOTE Confidence: 0.9781245

00:15:10.745 --> 00:15:12.105 then convert it back to

NOTE Confidence: 0.9781245

00:15:12.105 --> 00:15:13.305 real data. So you can

NOTE Confidence: 0.9781245

00:15:13.305 --> 00:15:15.225 really just beat that into

NOTE Confidence: 0.9781245

00:15:15.225 --> 00:15:16.605 an autoregressive model.

NOTE Confidence: 0.873317

00:15:16.985 --> 00:15:17.945 And so,

NOTE Confidence: 0.93582267

00:15:18.779 --> 00:15:19.980 you take the waveform. You

NOTE Confidence: 0.93582267

00:15:19.980 --> 00:15:21.200 have spectral representation.

NOTE Confidence: 0.9963702

00:15:21.580 --> 00:15:23.279 You take each spectral representation

NOTE Confidence: 0.9963702

00:15:23.420 --> 00:15:24.720 and convert it into

NOTE Confidence: 0.9433157

00:15:25.100 --> 00:15:26.720 a sequence of discrete tokens,

NOTE Confidence: 0.9990221

00:15:27.260 --> 00:15:27.760 and

NOTE Confidence: 0.968437

00:15:28.540 --> 00:15:29.580 voila. You can do next

NOTE Confidence: 0.968437

00:15:29.580 --> 00:15:30.700 step predictions. So you feed

NOTE Confidence: 0.968437

00:15:30.700 --> 00:15:32.459 in your history of tokens,

NOTE Confidence: 0.968437

00:15:32.459 --> 00:15:33.580 and then you can predict

NOTE Confidence: 0.968437

00:15:33.580 --> 00:15:35.015 the next token. So,

NOTE Confidence: 0.9897606

00:15:36.115 --> 00:15:37.715 it's basically a recipe that's,

NOTE Confidence: 0.9897606

00:15:38.035 --> 00:15:39.795 repeated all over. You just

NOTE Confidence: 0.9897606

00:15:39.795 --> 00:15:41.635 learn how to discretize your

NOTE Confidence: 0.9897606

00:15:41.635 --> 00:15:42.455 data into,

NOTE Confidence: 0.8945199

00:15:42.915 --> 00:15:44.115 some discrete bins, and then

NOTE Confidence: 0.8945199

00:15:44.115 --> 00:15:44.995 you learn how to address

NOTE Confidence: 0.8945199

00:15:44.995 --> 00:15:46.435 the model. And I think

NOTE Confidence: 0.8945199

00:15:46.435 --> 00:15:47.415 this could be applied.

NOTE Confidence: 0.96973443

00:15:48.710 --> 00:15:50.470 It's already applied to various

NOTE Confidence: 0.96973443

00:15:50.470 --> 00:15:51.530 things like speech,

NOTE Confidence: 0.99609625

00:15:52.310 --> 00:15:53.530 videos, images,

NOTE Confidence: 0.95932364

00:15:54.150 --> 00:15:55.430 and so on. And it's

NOTE Confidence: 0.95932364

00:15:55.430 --> 00:15:56.870 a pretty powerful technique that

NOTE Confidence: 0.95932364

00:15:56.870 --> 00:15:58.150 can be applied to other

NOTE Confidence: 0.95932364

00:15:58.150 --> 00:15:59.530 modalities as well.

NOTE Confidence: 0.9712601

00:16:02.665 --> 00:16:03.385 And now I want to

NOTE Confidence: 0.9712601

00:16:03.385 --> 00:16:04.265 talk a little bit about

NOTE Confidence: 0.9712601

00:16:04.265 --> 00:16:05.165 diffusion models.

NOTE Confidence: 0.9923687

00:16:05.945 --> 00:16:07.225 It's a set of new

NOTE Confidence: 0.9923687

00:16:07.225 --> 00:16:09.145 techniques that allows you to

NOTE Confidence: 0.9923687

00:16:09.145 --> 00:16:11.305 morph one probability distribution to

NOTE Confidence: 0.9923687

00:16:11.305 --> 00:16:11.805 another.

NOTE Confidence: 0.95028436

00:16:13.385 --> 00:16:14.025 And so,

NOTE Confidence: 0.99298847

00:16:15.170 --> 00:16:16.950 there's methods called optimal transport,

NOTE Confidence: 0.9803494

00:16:17.730 --> 00:16:19.730 flow matching, diffusion models. They're

NOTE Confidence: 0.9803494

00:16:19.730 --> 00:16:21.269 all trying to map

NOTE Confidence: 0.99975985

00:16:21.570 --> 00:16:22.070 distributions

NOTE Confidence: 0.9991649

00:16:22.610 --> 00:16:24.310 from one distribution to another.

NOTE Confidence: 0.8892764

00:16:24.850 --> 00:16:26.130 It might seem like a

NOTE Confidence: 0.8892764

00:16:26.130 --> 00:16:26.950 very arcane

NOTE Confidence: 0.98249334

00:16:27.595 --> 00:16:28.795 idea, but it's really a

NOTE Confidence: 0.98249334

00:16:28.795 --> 00:16:29.295 powerful,

NOTE Confidence: 0.99975014

00:16:29.995 --> 00:16:31.515 methodology when you want to

NOTE Confidence: 0.99975014

00:16:31.515 --> 00:16:32.175 think about

NOTE Confidence: 0.97722864

00:16:32.715 --> 00:16:34.095 how how to generate,

NOTE Confidence: 0.9961551

00:16:34.875 --> 00:16:36.555 data from noise. So in

NOTE Confidence: 0.9961551

00:16:36.555 --> 00:16:38.015 the case of diffusion models,

NOTE Confidence: 0.9961551

00:16:38.315 --> 00:16:39.295 you morph

NOTE Confidence: 0.9797806

00:16:39.710 --> 00:16:41.390 a Gaussian distribution, which is

NOTE Confidence: 0.9797806

00:16:41.390 --> 00:16:42.830 something that people know how

NOTE Confidence: 0.9797806

00:16:42.830 --> 00:16:43.410 to handle,

NOTE Confidence: 0.9928275

00:16:43.870 --> 00:16:45.070 and convert that into a

NOTE Confidence: 0.9928275

00:16:45.070 --> 00:16:46.910 real data distribution, which is

NOTE Confidence: 0.9928275

00:16:46.910 --> 00:16:47.650 really hard

NOTE Confidence: 0.9713933

00:16:48.110 --> 00:16:49.230 to handle. So if you

NOTE Confidence: 0.9713933

00:16:49.230 --> 00:16:51.230 give me images or or

NOTE Confidence: 0.9713933

00:16:51.230 --> 00:16:52.155 speech, I don't know what

NOTE Confidence: 0.9713933

00:16:52.235 --> 00:16:53.755 the data distribution itself is

NOTE Confidence: 0.9713933

00:16:53.755 --> 00:16:54.715 and how to model that

NOTE Confidence: 0.9713933

00:16:54.715 --> 00:16:56.955 distribution itself. Or multi omics

NOTE Confidence: 0.9713933

00:16:56.955 --> 00:16:57.915 data, like, how do you

NOTE Confidence: 0.9713933

00:16:58.075 --> 00:16:59.915 what is the actual distribution

NOTE Confidence: 0.9713933

00:16:59.915 --> 00:17:01.675 of data one doesn't know?

NOTE Confidence: 0.9713933

00:17:01.915 --> 00:17:03.675 And so the ability to

NOTE Confidence: 0.9713933

00:17:03.675 --> 00:17:04.955 to generate and sample from

NOTE Confidence: 0.9713933

00:17:04.955 --> 00:17:06.015 that is quite useful,

NOTE Confidence: 0.83288455

00:17:06.869 --> 00:17:07.369 and,

NOTE Confidence: 0.9985078

00:17:07.990 --> 00:17:09.850 mapping to a simple distribution

NOTE Confidence: 0.9985078

00:17:09.990 --> 00:17:11.369 allows us to do that.

NOTE Confidence: 0.9896939

00:17:11.910 --> 00:17:13.030 And so how does this

NOTE Confidence: 0.9896939

00:17:13.030 --> 00:17:14.570 actually work in practice?

NOTE Confidence: 0.9146771

00:17:15.109 --> 00:17:16.070 I'll show you an example

NOTE Confidence: 0.9146771

00:17:16.070 --> 00:17:16.810 with images.

NOTE Confidence: 0.97792745

00:17:17.910 --> 00:17:19.670 You have some image on

NOTE Confidence: 0.97792745

00:17:19.670 --> 00:17:20.790 the right hand side, x

NOTE Confidence: 0.97792745

00:17:20.790 --> 00:17:21.290 zero.

NOTE Confidence: 0.9652588

00:17:24.904 --> 00:17:25.865 I guess I don't see

NOTE Confidence: 0.9652588

00:17:25.865 --> 00:17:26.684 a mouse there.

NOTE Confidence: 0.9981302

00:17:27.144 --> 00:17:28.184 And you take that image

NOTE Confidence: 0.9981302

00:17:28.184 --> 00:17:29.404 on the right hand side

NOTE Confidence: 0.5950415

00:17:30.825 --> 00:17:31.325 and

NOTE Confidence: 0.9087844

00:17:38.150 --> 00:17:39.350 Maybe it's Never mind. It's

NOTE Confidence: 0.9087844

00:17:39.350 --> 00:17:39.429 okay.

NOTE Confidence: 0.99226874

00:17:41.109 --> 00:17:42.230 So you take an image

NOTE Confidence: 0.99226874

00:17:42.230 --> 00:17:43.289 on the right hand side.

NOTE Confidence: 0.9929583

00:17:43.750 --> 00:17:44.630 What you can do is

NOTE Confidence: 0.9929583

00:17:44.630 --> 00:17:45.450 you can just,

NOTE Confidence: 0.9767658

00:17:45.990 --> 00:17:47.530 scale it down in magnitude

NOTE Confidence: 0.9767658

00:17:47.590 --> 00:17:49.690 by multiplying by some compression

NOTE Confidence: 0.9767658

00:17:49.750 --> 00:17:50.984 term, and then you add

NOTE Confidence: 0.9767658

00:17:50.984 --> 00:17:52.505 some noise, which expands the

NOTE Confidence: 0.9767658

00:17:52.505 --> 00:17:54.265 data up again. And so

NOTE Confidence: 0.9767658

00:17:54.265 --> 00:17:55.625 you started with some data,

NOTE Confidence: 0.9767658

00:17:55.625 --> 00:17:56.585 and you can generate a

NOTE Confidence: 0.9767658

00:17:56.585 --> 00:17:57.865 whole bunch of data at

NOTE Confidence: 0.9767658

00:17:57.865 --> 00:17:59.005 different noise levels.

NOTE Confidence: 0.9612325

00:17:59.625 --> 00:18:00.125 And,

NOTE Confidence: 0.99967265

00:18:01.465 --> 00:18:02.425 what you really want to

NOTE Confidence: 0.99967265

00:18:02.425 --> 00:18:03.705 do is to learn a

NOTE Confidence: 0.99967265

00:18:03.705 --> 00:18:05.085 function that takes

NOTE Confidence: 0.99934196

00:18:05.450 --> 00:18:06.809 data at one noise level

NOTE Confidence: 0.99934196

00:18:06.809 --> 00:18:08.109 and cleans it up slightly

NOTE Confidence: 0.99558866

00:18:08.410 --> 00:18:09.869 to a slightly less,

NOTE Confidence: 0.99110234

00:18:10.410 --> 00:18:11.309 noisy level.

NOTE Confidence: 0.997515

00:18:11.690 --> 00:18:12.809 And so and you can

NOTE Confidence: 0.997515

00:18:12.809 --> 00:18:14.109 then apply that model.

NOTE Confidence: 0.9826151

00:18:14.490 --> 00:18:15.850 You start with noisy data,

NOTE Confidence: 0.9826151

00:18:15.850 --> 00:18:16.730 and then you clean it

NOTE Confidence: 0.9826151

00:18:16.730 --> 00:18:17.770 up a little bit. And

NOTE Confidence: 0.9826151

00:18:17.770 --> 00:18:18.809 then you clean it up

NOTE Confidence: 0.9826151

00:18:18.809 --> 00:18:19.690 a little bit, and you

NOTE Confidence: 0.9826151

00:18:19.690 --> 00:18:20.650 do this over and over

NOTE Confidence: 0.9826151

00:18:20.650 --> 00:18:21.150 again,

NOTE Confidence: 0.99317706

00:18:21.585 --> 00:18:23.025 till you are back to

NOTE Confidence: 0.99317706

00:18:23.025 --> 00:18:25.105 the clean cleanest level, which

NOTE Confidence: 0.99317706

00:18:25.105 --> 00:18:26.465 is where the data itself

NOTE Confidence: 0.99317706

00:18:26.465 --> 00:18:26.965 lies.

NOTE Confidence: 0.99713475

00:18:27.825 --> 00:18:28.865 And so that that's a

NOTE Confidence: 0.99713475

00:18:28.865 --> 00:18:30.945 very simplistic explanation of diffusion

NOTE Confidence: 0.99713475

00:18:30.945 --> 00:18:31.664 models. There's,

NOTE Confidence: 0.957609

00:18:32.225 --> 00:18:33.585 there's a whole range of,

NOTE Confidence: 0.9390027

00:18:34.279 --> 00:18:35.799 possibilities in this in this

NOTE Confidence: 0.9390027

00:18:35.799 --> 00:18:36.840 scheme. How do you add

NOTE Confidence: 0.9390027

00:18:36.840 --> 00:18:38.380 noise? How do you convert

NOTE Confidence: 0.9976297

00:18:38.840 --> 00:18:40.119 the noisy data back to

NOTE Confidence: 0.9976297

00:18:40.119 --> 00:18:40.940 clean data?

NOTE Confidence: 0.9725612

00:18:41.559 --> 00:18:42.519 And there's a whole bunch

NOTE Confidence: 0.9725612

00:18:42.519 --> 00:18:44.519 of techniques, that factor in

NOTE Confidence: 0.9725612

00:18:44.519 --> 00:18:45.559 different trade offs,

NOTE Confidence: 0.9973376

00:18:47.575 --> 00:18:48.715 in in these choices.

NOTE Confidence: 0.93135947

00:18:49.335 --> 00:18:51.355 There's also variants of diffusion

NOTE Confidence: 0.93135947

00:18:51.494 --> 00:18:51.994 that,

NOTE Confidence: 0.9277392

00:18:53.255 --> 00:18:54.295 don't look at it as

NOTE Confidence: 0.9277392

00:18:54.295 --> 00:18:55.655 a sequence of discrete steps,

NOTE Confidence: 0.9277392

00:18:55.655 --> 00:18:57.255 but it deal with this

NOTE Confidence: 0.9277392

00:18:57.255 --> 00:18:58.615 as, like, a continuous time

NOTE Confidence: 0.9277392

00:18:58.615 --> 00:18:59.735 step, which is almost like

NOTE Confidence: 0.9277392

00:18:59.735 --> 00:18:59.985 a,

NOTE Confidence: 0.97794175

00:19:03.110 --> 00:19:04.869 a diffusion process in continuous

NOTE Confidence: 0.97794175

00:19:04.869 --> 00:19:05.369 time.

NOTE Confidence: 0.9851712

00:19:05.750 --> 00:19:06.250 And,

NOTE Confidence: 0.98043406

00:19:06.869 --> 00:19:08.470 there's also techniques that apply

NOTE Confidence: 0.98043406

00:19:08.470 --> 00:19:09.670 this for discrete data. So

NOTE Confidence: 0.98043406

00:19:09.670 --> 00:19:11.110 I've been showing you continuous

NOTE Confidence: 0.98043406

00:19:11.110 --> 00:19:12.950 data. Even discrete data can

NOTE Confidence: 0.98043406

00:19:12.950 --> 00:19:14.390 work through diffusion models where

NOTE Confidence: 0.98043406

00:19:14.390 --> 00:19:15.050 you have,

NOTE Confidence: 0.9127442

00:19:15.705 --> 00:19:17.164 have categorical choices,

NOTE Confidence: 0.9271653

00:19:17.544 --> 00:19:18.984 kind of like maybe like

NOTE Confidence: 0.9271653

00:19:18.984 --> 00:19:20.744 mutations during evolution. It's,

NOTE Confidence: 0.9280713

00:19:21.385 --> 00:19:22.924 you know, just things

NOTE Confidence: 0.91761035

00:19:23.225 --> 00:19:24.205 mutate from,

NOTE Confidence: 0.9763899

00:19:26.105 --> 00:19:27.625 from signal down to noise,

NOTE Confidence: 0.9763899

00:19:27.625 --> 00:19:28.585 and then you learn a

NOTE Confidence: 0.9763899

00:19:28.585 --> 00:19:30.169 model on going backwards,

NOTE Confidence: 0.99806803

00:19:30.470 --> 00:19:32.009 to generate the data for,

NOTE Confidence: 0.9974429

00:19:32.629 --> 00:19:33.609 for real sequences.

NOTE Confidence: 0.9751269

00:19:34.710 --> 00:19:36.149 And there's even a continuous

NOTE Confidence: 0.9751269

00:19:36.149 --> 00:19:37.830 time version of this discrete

NOTE Confidence: 0.9751269

00:19:37.830 --> 00:19:38.789 diffusion process,

NOTE Confidence: 0.99756587

00:19:39.109 --> 00:19:40.250 if you can believe me.

NOTE Confidence: 0.9762351

00:19:41.190 --> 00:19:42.470 Okay. So these models work

NOTE Confidence: 0.9762351

00:19:42.470 --> 00:19:42.970 well.

NOTE Confidence: 0.9712226

00:19:44.355 --> 00:19:45.155 So I don't wanna leave

NOTE Confidence: 0.9712226

00:19:45.155 --> 00:19:46.195 you with the impression that

NOTE Confidence: 0.9712226

00:19:46.195 --> 00:19:47.635 everything just works right off

NOTE Confidence: 0.9712226

00:19:47.635 --> 00:19:48.835 the bat. So I wanna

NOTE Confidence: 0.9712226

00:19:48.835 --> 00:19:50.695 highlight an example just,

NOTE Confidence: 0.8980202

00:19:52.275 --> 00:19:53.315 just to leave with you

NOTE Confidence: 0.8980202

00:19:53.315 --> 00:19:54.055 with a vignette

NOTE Confidence: 0.9746112

00:19:55.690 --> 00:19:56.510 of how,

NOTE Confidence: 0.9954985

00:19:57.050 --> 00:19:58.250 the kinds of innovations you

NOTE Confidence: 0.9954985

00:19:58.250 --> 00:19:59.130 need to do to make

NOTE Confidence: 0.9954985

00:19:59.130 --> 00:20:00.570 some things work when you

NOTE Confidence: 0.9954985

00:20:00.570 --> 00:20:01.869 take on a new challenge.

NOTE Confidence: 0.90383697

00:20:02.330 --> 00:20:03.530 So if you use models

NOTE Confidence: 0.90383697

00:20:03.530 --> 00:20:04.330 work well, but if you

NOTE Confidence: 0.90383697

00:20:04.330 --> 00:20:05.770 get really large data, like

NOTE Confidence: 0.90383697

00:20:05.770 --> 00:20:07.690 high resolution images, it's a

NOTE Confidence: 0.90383697

00:20:07.690 --> 00:20:08.510 lot more tricky,

NOTE Confidence: 0.99846977

00:20:08.890 --> 00:20:09.850 to make it work right

NOTE Confidence: 0.99846977

00:20:09.850 --> 00:20:10.590 off the bat.

NOTE Confidence: 0.9983132

00:20:10.975 --> 00:20:12.734 And so what people will

NOTE Confidence: 0.9983132

00:20:12.734 --> 00:20:14.254 end up doing is I'm

NOTE Confidence: 0.9983132

00:20:14.254 --> 00:20:15.934 highlighting two different techniques in

NOTE Confidence: 0.9983132

00:20:15.934 --> 00:20:17.534 in literature. On the left

NOTE Confidence: 0.9983132

00:20:17.534 --> 00:20:18.274 hand side,

NOTE Confidence: 0.9061598

00:20:18.815 --> 00:20:19.934 what you can do is

NOTE Confidence: 0.9061598

00:20:19.934 --> 00:20:21.075 you first,

NOTE Confidence: 0.94382507

00:20:21.615 --> 00:20:22.734 learn an encoding of your

NOTE Confidence: 0.94382507

00:20:22.734 --> 00:20:24.595 data itself for high resolution

NOTE Confidence: 0.94382507

00:20:24.654 --> 00:20:26.400 images. You can learn a

NOTE Confidence: 0.94382507

00:20:26.400 --> 00:20:27.920 compression that compresses it to

NOTE Confidence: 0.94382507

00:20:27.920 --> 00:20:28.420 smaller

NOTE Confidence: 0.99971515

00:20:28.720 --> 00:20:29.220 images

NOTE Confidence: 0.9962982

00:20:29.680 --> 00:20:30.420 or smaller

NOTE Confidence: 0.87191826

00:20:30.720 --> 00:20:31.540 feature vectors,

NOTE Confidence: 0.9974254

00:20:32.000 --> 00:20:33.359 and then you can learn

NOTE Confidence: 0.9974254

00:20:33.359 --> 00:20:34.640 a diffusion model in that

NOTE Confidence: 0.9974254

00:20:34.640 --> 00:20:35.540 smaller space,

NOTE Confidence: 0.99922216

00:20:36.160 --> 00:20:38.320 and then generate everything in

NOTE Confidence: 0.99922216

00:20:38.320 --> 00:20:39.540 that compressed space.

NOTE Confidence: 0.99671453

00:20:39.965 --> 00:20:41.165 And from that compressed space,

NOTE Confidence: 0.99671453

00:20:41.165 --> 00:20:42.205 you can come back to

NOTE Confidence: 0.99671453

00:20:42.205 --> 00:20:43.025 the real data,

NOTE Confidence: 0.9977095

00:20:43.405 --> 00:20:44.365 from the model you first

NOTE Confidence: 0.9977095

00:20:44.365 --> 00:20:44.865 learned.

NOTE Confidence: 0.9946639

00:20:45.405 --> 00:20:46.285 On the right hand side

NOTE Confidence: 0.9946639

00:20:46.285 --> 00:20:47.885 is something called cascaded diffusion.

NOTE Confidence: 0.9946639

00:20:47.885 --> 00:20:48.605 So if you want to

NOTE Confidence: 0.9946639

00:20:48.605 --> 00:20:50.605 generate high resolution images, you

NOTE Confidence: 0.9946639

00:20:50.605 --> 00:20:51.805 generate things at a lower

NOTE Confidence: 0.9946639

00:20:51.805 --> 00:20:52.305 resolution

NOTE Confidence: 0.93544966

00:20:53.005 --> 00:20:54.720 and then use that as

NOTE Confidence: 0.93544966

00:20:54.720 --> 00:20:55.760 a seed for something that's

NOTE Confidence: 0.93544966

00:20:55.760 --> 00:20:56.820 at a higher resolution,

NOTE Confidence: 0.9875397

00:20:57.680 --> 00:20:59.280 and you expand it upwards

NOTE Confidence: 0.9875397

00:20:59.280 --> 00:20:59.680 up to,

NOTE Confidence: 0.99959856

00:21:01.280 --> 00:21:02.180 the full resolution.

NOTE Confidence: 0.98598504

00:21:05.760 --> 00:21:07.514 Okay. So I think I'm

NOTE Confidence: 0.98598504

00:21:07.514 --> 00:21:08.475 gonna run really out of

NOTE Confidence: 0.98598504

00:21:08.475 --> 00:21:09.754 time. So I'm gonna skip

NOTE Confidence: 0.98598504

00:21:09.754 --> 00:21:10.554 right to the end of

NOTE Confidence: 0.98598504

00:21:10.554 --> 00:21:11.514 my talk because I have

NOTE Confidence: 0.98598504

00:21:11.514 --> 00:21:12.955 only one minute, and I

NOTE Confidence: 0.98598504

00:21:12.955 --> 00:21:14.654 think this might be interesting.

NOTE Confidence: 0.9253347

00:21:17.674 --> 00:21:18.575 Okay. So

NOTE Confidence: 0.9992441

00:21:21.180 --> 00:21:22.220 I wanna talk a little

NOTE Confidence: 0.9992441

00:21:22.220 --> 00:21:23.820 bit quickly about how you

NOTE Confidence: 0.9992441

00:21:23.820 --> 00:21:25.260 might use this for predicting

NOTE Confidence: 0.9992441

00:21:25.260 --> 00:21:26.560 the structure of molecules.

NOTE Confidence: 0.9398308

00:21:27.020 --> 00:21:28.460 So as I mentioned with

NOTE Confidence: 0.9398308

00:21:28.460 --> 00:21:29.359 diffusion models,

NOTE Confidence: 0.94216084

00:21:29.740 --> 00:21:31.340 you have something you learned,

NOTE Confidence: 0.94216084

00:21:31.340 --> 00:21:32.619 which is the denoising model

NOTE Confidence: 0.94216084

00:21:32.619 --> 00:21:33.840 that takes in some noisy

NOTE Confidence: 0.94216084

00:21:34.060 --> 00:21:35.385 data and tries to clean

NOTE Confidence: 0.94216084

00:21:35.385 --> 00:21:36.045 it up.

NOTE Confidence: 0.9920825

00:21:36.345 --> 00:21:37.545 And you're given some features

NOTE Confidence: 0.9920825

00:21:37.545 --> 00:21:38.825 that describe the data as

NOTE Confidence: 0.9920825

00:21:38.825 --> 00:21:40.185 well, which help in the

NOTE Confidence: 0.9920825

00:21:40.185 --> 00:21:41.085 cleanup process.

NOTE Confidence: 0.94400984

00:21:41.545 --> 00:21:42.345 So you can do the

NOTE Confidence: 0.94400984

00:21:42.345 --> 00:21:43.085 same thing,

NOTE Confidence: 0.9553118

00:21:44.025 --> 00:21:44.525 with,

NOTE Confidence: 0.99466103

00:21:45.945 --> 00:21:46.685 with molecules.

NOTE Confidence: 0.9731785

00:21:47.305 --> 00:21:48.185 You can give it a

NOTE Confidence: 0.9731785

00:21:48.185 --> 00:21:50.125 mild representation of your molecule,

NOTE Confidence: 0.99033153

00:21:50.550 --> 00:21:52.150 which is for those that

NOTE Confidence: 0.99033153

00:21:52.150 --> 00:21:52.950 don't know, it's a way

NOTE Confidence: 0.99033153

00:21:52.950 --> 00:21:54.890 of, representing a compound

NOTE Confidence: 0.6345961

00:21:56.230 --> 00:21:56.730 sequence,

NOTE Confidence: 0.9814323

00:21:57.270 --> 00:21:58.650 that's used in in,

NOTE Confidence: 0.6910842

00:22:00.550 --> 00:22:01.930 chem informatics packages.

NOTE Confidence: 0.97638816

00:22:02.470 --> 00:22:03.734 And you can take the

NOTE Confidence: 0.97638816

00:22:03.734 --> 00:22:05.494 smiles feature and convert that

NOTE Confidence: 0.97638816

00:22:05.494 --> 00:22:07.275 to, features for a molecule,

NOTE Confidence: 0.97638816

00:22:07.494 --> 00:22:08.234 and then

NOTE Confidence: 0.97438097

00:22:08.535 --> 00:22:09.815 you have a denoising model

NOTE Confidence: 0.97438097

00:22:09.815 --> 00:22:10.855 that takes in the noisy

NOTE Confidence: 0.97438097

00:22:10.855 --> 00:22:11.815 coordinates for each of the

NOTE Confidence: 0.97438097

00:22:11.815 --> 00:22:13.015 molecule each of the atoms

NOTE Confidence: 0.97438097

00:22:13.015 --> 00:22:14.295 in the molecule and cleans

NOTE Confidence: 0.97438097

00:22:14.295 --> 00:22:15.575 up the coordinates. So, really,

NOTE Confidence: 0.97438097

00:22:15.575 --> 00:22:16.475 there's no

NOTE Confidence: 0.97416323

00:22:17.230 --> 00:22:19.070 no real information used. Don't

NOTE Confidence: 0.97416323

00:22:19.230 --> 00:22:20.509 you don't really bake in

NOTE Confidence: 0.97416323

00:22:20.509 --> 00:22:23.009 any sort of information about

NOTE Confidence: 0.97416323

00:22:23.070 --> 00:22:24.509 bond angles and any of

NOTE Confidence: 0.97416323

00:22:24.509 --> 00:22:26.029 any of that at all.

NOTE Confidence: 0.97416323

00:22:26.029 --> 00:22:27.309 You just basically train a

NOTE Confidence: 0.97416323

00:22:27.309 --> 00:22:28.830 model. You give you're given

NOTE Confidence: 0.97416323

00:22:28.830 --> 00:22:29.490 the compound,

NOTE Confidence: 0.98036623

00:22:30.065 --> 00:22:31.825 and you mutate you noise

NOTE Confidence: 0.98036623

00:22:31.825 --> 00:22:33.505 up its structure, and then

NOTE Confidence: 0.98036623

00:22:33.505 --> 00:22:34.785 you learn how to denoise

NOTE Confidence: 0.98036623

00:22:34.785 --> 00:22:35.365 the structure

NOTE Confidence: 0.9994246

00:22:35.744 --> 00:22:36.244 back.

NOTE Confidence: 0.9806048

00:22:36.945 --> 00:22:38.304 And so the way the

NOTE Confidence: 0.9806048

00:22:38.304 --> 00:22:39.585 the features are computed is

NOTE Confidence: 0.9806048

00:22:39.585 --> 00:22:40.645 you take a compound,

NOTE Confidence: 0.96830225

00:22:41.105 --> 00:22:43.430 and then you you label

NOTE Confidence: 0.96830225

00:22:43.430 --> 00:22:44.630 all the atoms. And from

NOTE Confidence: 0.96830225

00:22:44.630 --> 00:22:46.070 the atoms, you can compute

NOTE Confidence: 0.96830225

00:22:46.070 --> 00:22:46.730 a graph.

NOTE Confidence: 0.8680966

00:22:47.270 --> 00:22:48.970 This graph basically represents

NOTE Confidence: 0.99074894

00:22:49.270 --> 00:22:51.430 what atoms are connected to

NOTE Confidence: 0.99074894

00:22:51.430 --> 00:22:52.490 what other atoms.

NOTE Confidence: 0.924597

00:22:53.030 --> 00:22:53.530 And,

NOTE Confidence: 0.9690536

00:22:55.605 --> 00:22:56.405 I guess the detail is

NOTE Confidence: 0.9690536

00:22:56.405 --> 00:22:57.525 not important, but you can

NOTE Confidence: 0.9690536

00:22:57.525 --> 00:22:58.965 represent the structure of the,

NOTE Confidence: 0.9690536

00:22:59.125 --> 00:23:00.085 of a molecule in a

NOTE Confidence: 0.9690536

00:23:00.085 --> 00:23:02.005 graph. And then from that

NOTE Confidence: 0.9690536

00:23:02.005 --> 00:23:04.405 graph, you can actually compute

NOTE Confidence: 0.9690536

00:23:04.405 --> 00:23:05.785 something called a graph Laplacian,

NOTE Confidence: 0.9690536

00:23:05.845 --> 00:23:07.045 which allows you to compute

NOTE Confidence: 0.9690536

00:23:07.045 --> 00:23:08.244 features for each of the

NOTE Confidence: 0.9690536

00:23:08.244 --> 00:23:08.744 atoms

NOTE Confidence: 0.9898975

00:23:09.205 --> 00:23:10.025 in the graph.

NOTE Confidence: 0.97182846

00:23:10.619 --> 00:23:11.900 And you can then add

NOTE Confidence: 0.97182846

00:23:11.900 --> 00:23:13.420 on some descriptors for each

NOTE Confidence: 0.97182846

00:23:13.420 --> 00:23:14.700 of the atoms in the

NOTE Confidence: 0.97182846

00:23:14.700 --> 00:23:16.700 graph, things like the obvious

NOTE Confidence: 0.97182846

00:23:16.700 --> 00:23:18.400 but very basic things like,

NOTE Confidence: 0.97416455

00:23:19.020 --> 00:23:20.540 the atom type, the degree,

NOTE Confidence: 0.97416455

00:23:20.540 --> 00:23:21.820 the valence, and and so

NOTE Confidence: 0.97416455

00:23:21.820 --> 00:23:22.320 on.

NOTE Confidence: 0.9922282

00:23:22.905 --> 00:23:23.785 And then you just run

NOTE Confidence: 0.9922282

00:23:23.785 --> 00:23:25.725 a diffusion model. I won't

NOTE Confidence: 0.9922282

00:23:25.865 --> 00:23:26.825 go into the details of

NOTE Confidence: 0.9922282

00:23:26.825 --> 00:23:28.205 it, but, essentially,

NOTE Confidence: 0.9783787

00:23:28.665 --> 00:23:30.025 you get features for the

NOTE Confidence: 0.9783787

00:23:30.025 --> 00:23:31.785 atoms based on connectivity, some

NOTE Confidence: 0.9783787

00:23:31.785 --> 00:23:33.625 extra descriptors, and it's three

NOTE Confidence: 0.9783787

00:23:33.625 --> 00:23:34.605 d noisy coordinates.

NOTE Confidence: 0.9574274

00:23:34.984 --> 00:23:35.945 And then you just learn

NOTE Confidence: 0.9574274

00:23:35.945 --> 00:23:37.625 to predict the noisy coordinates

NOTE Confidence: 0.9574274

00:23:37.625 --> 00:23:39.040 at the next type. And

NOTE Confidence: 0.9574274

00:23:39.040 --> 00:23:40.880 you can then, during inference,

NOTE Confidence: 0.9574274

00:23:40.880 --> 00:23:42.240 just run that model. You

NOTE Confidence: 0.9574274

00:23:42.240 --> 00:23:44.020 start with the structure, some

NOTE Confidence: 0.9574274

00:23:44.320 --> 00:23:46.180 some random three d assignments

NOTE Confidence: 0.9574274

00:23:46.400 --> 00:23:47.140 for the,

NOTE Confidence: 0.97119284

00:23:47.440 --> 00:23:48.480 positions of each of the

NOTE Confidence: 0.97119284

00:23:48.480 --> 00:23:48.980 atom.

NOTE Confidence: 0.9996273

00:23:49.280 --> 00:23:50.580 And once you run them,

NOTE Confidence: 0.95090276

00:23:50.975 --> 00:23:52.115 the model starts,

NOTE Confidence: 0.9779156

00:23:52.895 --> 00:23:54.015 to clean up the three

NOTE Confidence: 0.9779156

00:23:54.015 --> 00:23:54.755 d positions.

NOTE Confidence: 0.99537826

00:23:55.215 --> 00:23:56.195 And at the end,

NOTE Confidence: 0.9145776

00:23:56.815 --> 00:23:58.255 it'll give you a a

NOTE Confidence: 0.9145776

00:23:58.255 --> 00:23:59.715 full structure for the molecule.

NOTE Confidence: 0.82697946

00:24:01.295 --> 00:24:02.595 Same for another one.

NOTE Confidence: 0.9316164

00:24:10.230 --> 00:24:12.010 And, yeah, the the TLDR

NOTE Confidence: 0.9316164

00:24:12.149 --> 00:24:13.190 is this works quite well.

NOTE Confidence: 0.9316164

00:24:13.190 --> 00:24:13.990 We got state of the

NOTE Confidence: 0.9316164

00:24:13.990 --> 00:24:15.770 art results on predicting structures,

NOTE Confidence: 0.9873488

00:24:16.470 --> 00:24:18.070 compared to prior works, although

NOTE Confidence: 0.9873488

00:24:18.070 --> 00:24:19.925 we actually didn't use any

NOTE Confidence: 0.9873488

00:24:19.925 --> 00:24:20.984 information about

NOTE Confidence: 0.9829673

00:24:21.365 --> 00:24:22.565 chemistry from it. All of

NOTE Confidence: 0.9829673

00:24:22.565 --> 00:24:24.244 that was just learned, by

NOTE Confidence: 0.9829673

00:24:24.244 --> 00:24:25.625 the model on its own.

NOTE Confidence: 0.99243706

00:24:26.725 --> 00:24:27.705 So to conclude,

NOTE Confidence: 0.99019647

00:24:28.565 --> 00:24:29.305 I hope,

NOTE Confidence: 0.9520272

00:24:29.925 --> 00:24:31.380 you'll take away that there's

NOTE Confidence: 0.9520272

00:24:31.540 --> 00:24:33.140 interesting ways to embed all

NOTE Confidence: 0.9520272

00:24:33.140 --> 00:24:34.260 kinds of data into a

NOTE Confidence: 0.9520272

00:24:34.260 --> 00:24:35.780 vector representation that you could

NOTE Confidence: 0.9520272

00:24:35.780 --> 00:24:37.160 use for your statistical models.

NOTE Confidence: 0.98749894

00:24:37.780 --> 00:24:39.060 Generative models will allow you

NOTE Confidence: 0.98749894

00:24:39.060 --> 00:24:40.340 to build a model where

NOTE Confidence: 0.98749894

00:24:40.340 --> 00:24:41.540 you can generate new data,

NOTE Confidence: 0.98749894

00:24:41.540 --> 00:24:42.660 which can be used for

NOTE Confidence: 0.98749894

00:24:42.660 --> 00:24:44.680 infilling or even finding correlations

NOTE Confidence: 0.98749894

00:24:44.740 --> 00:24:45.880 that you may not have

NOTE Confidence: 0.98749894

00:24:46.100 --> 00:24:46.600 expected.

NOTE Confidence: 0.77608454

00:24:47.825 --> 00:24:48.325 And,

NOTE Confidence: 0.9888279

00:24:48.945 --> 00:24:49.765 yeah, we're

NOTE Confidence: 0.87806916

00:24:51.665 --> 00:24:52.865 a group in Apple just

NOTE Confidence: 0.87806916

00:24:52.865 --> 00:24:54.705 doing fundamental machine learning research.

NOTE Confidence: 0.87806916

00:24:54.705 --> 00:24:55.205 So

NOTE Confidence: 0.66720533

00:24:55.905 --> 00:24:56.645 with that.