Investor Day

Company Participants

- Andrej Karpathy, Senior Director of AI
- Elon R. Musk, Co-Founder, Chief Executive Officer and Product Architect
- Pete Bannon, Vice President of Hardware Engineering
- Stuart Bowers, Vice President of Hardware Engineering
- Unidentified Speaker

Other Participants

- Adam Jonas, Analyst, Morgan Stanley
- Analyst
- Colin Langan, Analyst, UBS Investment Bank
- Colin Rusch, Analyst, Oppenheimer
- Jed Dorsheimer, Analyst, Canaccord Genuity
- Matt Joyce, Analyst, Loup Ventures
- Pradeep Ramani, Analyst, UBS Investment Bank
- Tripatinder Chowdhry, Analyst, Global Equities Research
- Unidentified Participant

Presentation

Unidentified Speaker

Hi, everyone. I'm sorry for being late. Welcome to our very first Analyst Day for Autonomy.

I really hope that this is something we can do a little bit more regularly now to keep you posted about the development we're doing with regards to autonomous driving. About three months ago we were getting prepped up for our Q4 earnings call with Elon and quite a few other executives, and one of the things that I told the group is that from all the conversations that I keep having with investors on a regular basis, the biggest gap that I see with what I see inside the Company and what the outside perception is, is our ability of autonomous driving. And it kind of makes sense because for the past couple of years we've been really talking about Model 3 ramp and a lot of the debate has revolved around Model 3. But in reality, a lot of things have been happening in the background. We've been working on the new full-self driving chip; we've had a complete overhaul of our neural net for vision recognition, et cetera.

So now that we finally started to produce our full self-driving computer, we thought it's a good idea to just open the veil, invite everyone in and talk about everything that we've been doing for the past two years. So about three years ago we wanted to use -- we wanted to find the best possible chip for full autonomy, and we found out that there's no chip that's been designed from ground up for neural nets. So we invited my colleague Pete Bannon, the VP of Silicon Engineering, to design such a chip for us. He's got about 35 years of experience of building chips and designing chips. About 12 of those years where for a company called PA Semi, which was later acquired by Apple. So he worked on dozens of different architectures and designs and he was the lead designer I think for Apple iPhone 5 just before joining Tesla.

And he's going to be joined on the stage by Elon Musk. Thank you.

Elon R. Musk {BIO 1954518 <GO>}

I was going to introduce Pete but Martin has done so. He's just the best chip and system architect that I know in the world and that's an honor to have you and your team at Tesla. And take it away just tell them about the incredible work that you and your team have done.

Pete Bannon {BIO 20590065 <GO>}

Thanks, Elon.

It's a pleasure to be here this morning and a real treat, really, to tell you about all the work that my colleagues and I have been doing here at Tesla for the last three years. I think we'll tell you a little bit about how the whole thing got started and then I'll introduce you to the full self-driving computer and tell you a little bit about how it works. We'll dive into the chip itself and go through some of those details. I'll describe how the custom neural network accelerator that we designed works. And then I'll show you some results. And hopefully, I'll still be awake by then.

I was hired in February of 2016. I asked Elon if he was willing to spend all the money it takes to do full custom system design, and he said, well, are we going to win, and I said well, yeah, of course. So he said I'm in. And so that got us started. We hired a bunch of people and started thinking about what a full -- what a custom designed chip for full autonomy would look like.

We spent 18 months doing the design and in August of 2017 we released the design for manufacturing. We got it back in December. It powered up and it actually worked very, very well on the first try. We made a few changes and released a B-Zero Rev in April of 2018. In July of 2018 the chip was qualified and we started full production -- production quality parts. In December of 2018 we had the autonomous driving stack running on the new hardware and we're able to start retrofitting employee cars and testing the hardware and software out in the real world.

Just last March, we started shipping the new computer in the Model S and X and just earlier in April we started production in Model 3. So this whole program, from the hiring of the first few employees to having it in full production in all three of our cars, is just a little over three years and it's probably the fastest system development program I've ever been associated with. And it really speaks a lot to the advantages of having a tremendous amount of vertical integration to allow you to do concurrent engineering and speed up deployment.

In terms of goals, we were totally focused exclusively on Tesla's requirements, and that makes life a lot easier. If you have one and only one customer, you don't have to worry about anything else. One of those goals was to keep the power under 100 watts so we could retrofit the new machine into the existing cars. We also wanted to lower park costs so we could enable full redundancy for safety. At the time we had a thumb in the wind estimate that it would take at least 50 trillion operations a second of neural network performance to drive a car, and so we wanted to get at least that much and really as much as we possibly could.

Batch size is how many items you operate on at the same time. So for example Google's TPU-1 has a batch size of 256. Then you have to wait around until you have 256 things to process before you can get started. We didn't want to do that, so we designed a machine with a batch size of one. So as soon as an image shows up, we process it immediately to minimize latency, which maximises safety. We needed a GPU to run some post-processing. At the time we were doing quite a lot of that, but we speculated that over time the amount of post-processing on the GPU would decline as the neural networks got better and better. And that has actually come to pass. So we took a risk by putting a fairly modest GPU in the design, as you'll see, and that turned out to be a good bet. Security is super-important. If you don't have a secure car, you can't have a safe car. So there's a lot of focus on security and then of course safety.

In terms of actually doing the chip design, as Elon alluded earlier, there was really no ground-up neural network accelerator in existence in 2016. Everybody out there was adding instructions to their CPU or GPU or DSP to make it better for inference but nobody was really just doing it natively. So we set out to do that ourselves. And then for other components on the chip, we purchased industry standard IP for CPUs and GPUs that allowed us to minimize the design time and also the risk to the program.

Another thing that was a little unexpected when I first arrived was our ability to leverage existing teams at Tesla. Tesla had a wonderful power supply design team, signal integrity analysis, package design, system software, firmware, board designs and a really good system validation program that we were able to take advantage of to accelerate this program.

Here's what it looks like. Over there on the right you see all connectors for the video that comes in from the eight cameras that are in the car. You can see the two self-driving computers in the middle of the board. And then on the left is the power supply and some control connections. And so I really love it when a solution is boiled down to its barest elements. You have video, computing and power, and it's straightforward and simple.

Here's the original hardware 2.5 enclosure that the computer went into. And we've been shipping for the last two years. Here's the new design for the FSD computer. It's basically the same, and that of course is driven by the constraints of having a retrofit program for the cars. I'd like to point out that this is actually a pretty small computer. It fits behind the glove box, between the glove box and the firewall in the car. It does not take up half of your trunk.

As I said earlier, there's two fully independent computers on the board. You can see them there highlighted in blue and green. To either side of the large SOC [ph], you can see the DRAM chips that we use for storage and then below left you see the flash chips that represent the file system. So these are two independent computers that boot up and run their own operating system...

Elon R. Musk {BIO 1954518 <GO>}

Yeah, if I can add something. The general principle here is that any part of this could fail, and the car will keep driving. So you could have cameras fail; you could have power circuits fail; you could have one of the Tesla full self-driving computer chips fail; car keeps driving. The probability of the -- of this computer failing is substantially lower than somebody losing consciousness. That's the key metric, at least in order of magnitude.

Pete Bannon {BIO 20590065 <GO>}

Yeah.

So one of the things that we -- an additional thing we do to keep the machine going is to have redundant power supplies in the car. So one machine is running on one power supply and the other one's on the other. The cameras are the same. So half of the cameras run on the blue power supply, the other half run on the green power supply, and both chips receive all of the video and process it independently.

So in terms of driving the car, the basic sequence is collect lots of information from the world around you. Not only do we have cameras, we also have radar, GPS, maps, the IMUs, ultrasonic sensors around the car; we have wheel tech [ph] steering angle. We know what the acceleration and deceleration of the car supposed to be. All of that gets integrated together to form a plan. And once we have a plan, the two machines exchange their independent version of the plan to make sure it's the same. And assuming that we agree, we then act and drive the car.

Now, once you've driven the car with some new controller, you of course want to validate it, so we validate that what we transmitted was what we intend to transmit to the other actuators in the car. And then you can use the sensor suite to make sure that it happens. So if you ask the car to accelerate or brake or steer right or left you can look at the accelerometers and make sure that you are in fact doing that. So there's a tremendous amount of redundancy and overlap in both our data acquisition and our data monitoring capabilities here.

Moving on to talk about the full self-driving chip a little bit. It's packaged in a 37.5 millimeter BGA with 1,600 balls. Most of those are used for power and ground, but plenty for signal as well. If you take the lid off, it looks like this. You can see the package substrate and you can see the dye sitting in the center there. If you take the dye off and flip it over, it looks like this. There's 13,000 C4 bumps scattered across the top of the die and then underneath that are 12 metal layers, and if you -- which is obscuring all the details of the design. So if you strip that off, it looks like this. This is a 14 nanometer FinFET CMOS process. It's 260 millimeters in size, which is a modest sized die. So for comparison, a typical cellphone chip is about 100 millimeters square, so quite a bit bigger than that but a high-end GPU would be more like 600 to 800 millimeter square. So we're sort of in the middle, I would call it the sweet spot. It's a comfortable size to build. There's 250 million logic gates on there and a total of 6 billion transistors, which even though I work on this all the time, that's mind boggling to me. The chip is manufactured and tested to AEC-Q100 standards, which is a standard automotive criteria.

Next I'd like to just walk around the chip and explain all the different pieces to it and I'm sort of going to go in the order that a pixel coming in from the camera would visit all the different pieces. So up there in the top left, you can see the camera serial interface. We can ingest 2.5 billion pixels per second, which is more than enough to cover all the sensors that we know about. We have an on-chip network that distributes data from the memory system, so the pixels would travel across the network to the memory controllers on the right and left edges of the chip. We use industry standard LPDD, they have four memory running at 4266 gigabits per second, which gives us a peak bandwidth of 68 gigabytes a second, which is a pretty healthy bandwidth. But again this is not like ridiculous, so we're sort of trying to stay in the comfortable sweet spot for cost reasons.

The image signal processor has a 24-bit internal pipeline that allows us to take full advantage of the HDR sensors that we have around the car. It does advanced tone mapping, which helps to bring out details in the shadows and then it has advanced noise reduction, which just improves your overall quality of the images that we're using in the neural network.

The neural network accelerator itself, there's two of them on the chip. They each have 32 megabytes of SRAM to hold temporary results and minimize the amount of data that we have to transmit on and off the chip, which helps reduce power. Each array has a 96/96 multiply add array with in-place accumulation, which allows us to do almost 10,000 multiply adds per cycle. Has ReLU hardware, dedicated pooling hardware and each of these deliver 306 -- excuse me, each one delivers 36 trillion operations per second and they operate at 2 gigahertz.

The two of them together on a die deliver 72 trillion operations a second. So we've exceeded our goal of 50 teraops by a fair bit. There's also a video encoder, we encode video and use it in a variety of places in the car including the backup camera display. There's optionally a user feature for a dash cam and also for a clip logging data to the cloud, which Stuart and Andrej will talk about more later. There's a GPU on the chip, it's modest performance, it has a support for both 32 and 16-bit floating point and then we have 12 A72, 64-bit CPUs for general purpose processing. They operate at 2.2 gigahertz and this represents about 2.5 times the performance available in the current solution. There's a safety system that contains two CPUs that operate in lock step. This system is the final arbiter of whether it's safe to actually drive the actuators in the car. So this is where the two plans come together and we decide whether it's safe or not to move forward.

And lastly, there's the safety system and then basically the job of the safety system is to ensure that this chip only runs software that's been cryptographically signed by Tesla. If it's not been signed by Tesla, then the chip does not operate.

Now I've told you a lot of the different performance numbers and I thought it'd be helpful maybe to put it into perspective a little bit. So throughout this talk, I'm going to talk about a neural network from our narrow camera. It uses 35 billion operations, 35 gigaops and if we use all 12 CPUs to process that network, we could do 1.5 frames per second, which is super slow, not nearly adequate to drive a car. If we use the 600 GigaFLOPS GPU, the same network we'd get 17 frames per second, which is still not good enough to drive a car with eight cameras. The neural network accelerators on the chip can deliver 2,100 frames per second and you can see from the scaling as we moved along, that the amount of computing in the CPU and GPU are basically insignificant to what's available in the neural network accelerator, it really is night and day.

So moving on to talk about the neural network accelerator, I'm just going to stop for some water. On the left, there's a cartoon of a neural network just to give you an idea what's going on. The data comes in at the top and visits each of the boxes and the data flows along the arrows to the different boxes. The boxes are typically convolutions or de-convolutions with ReLUs, the green boxes are pooling layers and the important thing about this is that, the data produced by one box is then consumed by the next box and then you don't need it anymore, you can throw it away. So all of that temporary data that gets created and destroyed as you flow through the network. there's no need to store that off-chip and DRAM. So we keep all that data in SRAM and I'll explain why that is super important in a few minutes.

If you look over on the right side of this, you can see that in this network, that's 35 billion operations, almost all of them are convolution, which is based on dot products. The rest are de-convolution, also based on dot product and then ReLU and pooling, which are relatively simple operations. So if you were designing some hardware, you'd clearly target doing dot products, which are based on multiply add and really kill that. But imagine that you sped it up by a factor of 10,000, so 100% of it sudden turns into 0.1% -- 0.01% and suddenly the ReLU and pooling operations are going to be quite significant. So our hardware design includes dedicated resources for processing ReLU and pooling as well.

Now this chip is operating in a thermally constrained environment. So we had to be very careful about how we burn that power. We want to maximize the amount of arithmetic we can do. So we picked integer add, it's nine times less energy than a corresponding floating point add and we picked 8-bit by 8-bit integer multiplied, which is significantly less power than the other multiply operations and there's probably enough accuracy to get good results. In terms of memory, we chose to use SRAM as much as possible and you can see there that going off chip to DRAM is approximately 100 times more expensive in terms of energy consumption than using local SRAM. So clearly, we want to use local SRAM as much as possible.

In terms of control, this is data that was published in a paper by Mark Horowitz at ISSCC, where he sort of critiqued how much power it takes to execute a single instruction on a regular integer CPU and you can see that the ADD operation is only 0.15% of the total power, all the rest of the power is control overhead and bookkeeping. So in our design, we start to basically get rid of all that as much as possible, because what we're really interested in is arithmetic.

So here's the design that we've finished, you can see that it's dominated by the 32 megabytes SRAM, those big banks on the left and right and in the center bottom and then all the computing is done in the upper middle. Every single clock will read 256 bytes of activation data out of the SRAM array, 128 bytes of weight data out of the SRAM array and we combine it in a 96/96 mul/add array, which performs 9,000 multiply adds per clock, at 2 gigahertz, that's a total of 36.8 teraops.

Now when we're done with the dot product, we unload the engine, so that we shift the data out across the dedicated ReLU unit, optionally across a pooling unit and then finally into a write buffer, where all the results get aggregated up and then we write out 128 bytes per cycle back into the SRAM and this whole thing cycles along all the time continuously. So we're doing dot products while we're unloading previous results, doing pooling and writing back into memory. If you add it all up to your [ph] hertz, you need 1 terabyte per second of SRAM bandwidth to support all that work and so the hardware supplies that. So 1 terabyte per second of bandwidth per engine, there's two on the chip, 2 terabytes per second.

The chip has a -- the accelerator has a relatively small instruction set, we have a DMA read operation to bring data in from memory, we have a DMA write operation to push results back out to memory. We have three dot product based instructions, convolution, de-convolution and inner product and then two relatively simple, scale is a one input/one output operation, and Eltwise is two inputs and one output and then, of course, stop when you're done. We had to develop a neural network compiler for this. So we take the neural network that's been trained by our Vision team as it would be deployed in the older cars and when you take that and compile it for use on the new accelerator, the compiler does layer fusion, which allows us to maximize the computing each rime we read data out of the SRAM and put it back. That's been trained by our vision team, as it would be deployed in the older cars and we take that and compile it for use on the new accelerator. The compiler does layer fusion which allows us to maximize the computing, each time we read data out of the SRAM and put it back. It also does some smoothing, so that the demand is on the memory system aren't too lumpy. And then we also do channel padding to reduce bank conflicts and we do bank aware SRAM allocation. And this is a case where we could have put more hardware in the design to handle bank conflicts. But by pushing it into software, we save hardware and power at the cost of some software complexity.

We also automatically insert DMAs into the graph, so that data arrives just in time for computing without having to stall the machine. And then at the end, we generate all the code, we generate all the weight data, we compress it and we add a CRC checksum for reliability. To run a program, all the neural network descriptions are programs that are loaded into SRAM at the start, and then they sit there ready to go all the time. So to run a network, you have to program the address at the input buffer which presumably is a new image that just arrived from a camera. You set the output buffer address, you set the pointer to the network weights and then you set go. And then the machine goes off and we'll sequence through the entire neural network all by itself, usually running for 1 million or 2 million cycles and then when it's done you get an interrupt and can post process the results.

So moving on to results, we had a goal to stay under 100 watts. This is measured data from cars driving around running the full autopilot stack. We're dissipating 72 watts, which is a little bit more power than the previous design, but with the dramatic improvement in performance it's still a pretty good answer. Of that 72 watts, about 15 watts is being consumed running the neural networks. In terms of cost, the silicon cost of this solution is about 80% of the what we were paying before. So we are saving money by switching to this solution. And in terms of performance, we took the narrow camera neural network, which I've been talking about, that has 35 billion operations in it. We ran it on the old hardware in a loop as quick as possible and we delivered a 110 frames per second. We took the same data, the same network, compiled it for hardware for the new FSD computer and using all four accelerators, we can get 23,00 frames per second processed. So a factor of 21.

Elon R. Musk {BIO 1954518 <GO>}

I think this is perhaps the most significant slide. It's night and day.

Pete Bannon {BIO 20590065 <GO>}

I've never worked on a project, where the performance increase was more than three. So this was pretty funny. If you compare it to say NVIDIA's DRIVE Xavier solution, a single chip delivers 21 TeraOPS. Our full self driving computer with two chips is a 144 TeraOPS. So to conclude, I think we have created a design that delivers outstanding performance, a 144 TeraOPS for neural network processing, it has outstanding power performance, we managed to jam all of that performance into the thermal budget that we had, that enables a fully redundant computing solution, it has a modest cost. And really the important thing is that this FSD computer will enable a new level of safety and autonomy in Tesla's vehicles without impacting their cost or range. Something that I think we're all looking forward too.

Elon R. Musk {BIO 1954518 <GO>}

I think, why don't we do a Q&A after each segments? So if you have questions about the hardware, they can ask right now. The reason I ask Pete to do just a detailed -- far more detail than perhaps most people would appreciate -- dive into the Tesla full-self driving computer is because, at first it seems improbable, how could it be that Tesla, who has never designed a chip before would design the best chip in the world? But that is objectively what has occurred, not best by a small margin, best by a huge margin. It's in the cars right now. All Tesla is being produced right now have this computer.

We switched over from the NVIDIA solution for S and X about a month ago. And I first switched over model 3 about 10 days ago. All cars being produced have the -- have all the hardware necessary, compute and otherwise for full self-driving. I'll say that again. All Tesla cars being produced right now have everything necessary for full self-driving. All you need to do is improve the software. And later today, you will drive the cars with the development version of the improved software and you will see for yourselves. Question for Pete?

Pete Bannon {BIO 20590065 <GO>}

Yeah.

Questions And Answers

Q - Tripatinder Chowdhry {BIO 5306842 <GO>}

Very impressive. I think -- Trip Chowdhry, Global Equities Research. Very, very impressive in every shape and form. I was wondering like I took some notes, you are using activation function, ReLU, the rectified linear unit. But if you think about the deep neural network, it has multiple layers and some algorithms may use different activation functions for different hidden layers like Softmax or tanh or 10 each. Do you have flexibility for incorporating different activation functions rather than LU in your platform, then I have a follow-up.

A - Pete Bannon {BIO 20590065 <GO>}

Yes, we do. We have implementations of tanh and sigmoid, for example.

Q - Tripatinder Chowdhry {BIO 5306842 <GO>}

Beautiful. One last question. Like in the nanometers you will mention 40 nanometers, as -- I was wondering what didn't make sense to come little lower and maybe 10 nanometers, two years down or maybe seven?

A - Pete Bannon {BIO 20590065 <GO>}

At the time I started the design not all of the IP that we wanted to purchase was available in 10 nanometer, so we had finished the design in 14.

A - Elon R. Musk {BIO 1954518 <GO>}

It's maybe worth pointing out that we finished this design like maybe 1.5 years 2 years ago and began design of the next generation. We're not talking about the next generation today, but we're about halfway through it. That will -- all the things that are obvious for next generation chip we're doing.

A - Pete Bannon {BIO 20590065 <GO>}

Yeah.

Q - Tripatinder Chowdhry {BIO 5306842 <GO>}

Hi. You talked about the software, is the piece now you did a great job, I was blown away. Understood 10% of what you said but I trust that it's in good hands.

A - Pete Bannon {BIO 20590065 <GO>}

Thanks.

Q - Tripatinder Chowdhry {BIO 5306842 <GO>}

So, it feels like you've got the hardware pieces done and that was really hard to do. And now you had to do the software piece. Now maybe that's outside of your expertise. How should we think about that software piece?

A - Elon R. Musk {BIO 1954518 <GO>}

What can ask for better introduction to Andrej and Stuart. Are they any questions for the chip part before -- the next part of the presentation is neural nets and software.

Q - Tripatinder Chowdhry {BIO 5306842 <GO>}

So maybe on the chip side, the last slide was 144 trillions of operations per second versus was it in video at 21.

A - Pete Bannon {BIO 20590065 <GO>}

That's right.

Q - Tripatinder Chowdhry {BIO 5306842 <GO>}

And maybe can you just contextualize that for a finance person, why that's so significant that gap? Thank you.

A - Pete Bannon {BIO 20590065 <GO>}

Well, I mean it's a factor of seven in performance Delta, so that means you can do seven times as many frames, you can run neural networks that are seven times larger and more sophisticated. So it's a very big currency that you can spend on lots of interesting things to make the car better.

Q - Tripatinder Chowdhry {BIO 5306842 <GO>}

I think that a Xavier power usage is higher than ours? Xavier powers is higher than ours, I think?

Q - Unidentified Participant

Or comparable?

A - Pete Bannon {BIO 20590065 <GO>}

I don't know that. I believe it's like that the -- at the best of my knowledge the power requirements would increase at least to the same degree of factor seven and costs would also increase by a factor of seven.

Q - Tripatinder Chowdhry {BIO 5306842 <GO>}

Great. So yeah...

A - Elon R. Musk {BIO 1954518 <GO>}

Power is a real problem because it also reduces range. So it has the pounds [ph] for power is very high. And then you have to get rid of that power by -- the thermal problem becomes really significant because you got to get rid of all that power.

Q - Tripatinder Chowdhry {BIO 5306842 <GO>}

Okay. Thank you very much. I think we have a lot of -- quite a...

A - Elon R. Musk {BIO 1954518 <GO>}

Just ask the questions. If you guys don't mind the day running a bit long? We're going to do that the drive demos afterwards. So if you got, if anybody needs to pop out and do the drive demos, a little sooner, you're welcome to do that. Do you want to make sure we answer your questions.

A - Unidentified Speaker

Yeah.

Q - Pradeep Ramani {BIO 19683324 <GO>}

Pradeep Ramani from UBS. Intel and AMD to some extent have started moving towards a chiplet-based architecture. I did not notice a chiplet based design here. Do you think that looking forward that would be something that might be of interest to you guys from an architecture standpoint?

A - Pete Bannon {BIO 20590065 <GO>}

A chiplet based architecture?

Q - Pradeep Ramani {BIO 19683324 <GO>}

Yes.

A - Elon R. Musk {BIO 1954518 <GO>}

We're not currently considering anything like that. I think that's mostly useful when you need to use different styles of technology. So if you want to integrate silicon germanium or DRAM technology on the same silicon substrate, that gets pretty interesting. But until the die size gets obnoxious, I wouldn't go there. to clear, the strategy here -- this started basically three years ago was designed to build a computer that is fully optimized and aiming for full self-driving. Then write-software that is designed to work specifically on that computer and get the most out of that computer. So you have tailored-hardware that is a master of one trade self-driving and NVIDIA is a great company but they have many customers. And so -- as they apply their resources, they need to do a generalized solution.

We care about one thing self-driving. So, it was designed to do that incredibly well. The software is also designed to run on that hardware incredibly well and the combination of the software and the hardware, I think is unbeatable.

Q - Unidentified Participant

Hi. The chip is designed to process video input, in case you use -- let's say lidar. Would it be able to process that as well or is it primarily for video?

A - Unidentified Speaker

What we're going to explain to you today is that LIDAR is a fool's errand. And anyone relying on LIDAR is doomed. Expensive sensors that are unnecessary. It's like having a whole bunch of expensive appendices, like one appendix is bad or now a whole bunch of them. That's ridiculous. You'll see.

Q - Unidentified Participant

Hi, Oh there's a gentlemen Mike.

A - Unidentified Speaker

Hi.

Q - Unidentified Participant

So just two questions on -- just on the power consumption, is there a way to maybe give us like a rule of thumb on every watt is reduce it range by certain percent or a certain amount. Just so we can get a sense of how much of --.

A - Unidentified Speaker

Model three, the target consumption is 250 watts per mile. It depends on the nature of the driving as to how many miles that affects in city would have a much bigger effect than on highway. So you know, if you're driving for an hour in a city and you had a solution hypothetically that, you know, it was a kilowatt, you'd lose for miles on a Model Three.

So if you're only going, say 12 miles an hour, then that's like there wouldn't be a 25% impact on a range and city. It's basically, the power of the system, it has a massive impact on city range which is where we think most of the robotaxi market will be. So power is extremely important.

Q - Unidentified Participant

(Technical Difficulty)

A - Unidentified Speaker

I'm sorry I couldn't hear you.

Q - Unidentified Participant

Sorry. Thank you. What's the primary design objective of the next generation chip?

A - Unidentified Speaker

When we don't talk too much about the next generation chip, but it's -- it'll be at least let's say three times better than the current system.

Q - Unidentified Participant

Go ahead.

A - Unidentified Speaker

About two years away.

Q - Unidentified Participant

To develop this chip, is the chip mean, you don't manufacture the chip. You contract that out and how much cost reduction does that save in the overall vehicle costs.

A - Unidentified Speaker

The 20% cost reduction, I cited was the piece cost per vehicle reduction. Not that wasn't in development costs, I was just the actual.

Q - Unidentified Participant

No, I'm saying that like if I'm manufacturing these in mass, is this saving money in doing it yourself?

A - Unidentified Speaker

Yes. A little bit.

I mean most chips are made -- most people don't make chips with their own fab, it's pretty unusual.

Q - Unidentified Participant

Yeah, I think that you don't see any supply issues without getting the chip-mass produced.

A - Unidentified Speaker

No.

Q - Unidentified Participant

The cost saving pays for the development. I mean, the basic strategy going to Elon was we're going to build this chip, it's going to reduce the cost and Elon's at at times a million cars a year deal.

A - Unidentified Speaker

That's correct.

Q - Unidentified Participant

It's really have specific questions we can ask them others. There will be a Q&A opportunity after Andrej talks and Stuart talks, so there will be two other Q&A opportunities. This is very specific, then.

A - Unidentified Speaker

I'll say I'll be here all afternoon.

Q - Unidentified Participant

Yeah. And second we'll be here at the end as well. So go ahead.

Yeah thanks. That die photo you had, there's the neural processor takes up quite a bit of the die, I'm curious is that your own design or is there some external IP there.

A - Unidentified Speaker

Yes. That was the custom design from by Tesla.

Q - Unidentified Participant

Okay. And then the -- I guess the follow-on, would be there's probably a fair amount of opportunity to reduce that footprint as you tweak the design?

A - Unidentified Speaker

It's actually quite dense. So in terms of reducing it, I don't think so. It'll will greatly enhance the functional capabilities in the next generation.

Q - Unidentified Participant

Okay. And then last question, can you share where you're fabbing this part?

A - Unidentified Speaker

Where we found that?

Q - Unidentified Participant

Samsung.

A - Unidentified Speaker

Samsung.

Yes. Austin, Texas.

Q - Unidentified Participant

Thank you.

A - Unidentified Speaker

There is one other back.

Q - Unidentified Participant

Hi, Grant (inaudible). I'm just curious how defensible your chip technologies and design is from it from a IP point of view and hoping that you won't be offering a lot of the IP to the outside for free. Thanks.

A - Pete Bannon {BIO 20590065 <GO>}

We have filed on the order of a dozen patents on this technology. Fundamentally, it's linear algebra, which I don't think you can patent. I'm not sure.

But I think, if somebody started today and they're really good they might have something like what we have right now in three years but in two years we'll have time something three times better.

Q - Unidentified Participant

Talking about the intellectual property protection, you have the best intellectual property and some people just steal it for the fun of it. I'm just wondering, if we look at few interactions with Aurora that companies took, the industry believe they stole your intellectual property. I think the key ingredient that you need to protect is the rates that associate to various parameters.

Do you think a chip can do something to prevent anybody, maybe encrypt all the rates so that even you don't know what the rates are at the chip level. So that your intellectual property remains inside it and nobody knows about it and nobody can just steal it.

A - Pete Bannon {BIO 20590065 <GO>}

I'd like to meet the person that could do that because they would I would hire them in a heartbeat. Yeah it's a real hard problem. Yeah, I enjoyed it. I mean we're doing encrypto, the -- it's a hard chip to crack. So if they can crack it, very good. So again it. And then also figure out the software and the neural net system and everything else, they can design it from scratch like that's all.

A - Elon R. Musk {BIO 1954518 <GO>}

It's our intention to prevent people from stealing all that stuff. I mean, if they do we hope it at least takes a long time.

A - Pete Bannon {BIO 20590065 <GO>}

It will definitely take them a long time. Yeah. I mean I feel like, if it was I'll go through that. How would you do it, very difficult. But the thing that's I think a very powerful sustainable advantage for us is the fleet. Nobody has the fleet, those weights are constantly being updated and improved based on billions of miles driven. Tesla has a 100 times more cars with the full self-driving hardware than everyone else combined.

You know, we have a -- quarter, we'll have 500,000 cars with the full eight camera setup, 12 ultrasonics. Someone will still be on hardware two, but we're still have the data gathering ability. And then year from now, we'll have over a 1 million cars with full soft driving, computer hardware everything.

A - Unidentified Speaker

It's just a massive data advantage. It's similar to like you know how like the Google search engine has a massive advantage because people use it and people are programming effectively programmed Google with the queries and the results.

A - Andrej Karpathy {BIO 20228714 <GO>}

Just press you on that. And please reframe the question and tackle him and if it's appropriate. But when we talk to Waymo or NVIDIA, they do speak with a equivalent conviction about their leadership because of their competence in simulating miles driven. Can you talk about the advantage of having real world miles versus simulated miles. Because I think they express that you know by the time you get 1 million miles they can simulate 1 billion and no Formula One race car driver for example could ever successfully complete a real world track with hard driving and a stimulator. Can you talk about the advantages? It sounds like that you perceived to have associated with having data ingestion coming from real world miles versus simulated miles.

A - Unidentified Speaker

Absolutely. The simulator we have is quite a good simulation too, but it just does not capture the long tail of weird things that happen in the real world. If the simulation fully capture the real world, well, will now would be proof that we're living in a simulation I think. Yeah. It doesn't. I wish. But it -- simulations to not capture the real world, and that the real was really weird and messy. You need that drive your cars in the road. We're seeing get it into that in Andrej and Stuart's presentation. So if you want to, we move on to Andrej.

Q - Unidentified Participant

Great. Thanks. Thank you everybody. Thank you very much.

A - Unidentified Speaker

The last question was actually a very good segue, because one thing to remember about our FSD computer is that it can run much more complex neural nets for much more precise image recognition. And to talk to you about how we actually get that image data and how we analyze them, we have our Senior Director of AI, Andrej Karpathy, who's going to explain all of that to you. Andrej has a Ph.D. from Stanford University, where he studied computer science focusing on education recognition and deep learning.

Andrej, why don't you just do your own intro?

There's a lot of Ph.Ds from Stanford. That's not important.

Yes.

Okay. Come on.

A - Andrej Karpathy {BIO 20228714 <GO>}

Thank you.

A - Unidentified Speaker

Andrej started the Computer Vision class at Stanford. That's much more significant. That's what matters. Just as -- can you please talk about your background in a way that is not bashful. Just tell the story about what exactly done.

A - Andrej Karpathy {BIO 20228714 <GO>}

Yeah, I mean -- yeah, see, I think I've been training neural networks basically for what is now a decade and these neural networks were not actually really used in the industry until maybe five or six years ago. So it's been some time that I've been trained these neural networks and that included situations at Stanford at opening Eye at Google and really just training a lot of neural networks not just for images, but also for natural language and designing architectures that couple of those two modalities for my Ph.D. So --

A - Unidentified Speaker

And the computer science class.

A - Andrej Karpathy {BIO 20228714 <GO>}

Oh! Yeah. And at Stanford actually taught the computer for neural networks class. And so I was the primary instructor for that class. I actually started the course and designed the entire curriculum. So in the beginning it was about 150 students and then it grew to 700 students over the next two or three years. So it's a very popular class, it's one of the largest classes at Stanford right now. So that was also really successful.

A - Unidentified Speaker

I mean, Andrej, is like really one of the best computer vision people in the world, arguably the best.

A - Andrej Karpathy {BIO 20228714 <GO>}

Okay. Thank you. Yeah. So hello, everyone. So Pete told you all about the chip that we've designed that runs neural networks in the car. My team is responsible for training of these neural networks and that includes all of data collection from the fleet, neural network training and then some of the deployment on to that chip. So what do the neural networks do exactly in the car. So what we're seeing here is a stream of videos from across the vehicle, across the car. These are eight cameras that send us videos and then these neural networks are looking at those videos and are processing them and making predictions about what they're seeing.

And so, some of the things that we're interested in and there are some of the things we're seeing in this visualization here are lean-line markings, other objects, the distances to those objects, what we call drivable space shown in blue, which is where the car is allowed to go and a lot of other predictions like traffic lights, traffic signs and so on.

Now for my talk I will talk roughly into -- in three stages. So first I'm going to give you a short primer on neural networks and how they work and how they're trained. And I need to do this because I need to explain in the second part why it is such a big deal that we have the fleet and why it's so important and why that's a key enabling factor to really train these neural networks and making them work effectively on the roads. And in the third stage, I'll talk about the vision in LIDAR and how we can estimate depth just from vision alone.

So the core problem that these networks are solving in the car is that a visual recognition. So for you and I these are very -- this is a very simple problem. You can look at all of these four images and you can see that they contain a cello, a boat, an iguana or a scissors. So this is very simple and effortless for us. This is not the case for computers. And the reason for that is that these images are to a computer really just a massive grid of pixels and that each pixel you have the brightness value at that point. And so instead of just seeing an image a computer really gets a million numbers in a grid that tell you the brightness values at all the positions.

A - Unidentified Speaker

You may tell us, if you will, that really is the matrix.

A - Andrej Karpathy {BIO 20228714 <GO>}

Yeah. And so we have to go from that grid of pixels and brightness values into high level concepts like iguana and so on. And as you might imagine this iguana has a certain pattern of brightness values. But iguanas actually can take on many appearances. So they can be in many different appearances, different poses and different brightness conditions against the different backgrounds. You can have a different crops of that iguana. And so we have to be robust across all those conditions and we have to understand that all those different brightness patterns actually correspond to iguanas.

Now the reason you and I are very good at this is because we have a massive neural network inside our heads that is processing those images. So light hits the retina, travels to the back of your brain to the visual cortex and the visual cortex consists of many neurons that are wired together and that are doing all the pattern recognition on top of those images. And really over the last, I would say about five years, the state-of-the-art approach is to processing images using computers have also started to use neural networks. But in this case artificial neural networks. But these artificial neural networks, and this is just a cartoon diagram of it, are a very rough mathematical approximation to your visual cortex. We'll really do have neurons and they are connected together. And here I'm only showing three or four neurons in three or four -- in four layers. But a typical neural network will have tens to hundreds of millions of neurons and each neuron will have a thousand connections. So these are really large pieces of almost simulated tissue.

And then what we can do is we can take those neural networks and we can show them images. So for example, I can feed my iguana into this neural network and the network will make predictions about what it's seeing. Now in the beginning these neural networks are initialized completely randomly. So the connection strengths between all those different neurons are completely random, and therefore the predictions of that network are also going to be completely random.

So it might think that you're actually looking at a boat right now and it's very unlikely that this is actually an iguana. And during the training -- during a training process really what we're doing is, we know that that's actually an iguana. We have a label. So what we're doing is we're basically saying we'd like the probability of iguana to be larger for this image and the probability of all the other things to go down.

And then there's a mathematical process called backpropagation and stochastic gradient descent that allows us to back propagate that signal through those connections and update every one of those connections just a little amount. And once the update is complete, the probability of iguana for this image will go up a little bit. So it might become 14% and probability of the other things will go down.

And of course, we don't just do this for this single image. We actually have entire large data sets that are labeled. So we have lots of images typically you might have millions of images, thousands of labels or something like that and you are doing forward backward passes over and over again. So you're showing the computer here's an image. It has an opinion and then you're saying this is the correct answer and it tunes itself a little bit. You repeat this millions of times and you sometimes you show images, the same image to the computer hundreds of times as well. So the network training typically will take on the order of few hours or a few days depending on how big of a network you're training. And that's the process of training a neural network.

Now there's something very unintuitive about the way neural networks work that I have to really get into, and that is that they really do require a lot of these examples and they really do start from scratch. They know nothing. And it's really hard to wrap your head around this. So as an example, here's a cute dog and you probably may not know the breed of this dog, but the correct answer is that this is Japanese spaniel. Now all of us are looking at this and we're seeing Japanese spaniel, and we're like, okay, I got it. I understand kind of what this Japanese spaniel looks like and if I show you a few more images of other dogs, you can probably pick out other Japanese spaniels here. So in particular those three look like a Japanese spaniel and the other ones do not.

So you can do this very quickly and you need one example, but computers do not work like this. They actually need a ton of data of Japanese spaniels. So this is a grid of Japanese spaniels showing them, you need thousands of examples showing them in different poses, different brightness conditions, different backgrounds, different crops. You really need to teach the computer from all the different angles what this Japanese spaniel looks like and it really requires all that data to get that to work. Otherwise the computer can't pick up on that pattern automatically.

So what does all this imply about the setting of self-driving. Of course, we don't care about dog breeds too much. Maybe we'll roll at some point, but for now we really care about lane line markings, objects, where they are, where we can drive and so on. So the way we do this is we don't have labels like iguana for images, but we do have images from the fleet like this, and we're interested in, for example, lane line markings. So we, a human, typically goes into an image and using a mouse annotates the lane line markings. So here's an example of an annotation that a human could create a label for this image. And it's saying that that's what you should be seeing in this image. These are the lane line markings.

And then what we can do is, we can go to that fleet and we can ask for more images from the fleet. And if you ask the fleet, if you just do a neat job of this and you just ask for images at random, the fleet might respond with images like this, typically going forward on some highway. This is what you might just get like a random collection like this, and we would annotate all that data. Now if you're not careful and you only annotate a random distribution of this data, your network will kind of pick up on this random distribution on data and work only in that regime. So if you show it slightly different example, for example here is an image that actually -- the road is curving and it is a bit of a more residential neighborhood. Then if you show the neural network, this image, that network might make a prediction that is incorrect, it might say that okay, well I've seen lots of times on highways, lanes just go forward, so here's a possible prediction. And of course this is very incorrect.

But the neural network really can't be blamed, it does not know that the train -- the tree on the left whether or not it matters or not, it does not know if the car on the right matters or not towards the lane line. It does not know that the buildings in the background matter or not. It really starts completely from scratch. And you and I know that the truth is that none of those things matter, what actually matters is there are a few white lane line like markings over there in the vanishing point. And the fact that they curl a little bit should pull the prediction, except there is no mechanism by which we can just tell the neural network, hey those lane line markings actually matter.

The only tool in the toolbox that we have is labeled data. So what we do is, we need to take images like this when the network fails and we need to label them correctly. So in this case we will turn the lane to the right. And then we need to feed lots of images of this to the neural net and neural net over time will accumulate -- will basically pick up on this pattern that those things there don't matter, but those lane line markings do and we learn to predict the correct lane.

So what's really critical is not just the scale of the dataset, we don't just want millions of images, we actually need to do really good job of covering the possible space of things that the car might encounter on the roads. So we need to teach the computer how to handle scenarios where it's night and wet. You have all these different specular reflections and as you might imagine the brightness patterns in these images will look very different.

We have to teach the computer how to deal with shadows, how to deal with forks in the road, how to deal with large objects that might be taking up most of that image, how to deal with tunnels or how to deal with construction sites. And in all these cases, there's no again explicit mechanism to tell the network what to do, we only have massive amounts of data, we want to source all those images and we want to annotate the correct lines and the network will pick up on the patterns of those.

Now large and varied data sets make -- basically make these networks work very well. This is not just a finding for us here at Tesla, this is a ubiquitous finding across the entire industry. So experiments and research from Google, from Facebook, from Baidu, from Alphabet's DeepMind all show similar plots where neural networks really love data and love scale and variety, as you add more data, these neural networks start to work better and get higher accuracies for free. So more data just makes them work better.

Now a number of companies have -- number of people have kind of pointed out that potentially we could use simulation to actually achieve the scale of the data sets and we're in charge of a lot of the conditions here, maybe we can achieve some variety in a simulator. Now at Tesla and that was also kind of brought up in the questions just just before this. Now at Tesla this is actually a screenshot of our own simulator. We use simulation extensively, we use it to develop and evaluate the software. We've also even use it for training quite successfully, So -- but really when it comes to training data from neural networks, there really is no substitute for real data. The simulator -- simulations have a lot of trouble with modeling appearance, physics and the behaviors of all the agents around you.

So here are some examples to really drive that point across, the real world really throws a lot of crazy stuff at you. So in this case for example we have very complicated environments with snow, with trees, with wind. We have various visual artifacts that are hard to simulate potentially, we have complicated construction sites, bushes and plastic bags that can go in that can kind of go round with the wind. Complicated construction sites that might feature lots of people, kids, animals all mixed in and simulating how those things interact and flow through this construction zone might actually be completely intractable. It's not about the movement of any one pedestrian in there, it's about how they respond to each other and how those cars respond to each other and how they respond to you driving in that setting and all of those are actually really tricky to simulate.

It's almost like you have to solve the self-driving problem to just simulate other cars in your simulation. So it's really complicated. So we have dogs, exotic animals and in some cases it's not even that you can simulate, it's that you can't even come up with it. So for example I didn't know that you can have truck-on-truck-on-truck like that, but in the real world you find this and you find lots of other things that they are very hard to really even come up with. So really that the variety that I'm seeing in the data coming from the fleet is just crazy, with respect to what we have in the simulator, we have a really good simulator.

Q - Unidentified Participant

Yeah, it's and I think like simulation -- you're fundamentally grading your own homework. So you know -- if you know that you're going to simulate it, okay, you kind of definitely solve for it, but as under saying, you don't know what -- you don't know the world is very weird and has millions of corner cases. And if somebody can produce a self-driving simulation that accurately matches reality, that in itself would be an a monumental achievement of human capability, they can't, there's no way.

A - Unidentified Speaker

Yeah. So I think the three points that I really try to drive home until now are to get neural networks to work well, you require these three essentials, you require a large dataset, a varied dataset and a real dataset. And if you have those capabilities, you can actually train neural networks and make them work very well. And so why is Tesla in such a unique and interesting position to really get all these free essentials right? And the answer to that of course is the fleet. We can really source data from it and make our neural network systems work extremely well.

So let me take you through a concrete example of -- for example making the object detector work better, to give you a sense of how we develop these neural networks, how we iterate on them and how we actually get them to work over time? So object detection is something we care a lot about. We'd like to put bounding boxes around say the cars and the objects here because we need to track them and we need to understand how they might move around. So again, we might ask human annotators to give us some annotations for these and humans might go in and might tell you that, okay, those patterns over there are cars and bicycles and so on.

And you can train your neural network on this, but if you're not careful, the neural network will make mispredictions in some cases. So as an example, if we stumble by a car like this, that has a bike on the back of it, then the neural network actually went when I joined -- would actually create two detections, it would create a car detection and a bicycle detection. And that's actually kind of correct because I guess both of those objects actually exist, but for the purposes of the controller and a planner downstream, you really don't want to deal with the fact that this bicycle can go with the car. The truth is that that bike is attached to that car. So in terms of like just objects on the road, there's a single object, a single car. And so what you'd like to do now is you'd like to just potentially annotate lots of those images as this is just a single car. So the process that that we go through internally in the team is that we take this image or a few images that show this pattern and we have a mechanism, a machine learning mechanism by which we can ask the fleet to source us examples that look like that. And the fleet might respond with images that contains those patterns.

So as an example these six images might come from the fleet, they all contain bikes on the backs of cars. And we would go in and we would annotate all those as just a single car. And then the performance of that detector actually improves and the network internally understands that, hey, when the bike is just attached to the car that's actually just a single car. And it can learn that given enough examples and that's how we sort of fix that problem.

I will mention that I talk quite a bit about sourcing data from the fleet. I just want to make a quick point that we've designed this from the beginning with privacy in mind and all the data that we use for training is unanalyzed. Now the fleet doesn't just respond with bicycles on backs of cars, we look for all the thing, we look for lots of things all the time. So for example, we look for boats and the fleet can respond with boats. We look for construction sites and the fleet can send us lots of construction sites from across the road. We look for even slightly more rare cases. So for example finding debris on the road is pretty important to us. So these are examples of images that have streamed to us from the fleet that show tires, cones, plastic bags and things like that.

If we can source these at scale, we can annotate them correctly and the neural network can learn how to deal with them in the world. Here's another example, animals of course also a very rare occurrence and event. But we want the neural network to really understand what's going on here that these are animals and we want to deal with that correctly.

So to summarize the process by which we iterate on neural network predictions looks something like this. We start with a SEED dataset that was potentially sourced at random. We annotate that dataset and then we train neural networks on that dataset and put them in the car. And then we have mechanisms by which we notice inaccuracies in the car when this detector maybe misbehaving. So for example if we detect that the neural network might be uncertain or if we detect that or if there's a driver intervention on any of those settings, we can create this trigger infrastructure that sends us data of those inaccuracies. And so for example if we don't perform very well on lane line detection on tunnels, then we can notice that there's a problem in tunnels, that image would enter our unit tests, so we can verify that we're actually fixing that problem over time. But now what you do is to fix this inaccuracy, you need to source many more examples that look like that.

So we ask the fleet to please send us many more tunnels and then we label all those tunnels correctly. We incorporate that into the training set and we retrain the network, redeploy and iterate the cycle over and over again. And so we refer to this iterative process by which we improve these predictions as the data engine. So iteratively, deploying something potentially in shadow mode, sourcing inaccuracies and incorporating into the training set over and over again. And we do this basically for all the predictions of these neural networks.

Now so far I've talked about lot of explicit labeling. So like I mentioned, we ask people to annotate data. This is an expensive process in time and also with respect, yeah, it's just an expensive process in time and also respect to -- yes, this is an expensive process. And so these annotations, of course, can be very expensive to achieve.

So, what I want to talk about also is really to utilize the power of the fleet. You don't want to go through this human annotation bottleneck. You want to just stream in data and automate it automatically. And we have multiple mechanisms by which we can do this. So as one example of a project that we recently worked on is the detection of cut-in. So, you're driving down the highway, someone is on the left or on the right, and they cut-in in front of you into your lane. So, here's a video showing the autopilot detecting that this car is intruding into your lane.

Now, of course, we'd like to detect a cut-in as fast as possible. So the way we approach this problem is we don't write explicit a code for is the left blinker on, is the right blinker on, track the keyboard over time and see if it's moving horizontally. We actually use a fleet learning approach. So the way this works is, we ask the fleet to please send us data whenever they see a car transition from a right lane to the center lane or from left to center. And then what we do is we rewind time backwards, and we automatically can annotate that, hey, that car will turn, well, in 1.3 seconds cut-in in front of the -- in front of you. And then we can use that for training the neural net.

And so the neural net will automatically pick up on lot of these patterns. So, for example, the cars are typically yard. They're moving this way, maybe the blinker is on, all that stuff happens internally inside the neural net just from these examples. So, we as a fleet automatically send out all these data, we can get 0.5 million or so images, and all of these would be annotated for cut-ins, and then we train network and then we took this cut-in network and we deployed it to the fleet, but we don't turn it on yet. We run it in shadow mode. And in shadow mode, the network is always making predictions. Hey, I think this vehicle is going to cut-in from the way it looks. This vehicle is going to cut-in. And then we look for mis-predictions.

So, as an example, this is an clip that we had from shadow mode of the cut-in network, and it's kind of hard to see, but the network thought that the vehicle right ahead of us on the right was going to cut-in, and you can sort of see that it's slightly flirting with the lane line. It's trying to -- it's sort of encroaching a little bit, and the network got excited and it thought that was going to be cut in, that vehicle will actually end up in our center lane. That turns out to be incorrect and the vehicle did not actually do that.

So, what we do now is we just churn the data engine. We source that ran in the shadow mode is making predictions, it makes some false positives and there are some false negative detections. So, we got overexcited, and sometimes, and sometimes we miss the cut-in when it actually happened. All those create a trigger that streams to us and that gets incorporated now for free, there's no humans harmed in the process of labeling this data incorporated for free into our training set. We retrained the network and redeployed the shadow mode. And so we can spin this a few times and we always look at the false positives and negatives coming from the fleet. And once we're happy with the false positive, false negative ratio, we actually flipped a bit and actually let the car control to that network.

And so, you may have noticed, we actually shipped one of our first versions of the cut-in detector, approximately, I think, three months ago. So if you've noticed that the car is much better at detecting cut-ins, that's fleet learning operating at scale. Yes, it actually works quite nicely. So that's fleet learning, no humans were harmed in the process, it's just a lot of neural network training based on data and a lot of shadow mode and looking at those results. In other...

A - Elon R. Musk {BIO 1954518 <GO>}

Essentially like everyone's training the network all the time is what it amounts to, whether autopilot is on or off, the network is being trained every mile that's driven for the car. That's harder to or above is training the network.

Q - Unidentified Participant

Yes, another interesting way that we use this in the scheme of fleet learning at the other project that I'll talk about is a path prediction. So, while you are driving a car which you're actually doing as you are emitting the data because you are steering the wheel, you're telling us how to traverse different environments. So, what we're looking at here is a some person in the fleet who took a left through an intersection. And what we do here is we have the full video of all the cameras and we know that the path that this person took because of the GPS, the initial measurement unit, the wheel angle, the wheel ticks. So we put all that together and we understand the path that this person took through this environment.

And then, of course, this -- we can use this for supervision for the network. So, we just sourced a lot of this in the fleet. We trained the neural network on those trajectories, and then the neural network predicts paths, just from that data. So, really what this is referred to typically is called imitation learning. We're taking human trajectories from the real world. I'm just trying to imitate how people drive in real worlds, and we can also apply the same data engine crank to all of this and make this work over time.

So, here's an example of path prediction going through a kind of a complicated environment. So what you're seeing here is a video and we are overlaying the predictions of the network. So this is a path that the network would follow in green, and some -- yeah.

A - Elon R. Musk {BIO 1954518 <GO>}

The crazy thing is the network is predicting paths it can't even see with incredibly high accuracy. It can't see around the corner, but it is saying the probability of that curve is extremely high. So that's the path. And it nails it. You will see that in the cars today. We are going to turn on augmented vision, so you can see the lane lines and the path predictions of the cars overlaid on the video.

A - Pete Bannon {BIO 20590065 <GO>}

Yeah. There's actually more going on under the hood that you can even tell...

A - Elon R. Musk {BIO 1954518 <GO>}

It sounds scary.

A - Pete Bannon {BIO 20590065 <GO>}

And of course there is lot of details I'm skipping over. You might not want to annotate all the drivers. You might annotate just -- you might want to just imitate the better drivers, and there's many technical ways that we actually slice and dice that data. But the interesting thing here is that this prediction is actually a 3D prediction, that we project back to the image here. So the path here forward is a three dimensional thing that we're just rendering in 2D, but we know about the slope of the ground from all of this and that's actually extremely valuable for driving.

So our prediction actually is life in the fleet today by the way, so if you're driving Cloverleaf, if you're in a Cloverleaf on a highway, until maybe five months ago or so, your car would not be able to do cloverleaf. Now it can. That's our prediction running live on your cars. We've said this a while ago. And today you are going to get to experience this for traversing intersections, a large component of how we go through intersections in your drives today is all sourced from our prediction from automatic labels.

So what I talked about so far is really the three key components of how we iterate on the predictions of the network and how we make it work over time. You require large, varied and real data set. We can really achieve that here at Tesla. And we do that through the scale the fleet, the data engine, shipping things in shadow mode, iterating that cycle and potentially even using fleet learning where no human annotators are harmed in the process, and just using it automatically and we can really do that at scale.

So in the next section of my talk, I'm going to especially talk about depth perception using vision only. So you might be familiar that there are at least two sensors in the car. One is vision, the cameras just getting pixels and the other is Lidar that a lot of -- that a lot of companies also use. And Lidar gives you these point measurements of distance around you. Now, one thing I'd like to point out first of all is you all came here -- you drove here, many of you, and you used your NeuroMet and vision, you were not shooting lasers out of your eyes and you still ended up here.

A - Unidentified Speaker

We might have, I mean, (multiple speakers).

A - Pete Bannon {BIO 20590065 <GO>}

So clearly the human NeuroMet derives distance and all the measurements in the 3D understanding of the world just from vision. It actually uses multiple cues to do so. I'll just briefly go over some of them, just to give you a sense of roughly what's going on inside. As an example, we have two eyes pointed out. So you get two independent measurements at every single time step of the role ahead of you, and your brain stitches this information together to arrive at some depth estimation because you can triangulate any points across those two viewpoints.

A lot of animals, instead, have eyes that are positioned on the sides. So they have very little overlap in their visual fields. So they will typically use structure for motion and the idea is that they bob their heads and because of the movement they actually get multiple observations of the world and you can triangulate again depths.

And even with one eye closed and completely motionless, you can still have some sense of depth perception if you did this, I don't think you would notice me coming two meters towards you or 100 meters back. And that's because there are a lot of very strong monocular cues that your brain also takes into account. This is an example of a pretty common visual illusion where you have these two blue bars are identical, but your brain, the way it stitches up the scene, is it just expects one of them to be larger than the other, because of the vanishing lines of this image. So your brain does a lot of this automatically. And NeuroMet, artificial NeuroMet can as well.

So let me give you three examples of how you can arrive at depth perception from vision alone. A classical approach and two that rely on your own networks. So here's a video going down, I think this is San Francisco of a Tesla. So these are our cameras, our sensing, and we're looking at all -- I'm only showing the main camera, but all the cameras are turned on, the eight cameras of the autopilot. And if you just have this six second clip, what you can do is you can stitch up this environment in 3D using multiview stereo techniques. So this is supposed to be a video. Is not a video?

I know it's -- here we go. So this is the 3D reconstruction of those six seconds of that car driving through that path. And you can see that this information is purely -- is very well recoverable from just videos, and roughly that's through process of triangulation and as I mentioned multiview stereo, and we've applied similar techniques, slightly more sparse and approximate also in the car. So it's remarkable all that information is really there in the sensor, and just a matter of extracting it.

The other project that I want to briefly talk about is as I mentioned there's nothing about neural network -- neural network is a very powerful visual recognition engines. And if you want them to predict depth, then you need to, for example, look for labels of depth and then they can actually do that extremely well. So there's nothing limiting networks from predicting this monocular depth except for label data. So one example project that we've actually looked at internally, is we use the forward facing radar which is shown in blue and that radar is looking out and measuring depths of objects. And we use that radar to annotate the -- what vision is seeing. The bounding boxes that come out of the neural networks.

So instead of human annotators telling you, okay this car and this bounding box is roughly 25 meters away, you can annotate that data much better using sensors. So you sensor annotation. So as an example of radar is quite good at that distance, you can annotate that and then you can train your network on it. And if you just have enough data of it this neural network is very good at predicting this patterns.

So here's an example of predictions of that. So in circles I'm showing radar objects and in -- and the keyboards that are coming out here are purely from vision. So the keyboards here are just coming out of vision and the depth of those keyboards is learned by a sensor annotation from the radar. So if this is working very well, then you would see that the circles in the top-down view would agree with the keyboards and they do, and that's because neural networks are very competent at predicting depths. They can learn the different sizes of vehicles internally and they know how big those vehicles are, and you can actually derive depth from that quite accurately.

The last mechanism I will talk about very briefly is slightly more fancy and gets a bit more technical, but it is a mechanism that has recently -- there's a few papers basically over the last year or two on this approach, it's called self-supervision. So what you do in a lot of these papers is you only feed raw videos into neural networks with no labels whatsoever, and you can still learn, you can still get in neural networks to learn depth. And it's a little bit technical, so I can't go into the full details but the idea is that neural network predicts depth at every single frame of that video and then there are no explicit targets that the neural network is supposed to regress to with the labels, but instead the objective for the network is to be consistent over time. So whatever depth you predict should be consistent over the duration of that video. And the only way to be consistent is to be right as the neural network automatically predicts the correct depths for all the pixels and we reproduce some of these results internally. So this also works quite well.

So in summary people drive with vision only no, no lasers are involved. This seems to work quite well. The point that I'd like to make is, that visual recognition and very powerful visual recognition is absolutely necessary for Autonomy. It's not nice to have, like we must have neural networks that actually really understand the environment around you, and LIDAR points are much less information rich environment. So vision really understands the full details just a few points around are much -- there's much less information in those. So as an example on the left here, is that a plastic bag or is that a tire? LIDAR might just give you a few points on that, but vision can tell you which one of that two is true and that impacts your control. Is that person who is slightly looking backwards, are they trying to merge into your lane on the bike or are they just going forward?

In the construction sites, what do those signs say. How should I behave in this world? The entire infrastructure that we have built up for roads is all designed for human visual consumption. So all of the signs, all the traffic lights, everything is designed for vision. And so that's where all that information is, and so you need that ability. Is that person distracted and on their phone? Are they going to walk into your lane? Those answers to all these questions are only found in vision and are necessary for level 4, level 5 Autonomy. And that is the capability that we are developing at Tesla. And through -- this is done through combination of large scale neural training through data engine and getting that to work over time, and using the power of the fleet.

And so in this sense LIDAR is really a shortcut. It sidesteps the fundamental problems, the important problem visual recognition that is necessary for Autonomy. And so it gives a full sense of progress and is ultimately crotch [ph]. It does give like really fast demos.

So, if I was to summarize the entire -- my entire talk in one slide it would be this. All of Autonomy, because you want level four, level five systems that can handle all the possible situations in 99.99% of the cases, and chasing some of the last few nights is going to be very tricky and very difficult and it's going to require a very powerful visual system. So I'm showing you some images of what you might encounter in any one slice of that nine. So in the beginning you just have very simple cars going forward. Then those cars start to look a little bit funny, then maybe you have bikes and cars, then maybe you have cars and cars, then maybe you start to get into really rare events like cars turned over or even cars airborne. We see a lot of things coming from the fleet, and we see them at some rate, at -- like a really good rate compared to all of our competitors.

And so the rate of progress at which you can actually address these problems iterate on the software and really feed the neural networks with the right data. That rate of progress is really just proportional to how often you encounter these situations in a while, and we encounter them significantly more frequently than anyone else which is why we're going to do extremely well. Thank you.

Q - Unidentified Participant

It's all super impressive. Thank you so much. How much data -- how many pictures are you collecting on average from each car per period of time? And then it sounds like the new hardware with the dual-dual active-active computers gives you some really interesting opportunities to run in full simulation, one copy of the neural net while you're running the other one -- the other one drive the car and compare the results to do quality assurance. And then I was also wondering if there are other opportunities to use the computers for training when they're parked in the garage for the 90% of the time that I'm not driving my Tesla around. Thank you very much.

A - Pete Bannon {BIO 20590065 <GO>}

Yes. So for the first question, how much data do we get from the fleet. It's really important to point out. It's not just the scale of the data set. It really is the variety of data set that matters. If you just have lots of images of something going forward on the highway at some point neural networks gets it. You don't know need that data. So we are really strategic in how we pick and choose, and the trigger infrastructure that we've built-up is quite sophisticated and allows us to get just the data that we need right now. And so it's not a massive amount of data. It's just very well big data.

For the second question, with respect to redundancy. Absolutely, you can run basically the copy of the network on both, and that is actually how its designed to achieve a level 4, level 5 system that is redundant. So that's absolutely the case.

And your last question, I'm sorry I did not...

A - Unidentified Speaker

Training.

A - Elon R. Musk {BIO 1954518 <GO>}

The car is an inference optimized computer. We do have a major program at Tesla which we don't have enough time to talk about today called Dojo. That's a super powerful training computer. The goal Dojo will be -- to be able to take in vast amounts of data and train at a video level, and do unsupervised massive training of vast amounts of video with the Dojo program -- Dojo computer. But that's for another day.

Q - Unidentified Participant

(inaudible) like a test pilot in a way because I drive the four, five, ten and all these really tricky really long tail things happen every day. But the one challenge that I'm curious to how you're going to solve is changing lanes, because whenever I try to get into a lane with traffic, everybody cuts you off. And so human behavior is very irrational when you're driving in L.A. and the car just wants to do it safely and you almost have to do it unsafely. So I was wondering how you're going to solve that problem?

A - Pete Bannon {BIO 20590065 <GO>}

Yeah. So one thing I will point out is I spoke about the data engine as iterating on neural networks. But we do the exact same thing on level of software and all the hyper parameters that go in to the choices of when we actually might change, how aggressive we are? We're always changing those, potentially run them in shadow mode and seeing how well they work. And so to tune our heuristics around when it's okay to lane change, we would also potentially utilize the data engine in a shadow mode and so on.

Ultimately, actually designing all the different heuristics for when it's okay to lane change is actually a little bit intractable I think in the general case. And so ideally you actually want to use fleet learning to guide those decisions. So when do humans lane change in what scenarios, and when do they feel it's not safe for lane change and let's just look at a lot of the data and train machine-learning classifiers for distinguishing when it is safe to do so. And those machine-learning classifiers can write much better code than humans because they have the maximum amount data backing that. So they can really tune all the right thresholds and agree with humans an do something safe.

Q - Unidentified Participant

We'll probably have a mode that goes beyond Mad Max mode to L.A. traffic mode.

A - Elon R. Musk {BIO 1954518 <GO>}

Yeah. Well you know Mad Max would have a hard time in L.A. traffic. I think.

A - Pete Bannon {BIO 20590065 <GO>}

Yes. So really it's a trade-off like you do want to create unsafe situations but you want to be assertive. But that little dance of how you make that work as a human is actually very complicated and it's very hard to write in code. But I think we really do -- it really does seem like machine-learning approach is kind of like the right way to go about it where we just look at a lot of ways that people do this and try to imitate that.

A - Andrej Karpathy {BIO 20228714 <GO>}

We are just being like more conservative right now and then as we gain higher confidence we'll allow users to select a more aggressive mode. That'll be up to the user. But in the more aggressive modes and trying to merge in traffic, there is a slight -- no matter how many knew if there is a slight chance of like a fender bender not a serious accident but you basically will have a choice of, do you want to have a non zero-chance of a fender bender on freeway traffic which is unfortunately the only way to navigate LA traffic. Yes. (Multiple Speakers) like LA story. That movie is a great movie.

Q - Unidentified Participant

(Multiple Speakers) this is a game of chicken that's going on.

A - Unidentified Speaker

(Multiple Speakers) And it will go after more aggressive options over time that how the user is specified. (Multiple Speakers) Yes (inaudible). Exactly.

Q - Jed Dorsheimer {BIO 6360573 <GO>}

Hello. Hi, Jed Dorsheimer from Canaccord Genuity. Thank you and congratulations on everything that you've developed. When we look at the Alpha Zero project, it was a very defined and limited variable in terms of the parameters on that which allowed for the learning curve to be so quick. The risk or want to -- what you're trying to do here is almost develop consciousness in the cars through the neural network and so I guess the challenge is how do you not create a circular reference in terms of the pulling from the centralized model of the fleet to that handoff where the car has enough information -- where is that line, I guess in terms of the point of the learning process to handing it off where there's enough information in the car and not having to pull from the fleet?

A - Unidentified Speaker

Well, the car can operate if it's completely disconnected from the fleet. It just -- it uploads the training that's better and better as the fleet gets better and better. So simply, if you're just going actually to fund the fleet from that point onwards, it would stop getting better but it will so function fine.

Q - Jed Dorsheimer {BIO 6360573 <GO>}

But I guess, in the hardware portion of your share -- in the hard -- the previous version, you talked about a lot of the power benefits of not storing a lot of the images. And so in this portion, you're talking about the learning that's going on by pulling from the fleet. I guess I'm having a hard time reconciling how if there was a situation where I'm driving up the hill as you showed and I'm predicting where the road is going to go, that's coming from all of the other fleet variables that led to that intelligence. How -- I'm not -- how I'm getting the benefit of the low power using the cameras with the neural network that's where I'm losing the two. Maybe it's just me but I guess that's --

A - Unidentified Speaker

I mean the compute power in the full self-driving computer is incredible and maybe we should've mentioned that if it had never seen that road before, it would still have made those predictions provided it was a road in the United States.

Q - Unidentified Participant

In the case of light or the march of nines, isn't there an example, I want to just get to your slam on LIDAR, because it's pretty clear you don't like LIDAR, In this --

A - Unidentified Speaker

LIDAR is lame -- LIDAR is lame.

Q - Unidentified Participant

Isn't there like a case where at some point 99999 down the road, where actually LIDAR may be helpful and why not have it as some sort of a redundancy or backup? So that's my first question. And the second -- so you can still have your focus on computer vision but just have it as a redundant. And my second question is, if that is true, what happens to the rest of the industry that's building their autonomy solutions on LIDAR?

A - Unidentified Speaker

They're all going to dump LIDAR. That's my prediction. Mark my words. I should point out that I don't actually super hate LIDAR as much as i may sound but at SpaceX -- SpaceX Dragon uses LIDAR to navigate to the space station and dock. Not only that we -- SpaceX developed its own LIDAR from scratch to do that and I've spearheaded that effort personally because in that scenario LIDAR makes sense and in cars it's freaking stupid. It's expensive and unnecessary and as Andrej was saying, once you saw a vision, it's worthless. So you have expensive hardware that's worthless on the car. But we do have a forward radar which is low cost and is helpful especially for occlusion situations. So if there's like fog or dust or snow, the radar can see through that. If you're gonna use active photon generation, don't use visible wavelength because once you -- with passive optical you've taken care of all visible wavelength and stuff, if you want to use a wavelength that is occlusion penetrating like radar. So LIDAR is just active photon generation in the visual spectrum.

We can do active photon generation, do it outside the visual spectrum in the radar spectrum. So like a 3.8 millimeters versus 400 nanometers to 700 nanometers, going to be a much better occlusion penetration and that's why we have a forward radar and then we also have 12 ultrasonics for near-field information in addition to the eight cameras and the forward radar. Only the radar in the forward direction because that's the only direction you are going real fast. So that's -- we've gone over this multiple times like always show we have the right size of sweet, should we add anything more? No.

Q - Unidentified Participant

Hi. Right here. So you had mentioned that you asked the fleet for the information that you're looking for some of the vision. I have two questions about that. It sounds like the cars are doing some computation to determine what kind of information to send back to you, is that a correct assumption? And are they doing that in real time or are they doing based on stored information?

A - Unidentified Speaker

Yes. So they absolutely do the computation in real time on the car and with it and we wait to basically specify condition that we're interested in and then those cars do that computation there. If they did not, then we'd have to send all the data and do that offline in our back-end, we don't want to do that. So all that computation we have is on the car.

Q - Unidentified Participant

So, it's -- based on that question, it sounds like you guys are in a really good position to have currently half a million cars in the future potentially millions of cars that are essentially computers representing free -- almost free data centers for you to do computational, is that a huge future opportunity for Tesla?

A - Unidentified Speaker

It's current.

Q - Unidentified Participant

Current opportunity? And that's not really factored in for anything yet. That's incredible. Thank you.

A - Unidentified Speaker

We've 425,000 cars with hardware tuned and beyond which is it means they got all eight cameras the right of the rador and ultrasonics. And they've got at least a video computer which is enough to essentially figure out what information is important, what is not, compress the information as important to the most salient elements and upload it to the network for training. That's a massive compression of real world data.

Q - Unidentified Participant

You have these sort of network of millions of computers which is like massive data centers essentially that are distributed data centers for computational capacity. Do you see it being used for other things besides self-driving in the future?

A - Unidentified Speaker

I suppose it could possibly be used for something besides self-driving. We're into focused on self-driving. So as we get that really nailed maybe there's going to be some other use for millions and then tens of millions of computers with hardware three or four self-driving computer, yeah, maybe there would be.

Q - Unidentified Participant

(Technical Difficulty)

A - Unidentified Speaker

It could be -- it could be just like some sort of AWS angle here, it's possible.

Q - Matt Joyce {BIO 15059467 <GO>}

Hello. Hi, Elon, Matt Joyce, Loup Ventures. I own a Model 3 in Minnesota where it snows a lot. Since camera and radar cannot see road markings through snow, what is your technical strategy to solve this challenge? Is that involve high precision GPS at all?

A - Elon R. Musk {BIO 1954518 <GO>}

So, actually, like today, actually autopilot will do a decent job in snow even when lane markings are covered, even when lane markings are faded, covered or when there's lots of rain on them, we still seem to drive relatively well. We didn't specifically go after snow yet with our data engine, but I actually think this is completely tractable because a lot of those images are -- even when things are snowy, when you ask a human annotator, where are the lane lines, they actually could tell you. They actually are relatively consistent in (inaudible) lane lines. As long as the annotators are consistent on your data, then I have -- there's -- then your network will pick up on those patterns and they will do just fine. So it's really just about it's the signal there even for the human annotator if that is -- and the answer to that is yes then you know that we can do it just fine.

A - Unidentified Speaker

Yeah, there's actually -- there are a number of important signals as Andrej was saying. So lane lines are one of those things. But one of the most important signals is drive space. So what is drivable space and what is not drivable space? And what actually really matters the most is drivable space more than lane lines and the prediction of drivable space is extremely good. And I think especially after this upcoming winter will be incredible. It's like it will be like how could it possibly be that good. That's crazy.

The other thing to point out is, maybe it's not even only about human annotators, as long as you as a human can drive through that environment that through fleet learning, we actually know the path you took and you obviously use vision to guide you through that path, you did not just use the lane line markings, you used the entire geometry of the entire scene, so you see -- you see how the road is roughly curling, you see how the cars are positioned around you, network will pick up on all those patterns automatically inside it if you just have enough of the data people traversing those environments.

It's actually extremely important that things not be rigidly tied to GPS, because GPS era can vary quite a bit and the actual situation for a road can vary quite a bit. So there could be reconstruction, there could be a detour and if the car is using GPS as primary, this is a real bad situation, it's asking for trouble. It's fine to use GPS for like tips and tricks. So it's like you can drive your your home neighborhood better than a neighborhood enough and like some other country or some other part of the country. So you know your own neighborhood well and you use kind of like the knowledge of your neighborhood to drive with more confidence to maybe have counterintuitive shortcuts and that kind of thing. But it's -- the GPS overlay data should only be helpful but never primary, if it's ever primary, it's problem.

Q - Unidentified Participant

So question back here in the back corner, I just wanted to follow-up partially on that because several of your competitors in the space over the past few years have made -- have talked about how they are augmenting all of their perception and path planning capabilities that are kind of on the car platform with high definition maps of the areas that they are driving. Does that play a role in your system? Do you see it adding any value? Are there areas where you would like to get more data that is not collected from the fleet but is more kind of mapping style data?

A - Unidentified Speaker

I think the high precision -- the high type -- high precision GPS maps and lanes are a really bad idea. The system becomes extremely brittle. So any change like this -- this might -- any change to the system makes it, it can't adapt. So if it locks onto a GPS and high precision lane lines and does not allow vision override, in fact this vision should be the thing that does everything and then like lane lines that are a guideline but they're not the main thing. And we briefly barked up the tree of high precision lane lines and then realized that was a huge mistake and reversed it out. Isn't that good?

Q - Unidentified Participant

So this is very helpful for understanding annotation where the objects are and how the car drives. But what about the negotiation aspect for parking and roundabouts and other things where there are other cars on the road that are human-driven where it's more art than science?

A - Unidentified Speaker

Does pretty good. Actually, like, with cut into stuff, it's doing really well.

Yeah, so, (inaudible) we're using a lot of machine learning right now in terms of predicting kind of creating an explicit representation of what the road looks like and then there's an explicit planner and a controller on top of that representation and there's a lot of heuristics for how to traverse and negotiate and so on. There is a long tail just like in what the visual environments look like, there's a long tail and just those negotiations and all game of chicken that you play with other people and so on. And so, I think we have a lot of confidence that eventually there must be some kind of a fleet learning components to how you actually do that because writing all those rules by hand is going to -- it's going to quickly plateau, I think.

Yeah, we've dealt with this issue with cut-ins and it's like we'll allow gradually more aggressive behavior on the part of the user, they can just dial the setting up and say, be more aggressive, you're less aggressive, drive easy, chill mode, aggressive.

Q - Unidentified Participant

Incredible progress, phenomenal. Two questions. First, in terms of platooning, do you think the system is geared because somebody asked about when there is snow on the road, but if you have platooning feature, you can just follow the car in front. Does your system -- is your system capable of doing that? Then I have two follow-ups.

A - Unidentified Speaker

So you've asked about platooning. So I think like we could absolutely build those features. But again if you just use -- if you just train your networks, for example, on imitating humans, humans already like follow the car ahead and so that neural network actually incorporates those patterns internally. It's just it figures out that there's a correlation between the way the car ahead of you faces and the path that you are going to take. But that's all done internally in the net. So you're just concerned with getting enough data and the tricky data and the neural network training process, actually, it's quite magical, does all the other stuff automatically. And so you turn all the different problems into just one problem, just collect your data set and using a lot to train.

There's three steps to self-driving, there's being feature complete then there's being future complete to agree that where we think that the person in the car does not need to pay attention. And then there's the reliability level where we also convince regulators that that is true. So there's kind of like three levels. We expect to be set feature complete in self-driving this year and we expect to be confident enough from our standpoint to say that we think people do not need to touch the wheel, look out of the window. Sometime probably around I don't know second quarter of next year, and then we start to expect to get regulatory approval at least in some jurisdictions for that towards the end of next year. That's roughly the timeline that I expect things to go on. And probably for trucks, the tuning will be approved by regulators before anything else. And you can have like maybe, if you're a long haul -- doing long haul freight you can have one driver in the front and then have four semis trailing behind in a platooning manner. And I think that probably the regulators would be quicker to approve that than other things.

Q - Unidentified Participant

Regarding, of course, you don't have to convince us LIDAR is a technology, in my opinion, which has an answer looking for a question probably dead. I mean this is very impressive what we saw today and probably demo could show something more. I was wondering what is the maximum dimension of a matrix that you may be having in your training or in your deep learning pipeline, a ballpark figure?

A - Unidentified Speaker

Maximum dimension of the matrix. So you know (inaudible) apply operations inside (inaudible) asking about them. And there's many different ways to answer that question but I'm not 100% sure if they're useful, they're useful answers, these neural (inaudible) typically have like I mentioned about tens to hundreds of millions of neurons, each of them will on average have about a 1,000 connections to neurons below. So these are the typical scales that are kind of used across the industry and also that we would use as well.

Q - Unidentified Participant

Yeah I've been actually very impressed by the rate of improvement on autopilot the past year on my Model 3, and the two scenarios I wanted your feedback on. Last week, first scenario was I was on the right hand most lane of the freeway and there was a highway on-ramp and then my Model 3 actually was able to detect the two cars on the side, slow down and let the car go in front of me and one car go behind me and I was like, oh! my gosh, this is like insane. Like I didn't think my Model 3 could do that. So that was like super impressive. But the same week, another scenario which is I was on the right hand lane again, but my right hand lane was merging with the left lane and it wasn't a on-ramp it's just a normal highway -- freeway lane. And my Model 3 wasn't able to detect really that situation and I wasn't able to slow down or speed up and I had to intervene, kind of. So can you -- from your perspective kind of share -- kind of the background on how a neural net would -

- how Tesla might adjust for that? And how that could be improved over the -- over time?

A - Unidentified Speaker

Yeah. So like I mentioned, we have a very sophisticated trigger infrastructure. If you have intervened, it's actually potentially likely that we receive the clip and that we can actually analyze it and see what happened and tune the system. So probably enter some statistics over, okay, at what rate are we correctly merging the traffic. And we look at those numbers and we look at the clips and we see what's wrong and we try to fix those clips and make progress against those benchmarks. So yes.

Q - Unidentified Participant

(Technical Difficulty)

A - Unidentified Speaker

Yes. So we would potentially go through a phase of categorization and then we look at some of the biggest kind of categories that actually seem to semantically be related to a simple problem. And then we would look at some of those and then try to develop software against that.

Okay, we do have one more presentation which is that the software is like essentially the what about hardware with Stuart, there's the sort of neural net vision with Andrej. And then there's the software engineering at scale that's again presented by Stuart. So next we'll be up shortly afterwards to ask questions, so yeah, thanks.

I just wanted to very briefly say if you have an early flight and you want to do a test ride with our latest development software if you could please speak to my colleague and or drop her an email and we can take you out for a test ride. And Stuart, over to you.

A - Stuart Bowers {BIO 20627575 <GO>}

(Video Presentation) So that's actually a clip of a longer than 30-minute uninterrupted drive with no interventions navigating on upon the highway system which is in production today in hundreds of thousands of cars. So I'm Stuart and I'm here to talk about how we build so many systems at scale, just like a really short induction I'm kind of where I'm coming from, and what I do. So I've been in a couple of companies or less I've been in writing software profession for about 12 years. The thing that excites me most and I'm really passionate about is taking the cutting edge of machine learning and actually connecting that with customers through robustness and scale. So at Facebook, I worked initially inside of our ads infrastructure to build some of the machinery learning, some really really smart people. And we actually tried to build that into a single platform that we could then scale to all the other aspects of the business from how we rank the News Feed to how we deliver search results to how we make every recommendation across the platform. And that became the applied machine learning group, that is something I was incredibly proud of and a lot of that wasn't just the core algorithms and the really important improvements that happened there those that matters, a lot of actually the engineering practices to build these systems at scale and the same thing is true it's Snap where I went where we were really, really excited to sort of actually help to monetize this product.

But the hardest part is we were using Google at the time and they were effectively running us on a fairly small scale and we wanted to build that same infrastructure, we take an understanding of these users, connect that with cutting edge machine learning, build that at massive scale and handed billions and trillions of predictions and auctions every day in which is really robust. And so when opportunity came to come to Tesla that's something I'm really incredibly excited to do which is specifically take the amazing things that are happening both in the hardware side and the computer vision and AI side and actually package that together with all the planning, the controls, the testing, the kernel patching of the operating system all of our continuous integration, our simulation actually build that into a product we get onto people's cars in production today. And so I want to talk about the timeline for how we did that with Navigate on Autopilot and how we're going to do that as we get Navigate on Autopilot off the highway and on to city streets.

So we're at 770 million miles already for Navigate on Autopilot. Something really, really, really cool and I think one thing that is worth kind of calling out on this is that we're continuing to accelerate and keep learning from this data like Andrej talked about this data engine as this accelerates up, we actually do make more and more assertive lane changes, we are learning from these cases where we will intervene either because they failed to detect a merge correctly or because they wanted the car to be a little more peppy in different environments. And we just want to keep making that progress. So to start all of this, we begin with trying to understand the world around us. And we talked about the different sensors in the vehicle. But I want to like dig in a little bit more here. We have eight cameras but then we also have additionally 12 ultrasonic sensors, a radar, an inertial measurement unit, GPS, and then one thing we forget about also the (inaudible) and steering actions. So not only can we look at what's happening around the vehicle, we can look at how humans chose to interact with that environment.

And so I'll talk in this clip right now. This basically is showing what's happening today in the car and we're continuing to push this forward. So we start with a single neural network. We see the detections around it, we then build all that together with multiple neural networks in multiple detections. We bring in the other sensors and we convert that into what it calls a vector space an understanding of the world around us. And this is something where as we continue to get better and better at this we're moving more and more of this logic into the neural networks themselves. And the obvious end game here is that the neural network looks across all the cars, brings in all the information together and just ultimately outputs a source of truth for the world around us. And this is actually not like an artist rendering of me since this is actually the output of one of the debugging tools that we use on a team everyday to understand what the world looks like around us.

So another thing I think is really exciting to me I think when I do hear about sensors like LIDAR, a common question is around just having extra sensor modalities like why not have some redundancy in the vehicle and I want to dig in on one thing that's not -- it's not always obvious with neural networks themselves. So we have a neural network running on our say wide fisheye camera, that neural network is not making one prediction about the world. It's making many separate predictions some of which actually audit each other. So as a real example, we have the ability to detect a pedestrian. That's something we train very very carefully on and put a lot of work into. We also have the ability to detect obstacles in the roadway and a pedestrian is an obstacle. And it's shown differently to the neural network. It says, oh, there is a thing I can't drive through and these together combined to give us an increased sense of what we can and can't do in front of a vehicle and how to plan for that.

We then do this across multiple cameras because we have overlapping fields of view in many places around the vehicle. In front, we have a particularly large number of overlapping fields of view. Lastly, we can combine that with things like the radar and the ultrasonics to build this extremely precise understanding what's happening front of the car. We can use that both to learn future behaviors that are very accurate, we can also build very accurate predictions of how things will continue to happen in front of us. So one example is really exciting is we actually look at bicyclists and people and not just ask where are you now but where are you going. And this is actually the heart of our next generation automatic emergency braking system which will not just stop the people in your path, will separate who are going to be in your path and that's running a shadow mode right now. We'll go to the fleet this quarter. I'll talk about shadow mode in a second.

So when you want to start a feature like this for Navigate on Autopilot on the highway system, you can start by learning from data and you can just look at how humans do things today, what is their assertiveness profile, how do they change lanes, what causes them to either abort or change like their maneuvers and you can see things that are not immediately obvious like, oh yes simultaneous merging is rare, but very complicated and very important and you can start to build opinions about different scenarios such as a fast overtaking vehicle. So this is what we do when you initially have some algorithm you want to try out we can put them on the fleet and we can see what they would have done in a real world scenario such as this car is overtaking us very quickly, this is taken from our actual simulation environment showing different paths that we have considered taking and how those overlay on the real world behavior of a user.

When you get those algorithms tuned up when you feel good about them specifically. This is really taking that out within neural own network putting it in that vector space and building and tuning these parameters on top of it. Ultimately I think we can do through more and more machine learning, you go out to a controlled deployment which for us is our early access program. And as you get this out to a couple of thousand people who are really excited to give you highly vigilant but useful feedback about how this behaves not an open loop and a closed loop way in the real world and you watch their interventions and we talk about it like when somebody takes over we can actually get that clip, try to understand what happens. And one thing we can really do is we can actually play this back again in an open loop way and ask. As we build our software, are we getting closer or farther from how humans behave in the real world. And one thing which is super cool with the full self-driving computers we're actually building our own racks and infrastructure to basically compete for a full self-driving computers fully wrapped up build these into our own cluster and actually run this very sophisticated data infrastructure to actually understand over time as we tune and fix these algorithms are we getting closer and closer to how humans behave and ultimately can we exceed their capabilities and so once we have this we feel really good about it, we want to do our wide rollout.

But to start we actually asked everybody to confirm the car's behavior via stock confirm, and so we started making lots and lots of predictions about how we should be navigating the highway. We asked people to tell us, is this right or is this wrong. And this is again a chance to turn that data engine. And we did spot some really tricky and interesting long tails of -- in this case so I think really for example like this are these very interesting cases of simultaneous merging where you start going and then somebody moves either behind or before you not noticing you. And what is the appropriate behavior here and what are the tunings of the neural network we need to do to be super precise about the appropriate behaviors here. We worked, we tuned these in the background, we made them better. And over the course of time we've got nine million successfully accepted lane changes and we use these again with our continuous integration infrastructure to actually understand how do we think we're ready and this is one thing where full self-driving is really exciting to me since we own the entire software stack, straight from the kernel patching, all the way to the ice bucket -- the tuning on the image signal processor, we can start to collect even more data that is even more accurate and this allows us to do even better and better tuning this faster iteration cycles.

And so earlier this month we were kind of like -- we're ready to deploy and even more seamless version of Navigate on Autopilot on the highway system. And that seamless version does not require stock confirms. So you can sit there, relax put your hand on the wheel and just oversee what the car is doing. And in this case we're actually seeing over 100,000 automated lane changes every single day on the highway system. It is something just like super cooled us to the point scale and the thing that kind of most excited up from all this is the actual lifecycle of this and how we actually turn that data engine crank faster and faster and faster with time. And I think one thing is really becoming very clear is the combination of the infrastructure we have built, the tooling we built on top of that and the combined power of the full self-driving computer. I believe we can do this even faster as we move now to be an Autopilot from the highway system onto city streets.

And so with that I'll hand off to Elon.

A - Elon R. Musk {BIO 1954518 <GO>}

To the best of my knowledge, all those lane changes have occurred with zero accidents.

A - Stuart Bowers {BIO 20627575 <GO>}

That is correct.

Yes I watch every single accidents.

A - Elon R. Musk {BIO 1954518 <GO>}

So it's conservative obviously. But it's hundreds of thousands going to millions of lane changes and zero accidents is I think a great achievement by the team.

A - Stuart Bowers {BIO 20627575 <GO>}

Thank you.

A - Elon R. Musk {BIO 1954518 <GO>}

So let's see a few other things that are for me worth mentioning, in order to have a self-driving car or robot taxi, you really need redundancy throughout the vehicle at the hardware level. So starting in, maybe it was October 2016, all cars made by Tesla have redundant power steering, so redundant motors in the power steering, so any one failure of the -- If the motor fails, the car can still steer, all of the power and data lines have redundancy, so you can sever any given power line or any data line and the car will keep driving. The auxiliary power system even if the main pack, you lose complete power in the main pack, the car's capable of steering and braking using the auxiliary power system, you can completely lose the main pack and the car is safe.

The whole system from a hardware standpoint has been designed for -- to be a robot taxi since basically October 2016, so when we rolled out hardware autopilot Version two, we do not expect to upgrade cars made before that, we think it would actually cost more to make a new car than to upgrade the cars, just to give a sense of how hard it is to do this. Unless this is designed then, it's not worth it.

So we've gone through a future of self-driving where it's glitz, it's hardware, it's vision and then there's a lot of software and there's -- the software problem here should not be minimized some massive software problem that -- yes managing vast amounts of data, training against the data, how do you control the car based on the vision, it's a very difficult software problem.

So going after -- going over just like Tesla master plan, obviously we've made a bunch of forward looking statements as they call it and let's go through some of our other forward-looking statements that we've made, you know way back, when we created the company, we separate both Tesla Roadster, they said it was impossible and that even if we did build it, nobody would buy it. This is like a universal opinion was that building an electric car was extremely dumb and would fail. I agreed with them that probability of failure was high but that this was important. So we built the Tesla Roadster (inaudible) in 2008 and shipping that car. It's not collector's item. Then (inaudible) the more affordable car with the Model S, we did that, again, we were told that's impossible. I was called a fraud and a liar and it was not going to happen, it is all untrue. Okay. Famous last words now is we're in production with the Model S in 2012, exceeded all expectations. There is still in 2019 no car that can compete with Model S of 2012.

It's seven years later, still waiting as it would build a affordable car very highly affordable, as an affordable, more affordable with the Model 3. We bought the Model 3. We're in production. I said we'd get over 5000 cars of Model 3, at this point 5000 cars a week is a walk in the park for us. It's not even hard. So we do large scale solar. We did through the (inaudible) acquisition and that would develop into place solar roof which is going really well. We're now in Version 3 of the solar tile roof and we expect to spool up production of the solar tile roof significantly later this year. I have it on my house and it's great and I sort of make the power roll and the power pack. We made the power roll and power pack, in fact the power pack is now deployed in massive grid scale utility systems around the world including the largest operating battery projects in the world that above 100 megawatts and in the next or probably by next year or two years at the most we expect to have a gigawatt scale battery project completed.

So all these things I said would do them, we did it. So we do it. We did it. We're going to do the robot taxi thing too. Only criticism and it's a fair one and sometimes I'm not on time. But I get it done and the Tesla team gets it done. So what we're going to do this year is we're going to reach combined production of 10,000 a week between our S 6 and 3. I feel very confident about that and we feel very confident about being future complete with self-driving. Next year, we'll expand the product line with model Y and Semi, and we expect to have the first operating robot taxis next year with no one in them, next year.

It's always difficult to like when things are an exponential -- at an exponential rate of improvement, it's very difficult to correct one's mind around it because we're used to extrapolating on a linear basis but when you've got massive amounts of like as the hardware -- massive as a hardware on the road that the cumulative data is increasing exponentially the software is getting better at an exponential rate, I feel very confident predicting autonomous robot taxis for Tesla next year not in all jurisdictions -- not in all jurisdictions because we won't have regulatory approval everywhere but I'm confident we'll have least regulatory approvals somewhere literally next year.

So any customer will be able to add or remove their car to the Tesla network. So expect this to operate it is somewhat sort of like a combination of maybe the Uber and Airbnb model. So if you own the car you can add or subtract it to the Tesla network and Tesla would take 25% or 30% of the revenue and then in places where there aren't enough people sharing their cars we would just have dedicated Tesla vehicles. So, when you use the car we'll show you our ride sharing app, you will just be able to summon the car from the parking lot, get in and go for a drive.

It's really simple to just take the same Tesla app that you currently have, we'll just update the app and add as summon Tesla or commit your car to the fleet. So I see that summon your car or add at -- summon a Tesla or add your -- add or subtract your car to the fleet, you'll be able to do that from your phone. So we see the potential for smoothing out the demand just pushing curve and having a car operate at a much higher utility than an old car operator, so like typically the use of a car is about 10 hours to 12 hours a week. So most people will drive 1.5 hours to 2 hours a day, typically 10 hours to 12 hours a week of total driving. But if you have a car that can operate autonomously, then most likely you could probably -- most likely you could have that car operate for a third of the week or longer. So there are 168 hours in a week. So probably you've got something on the order of 55 hours to 60 hours a week of operation, maybe a bit longer, so that the fundamental utility of vehicle increases by a factor of 5. So you can look at this from a macroeconomic standpoint and say just if this was like some if we were operating some vague simulation, if you could upgrade your simulation to increase the utility of cars by a factor of 5, that would be a massive increase in the economic efficiency of the simulation, just gigantic. So we'll do Model 3 -- S3 and X as taxis, but we've made an important change to our leases. So if you lease a Model 3, you don't have the option of buying it at the end of the lease. We want them back. If you buy the car, you can keep it, but if you lease it, you have to give it back.

And as I said, where -- in any locations where there's not enough to supply for sharing, Tesla will just make its own cars and add them to the network in that place. So the current cost of Model 3 Robotaxi is less than $38,000. We expect that number to improve over time and resigning the cars, the cars currently being built are all designed for a million miles of operation. So the drive units it's line and test and validated for million miles of operation. The current battery pack is about maybe 300,000 miles to 500,000 miles. The new battery pack that probably go into production next year is designed exclusively for a million miles of operation. The entire vehicle battery pack inclusive is designed to operate for a million miles with minimal maintenance to actually adjusting tyre design and really optimizing the car for a high proficient Robotaxi.

And at some point, you won't need steering wheels or pedals and we'll just leave those. So as these things become less and less important, we just leave parts, just they won't be there. If you say like, probably, two years from now, we will make a car that has no steering wheels or pedals. And if we need to accelerate that time, we can always just delete parts, easy. And probably say long term, three years, Robotaxis with eliminated parts, maybe it ends up being $25,000 or less, and you want a super efficient car, so the electricity consumption is very low. So we're currently at 1.5 miles per kilowatt hour but we can improve that to five and beyond.

And there is just really no company that has the full stack integration. We've got the vehicle design and manufacturing, we've got the computer hardware in-house, we've got the in-house software development and AI, and we've got by far the biggest suite. It's extremely difficult, not impossible perhaps but extremely difficult to catch up when Tesla has 100 times more miles per day than everyone else combined. This is the cost of running a gasoline car or the average cost of running a car in the US is taken from AAA. So it's currently about $0.62 a mile. It doesn't have 1,000 miles or 15 million vehicles. It adds up to 2 trillion a year. These are literally just taken from the AAA website.

The cost of ridesharing is according to Uber and Lyft is $2 to $3 a mile. The cost to run a Robotaxi, we think less than $0.18 a mile, and dropping. Like this is -- this would be current cost, future cost will be lower. If you say, what would be the probable gross profit from a single Robotaxi. We think probably something on order of $30,000 per year, and we expect that we're designing the cars the same way that commercial semi-trailer -- semi trucks are designed. Commercial semi-trucks are all designed for a million-mile life, and we're designing the cars for a million-mile life as well. So in nominal dollars that would be a little over $300,000 over the course of 11 years, maybe higher. I think this consumption is actually relatively conservative, and this assumes that 50% of the miles driven are -- there's nothing or not useful. So this is only at 50% utility.

By the middle of next year we'll have over a million Tesla cars on the road with full self-driving hardware, feature complete at a reliably level that we would consider that no one needs to pay attention. Meaning if you could go to sleep and you -- from our standpoint, if you fast forward a year, to look maybe a year, maybe a year and three months, but next year for sure. We will have over a million Robotaxis on the road.

The fleet wakes up within over the year update, that's all it takes. And say what is net present value of Robotaxi, probably, on the order of a $200,000. So buying a Model 3 is a good deal.

Any questions?

Q - Unidentified Participant

Well, I mean, you want your fleet, your our own fleet.

A - Unidentified Speaker

In our own fleet, I don't know, I guess long-term, we have probably on the order of 10 million vehicles. I mean our production rates generally -- if you look at a compound annual production rate since 2012 which is like the -- that's our first full year of more or less production. We went from 23,000 vehicles produced in 2013 to around 250,000 vehicles produced last year. So in the course of five years, we increased output by a factor of 10. I would expect that something similar occurs over the next five or six years. As for sharing, sharing versus I don't know. The nice thing is that essentially customers are fronting us the money for the car. It's great.

Q - Analyst

So in terms of that one thing is the snake charger, I'm curious about that and also how did you determine the pricing? It looks like you're undercutting the average Lyft or Uber ride by about 50%. So I'm curious if you could talk a little bit about the pricing strategy?

A - Elon R. Musk {BIO 1954518 <GO>}

Sure. We would expect the -- solving for the snake charger, it's pretty straightforward from a vision standpoint. It's like a known situation. Any kind of known situation with vision is like a charge port is trivial. So yes, the cars would just automatically park and automatically plug in. There would be no one -- no human supervision required. Yes, so that was a pricing. We just threw some numbers on there. I mean I think definitely plug in whatever pricing you think makes sense. We're just kind of randomly said okay maybe a $1.

And the things like it's -- there's like on order of 2 billion cars and trucks in the world. So Robotaxis will be in extremely high demand for a very long time. And from my observation, as far as the auto industry is very slow to adapt. I mean like I said there's still not a car on the road that you can buy today that is as good as the Model S was in 2012. So that suggests a pretty slow rate of adaptation for the car industry and so probably $1 is conservative for the next 10 years, because if you also sort of think like -- there's like actually not enough appreciation for the difficulty of manufacturing. Manufacturing is insanely difficult.

A lot of people I talk to think like, if you just have the right design you can like instantly make as much of that thing as the world wants. This is not true. It's extremely hard to design a new manufacturing system for new technology. I mean Audi is having major problems manufacturing E-tron and they are extremely good at manufacturing. And if they're having problems what about others. So there is on the order of 2 billion cars and trucks in the world, on the order of about 100 million units per year of production capacity of vehicles, but only if the old design, it will take a very long time to convert all of that to full self-driving cars, and they really need to be electric because the cost of operation of a gasoline diesel car is much higher than the electric car. So any Robotaxi that isn't electric will absolutely not be competitive.

Q - Colin Rusch {BIO 15823117 <GO>}

Elon, it's Colin Rusch from Oppenheimer over here. Obviously, we appreciate that the customers are fronting some of the cash for this fleet and getting built up, but it sounds like a massive balance sheet commitment from the organization over the course of time. Can you talk a little bit about what that looks like? What are your expectations are in terms of financing over the next call it three years, three or four years for building up this fleet and starting to monetize it with your customer base?

A - Elon R. Musk {BIO 1954518 <GO>}

We're aiming to beat approximately cash flow neutral during the fleet build up phase, and then are expected to be extremely cash flow positive once the Robotaxis are enabled. But I don't want to talk about financing rounds, It would be difficult about financing rounds in this venue, but I think we'll make the right move. I think we'll make the moves we think we should make.

Q - Unidentified Participant

Other question, if I'm Uber, why wouldn't I just buy all your cars? Why would I let you put me out of business?

A - Elon R. Musk {BIO 1954518 <GO>}

There's a clause that we put into our cars. I think it was about three or four years ago. They can only be used in the Tesla network.

Q - Unidentified Participant

So even a private person like if I go out and buy 10 Model 3s, I can't -- I can run on the network that's a business now, right.

A - Elon R. Musk {BIO 1954518 <GO>}

You're only allowed to use Tesla network.

Q - Unidentified Participant

Right. But if I use the Tesla network in theory I could run a car showing Robotaxi business with my 10 Model 3s.

A - Elon R. Musk {BIO 1954518 <GO>}

Yes, but it's like the app store that you can only -- you can add -- only add or remove them through the Tesla network and then Tesla gets revenue share.

Q - Unidentified Participant

But that's similar to Airbnb though in that I have this home my car and now I can just rent them out, so I can make an extra income from owning multiple cars and just renting them out. Like I have a Model 3, I aspire to get this roadster here next, when you build it and I'm going to just rent my Model 3 out, why would I give it back to you?

A - Elon R. Musk {BIO 1954518 <GO>}

I guess you could operate a rental car fleet, but I think this is very unwieldy. Yes.

Q - Unidentified Participant

I don't know, it seems easy.

A - Elon R. Musk {BIO 1954518 <GO>}

Okay. Try it.

Q - Unidentified Participant

In order to operate a Robotaxi network, it sounds like you have to solve certain problems like for example auto pilot today if you over steer it lets you take over. But if it's a ridesharing product that someone else is getting in the passenger seat, like moving the steering can't let that person take over the car for example, because they might not even be in the driver seat. So is the hardware already there for it to be a Robotaxi and it might get into situations such as a cop pulling it over where some human might need to intervene. Like using central fleet of operators that remotely sort of interact with humans or I mean is all of that type of infrastructure already built in to each of the cars? Does that make sense?

A - Elon R. Musk {BIO 1954518 <GO>}

I think there will be sort of a phone home thing, where if the car gets stuck it will just phone home to Tesla and ask for a solution. Things like being pulled over for -- by a police officer, that's easy for us to program in that's not a problem. But it will be possible for somebody to take over using the steering wheel at least for some period of time and then probably down the road, we'll just cap this steering wheel, so there's no steering control. We'll just take steering wheel put a cap on and give it like a couple of years.

Q - Unidentified Participant

Hardware modification to the car in order for it to enable that or?

A - Pete Bannon {BIO 20590065 <GO>}

Yes. We'll literally just unbolt the steering wheel and put a cap on where the steering wheel handle currently is.

Q - Unidentified Participant

But that is a like future car that you would put out. But what about today's cars where the steering wheel is a mechanism to take over autopilot. Like, so if it's in a Robotaxi mode would someone be able to take it over by just simply moving the steering wheel type?

A - Elon R. Musk {BIO 1954518 <GO>}

Yes. I think there'll be a transition period where people will be able to take over and should be able to take over from the Robotaxi. And then once regulators are comfortable with us not having a steering wheel, we'll just delete that. And for cars that are on -- there in the fleet, obviously with the permission of the owner, if it's owned by somebody else, we would just take the steering wheel off and put a cap where the steering wheel currently touches.

Q - Unidentified Participant

So there might be like two phases to Robotaxi. One, where the service is provided and you come in as the driver but could potentially take over and then in the future there might not be a driver option. Is that how you see it as well or like...

A - Elon R. Musk {BIO 1954518 <GO>}

In the future, there won't -- in future will -- the probability of the steering wheel being taken away in the future is 100%. Consumers will demand it.

Q - Unidentified Participant

But initially you would...

A - Elon R. Musk {BIO 1954518 <GO>}

This is not -- this is not -- I want to be clear. This is not me proscribing a point of view about the world. This is me predicting what consumers will demand. Yes consumers will demand in the future that people are not allowed to drive these two ton death machines.

Q - Unidentified Participant

I totally totally agree with that. But in order for a Model 3 today to be part of the Robotaxi network when you call it, you would then get into the driver seat essentially because just to be on the safe.

A - Elon R. Musk {BIO 1954518 <GO>}

That's right.

Q - Unidentified Participant

Does that make sense. Thank you , bye.

A - Elon R. Musk {BIO 1954518 <GO>}

Exactly.

Q - Analyst

Thank you.

A - Elon R. Musk {BIO 1954518 <GO>}

Which is a sort of like there were amphibians but then pretty much that things just become like land creatures. There will be a little bit of surviving amphibian phase.

Q - Unidentified Participant

Hi. The strategy we've heard from other players in the Robotaxi space is to select a certain municipal area to create a geo-fenced self-driving that way you're using an HD map to have a more confined area with a bit more safety. a) we didn't hear much today around the importance of HD maps. To what extent is an HD map necessary for you. And the second, we also didn't hear much about deploying this into specific municipalities where you're working with the municipality to get the buy in from them and you're also getting a more defined area. So what's the importance of HD maps and to what extent are you looking at specific municipalities for rollout?

A - Elon R. Musk {BIO 1954518 <GO>}

I think HD maps are a mistake. We actually had HD maps for a while. I actually can't count that, because you either need HD maps in which case if anything changes about the environment, the car will break down or you don't need HD maps in which case why you're wasting your time doing HD maps. So the HD maps thing like the two main crutches that are -- that should not be used and won't -- in retrospect, retrospect the obviously false and foolish are LIDAR and HD maps. Mark my words.

Q - Unidentified Participant

Hello.

A - Elon R. Musk {BIO 1954518 <GO>}

If you need a geo-fenced area, you don't have real self-driving.

Q - Unidentified Participant

Just, it sounds like maybe battery supply could be the only bottleneck left towards this vision. And also could you just clarify how you get the battery packs to last a million miles?

A - Elon R. Musk {BIO 1954518 <GO>}

I think cells will be a constraint, that's a subject for a whole separate -- that's a whole separate subject. And I think we're actually going to want to push us sort of standard range plus battery more than our long range battery because the energy content in the long range pack is 50% higher kilowatt hours. So, essentially, you can make a third more cars if you just -- if they all sort of stand range plus instead of the long range pack, it's the ones like around 50 kilowatt hours, the older ones around 75 kilowatt hours. So we're actually calling it a bias, our sales intentionally towards the smaller battery pack in order to have higher volume of what -- basically -- but the obvious next thing I think here is to maximize the number of autonomous units or the number of -- maximize the output that will subsequently result in the biggest autonomously down the road.

So we are making -- doing a number of things in that regard, but it's just not for today's meeting.

Q - Unidentified Participant

And the million mile life, is that ...

A - Elon R. Musk {BIO 1954518 <GO>}

And the million mile-life is basically just of not getting the cycle life of the pack to -- you need basically, in order, like we say we've got basic math, if you got a 250 mile range pack, you're going to need 4,000 cycles. Very achievable. We already do that with our stationary storage, stationary storage solutions like power pack, we are ready to play power pack with 4,000 cycle life capability.

Q - Unidentified Participant

Can I ask a...

A - Elon R. Musk {BIO 1954518 <GO>}

Sorry.

Q - Unidentified Participant

I want to ask...

A - Elon R. Musk {BIO 1954518 <GO>}

Its like ventriloquism.

Q - Unidentified Participant

It's obviously significantly very constructive margin implications to the extent you can drive attach rates much higher over the full self-driving option. I'd just be curious if you can level us that kind of where you are in terms of those attach rates and how you expect to educate consumers about the Robotaxi scenario so that attach rates do materially improve over time?

A - Elon R. Musk {BIO 1954518 <GO>}

Sorry it's a bit hard to hear your question.

Q - Unidentified Participant

Yeah, do -- just curious where we are today in terms of full self-driving attach rates in terms of the financial implications? I think it's hugely beneficial if those attach rates materially increase because of the higher gross margin dollar that flow through to the extent people do sign up for full FST. Just curious how you see that ramping? Or what the attach rates are today versus when do you expect -- how do you expect to educate consumers and get them aware that they should be attaching FST to their vehicle purchases?

A - Elon R. Musk {BIO 1954518 <GO>}

We are going to ramp that up massively after today. Yes. I mean if the fundamental -- really fundamental message that consumers should be taking today is that it's financially insane to buy anything other than a Tesla. It will be like owning a horse in three years. I mean, fine if you don't own a horse but you should go into it with that expectation. If you buy a car that does not -- that does not have the hardware necessary for full self-driving, it is like buying a horse. And the only car that has the hardware necessary for full self-driving is Tesla.

Like you should really think about their purchase any other vehicle. That's basically crazy to buy any other car than Tesla. We need to make that

-- convey that augment clearly and we will have today.

Q - Unidentified Participant

Perfect. Thanks for bringing the future to present very informational time today. I was wondering like you did not talk much about Tesla pickup and let me give a context for that. I could be wrong but the way I'm looking at Tesla network, it will as a early adopter and something as a test bread, I think Tesla's pick up maybe the first phase of putting the vehicles in network, because the utility of Tesla pickup would be pretty much people who are either loading a lot of stuff or are in the profession of construction or little here and there odd items, like picking up stuff from Home Depot. I would say that maybe it needs to have a two stage process, pickup trucks, exclusively for Tesla network as a starting point and people like me can buy them later. But what are your thoughts on that?

A - Elon R. Musk {BIO 1954518 <GO>}

Well today was really just about Autonomy. There's a lot that we could talk about such as cell production, pickup truck in future vehicles, but today was to just focus on Autonomy. I agree it's a major thing. I'm excited for the Tesla pickup truck unveiling this year. It's going to be great.

Q - Colin Langan {BIO 15908877 <GO>}

Colin Langan, UBS. Just so we understand the definitions, when you refer to feature complete self-driving it sounds like you're talking Level 5 in geo- fence is that what's expected by the end of the year just so more all the same thing. And then the regulatory process. I mean have you talked to regulators about this? This seems quite an aggressive timeline from what other people have put out there. I mean are they -- what are the hurdles that are needed and what is the timeline to get approval and do you need things like in California, they're tracking miles that -- with an operator line that you need those things. What does that process going to look like?

A - Elon R. Musk {BIO 1954518 <GO>}

Yes. We talked to regulators around the world all the time, as we introduce additional features like navigate and autopilot, we -- this requires like a regulatory approval on a project jurisdiction basis. So, but I think fundamentally, regulators in my experience are convinced by data. So if you have a massive amount of data that shows that autonomy is safe, they listen to it. They may take time to digest the information. That process may take a bit of time, but they have always come to the right conclusion from what I've seen.

Q - Unidentified Participant

Now I have a question over here.

A - Elon R. Musk {BIO 1954518 <GO>}

And it's got lights in my eyes and a pillar. Okay.

Q - Unidentified Participant

I just wanted just to -- some of the work we've done trying to better understand the ride hail market. It looks like it's very concentrated in major dense urban centers. So is the way to think about this that the robo taxis would probably deploy more into that area, and the additional full self driving for personally own vehicles would be in these suburban areas?

A - Elon R. Musk {BIO 1954518 <GO>}

I think like probably, yes, like Tesla owned robo taxis would be in dense urban areas along with customer vehicles. And then as you get to medium and low density areas it would tend to be more that people own the car and occasionally lend it out.

Q - Unidentified Participant

(Inaudible)

A - Elon R. Musk {BIO 1954518 <GO>}

Yes. There are a lot of edge cases in Manhattan and say downtown San Francisco, but those are -- and there are various areas around the world that that have a challenging urban environments. But we do not expect this to be a significant issue. When I say feature complete I mean it will work in downtown San Francisco and downtown Manhattan this year.

Q - Unidentified Participant

Hi. I have a neural net architecture question. Do you use different models for say path planning and perception or different types of AI and sort of how do you split up that problem across the different pieces of autonomy.

A - Elon R. Musk {BIO 1954518 <GO>}

Well essentially -- right now AI neural nets we use it really for object recognition and we're still basically just using it as still frames. So identifying objects in still frames and tying it together in a perception path planning layer thereafter. But what's happening is steadily is that the neural net is kind of heating into the software base more and more. And so over time we expect the neural net to do more and more now from a computational cost standpoint there are some things that are very simple for heuristics and very difficult for neural net. And so it probably makes sense to maintain some level of heuristics in the system because they're just computationally a thousand times easier than a neural net. Neural net is like a cruise missile. And if you try to swat a fly, just use a fly swatter, not a cruise missile.

So but overtime I would expect that it moves really to just training it on against video, and then video-in car steering and pedals out or basically video in that lateral and longitudinal acceleration out, almost entirely, At last that's what we're going to use the Dojo system for. There's no system that can currently do that.

Q - Unidentified Participant

And maybe over here. Just going back to the sensor suite discussion on the -- one area I'd like to talk about is a lack of side radars in a situation where you have an intersection with a stop sign where there's maybe a 35 mile or 40 mile per hour cross traffic. Are you comfortable with the sensor suite the side cameras being able to handle that? Just maybe talk about that?

A - Elon R. Musk {BIO 1954518 <GO>}

Yes. No problem. It is essentially the car is going to do kind of what a human would do. Like think humans like basically a camera on a slow gimbal. And it's quite remarkable that people are able to drive the car in the way that they are. Because if -- you can't look in all directions at once. The car can literally look in all directions at once with multiple cameras. So humans are able to drive just by sort of looking this way or looking that way they're actually stuck in their driver's seat. They can't really get out the driver seat. So it's kind of one camera on a gimbal and is able to drive, a conscientious driver and drive with very high safety.

The cameras in the cars have a better advantage point than the person. So they're like up in the B pillar or at in front of the rear view mirror. They've really got a great vantage point. So if you turning on to a road that's got a lot of high speed traffic, you can just do a quest there's just like gradual like turn a little bit then go fully into the road let the camera see what's going on. And if things look good and then the rear cameras don't show any oncoming traffic off you go, and if it looks sketchy you can pull back a little bit just like a person. The behaviors like remarkably, starts to become remarkably lifelike. It's like quite eerie actually. A good car just starts behaving like a person.

Q - Unidentified Participant

Over here. Here you go. Ventriloquist right here.

A - Elon R. Musk {BIO 1954518 <GO>}

Okay.

Q - Unidentified Participant

Given all the value you're creating in your auto business by wrapping all of this technology around yourselves, I guess I'm curious as to why you would still be taking some of your cell capacity and putting it into power wall and Power Pack wouldn't it make sense to put every single unit you can make into this part of your business?

A - Elon R. Musk {BIO 1954518 <GO>}

We've already stolen almost all the cell lines for that we're meant to go to power will and power pack and use them for Model 3. I mean last year in order to make our Model 3 production and not be cell starved, we had to convert all of the 2170 lines at the Gigafactory to car cells. The -- so actual output in total gigawatt hours of stationary storage compared to vehicles is an order of magnitude different. And for stationary storage, we can basically use a whole bunch of miscellaneous cells out there. So we can just gather cells of -- from multiple suppliers all around the world, and you don't have a homologation issue or a safety issue like you have with cars. That's basically -- our stationary battery business has been just kind of beating over scraps for quite a while. But like -- everything like the production is being, there are many, many constraints of a massive production system. It's like the degree to which manufacturing a supply chain is underappreciated is amazing. There are a whole series of constraints. And what is the constraint in one week, may not be the constraint in another week. It's insanely difficult to make a car, especially one which is rapidly evolving. So yeah. But I'll just take a few more questions, and then, I think we should just break for, so you can cut out the cars.

Q - Adam Jonas {BIO 3339456 <GO>}

Hi Elon, Adam Jonas. Questions on safety. What data can you share with us today, how safe this technology is, which should obviously be important in a regulatory or insurance discussion.

A - Elon R. Musk {BIO 1954518 <GO>}

We published the accidents per mile every quarter, and what we see right now is that our autopilot is about twice as safe as a normal driver on average, and we expect that to increase quite a bit over time. Like I said, in the future, it will be consumers will want to outlaw, I don't think they will succeed or am I saying I agree with this position. But in the future, consumers will want to outlaw people driving their own cars, because there's unsafe. If things like elevators, elevators use to be operated on a big lever, like go up and down the floor, there's like a big relay, and yet elevator operators, but then periodically they would get tired or drunk or something and then they would turn the lever at the wrong time and sever somebody in half. So now you do not have tell elevator operators, and you would be quite alarming if you went into an elevator that had a big lever that could just move between floors arbitrarily. So there's just buttons. And in the long-term, again not a value judgment, I'm not saying I want the world to be this way, I'm saying consumers will most likely demanded that -- the people in our live broadcast.

Q - Adam Jonas {BIO 3339456 <GO>}

And Elon, my follow up. Can you share with us how much Tesla's spending on autopilot or autonomous technology, by order of magnitude on an annual basis? Thank you.

A - Elon R. Musk {BIO 1954518 <GO>}

It's basically our entire expense structure.

Q - Unidentified Participant

Question on the economics of the Tesla network. Just so understand, it look like, so you get a Model 3 off lease. $25,000 goes on the balance sheet would be an asset. And then you -- it would cash flow $30,000 a year, roughly is that the way to think about?

A - Elon R. Musk {BIO 1954518 <GO>}

Yes. (Multiple Speakers) Something like that, yeah.

Q - Unidentified Participant

And then just in terms of financing of it, there's a question earlier you mentioned you would do it. Is it cash flow neutral to the robo-taxi program? Or cash flow neutral to

Tesla as a whole.

A - Elon R. Musk {BIO 1954518 <GO>}

Sorry. The cash flow neutral to...

Q - Unidentified Participant

In terms of -- he asked a question about financing the robo-taxi yet. It looks to me like they're self financing. But, you mentioned there would be basically cash flow neutral. Is that what you're referring to?

A - Elon R. Musk {BIO 1954518 <GO>}

I just think between now and when the robo-taxis are fully deployed throughout the world, the sensible thing for us is to maximize rate and drive the company to cash flow neutral.

Q - Unidentified Participant

Okay, In items of...

A - Elon R. Musk {BIO 1954518 <GO>}

Once the robo-taxi fleet is active, I would expect it to be extremely cash flow positive.

Q - Unidentified Participant

In this -- so you were talking about production.

A - Elon R. Musk {BIO 1954518 <GO>}

Yeah.

Q - Unidentified Participant

To produce. Okay thanks.

A - Elon R. Musk {BIO 1954518 <GO>}

Maximize the number of autonomous units made.

Q - Unidentified Participant

Thank you.

A - Elon R. Musk {BIO 1954518 <GO>}

I guess, maybe one last question.

Q - Unidentified Participant

Hi, if I add my Tesla to the robo-taxi network, who is liable for an accident? Is it Tesla or is it me? If the vehicle has an accident and harms somebody.

A - Elon R. Musk {BIO 1954518 <GO>}

I mean, it's probably Tesla. It's probably Tesla. I think the right thing to do is to make sure there are very, very few accidents.

Alright. Thanks everyone. Please enjoy the price. Thank you.

A - Pete Bannon {BIO 20590065 <GO>}

Thank you, very much.