0:00

[Music]

0:05

ever since computers were invented

0:07

they've really just been glorified

0:08

calculators machines that execute the

0:11

exact instructions given to them by the

0:13

programmers but something incredible is

0:15

happening now computers have started

0:16

gaining the ability to learn and think

0:19

and communicate just like we do they can

0:21

do creative intellectual work that

0:23

previously only humans could do we call

0:25

this technology generative Ai and you

0:27

may have encountered it already through

0:29

products like GPT basically intelligence

0:32

is now available as a service kind of

0:34

like a giant brain floating in the sky

0:36

that anyone can talk to it's not perfect

0:39

but it is surprisingly capable and it is

0:40

improving at an exponential rate this is

0:43

a big deal it's going to affect just

0:45

about every person and Company on the

0:47

planet positively or negatively this

0:49

video is here to help you understand

0:51

what generative AI is all about in

0:53

Practical terms beyond the hype the

0:54

better you understand this technology as

0:56

a person team or company the better

0:58

equipped you will be to survive and

1:00

thrive in the age of AI so here's a

Einstein in your basement

1:03

silly but useful mental model for this

1:05

you have Einstein in your basement in

1:07

fact everyone does and by Einstein I

1:10

really mean the combination of every

1:12

smart person who ever lived you can talk

1:14

to Einstein whenever you want he has

1:16

instant access to the sum of all human

1:18

knowledge and will answer anything you

1:20

want within seconds never running out of

1:21

patience he can also take on any role

1:23

you want a comedian poet doctor coach

1:27

and will be an expert within that field

1:29

he has some human-like limitations

1:31

though he can make mistakes he can jump

1:33

to conclusions he can misunderstand you

1:35

but the biggest limitation is actually

1:37

your imagination and your ability to

1:39

communicate effectively with them this

1:41

skill is known as prompt engineering and

1:43

in the age of AI this is as essential as

1:46

reading and writing most people vastly

1:49

underestimate what this Einstein in your

1:51

basement can do it's like going to the

1:53

real Einstein and asking him to proof

1:55

read a high school report or hiring a

1:56

worldclass five-star chef and having him

1:59

chop onion the more you interact with

2:01

Einstein the more you will discover

2:02

surprising and Powerful ways for him to

2:04

help you or your company okay enough

What is AI

2:07

fluffy metaphors let's clarify some

2:08

terms AI as you probably know stands for

2:11

artificial intelligence AI is not new FS

2:14

like machine learning and computer

2:15

vision have been around for decades

2:17

whenever you see a YouTube

2:18

recommendation or a web search result or

2:21

whenever you get a credit card

2:22

transaction approved that's traditional

2:24

AI in action generative AI is AI that

2:27

generates new original content rather

2:29

than just finding or classifying

2:30

existing content that's the G in GPT for

2:33

example large language models or llms

2:36

are a type of generative AI that can

2:38

communicate using normal human language

2:41

chat GPT is a product by the company

2:43

open AI it started as an llm essentially

2:46

an advanced chatbot using a new

2:47

architecture called the Transformer

2:49

architecture which by the way is the T

2:51

in GPT it is so fluent at human language

2:54

that anyone can use it you don't need to

2:55

be an AI expert or programmer and that's

2:57

kind of what triggered the whole

2:58

Revolution so how does it actually work

How does it work

3:02

well a large language model is an

3:03

artificial neural network basically a

3:06

bunch of numbers or or parameters

3:08

connected to each other similar to how

3:09

our brain is a bunch of neurons or brain

3:11

cells connected to each other neural

3:12

networks only deal with numbers you send

3:15

in numbers and depending on how the

3:16

parameters are set all the numbers come

3:18

out but any kind of content such as text

3:20

or images can be represented as numbers

3:22

so let's say I write dogs are when I

3:25

send that to a large language model that

3:27

gets converted to numbers processed by

3:29

the neural network and then the

3:30

resulting numbers are converted back

3:31

into text in this case the word animals

3:34

dogs are animals so yeah this is

3:36

basically a guest to next word machine

3:39

the interesting part is if we take that

3:41

output and combine it with the input and

3:43

send it through the model again then it

3:45

will continue adding new words that's

3:46

what's going on behind the scenes when

3:48

you type something in chat GPT in this

3:50

case for example it generated a whole

3:51

story and I can continue this

3:53

indefinitely by adding more

3:55

prompts a large language model may have

3:58

billions or even trillions of of

3:59

parameters that's why they're called

4:01

large so how are all these numbers set

Training

4:04

well not through manual programming that

4:06

would be impossible but through training

4:09

just like babies learning to speak a

4:10

baby isn't told how to speak she doesn't

4:13

get an instruction manual instead she

4:15

listens to people speaking around her

4:16

and when she's heard enough she starts

4:18

seeing the pattern she speaks a few

4:20

words at first to the Delight of her

4:21

parents and then later on full sentences

4:24

similarly during a training period the

4:26

language model is fed a mindboggling

4:28

amount of text to learn from Mostly from

4:31

internet sources it then plays guess the

4:33

next word with all of this over and over

4:35

again and the parameters are

4:37

automatically tweaked until it starts

4:38

getting really good at predicting the

4:40

next word this is called back

4:41

propagation which is a fancy term for oh

4:44

I guessed wrong I better change

4:45

something however to become truly useful

4:47

a model also needs to undergo human

4:49

training this is called reinforcement

4:51

learning with human feedback and it

4:53

involves thousands of hours of humans

4:55

painstakingly testing and evaluating

4:57

output from the model and giving

4:58

feedback kind of like training a a dog

5:01

with a clicker to reinforce good

5:02

behavior that's why a model like GPT

5:04

won't tell you how to rob a bank it

5:06

knows very well how to rob a bank but

5:08

through human training it has learned

5:09

that it shouldn't help people commit

5:11

crimes when training is done the model

5:13

is mostly Frozen other than some fine

5:15

tuning that can happen later that's what

5:17

the P stands for in GPT pre-trained

5:19

although in the future we will probably

5:20

have models that can learn continuously

5:22

rather than just uh during training and

5:24

fine-tuning now although chat GPT kind

Models

5:26

of got the ball rolling GPT isn't the

5:29

only model out there in fact new models

5:31

are sprouting like

5:32

mushrooms they vary a lot in terms of

5:34

speed capability and cost some can be

5:37

downloaded and run locally others are

5:38

only online some are free or open source

5:41

others are commercial products some are

5:43

super easy to use While others require

5:45

complicated technical setup some are

5:47

specialized for certain use cases others

5:50

are more General and can be used for

5:51

almost anything and some are baked into

5:54

products in the form of co-pilots or or

5:56

chat windows it's it's the Wild West

6:00

just keep in mind that you generally get

6:01

what you pay for so with a free model

6:04

you may just be getting a smart high

6:06

school student in your basement rather

6:08

than Einstein the difference between for

6:11

example GPT 3.5 and GPT 4 is massive

Different Models

6:15

note that there are different types of

6:17

generative AI models that generate

6:18

different types of content texto text

6:20

models like gpc4 take text as input and

6:23

generate text as output the text can be

6:25

natural language but it can also be

6:27

structured information like code Json

6:29

HTML I use this a lot myself to generate

6:32

code when programming uh it saves an

6:34

incredible amount of time and I also

6:35

learn a lot from the code it generates

6:37

text to image models will generate

6:39

images describe what you want and an

6:41

image gets generated for you you can

6:42

even pick a style image to image models

6:45

can do things like transforming or

6:47

combining images and we have image to

6:50

text models which describe the contents

6:52

of a given image and speech to text

6:54

models create voice transcriptions which

6:56

is useful for things like uh meeting

6:58

notes text Audio models they generate

7:00

music or sounds from a prompt for

7:02

example here is some sound generated

7:04

from The Prompt people talking in a

7:08

busy okay guys enough stop now thank you

7:13

and there are even text to video models

7:15

that generate videos from a prompt

7:17

sooner or later we'll have infinite

7:18

movie series that autog generate the

7:20

next episode tailored to your tastes as

7:22

you're watching kind of scary if you

7:24

think about it one Trend now is

7:26

multimodal AI products meaning they

7:28

combine different models into one

7:30

product so you can work with text images

7:32

audio Etc without switching tools the

7:35

chat GPT mobile app is a good example of

7:37

this just for fun I took a photo of this

7:40

room and I asked where I could hide

7:41

stuff I kind of like that it mentioned

7:44

the stove but warned that that it could

7:46

get hot there when I have things to

7:48

figure out such as the contents of this

7:50

video I like to take walks using chat

7:52

GPT as as a sounding board I start by

7:55

saying always respond with the word okay

7:57

unless I ask you for something that way

7:59

it'll just listen and not interrupt

8:01

after I finish dumping my thoughts I ask

8:03

for feedback we have some discussion and

8:06

then I ask it to summarize and text

8:07

afterwards I really recommend trying

8:09

this it's it's a really useful way to

8:11

use tools like this turns out Einstein

8:13

isn't stuck in the basement after all

8:15

you can take him out for a walk

8:17

initially language models were just word

8:19

predictors statistical machines with

8:22

limited practical use but as they became

8:24

larger and were trained on more data

8:26

they started gaining emergent

8:28

capabilities expected capabilities that

8:30

surprised even the developers of the

8:31

technology they could roleplay write

8:34

poetry write high quality code discuss

8:36

company strategy provide legal and

8:38

medical advice coach teach basically

8:41

creative and intellectual things that

8:43

only humans could do previously it turns

8:45

out that when a model has seen enough

8:47

text and images it starts to see

8:49

patterns and understand higher level

8:51

Concepts just like a baby learning to

8:53

understand the world let's take a simple

8:55

example I'll give gpc4 this little

8:57

drawing that involves a string a pair of

9:00

scissors an egg a pot and a fire what

9:03

will happen if I use the scissors the

9:05

model has most likely not been trained

9:07

on this exact scenario yet it gave a

9:10

pretty good answer which demonstrates a

9:11

basic understanding of the nature of

9:13

scissors eggs gravity and heat when GPT

9:16

4 was released I started using it as a

9:18

coding assistant and I was blown away

9:20

when prompted effectively it was a

9:22

better programmer than anyone I've

9:23

worked with same with article writing

9:25

product design Workshop planning and

9:27

just about anything I used it for for

9:29

the main bottleneck was my prompt

9:32

engineering skills so I decided to make

9:33

a career shift and focus entirely on

9:35

learning and teaching how to make this

9:37

technology useful hence this video now

9:40

let's take a step back and look at the

9:41

implications for 300,000 years or so we

9:44

homosapiens have been the most

9:46

intelligent species on Earth depending

9:48

of course on how you define intelligence

9:50

but the thing is our intellectual

9:51

capabilities aren't really improving

9:53

that much our brains are about the same

9:55

size same weight as they've been for

9:56

thousands of years computers on the

9:58

other hand have been around for only 80

10:00

years or so and now with generative AI

10:02

they are suddenly capable of speaking

10:04

human languages fluently and carrying

10:06

out an increasing number of intellectual

10:08

creative tasks that previously only

10:10

humans could do so we are right here at

10:12

the Crossing Point where AI is better at

10:14

some things and humans are better at

10:15

some things but ai's capabilities are

10:17

improving at an exponential rate while

10:19

ours aren't we don't know how long that

10:22

exponential Improvement will continue or

10:24

if it will level off at some point but

10:25

we're definitely entering a new world

10:27

order now this isn't the first

10:29

Revolution we've experienced we tamed

10:31

fire we learned how to do agriculture we

10:33

invented the printing press steam power

10:35

Telegraph these were all revolutionary

10:37

changes but they took decades or

10:39

centuries to become widespread in the AI

10:42

Revolution new technology spreads

10:44

worldwide almost instantly dealing with

10:46

this rate of change is a huge challenge

10:48

for both individuals and

The AI Mindset

10:50

companies I've noticed that people and

10:52

companies tend to fall into different

10:54

kind of mindset categories when it comes

10:56

to AI on one side we have denial the

10:59

belief that AI cannot do my job or we

11:02

don't have time to look into this

11:03

technology this is a dangerous place to

11:05

be a common saying is AI might not take

11:08

your job but people using AI will and

11:11

this is true for both individuals and

11:13

companies on the other side of the scale

11:15

we have panic and despair the belief

11:16

that AI is going to take my job no

11:18

matter what AI is going to make my

11:19

company go bankrupt neither of these

11:21

mindsets are helpful so I propose a

11:24

middle ground a balanced positive

11:26

mindset AI is going to make me my team

11:28

my company insanely productive

11:31

personally with this mindset I feel like

11:33

I've gained superpowers I can go from

11:35

idea to result in so much shorter time I

11:38

can focus more on what I want to achieve

11:40

and less on the grunt workk of building

11:41

things and I'm learning a lot faster too

11:43

it's like having an awesome Mentor with

11:45

me at all times this mindset not only

11:47

feels good but it also equips you for

11:49

the future makes you less likely to lose

11:51

your job or your company and more likely

11:53

to thrive in the age of AI despite all

11:55

the

Is human role needed

11:56

uncertainty so one important question is

11:59

is human role X needed in the age of AI

12:02

for example are doctors needed

12:03

developers lawyers CEOs uh whatever so

12:06

this question becomes more and more

12:08

relevant as the AI capabilities improve

12:11

well some jobs will disappear for sure

12:13

but for most roles I think we humans are

12:15

still needed someone with domain

12:17

knowledge still needs to decide what to

12:19

ask the AI how to formulate The Prompt

12:21

what context needs to be provided and

12:23

how to evaluate the result AI models

12:25

aren't perfect they can be absolutely

12:27

brilliant sometimes but sometimes also

12:30

terribly stupid they can sometimes

12:32

hallucinate and provide bogus

12:33

information in a very convincing way so

12:36

when should you trust the AI response

12:38

when should you double check or do the

12:39

work yourself what about legal

12:41

compliance data security what

12:43

information can we send to an AI model

12:45

and where is that data stored a human

12:48

expert is needed to make these judgment

12:50

calls and compensate for the weaknesses

12:51

of the AI model so I recommend thinking

12:54

of AI as your colleague a genius but

12:56

also an oddball with some personal

12:58

quirks that you need to learn to work

12:59

with you need to recognize when your

13:01

Genius colleague is drunk as a doctor my

13:04

AI colleague can help diagnose rare

13:06

diseases that I didn't even know existed

13:08

as a lawyer my AI colleague could do

13:10

legal research and review contracts

13:12

allowing me to spend more time with my

13:13

client or as a teacher my AI colleague

13:16

could grade tests help generate course

13:18

content provide individual support to

13:20

students Etc and if you're not sure how

13:22

it can help you just ask it I work as X

13:26

how can you help me overall I find that

13:29

the combination of human plus AI That's

13:31

where the magic lies it's important to

Models vs products

13:34

distinguish between the models and the

13:36

products that build on top of them as a

13:38

user you don't normally interact with

13:39

the model directly instead you interact

13:41

with a product website or a mobile app

13:43

which in turn talks to the model behind

13:45

the scenes products provide a user

13:47

interface and add capabilities and data

13:49

that aren't part of the model itself for

13:51

example the chat gpt product keeps track

13:54

of your message history while the GPT 4

13:56

model itself doesn't have any message

13:58

history history as a developer you can

14:00

use these models to build your own AI

14:02

powered products and features for

14:04

example let's say you have an e-learning

14:06

site you could add a chat bot to answer

14:08

questions about the courses or as a

14:10

recruitment company you might build AI

14:12

powered tools to help evaluate

14:13

candidates in both these cases your

14:15

users interact with your product and

14:17

then your product interacts with the

14:18

model this is done via apis or

14:20

application programming interfaces which

14:22

allow your code to talk to the model so

14:24

here's a simple example of using open AI

14:27

API to talk to GP

14:29

not a lot of code needed and here's

14:31

another example of the automatic

14:33

candidate evaluation thing I talked

14:34

about it takes a job description and a

14:37

bunch of CVS in a folder and evaluates

14:39

each candidate automatically and

14:41

incidentally the code itself is mostly

14:43

AI written as a product developer you

14:46

can use AI models kind of like an

14:48

external brain to insert intelligence

14:50

into your product very powerful in order

Prompt engineering

14:54

to use generative AI effectively you

14:56

need to get good at prompt engineering

14:58

or prompt design as I prefer to call it

15:00

this skill is needed both as a user and

15:02

as a product developer because in both

15:04

cases you need to be able to craft

15:06

effective prompts that produce useful

15:08

results from an AI model here's an

15:10

example let's say I want help planning a

15:12

workshop this prompt is unlikely to give

15:15

useful results because no matter how

15:17

smart the AI is if it doesn't know the

15:19

context of my workshop it can only give

15:21

fague high level recommendations the

15:23

second prompt is better now I provided

15:25

some context this is normally done

15:27

iteratively write a prompt look at the

15:29

result add a follow-up prompt to provide

15:31

more information or edit the original

15:33

prompt and rinse and repeat until you

15:35

get a good result in this third approach

15:37

I ask it to interview me so instead of

15:39

me providing a bunch of context up front

15:42

I'm basically saying what do you need to

15:43

know in order order to help me and then

15:45

it will propose a workshop agenda after

15:47

I often combine these two I provide a

15:49

bit of context and then I tell it to ask

15:51

me if it needs any more information

15:53

these are just some examples of prompt

15:54

engineering techniques so overall the

15:57

better you get at prompt engineering the

15:58

faster and better results you will get

16:00

from AI there are plenty of courses

16:02

books videos articles to help you learn

16:04

this but the most important thing is is

16:06

to practice and Learn by doing a nice

16:08

side effect is that you will become

16:09

better at communicating in general since

16:11

prompt engineering is really all about

16:13

Clarity and effective

16:15

communication I think the next Frontier

Autonomous agents

16:17

for generative AI is autonomous agents

16:19

with tools these are AI powerered

16:21

software entities that run on their own

16:23

rather than just sitting around waiting

16:24

for you to prompt them all the time so

16:26

you go down to Einstein in your basement

16:28

and do what a good leader would do for a

16:29

team you give him a high level Mission

16:31

and the tools needed to accomplish it

16:33

and then open the door and let him out

16:35

to run his own show without

16:36

micromanagement the tools could be

16:38

things like access to the internet

16:40

access to money ability to send and

16:42

receive messages order pizza or whatever

16:45

for this prompt engineering becomes even

16:47

more important because your autonomous

16:49

tool wielding agent can do a lot of good

16:51

or a lot of harm depending on how well

16:54

you craft that mission

16:55

statement all right let's wrap it up

16:58

here are the key things I hope you will

17:00

remember from this video generative AI

17:02

is a super useful tool that can help

17:04

both you your team and your company in a

17:06

big way the better you understand it the

17:08

more likely it is to be an opportunity

17:10

rather than a threat generative AI is

17:12

more powerful than you think the biggest

17:14

limitation is not the technology but

17:17

your imagination like what can I do and

17:19

your prompt engineering skills how do I

17:21

do it prompt engineering design is a

17:24

crucial skill like all new skills just

17:27

accept that you will kind of suck at it

17:29

at first but you'll improve over time

17:31

with deliberate practice so my best tip

17:33

is experiment make this part of your

17:36

day-to-day life and the Learning Happens

17:38

automatically hope this video was

17:40

helpful thanks for

17:41

[Music]

17:55

watching